Copyrighted Material

MATHEMATICAL Methods for Physicists

SIXTH EDITION

ARFKEN & WEBER

MATHEMATICAL METHODS FOR PHYSICISTS

SIXTH EDITION

George B. Arfken
Miami University
Oxford, OH

Hans J. Weber
University of Virginia
Charlottesville, VA

Amsterdam Boston Heidelberg London New York Oxford
Paris San Diego San Francisco Singapore Sydney Tokyo

This page intentionally left blank

MATHEMATICAL METHODS FOR PHYSICISTS

SIXTH EDITION

This page intentionally left blank

Acquisitions Editor	Tom Singer
Project Manager	Simon Crump
Marketing Manager	Linda Beattie
Cover Design	Eric DeCicco
Composition	VTEX Typesetting Services
Cover Printer	Phoenix Color
Interior Printer	The Maple–Vail Book Manufacturing Group

Elsevier Academic Press
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
84 Theobald's Road, London WC1X 8RR, UK

This book is printed on acid-free paper. ♾

Copyright © 2005, Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone: (+44) 1865 843830, fax: (+44) 1865 853333, e-mail: permissions@elsevier.co.uk. You may also complete your request on-line via the Elsevier homepage (http://elsevier.com), by selecting "Customer Support" and then "Obtaining Permissions."

Library of Congress Cataloging-in-Publication Data
Appication submitted

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 0-12-059876-0 Case bound
ISBN: 0-12-088584-0 International Students Edition

For all information on all Elsevier Academic Press Publications
visit our Web site at www.books.elsevier.com

Printed in the United States of America
05 06 07 08 09 10 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries
www.elsevier.com | www.bookaid.org | www.sabre.org
ELSEVIER BOOK AID International Sabre Foundation

CONTENTS

Preface **xi**

1 Vector Analysis **1**
1.1 *Definitions, Elementary Approach* . . . 1
1.2 *Rotation of the Coordinate Axes* . . . 7
1.3 *Scalar or Dot Product* . . . 12
1.4 *Vector or Cross Product* . . . 18
1.5 *Triple Scalar Product, Triple Vector Product* . . . 25
1.6 *Gradient,* ∇ . . . 32
1.7 *Divergence,* ∇ . . . 38
1.8 *Curl,* $\nabla\times$ . . . 43
1.9 *Successive Applications of* ∇ . . . 49
1.10 *Vector Integration* . . . 54
1.11 *Gauss' Theorem* . . . 60
1.12 *Stokes' Theorem* . . . 64
1.13 *Potential Theory* . . . 68
1.14 *Gauss' Law, Poisson's Equation* . . . 79
1.15 *Dirac Delta Function* . . . 83
1.16 *Helmholtz's Theorem* . . . 95
Additional Readings . . . 101

2 Vector Analysis in Curved Coordinates and Tensors **103**
2.1 *Orthogonal Coordinates in* $\mathbb{R}^3$ . . . 103
2.2 *Differential Vector Operators* . . . 110
2.3 *Special Coordinate Systems: Introduction* . . . 114
2.4 *Circular Cylinder Coordinates* . . . 115
2.5 *Spherical Polar Coordinates* . . . 123

2.6 *Tensor Analysis* . . . 133
2.7 *Contraction, Direct Product* . . . 139
2.8 *Quotient Rule* . . . 141
2.9 *Pseudotensors, Dual Tensors* . . . 142
2.10 *General Tensors* . . . 151
2.11 *Tensor Derivative Operators* . . . 160
Additional Readings . . . 163

3 Determinants and Matrices 165
3.1 *Determinants* . . . 165
3.2 *Matrices* . . . 176
3.3 *Orthogonal Matrices* . . . 195
3.4 *Hermitian Matrices, Unitary Matrices* . . . 208
3.5 *Diagonalization of Matrices* . . . 215
3.6 *Normal Matrices* . . . 231
Additional Readings . . . 239

4 Group Theory 241
4.1 *Introduction to Group Theory* . . . 241
4.2 *Generators of Continuous Groups* . . . 246
4.3 *Orbital Angular Momentum* . . . 261
4.4 *Angular Momentum Coupling* . . . 266
4.5 *Homogeneous Lorentz Group* . . . 278
4.6 *Lorentz Covariance of Maxwell's Equations* . . . 283
4.7 *Discrete Groups* . . . 291
4.8 *Differential Forms* . . . 304
Additional Readings . . . 319

5 Infinite Series 321
5.1 *Fundamental Concepts* . . . 321
5.2 *Convergence Tests* . . . 325
5.3 *Alternating Series* . . . 339
5.4 *Algebra of Series* . . . 342
5.5 *Series of Functions* . . . 348
5.6 *Taylor's Expansion* . . . 352
5.7 *Power Series* . . . 363
5.8 *Elliptic Integrals* . . . 370
5.9 *Bernoulli Numbers, Euler–Maclaurin Formula* . . . 376
5.10 *Asymptotic Series* . . . 389
5.11 *Infinite Products* . . . 396
Additional Readings . . . 401

6 Functions of a Complex Variable I Analytic Properties, Mapping 403
6.1 *Complex Algebra* . . . 404
6.2 *Cauchy–Riemann Conditions* . . . 413
6.3 *Cauchy's Integral Theorem* . . . 418

6.4 *Cauchy's Integral Formula* . . . 425
6.5 *Laurent Expansion* . . . 430
6.6 *Singularities* . . . 438
6.7 *Mapping* . . . 443
6.8 *Conformal Mapping* . . . 451
Additional Readings . . . 453

7 Functions of a Complex Variable II **455**
7.1 *Calculus of Residues* . . . 455
7.2 *Dispersion Relations* . . . 482
7.3 *Method of Steepest Descents* . . . 489
Additional Readings . . . 497

8 The Gamma Function (Factorial Function) **499**
8.1 *Definitions, Simple Properties* . . . 499
8.2 *Digamma and Polygamma Functions* . . . 510
8.3 *Stirling's Series* . . . 516
8.4 *The Beta Function* . . . 520
8.5 *Incomplete Gamma Function* . . . 527
Additional Readings . . . 533

9 Differential Equations **535**
9.1 *Partial Differential Equations* . . . 535
9.2 *First-Order Differential Equations* . . . 543
9.3 *Separation of Variables* . . . 554
9.4 *Singular Points* . . . 562
9.5 *Series Solutions—Frobenius' Method* . . . 565
9.6 *A Second Solution* . . . 578
9.7 *Nonhomogeneous Equation—Green's Function* . . . 592
9.8 *Heat Flow, or Diffusion, PDE* . . . 611
Additional Readings . . . 618

10 Sturm–Liouville Theory—Orthogonal Functions **621**
10.1 *Self-Adjoint ODEs* . . . 622
10.2 *Hermitian Operators* . . . 634
10.3 *Gram–Schmidt Orthogonalization* . . . 642
10.4 *Completeness of Eigenfunctions* . . . 649
10.5 *Green's Function—Eigenfunction Expansion* . . . 662
Additional Readings . . . 674

11 Bessel Functions **675**
11.1 *Bessel Functions of the First Kind,* $J_\nu(x)$ . . . 675
11.2 *Orthogonality* . . . 694
11.3 *Neumann Functions* . . . 699
11.4 *Hankel Functions* . . . 707
11.5 *Modified Bessel Functions,* $I_\nu(x)$ *and* $K_\nu(x)$ . . . 713

11.6 *Asymptotic Expansions* . . . 719
11.7 *Spherical Bessel Functions* . . . 725
Additional Readings . . . 739

12 Legendre Functions **741**
12.1 *Generating Function* . . . 741
12.2 *Recurrence Relations* . . . 749
12.3 *Orthogonality* . . . 756
12.4 *Alternate Definitions* . . . 767
12.5 *Associated Legendre Functions* . . . 771
12.6 *Spherical Harmonics* . . . 786
12.7 *Orbital Angular Momentum Operators* . . . 793
12.8 *Addition Theorem for Spherical Harmonics* . . . 797
12.9 *Integrals of Three Y's* . . . 803
12.10 *Legendre Functions of the Second Kind* . . . 806
12.11 *Vector Spherical Harmonics* . . . 813
Additional Readings . . . 816

13 More Special Functions **817**
13.1 *Hermite Functions* . . . 817
13.2 *Laguerre Functions* . . . 837
13.3 *Chebyshev Polynomials* . . . 848
13.4 *Hypergeometric Functions* . . . 859
13.5 *Confluent Hypergeometric Functions* . . . 863
13.6 *Mathieu Functions* . . . 869
Additional Readings . . . 879

14 Fourier Series **881**
14.1 *General Properties* . . . 881
14.2 *Advantages, Uses of Fourier Series* . . . 888
14.3 *Applications of Fourier Series* . . . 892
14.4 *Properties of Fourier Series* . . . 903
14.5 *Gibbs Phenomenon* . . . 910
14.6 *Discrete Fourier Transform* . . . 914
14.7 *Fourier Expansions of Mathieu Functions* . . . 919
Additional Readings . . . 929

15 Integral Transforms **931**
15.1 *Integral Transforms* . . . 931
15.2 *Development of the Fourier Integral* . . . 936
15.3 *Fourier Transforms—Inversion Theorem* . . . 938
15.4 *Fourier Transform of Derivatives* . . . 946
15.5 *Convolution Theorem* . . . 951
15.6 *Momentum Representation* . . . 955
15.7 *Transfer Functions* . . . 961
15.8 *Laplace Transforms* . . . 965

15.9 *Laplace Transform of Derivatives* . . . 971
15.10 *Other Properties* . . . 979
15.11 *Convolution (Faltungs) Theorem* . . . 990
15.12 *Inverse Laplace Transform* . . . 994
Additional Readings . . . 1003

16 Integral Equations **1005**
16.1 *Introduction* . . . 1005
16.2 *Integral Transforms, Generating Functions* . . . 1012
16.3 *Neumann Series, Separable (Degenerate) Kernels* . . . 1018
16.4 *Hilbert–Schmidt Theory* . . . 1029
Additional Readings . . . 1036

17 Calculus of Variations **1037**
17.1 *A Dependent and an Independent Variable* . . . 1038
17.2 *Applications of the Euler Equation* . . . 1044
17.3 *Several Dependent Variables* . . . 1052
17.4 *Several Independent Variables* . . . 1056
17.5 *Several Dependent and Independent Variables* . . . 1058
17.6 *Lagrangian Multipliers* . . . 1060
17.7 *Variation with Constraints* . . . 1065
17.8 *Rayleigh–Ritz Variational Technique* . . . 1072
Additional Readings . . . 1076

18 Nonlinear Methods and Chaos **1079**
18.1 *Introduction* . . . 1079
18.2 *The Logistic Map* . . . 1080
18.3 *Sensitivity to Initial Conditions and Parameters* . . . 1085
18.4 *Nonlinear Differential Equations* . . . 1088
Additional Readings . . . 1107

19 Probability **1109**
19.1 *Definitions, Simple Properties* . . . 1109
19.2 *Random Variables* . . . 1116
19.3 *Binomial Distribution* . . . 1128
19.4 *Poisson Distribution* . . . 1130
19.5 *Gauss' Normal Distribution* . . . 1134
19.6 *Statistics* . . . 1138
Additional Readings . . . 1150
General References . . . 1150

Index **1153**

This page intentionally left blank

PREFACE

Through six editions now, *Mathematical Methods for Physicists* has provided all the mathematical methods that aspirings scientists and engineers are likely to encounter as students and beginning researchers. More than enough material is included for a two-semester undergraduate or graduate course.

The book is advanced in the sense that mathematical relations are almost always proven, in addition to being illustrated in terms of examples. These proofs are not what a mathematician would regard as rigorous, but sketch the ideas and emphasize the relations that are essential to the study of physics and related fields. This approach incorporates theorems that are usually not cited under the most general assumptions, but are tailored to the more restricted applications required by physics. For example, Stokes' theorem is usually applied by a physicist to a surface with the tacit understanding that it be simply connected. Such assumptions have been made more explicit.

PROBLEM-SOLVING SKILLS

The book also incorporates a deliberate focus on problem-solving skills. This more advanced level of understanding and active learning is routine in physics courses and requires practice by the reader. Accordingly, extensive problem sets appearing in each chapter form an integral part of the book. They have been carefully reviewed, revised and enlarged for this Sixth Edition.

PATHWAYS THROUGH THE MATERIAL

Undergraduates may be best served if they start by reviewing Chapter 1 according to the level of training of the class. Section 1.2 on the transformation properties of vectors, the cross product, and the invariance of the scalar product under rotations may be postponed until tensor analysis is started, for which these sections form the introduction and serve as

examples. They may continue their studies with linear algebra in Chapter 3, then perhaps tensors and symmetries (Chapters 2 and 4), and next real and complex analysis (Chapters 5–7), differential equations (Chapters 9, 10), and special functions (Chapters 11–13).

In general, the core of a graduate one-semester course comprises Chapters 5–10 and 11–13, which deal with real and complex analysis, differential equations, and special functions. Depending on the level of the students in a course, some linear algebra in Chapter 3 (eigenvalues, for example), along with symmetries (group theory in Chapter 4), and tensors (Chapter 2) may be covered as needed or according to taste. Group theory may also be included with differential equations (Chapters 9 and 10). Appropriate relations have been included and are discussed in Chapters 4 and 9.

A two-semester course can treat tensors, group theory, and special functions (Chapters 11–13) more extensively, and add Fourier series (Chapter 14), integral transforms (Chapter 15), integral equations (Chapter 16), and the calculus of variations (Chapter 17).

CHANGES TO THE SIXTH EDITION

Improvements to the Sixth Edition have been made in nearly all chapters adding examples and problems and more derivations of results. Numerous left-over typos caused by scanning into LaTeX, an error-prone process at the rate of many errors per page, have been corrected along with mistakes, such as in the Dirac γ-matrices in Chapter 3. A few chapters have been relocated. The Gamma function is now in Chapter 8 following Chapters 6 and 7 on complex functions in one variable, as it is an application of these methods. Differential equations are now in Chapters 9 and 10. A new chapter on probability has been added, as well as new subsections on differential forms and Mathieu functions in response to persistent demands by readers and students over the years. The new subsections are more advanced and are written in the concise style of the book, thereby raising its level to the graduate level. Many examples have been added, for example in Chapters 1 and 2, that are often used in physics or are standard lore of physics courses. A number of additions have been made in Chapter 3, such as on linear dependence of vectors, dual vector spaces and spectral decomposition of symmetric or Hermitian matrices. A subsection on the diffusion equation emphasizes methods to adapt solutions of partial differential equations to boundary conditions. New formulas have been developed for Hermite polynomials and are included in Chapter 13 that are useful for treating molecular vibrations; they are of interest to the chemical physicists.

ACKNOWLEDGMENTS

We have benefited from the advice and help of many people. Some of the revisions are in response to comments by readers and former students, such as Dr. K. Bodoor and J. Hughes. We are grateful to them and to our Editors Barbara Holland and Tom Singer who organized accuracy checks. We would like to thank in particular Dr. Michael Bozoian and Prof. Frank Harris for their invaluable help with the accuracy checking and Simon Crump, Production Editor, for his expert management of the Sixth Edition.

CHAPTER 1

VECTOR ANALYSIS

1.1 DEFINITIONS, ELEMENTARY APPROACH

In science and engineering we frequently encounter quantities that have magnitude and magnitude only: mass, time, and temperature. These we label **scalar** quantities, which remain the same no matter what coordinates we use. In contrast, many interesting physical quantities have magnitude and, in addition, an associated direction. This second group includes displacement, velocity, acceleration, force, momentum, and angular momentum. Quantities with magnitude and direction are labeled **vector** quantities. Usually, in elementary treatments, a vector is defined as a quantity having magnitude and direction. To distinguish vectors from scalars, we identify vector quantities with boldface type, that is, $\mathbf{V}$.

Our vector may be conveniently represented by an arrow, with length proportional to the magnitude. The direction of the arrow gives the direction of the vector, the positive sense of direction being indicated by the point. In this representation, vector addition

$$\mathbf{C} = \mathbf{A} + \mathbf{B} \tag{1.1}$$

consists in placing the rear end of vector $\mathbf{B}$ at the point of vector $\mathbf{A}$. Vector $\mathbf{C}$ is then represented by an arrow drawn from the rear of $\mathbf{A}$ to the point of $\mathbf{B}$. This procedure, the triangle law of addition, assigns meaning to Eq. (1.1) and is illustrated in Fig. 1.1. By completing the parallelogram, we see that

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}, \tag{1.2}$$

as shown in Fig. 1.2. In words, vector addition is **commutative**.

For the sum of three vectors

$$\mathbf{D} = \mathbf{A} + \mathbf{B} + \mathbf{C},$$

Fig. 1.3, we may first add $\mathbf{A}$ and $\mathbf{B}$:

$$\mathbf{A} + \mathbf{B} = \mathbf{E}.$$

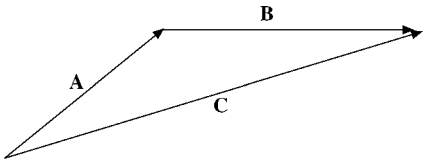

FIGURE 1.1 Triangle law of vector addition.

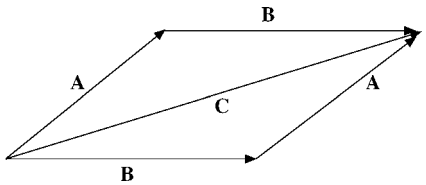

FIGURE 1.2 Parallelogram law of vector addition.

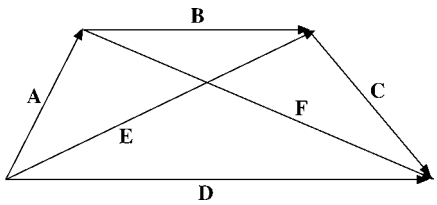

FIGURE 1.3 Vector addition is associative.

Then this sum is added to **C**:

$$\mathbf{D} = \mathbf{E} + \mathbf{C}.$$

Similarly, we may first add **B** and **C**:

$$\mathbf{B} + \mathbf{C} = \mathbf{F}.$$

Then

$$\mathbf{D} = \mathbf{A} + \mathbf{F}.$$

In terms of the original expression,

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}).$$

Vector addition is **associative**.

A direct physical example of the parallelogram addition law is provided by a weight suspended by two cords. If the junction point (*O* in Fig. 1.4) is in equilibrium, the vector

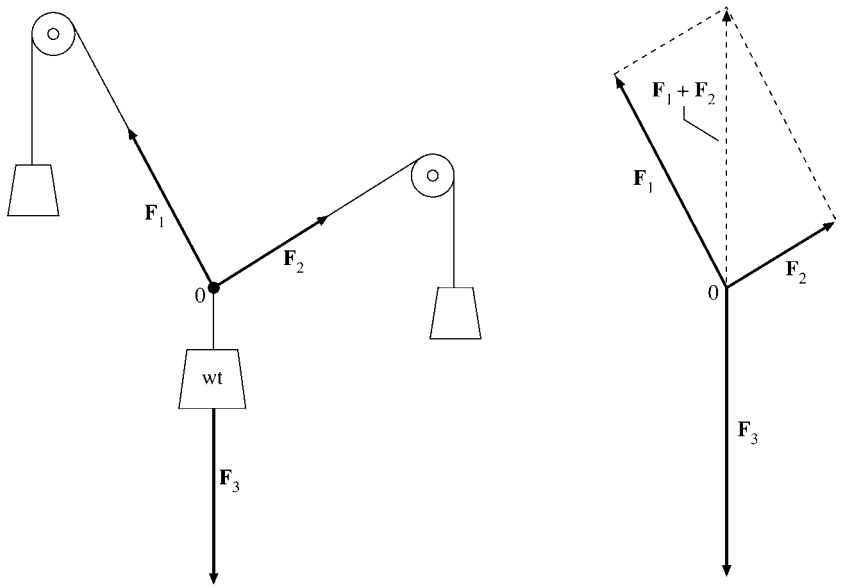

FIGURE 1.4 Equilibrium of forces: $\mathbf{F}_1 + \mathbf{F}_2 = -\mathbf{F}_3$.

sum of the two forces $\mathbf{F}_1$ and $\mathbf{F}_2$ must just cancel the downward force of gravity, $\mathbf{F}_3$. Here the parallelogram addition law is subject to immediate experimental verification.[1]

Subtraction may be handled by defining the negative of a vector as a vector of the same magnitude but with reversed direction. Then

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B}).$$

In Fig. 1.3,

$$\mathbf{A} = \mathbf{E} - \mathbf{B}.$$

Note that the vectors are treated as geometrical objects that are independent of any coordinate system. This concept of independence of a preferred coordinate system is developed in detail in the next section.

The representation of vector $\mathbf{A}$ by an arrow suggests a second possibility. Arrow $\mathbf{A}$ (Fig. 1.5), starting from the origin,[2] terminates at the point (A_x, A_y, A_z). Thus, if we agree that the vector is to start at the origin, the positive end may be specified by giving the Cartesian coordinates (A_x, A_y, A_z) of the arrowhead.

Although $\mathbf{A}$ could have represented any vector quantity (momentum, electric field, etc.), one particularly important vector quantity, the displacement from the origin to the point

[1] Strictly speaking, the parallelogram addition was introduced as a definition. Experiments show that if we assume that the forces are vector quantities and we combine them by parallelogram addition, the equilibrium condition of zero resultant force is satisfied.

[2] We could start from any point in our Cartesian reference frame; we choose the origin for simplicity. This freedom of shifting the origin of the coordinate system without affecting the geometry is called **translation invariance**.

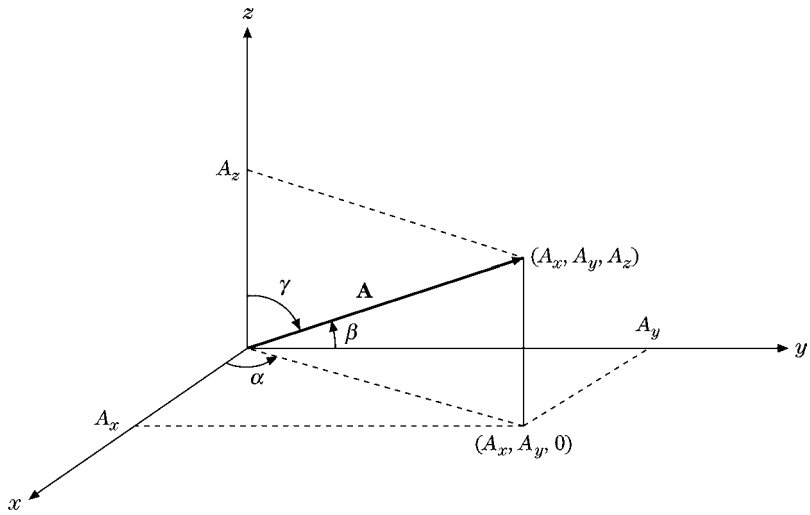

FIGURE 1.5 Cartesian components and direction cosines of **A**.

(x, y, z), is denoted by the special symbol **r**. We then have a choice of referring to the displacement as either the vector **r** or the collection (x, y, z), the coordinates of its endpoint:

$$\mathbf{r} \leftrightarrow (x, y, z). \tag{1.3}$$

Using r for the magnitude of vector **r**, we find that Fig. 1.5 shows that the endpoint coordinates and the magnitude are related by

$$x = r\cos\alpha, \qquad y = r\cos\beta, \qquad z = r\cos\gamma. \tag{1.4}$$

Here $\cos\alpha$, $\cos\beta$, and $\cos\gamma$ are called the **direction cosines**, α being the angle between the given vector and the positive x-axis, and so on. One further bit of vocabulary: The quantities A_x, A_y, and A_z are known as the (Cartesian) **components** of **A** or the **projections** of **A**, with $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$.

Thus, any vector **A** may be resolved into its components (or projected onto the coordinate axes) to yield $A_x = A\cos\alpha$, etc., as in Eq. (1.4). We may choose to refer to the vector as a single quantity **A** or to its components (A_x, A_y, A_z). Note that the subscript x in A_x denotes the x component and not a dependence on the variable x. The choice between using **A** or its components (A_x, A_y, A_z) is essentially a choice between a geometric and an algebraic representation. Use either representation at your convenience. The geometric "arrow in space" may aid in visualization. The algebraic set of components is usually more suitable for precise numerical or algebraic calculations.

Vectors enter physics in two distinct forms. (1) Vector **A** may represent a single force acting at a single point. The force of gravity acting at the center of gravity illustrates this form. (2) Vector **A** may be defined over some extended region; that is, **A** and its components may be functions of position: $A_x = A_x(x, y, z)$, and so on. Examples of this sort include the velocity of a fluid varying from point to point over a given volume and electric and magnetic fields. These two cases may be distinguished by referring to the vector defined over a region as a **vector field**. The concept of the vector defined over a region and

being a function of position will become extremely important when we differentiate and integrate vectors.

At this stage it is convenient to introduce unit vectors along each of the coordinate axes. Let $\hat{\mathbf{x}}$ be a vector of unit magnitude pointing in the positive x-direction, $\hat{\mathbf{y}}$, a vector of unit magnitude in the positive y-direction, and $\hat{\mathbf{z}}$ a vector of unit magnitude in the positive z-direction. Then $\hat{\mathbf{x}}A_x$ is a vector with magnitude equal to $|A_x|$ and in the x-direction. By vector addition,

$$\mathbf{A} = \hat{\mathbf{x}}A_x + \hat{\mathbf{y}}A_y + \hat{\mathbf{z}}A_z. \tag{1.5}$$

Note that if $\mathbf{A}$ vanishes, all of its components must vanish individually; that is, if

$$\mathbf{A} = 0, \qquad \text{then } A_x = A_y = A_z = 0.$$

This means that these unit vectors serve as a **basis**, or complete set of vectors, in the three-dimensional Euclidean space in terms of which any vector can be expanded. Thus, Eq. (1.5) is an assertion that the three unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ span our real three-dimensional space: Any vector may be written as a linear combination of $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$. Since $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ are linearly independent (no one is a linear combination of the other two), they form a **basis** for the real three-dimensional Euclidean space. Finally, by the Pythagorean theorem, the magnitude of vector $\mathbf{A}$ is

$$\boxed{|\mathbf{A}| = \left(A_x^2 + A_y^2 + A_z^2\right)^{1/2}.} \tag{1.6}$$

Note that the coordinate unit vectors are not the only complete set, or basis. This resolution of a vector into its components can be carried out in a variety of coordinate systems, as shown in Chapter 2. Here we restrict ourselves to Cartesian coordinates, where the unit vectors have the coordinates $\hat{\mathbf{x}} = (1, 0, 0)$, $\hat{\mathbf{y}} = (0, 1, 0)$ and $\hat{\mathbf{z}} = (0, 0, 1)$ and are all constant in length and direction, properties characteristic of Cartesian coordinates.

As a replacement of the graphical technique, addition and subtraction of vectors may now be carried out in terms of their components. For $\mathbf{A} = \hat{\mathbf{x}}A_x + \hat{\mathbf{y}}A_y + \hat{\mathbf{z}}A_z$ and $\mathbf{B} = \hat{\mathbf{x}}B_x + \hat{\mathbf{y}}B_y + \hat{\mathbf{z}}B_z$,

$$\mathbf{A} \pm \mathbf{B} = \hat{\mathbf{x}}(A_x \pm B_x) + \hat{\mathbf{y}}(A_y \pm B_y) + \hat{\mathbf{z}}(A_z \pm B_z). \tag{1.7}$$

It should be emphasized here that the unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ are used for convenience. They are not essential; we can describe vectors and use them entirely in terms of their components: $\mathbf{A} \leftrightarrow (A_x, A_y, A_z)$. This is the approach of the two more powerful, more sophisticated definitions of vector to be discussed in the next section. However, $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ emphasize the **direction**.

So far we have defined the operations of addition and subtraction of vectors. In the next sections, three varieties of multiplication will be defined on the basis of their applicability: a scalar, or inner, product, a vector product peculiar to three-dimensional space, and a direct, or outer, product yielding a second-rank tensor. Division by a vector is not defined.

Exercises

1.1.1 Show how to find **A** and **B**, given $\mathbf{A}+\mathbf{B}$ and $\mathbf{A}-\mathbf{B}$.

1.1.2 The vector **A** whose magnitude is 1.732 units makes equal angles with the coordinate axes. Find A_x, A_y, and A_z.

1.1.3 Calculate the components of a unit vector that lies in the xy-plane and makes equal angles with the positive directions of the x- and y-axes.

1.1.4 The velocity of sailboat A relative to sailboat B, $\mathbf{v}_{\text{rel}}$, is defined by the equation $\mathbf{v}_{\text{rel}} = \mathbf{v}_A - \mathbf{v}_B$, where $\mathbf{v}_A$ is the velocity of A and $\mathbf{v}_B$ is the velocity of B. Determine the velocity of A relative to B if

$$\mathbf{v}_A = 30 \text{ km/hr east}$$
$$\mathbf{v}_B = 40 \text{ km/hr north}.$$

ANS. $\mathbf{v}_{\text{rel}} = 50$ km/hr, 53.1° south of east.

1.1.5 A sailboat sails for 1 hr at 4 km/hr (relative to the water) on a steady compass heading of 40° east of north. The sailboat is simultaneously carried along by a current. At the end of the hour the boat is 6.12 km from its starting point. The line from its starting point to its location lies 60° east of north. Find the x (easterly) and y (northerly) components of the water's velocity.

ANS. $v_{\text{east}} = 2.73$ km/hr, $v_{\text{north}} \approx 0$ km/hr.

1.1.6 A vector equation can be reduced to the form $\mathbf{A} = \mathbf{B}$. From this show that the one vector equation is equivalent to **three** scalar equations. Assuming the validity of Newton's second law, $\mathbf{F} = m\mathbf{a}$, as a **vector** equation, this means that a_x depends only on F_x and is independent of F_y and F_z.

1.1.7 The vertices A, B, and C of a triangle are given by the points $(-1, 0, 2)$, $(0, 1, 0)$, and $(1, -1, 0)$, respectively. Find point D so that the figure $ABCD$ forms a plane parallelogram.

ANS. $(0, -2, 2)$ or $(2, 0, -2)$.

1.1.8 A triangle is defined by the vertices of three vectors **A**, **B** and **C** that extend from the origin. In terms of **A**, **B**, and **C** show that the **vector** sum of the successive sides of the triangle $(AB + BC + CA)$ is zero, where the side AB is from A to B, etc.

1.1.9 A sphere of radius a is centered at a point $\mathbf{r}_1$.

(a) Write out the algebraic equation for the sphere.
(b) Write out a **vector** equation for the sphere.

ANS. (a) $(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2 = a^2$.
(b) $\mathbf{r} = \mathbf{r}_1 + \mathbf{a}$, with $\mathbf{r}_1 =$ center.
(**a** takes on all directions but has a fixed magnitude a.)

1.1.10 A corner reflector is formed by three mutually perpendicular reflecting surfaces. Show that a ray of light incident upon the corner reflector (striking all three surfaces) is reflected back along a line parallel to the line of incidence.
Hint. Consider the effect of a reflection on the components of a vector describing the direction of the light ray.

1.1.11 *Hubble's law*. Hubble found that distant galaxies are receding with a velocity proportional to their distance from where we are on Earth. For the ith galaxy,

$$\mathbf{v}_i = H_0 \mathbf{r}_i ,$$

with us at the origin. Show that this recession of the galaxies from us does **not** imply that we are at the center of the universe. Specifically, take the galaxy at $\mathbf{r_1}$ as a new origin and show that Hubble's law is still obeyed.

1.1.12 Find the diagonal vectors of a unit cube with one corner at the origin and its three sides lying along Cartesian coordinates axes. Show that there are four diagonals with length $\sqrt{3}$. Representing these as vectors, what are their components? Show that the diagonals of the cube's faces have length $\sqrt{2}$ and determine their components.

1.2 ROTATION OF THE COORDINATE AXES[3]

In the preceding section vectors were defined or represented in two equivalent ways: (1) geometrically by specifying magnitude and direction, as with an arrow, and (2) algebraically by specifying the components relative to Cartesian coordinate axes. The second definition is adequate for the vector analysis of this chapter. In this section two more refined, sophisticated, and powerful definitions are presented. First, the vector field is defined in terms of the behavior of its components under rotation of the coordinate axes. This transformation theory approach leads into the tensor analysis of Chapter 2 and groups of transformations in Chapter 4. Second, the component definition of Section 1.1 is refined and generalized according to the mathematician's concepts of vector and vector space. This approach leads to function spaces, including the Hilbert space.

The definition of vector as a quantity with magnitude and direction is incomplete. On the one hand, we encounter quantities, such as elastic constants and index of refraction in anisotropic crystals, that have magnitude and direction **but** that are not vectors. On the other hand, our naïve approach is awkward to generalize to extend to more complex quantities. We seek a new definition of vector field using our coordinate vector $\mathbf{r}$ as a prototype.

There is a physical basis for our development of a new definition. We describe our physical world by mathematics, but it and any physical predictions we may make must be **independent** of our mathematical conventions.

In our specific case we assume that space is isotropic; that is, there is no preferred direction, or all directions are equivalent. Then the physical system being analyzed or the physical law being enunciated cannot and must not depend on our choice or **orientation** of the coordinate axes. Specifically, if a quantity S does not depend on the orientation of the coordinate axes, it is called a scalar.

[3]This section is optional here. It will be essential for Chapter 2.

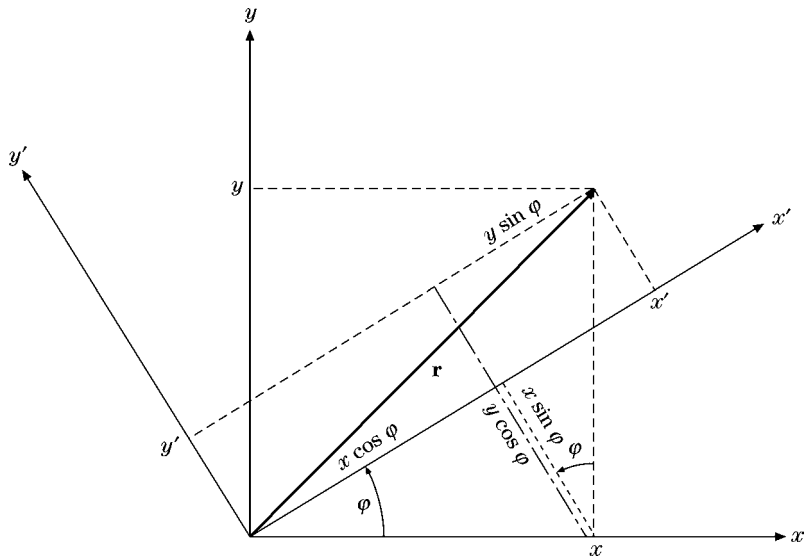

FIGURE 1.6 Rotation of Cartesian coordinate axes about the z-axis.

Now we return to the concept of vector **r** as a geometric object independent of the coordinate system. Let us look at **r** in two different systems, one rotated in relation to the other.

For simplicity we consider first the two-dimensional case. If the x-, y-coordinates are rotated counterclockwise through an angle φ, **keeping r**, **fixed** (Fig. 1.6), we get the following relations between the components resolved in the original system (unprimed) and those resolved in the new rotated system (primed):

$$\begin{aligned} x' &= x\cos\varphi + y\sin\varphi, \\ y' &= -x\sin\varphi + y\cos\varphi. \end{aligned} \tag{1.8}$$

We saw in Section 1.1 that a vector could be represented by the coordinates of a point; that is, the coordinates were proportional to the vector components. Hence the components of a vector must transform under rotation as coordinates of a point (such as **r**). Therefore whenever any pair of quantities A_x and A_y in the x-y coordinate system is transformed into (A'_x, A'_y) by this rotation of the coordinate system with

$$\begin{aligned} A'_x &\equiv A_x\cos\varphi + A_y\sin\varphi, \\ A'_y &\equiv -A_x\sin\varphi + A_y\cos\varphi, \end{aligned} \tag{1.9}$$

we **define**[4] A_x and A_y as the components of a vector **A**. Our vector now is defined in terms of the transformation of its components under rotation of the coordinate system. If A_x and A_y transform in the same way as x and y, the components of the general two-dimensional coordinate vector **r**, they are the components of a vector **A**. If A_x and A_y do not show this

[4] A scalar quantity does not depend on the orientation of coordinates; $S' = S$ expresses the fact that it is invariant under rotation of the coordinates.

form invariance (also called **covariance**) when the coordinates are rotated, they do not form a vector.

The vector field components A_x and A_y satisfying the defining equations, Eqs. (1.9), associate a magnitude A and a direction with each point in space. The magnitude is a scalar quantity, invariant to the rotation of the coordinate system. The direction (relative to the unprimed system) is likewise invariant to the rotation of the coordinate system (see Exercise 1.2.1). The result of all this is that the components of a vector may vary according to the rotation of the primed coordinate system. This is what Eqs. (1.9) say. But the variation with the angle is just such that the components in the rotated coordinate system, A'_x and A'_y, define a vector with the same magnitude and the same direction as the vector defined by the components A_x and A_y relative to the x-, y-coordinate axes. (Compare Exercise 1.2.1.) The components of **A** in a particular coordinate system constitute the **representation** of **A** in that coordinate system. Equations (1.9), the transformation relations, are a guarantee that the entity **A** is independent of the rotation of the coordinate system.

To go on to three and, later, four dimensions, we find it convenient to use a more compact notation. Let

$$\begin{aligned} x &\to x_1 \\ y &\to x_2 \end{aligned} \tag{1.10}$$

$$\begin{aligned} a_{11} &= \cos\varphi, & a_{12} &= \sin\varphi, \\ a_{21} &= -\sin\varphi, & a_{22} &= \cos\varphi. \end{aligned} \tag{1.11}$$

Then Eqs. (1.8) become

$$\begin{aligned} x'_1 &= a_{11}x_1 + a_{12}x_2, \\ x'_2 &= a_{21}x_1 + a_{22}x_2. \end{aligned} \tag{1.12}$$

The coefficient a_{ij} may be interpreted as a direction cosine, the cosine of the angle between x'_i and x_j; that is,

$$\begin{aligned} a_{12} &= \cos(x'_1, x_2) = \sin\varphi, \\ a_{21} &= \cos(x'_2, x_1) = \cos\left(\varphi + \frac{\pi}{2}\right) = -\sin\varphi. \end{aligned} \tag{1.13}$$

The advantage of the new notation[5] is that it permits us to use the summation symbol $\sum$ and to rewrite Eqs. (1.12) as

$$x'_i = \sum_{j=1}^{2} a_{ij}x_j, \qquad i = 1, 2. \tag{1.14}$$

Note that i remains as a parameter that gives rise to one equation when it is set equal to 1 and to a second equation when it is set equal to 2. The index j, of course, is a summation index, a dummy index, and, as with a variable of integration, j may be replaced by any other convenient symbol.

[5] You may wonder at the replacement of one parameter φ by four parameters a_{ij}. Clearly, the a_{ij} do not constitute a minimum set of parameters. For two dimensions the four a_{ij} are subject to the three constraints given in Eq. (1.18). The justification for this redundant set of direction cosines is the convenience it provides. Hopefully, this convenience will become more apparent in Chapters 2 and 3. For three-dimensional rotations (9 a_{ij} but only three independent) alternate descriptions are provided by: (1) the Euler angles discussed in Section 3.3, (2) quaternions, and (3) the Cayley–Klein parameters. These alternatives have their respective advantages and disadvantages.

The generalization to three, four, or N dimensions is now simple. The set of N quantities V_j is said to be the components of an N-dimensional vector $\mathbf{V}$ if and only if their values relative to the rotated coordinate axes are given by

$$V_i' = \sum_{j=1}^{N} a_{ij} V_j, \qquad i = 1, 2, \ldots, N. \tag{1.15}$$

As before, a_{ij} is the cosine of the angle between x_i' and x_j. Often the upper limit N and the corresponding range of i will not be indicated. It is taken for granted that you know how many dimensions your space has.

From the definition of a_{ij} as the cosine of the angle between the positive x_i' direction and the positive x_j direction we may write (Cartesian coordinates)[6]

$$a_{ij} = \frac{\partial x_i'}{\partial x_j}. \tag{1.16a}$$

Using the inverse rotation ($\varphi \to -\varphi$) yields

$$x_j = \sum_{i=1}^{2} a_{ij} x_i' \qquad \text{or} \qquad \frac{\partial x_j}{\partial x_i'} = a_{ij}. \tag{1.16b}$$

Note that these are **partial derivatives**. By use of Eqs. (1.16a) and (1.16b), Eq. (1.15) becomes

$$V_i' = \sum_{j=1}^{N} \frac{\partial x_i'}{\partial x_j} V_j = \sum_{j=1}^{N} \frac{\partial x_j}{\partial x_i'} V_j. \tag{1.17}$$

The direction cosines a_{ij} satisfy an **orthogonality condition**

$$\sum_i a_{ij} a_{ik} = \delta_{jk} \tag{1.18}$$

or, equivalently,

$$\sum_i a_{ji} a_{ki} = \delta_{jk}. \tag{1.19}$$

Here, the symbol δ_{jk} is the Kronecker delta, defined by

$$\begin{aligned} \delta_{jk} &= 1 \qquad \text{for} \qquad j = k, \\ \delta_{jk} &= 0 \qquad \text{for} \qquad j \neq k. \end{aligned} \tag{1.20}$$

It is easily verified that Eqs. (1.18) and (1.19) hold in the two-dimensional case by substituting in the specific a_{ij} from Eqs. (1.11). The result is the well-known identity $\sin^2\varphi + \cos^2\varphi = 1$ for the nonvanishing case. To verify Eq. (1.18) in general form, we may use the partial derivative forms of Eqs. (1.16a) and (1.16b) to obtain

$$\sum_i \frac{\partial x_j}{\partial x_i'} \frac{\partial x_k}{\partial x_i'} = \sum_i \frac{\partial x_j}{\partial x_i'} \frac{\partial x_i'}{\partial x_k} = \frac{\partial x_j}{\partial x_k}. \tag{1.21}$$

[6] Differentiate x_i' with respect to x_j. See discussion following Eq. (1.21).

The last step follows by the standard rules for partial differentiation, assuming that x_j is a function of x_1', x_2', x_3', and so on. The final result, $\partial x_j/\partial x_k$, is equal to δ_{jk}, since x_j and x_k as coordinate lines ($j \neq k$) are assumed to be perpendicular (two or three dimensions) or orthogonal (for any number of dimensions). Equivalently, we may assume that x_j and x_k ($j \neq k$) are totally independent variables. If $j = k$, the partial derivative is clearly equal to 1.

In redefining a vector in terms of how its components transform under a rotation of the coordinate system, we should emphasize two points:

1. This definition is developed because it is useful and appropriate in describing our physical world. Our vector equations will be independent of any particular coordinate system. (The coordinate system need not even be Cartesian.) The vector equation can always be expressed in some particular coordinate system, and, to obtain numerical results, we must ultimately express the equation in some specific coordinate system.
2. This definition is subject to a generalization that will open up the branch of mathematics known as tensor analysis (Chapter 2).

A qualification is in order. The behavior of the vector components under rotation of the coordinates is used in Section 1.3 to prove that a scalar product is a scalar, in Section 1.4 to prove that a vector product is a vector, and in Section 1.6 to show that the gradient of a scalar ψ, $\nabla\psi$, is a vector. The remainder of this chapter proceeds on the basis of the less restrictive definitions of the vector given in Section 1.1.

Summary: Vectors and Vector Space

It is customary in mathematics to label an ordered triple of real numbers (x_1, x_2, x_3) a **vector** $\mathbf{x}$. The number x_n is called the nth component of vector $\mathbf{x}$. The collection of all such vectors (obeying the properties that follow) form a three-dimensional real **vector space**. We ascribe five properties to our vectors: If $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{y} = (y_1, y_2, y_3)$,

1. Vector equality: $\mathbf{x} = \mathbf{y}$ means $x_i = y_i$, $i = 1, 2, 3$.
2. Vector addition: $\mathbf{x} + \mathbf{y} = \mathbf{z}$ means $x_i + y_i = z_i$, $i = 1, 2, 3$.
3. Scalar multiplication: $a\mathbf{x} \leftrightarrow (ax_1, ax_2, ax_3)$ (with a real).
4. Negative of a vector: $-\mathbf{x} = (-1)\mathbf{x} \leftrightarrow (-x_1, -x_2, -x_3)$.
5. Null vector: There exists a null vector $\mathbf{0} \leftrightarrow (0, 0, 0)$.

Since our vector components are real (or complex) numbers, the following properties also hold:

1. Addition of vectors is commutative: $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
2. Addition of vectors is associative: $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$.
3. Scalar multiplication is distributive:

$$a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}, \qquad \text{also} \qquad (a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}.$$

4. Scalar multiplication is associative: $(ab)\mathbf{x} = a(b\mathbf{x})$.

Further, the null vector **0** is unique, as is the negative of a given vector **x**.

So far as the vectors themselves are concerned this approach merely formalizes the component discussion of Section 1.1. The importance lies in the extensions, which will be considered in later chapters. In Chapter 4, we show that vectors form both an Abelian group under addition and a linear space with the transformations in the linear space described by matrices. Finally, and perhaps most important, for advanced physics the concept of vectors presented here may be generalized to (1) complex quantities,[7] (2) functions, and (3) an infinite number of components. This leads to infinite-dimensional function spaces, the Hilbert spaces, which are important in modern quantum theory. A brief introduction to function expansions and Hilbert space appears in Section 10.4.

Exercises

1.2.1 (a) Show that the magnitude of a vector, $A = (A_x^2 + A_y^2)^{1/2}$, is independent of the orientation of the rotated coordinate system,

$$\left(A_x^2 + A_y^2\right)^{1/2} = \left(A_x'^2 + A_y'^2\right)^{1/2},$$

that is, independent of the rotation angle φ.

This independence of angle is expressed by saying that A is **invariant** under rotations.

(b) At a given point (x, y), **A** defines an angle α relative to the positive x-axis and α' relative to the positive x'-axis. The angle from x to x' is φ. Show that $\mathbf{A} = \mathbf{A}'$ defines the **same** direction in space when expressed in terms of its primed components as in terms of its unprimed components; that is,

$$\alpha' = \alpha - \varphi.$$

1.2.2 Prove the orthogonality condition $\sum_i a_{ji} a_{ki} = \delta_{jk}$. As a special case of this, the direction cosines of Section 1.1 satisfy the relation

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1,$$

a result that also follows from Eq. (1.6).

1.3 Scalar or Dot Product

Having defined vectors, we now proceed to combine them. The laws for combining vectors must be mathematically consistent. From the possibilities that are consistent we select two that are both mathematically and physically interesting. A third possibility is introduced in Chapter 2, in which we form tensors.

The projection of a vector **A** onto a coordinate axis, which gives its Cartesian components in Eq. (1.4), defines a special geometrical case of the scalar product of **A** and the coordinate unit vectors:

$$A_x = A\cos\alpha \equiv \mathbf{A}\cdot\hat{\mathbf{x}}, \quad A_y = A\cos\beta \equiv \mathbf{A}\cdot\hat{\mathbf{y}}, \quad A_z = A\cos\gamma \equiv \mathbf{A}\cdot\hat{\mathbf{z}}. \tag{1.22}$$

[7] The n-dimensional vector space of real n-tuples is often labeled $\mathbb{R}^n$ and the n-dimensional vector space of complex n-tuples is labeled $\mathbb{C}^n$.

This special case of a scalar product in conjunction with general properties the scalar product is sufficient to derive the general case of the scalar product.

Just as the projection is linear in **A**, we want the scalar product of two vectors to be linear in **A** and **B**, that is, obey the distributive and associative laws

$$\mathbf{A}\cdot(\mathbf{B}+\mathbf{C})=\mathbf{A}\cdot\mathbf{B}+\mathbf{A}\cdot\mathbf{C} \tag{1.23a}$$

$$\mathbf{A}\cdot(y\mathbf{B})=(y\mathbf{A})\cdot\mathbf{B}=y\mathbf{A}\cdot\mathbf{B}, \tag{1.23b}$$

where y is a number. Now we can use the decomposition of **B** into its Cartesian components according to Eq. (1.5), $\mathbf{B}=B_x\hat{\mathbf{x}}+B_y\hat{\mathbf{y}}+B_z\hat{\mathbf{z}}$, to construct the general scalar or dot product of the vectors **A** and **B** as

$$\begin{aligned}\mathbf{A}\cdot\mathbf{B}&=\mathbf{A}\cdot(B_x\hat{\mathbf{x}}+B_y\hat{\mathbf{y}}+B_z\hat{\mathbf{z}})\\&=B_x\mathbf{A}\cdot\hat{\mathbf{x}}+B_y\mathbf{A}\cdot\hat{\mathbf{y}}+B_z\mathbf{A}\cdot\hat{\mathbf{z}}\qquad\text{upon applying Eqs. (1.23a) and (1.23b)}\\&=B_xA_x+B_yA_y+B_zA_z\qquad\text{upon substituting Eq. (1.22).}\end{aligned}$$

Hence

$$\boxed{\mathbf{A}\cdot\mathbf{B}\equiv\sum_i B_iA_i=\sum_i A_iB_i=\mathbf{B}\cdot\mathbf{A}.} \tag{1.24}$$

If $\mathbf{A}=\mathbf{B}$ in Eq. (1.24), we recover the magnitude $A=(\sum A_i^2)^{1/2}$ of **A** in Eq. (1.6) from Eq. (1.24).

It is obvious from Eq. (1.24) that the scalar product treats **A** and **B** alike, or is symmetric in **A** and **B**, and is commutative. Thus, alternatively and equivalently, we can first generalize Eqs. (1.22) to the projection A_B of **A** onto the direction of a vector $\mathbf{B}\neq 0$ as $A_B=A\cos\theta\equiv\mathbf{A}\cdot\hat{\mathbf{B}}$, where $\hat{\mathbf{B}}=\mathbf{B}/B$ is the unit vector in the direction of **B** and θ is the angle between **A** and **B**, as shown in Fig. 1.7. Similarly, we project **B** onto **A** as $B_A=B\cos\theta\equiv\mathbf{B}\cdot\hat{\mathbf{A}}$. Second, we make these projections symmetric in **A** and **B**, which leads to the definition

$$\boxed{\mathbf{A}\cdot\mathbf{B}\equiv A_BB=AB_A=AB\cos\theta.} \tag{1.25}$$

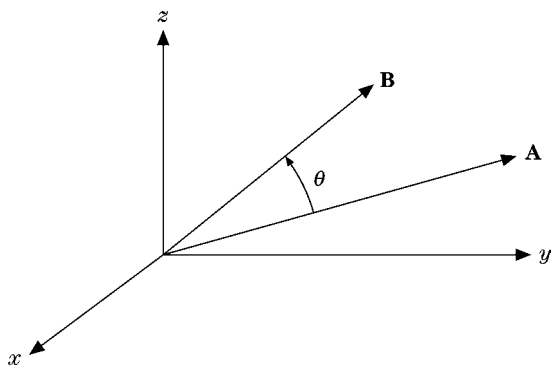

FIGURE 1.7 Scalar product $\mathbf{A}\cdot\mathbf{B}=AB\cos\theta$.

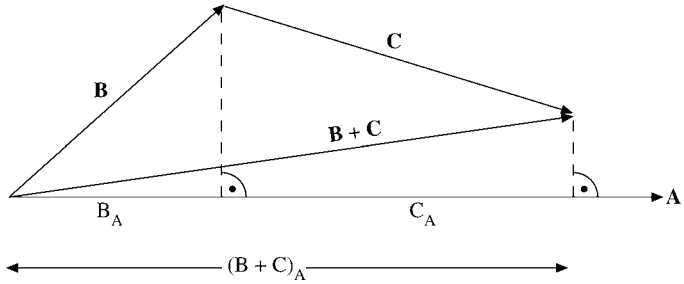

FIGURE 1.8 The distributive law
$\mathbf{A} \cdot (\mathbf{B} + \mathbf{C}) = AB_A + AC_A = A(\mathbf{B} + \mathbf{C})_A$, Eq. (1.23a).

The distributive law in Eq. (1.23a) is illustrated in Fig. 1.8, which shows that the sum of the projections of **B** and **C** onto **A**, $B_A + C_A$ is equal to the projection of $\mathbf{B} + \mathbf{C}$ onto **A**, $(\mathbf{B} + \mathbf{C})_A$.

It follows from Eqs. (1.22), (1.24), and (1.25) that the coordinate unit vectors satisfy the relations

$$\hat{\mathbf{x}} \cdot \hat{\mathbf{x}} = \hat{\mathbf{y}} \cdot \hat{\mathbf{y}} = \hat{\mathbf{z}} \cdot \hat{\mathbf{z}} = 1, \tag{1.26a}$$

whereas

$$\hat{\mathbf{x}} \cdot \hat{\mathbf{y}} = \hat{\mathbf{x}} \cdot \hat{\mathbf{z}} = \hat{\mathbf{y}} \cdot \hat{\mathbf{z}} = 0. \tag{1.26b}$$

If the component definition, Eq. (1.24), is labeled an algebraic definition, then Eq. (1.25) is a geometric definition. One of the most common applications of the scalar product in physics is in the calculation of **work** = **force**·**displacement**$\cdot \cos\theta$, which is interpreted as displacement times the projection of the force along the displacement direction, i.e., the scalar product of force and displacement, $W = \mathbf{F} \cdot \mathbf{S}$.

If $\mathbf{A} \cdot \mathbf{B} = 0$ and we know that $\mathbf{A} \neq 0$ and $\mathbf{B} \neq 0$, then, from Eq. (1.25), $\cos\theta = 0$, or $\theta = 90°, 270°$, and so on. The vectors **A** and **B** must be perpendicular. Alternately, we may say **A** and **B** are orthogonal. The unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ are mutually orthogonal. To develop this notion of orthogonality one more step, suppose that **n** is a unit vector and **r** is a nonzero vector in the xy-plane; that is, $\mathbf{r} = \hat{\mathbf{x}}x + \hat{\mathbf{y}}y$ (Fig. 1.9). If

$$\mathbf{n} \cdot \mathbf{r} = 0$$

for **all** choices of **r**, then **n** must be perpendicular (orthogonal) to the xy-plane.

Often it is convenient to replace $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ by subscripted unit vectors $\mathbf{e}_m, m = 1, 2, 3$, with $\hat{\mathbf{x}} = \mathbf{e}_1$, and so on. Then Eqs. (1.26a) and (1.26b) become

$$\mathbf{e}_m \cdot \mathbf{e}_n = \delta_{mn}. \tag{1.26c}$$

For $m \neq n$ the unit vectors $\mathbf{e}_m$ and $\mathbf{e}_n$ are orthogonal. For $m = n$ each vector is normalized to unity, that is, has unit magnitude. The set $\mathbf{e}_m$ is said to be **orthonormal**. A major advantage of Eq. (1.26c) over Eqs. (1.26a) and (1.26b) is that Eq. (1.26c) may readily be generalized to N-dimensional space: $m, n = 1, 2, \ldots, N$. Finally, we are picking sets of unit vectors $\mathbf{e}_m$ that are orthonormal for convenience – a very great convenience.

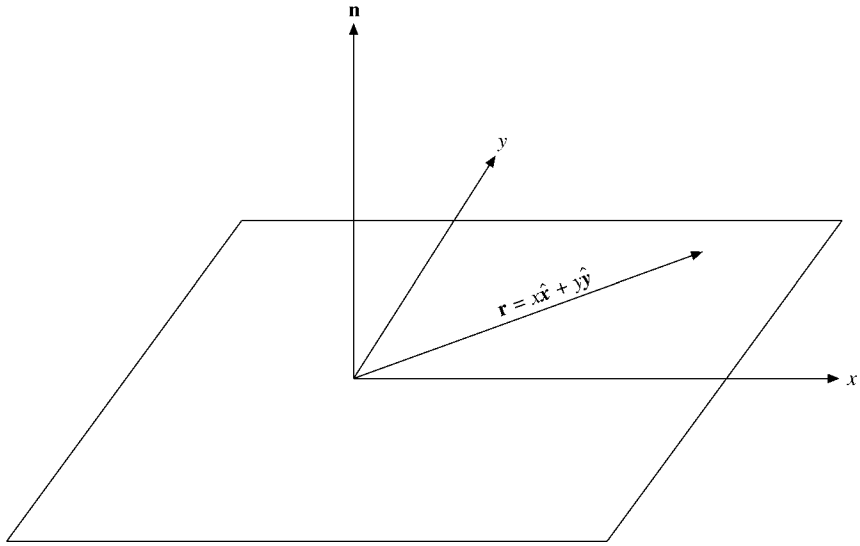

FIGURE 1.9 A normal vector.

Invariance of the Scalar Product Under Rotations

We have not yet shown that the word *scalar* is justified or that the scalar product is indeed a scalar quantity. To do this, we investigate the behavior of $\mathbf{A} \cdot \mathbf{B}$ under a rotation of the coordinate system. By use of Eq. (1.15),

$$A'_x B'_x + A'_y B'_y + A'_z B'_z = \sum_i a_{xi} A_i \sum_j a_{xj} B_j + \sum_i a_{yi} A_i \sum_j a_{yj} B_j + \sum_i a_{zi} A_i \sum_j a_{zj} B_j. \tag{1.27}$$

Using the indices k and l to sum over x, y, and z, we obtain

$$\sum_k A'_k B'_k = \sum_l \sum_i \sum_j a_{li} A_i a_{lj} B_j, \tag{1.28}$$

and, by rearranging the terms on the right-hand side, we have

$$\sum_k A'_k B'_k = \sum_l \sum_i \sum_j (a_{li} a_{lj}) A_i B_j = \sum_i \sum_j \delta_{ij} A_i B_j = \sum_i A_i B_i. \tag{1.29}$$

The last two steps follow by using Eq. (1.18), the orthogonality condition of the direction cosines, and Eqs. (1.20), which define the Kronecker delta. The effect of the Kronecker delta is to cancel all terms in a summation over either index except the term for which the indices are equal. In Eq. (1.29) its effect is to set $j = i$ and to eliminate the summation over j. Of course, we could equally well set $i = j$ and eliminate the summation over i.

Equation (1.29) gives us

$$\boxed{\sum_k A'_k B'_k = \sum_i A_i B_i,} \tag{1.30}$$

which is just our definition of a scalar quantity, one that remains invariant under the rotation of the coordinate system.

In a similar approach that exploits this concept of invariance, we take $\mathbf{C} = \mathbf{A} + \mathbf{B}$ and dot it into itself:

$$\begin{aligned}\mathbf{C}\cdot\mathbf{C} &= (\mathbf{A}+\mathbf{B})\cdot(\mathbf{A}+\mathbf{B})\\ &= \mathbf{A}\cdot\mathbf{A}+\mathbf{B}\cdot\mathbf{B}+2\mathbf{A}\cdot\mathbf{B}.\end{aligned} \tag{1.31}$$

Since

$$\mathbf{C}\cdot\mathbf{C} = C^2, \tag{1.32}$$

the square of the magnitude of vector $\mathbf{C}$ and thus an invariant quantity, we see that

$$\mathbf{A}\cdot\mathbf{B} = \frac{1}{2}\left(C^2 - A^2 - B^2\right), \qquad \text{invariant.} \tag{1.33}$$

Since the right-hand side of Eq. (1.33) is invariant — that is, a scalar quantity — the left-hand side, $\mathbf{A}\cdot\mathbf{B}$, must also be invariant under rotation of the coordinate system. Hence $\mathbf{A}\cdot\mathbf{B}$ is a scalar.

Equation (1.31) is really another form of the law of cosines, which is

$$C^2 = A^2 + B^2 + 2AB\cos\theta. \tag{1.34}$$

Comparing Eqs. (1.31) and (1.34), we have another verification of Eq. (1.25), or, if preferred, a vector derivation of the law of cosines (Fig. 1.10).

The dot product, given by Eq. (1.24), may be generalized in two ways. The space need not be restricted to three dimensions. In n-dimensional space, Eq. (1.24) applies with the sum running from 1 to n. Moreover, n may be infinity, with the sum then a convergent infinite series (Section 5.2). The other generalization extends the concept of vector to embrace functions. The function analog of a dot, or inner, product appears in Section 10.4.

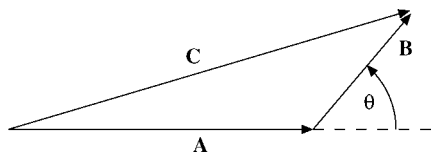

FIGURE 1.10 The law of cosines.

Exercises

1.3.1 Two unit magnitude vectors $\mathbf{e}_i$ and $\mathbf{e}_j$ are required to be either parallel or perpendicular to each other. Show that $\mathbf{e}_i \cdot \mathbf{e}_j$ provides an interpretation of Eq. (1.18), the direction cosine orthogonality relation.

1.3.2 Given that (1) the dot product of a unit vector with itself is unity and (2) this relation is valid in all (rotated) coordinate systems, show that $\hat{\mathbf{x}}' \cdot \hat{\mathbf{x}}' = 1$ (with the primed system rotated 45° about the z-axis relative to the unprimed) implies that $\hat{\mathbf{x}} \cdot \hat{\mathbf{y}} = 0$.

1.3.3 The vector $\mathbf{r}$, starting at the origin, terminates at and specifies the point in space (x, y, z). Find the surface swept out by the tip of $\mathbf{r}$ if

(a) $(\mathbf{r} - \mathbf{a}) \cdot \mathbf{a} = 0$. Characterize $\mathbf{a}$ geometrically.
(b) $(\mathbf{r} - \mathbf{a}) \cdot \mathbf{r} = 0$. Describe the geometric role of $\mathbf{a}$.
The vector $\mathbf{a}$ is constant (in magnitude and direction).

1.3.4 The interaction energy between two dipoles of moments $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ may be written in the vector form

$$V = -\frac{\boldsymbol{\mu}_1 \cdot \boldsymbol{\mu}_2}{r^3} + \frac{3(\boldsymbol{\mu}_1 \cdot \mathbf{r})(\boldsymbol{\mu}_2 \cdot \mathbf{r})}{r^5}$$

and in the scalar form

$$V = \frac{\mu_1 \mu_2}{r^3}(2\cos\theta_1 \cos\theta_2 - \sin\theta_1 \sin\theta_2 \cos\varphi).$$

Here θ_1 and θ_2 are the angles of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ relative to $\mathbf{r}$, while φ is the azimuth of $\boldsymbol{\mu}_2$ relative to the $\boldsymbol{\mu}_1$–$\mathbf{r}$ plane (Fig. 1.11). Show that these two forms are equivalent.
Hint: Equation (12.178) will be helpful.

1.3.5 A pipe comes diagonally down the south wall of a building, making an angle of 45° with the horizontal. Coming into a corner, the pipe turns and continues diagonally down a west-facing wall, still making an angle of 45° with the horizontal. What is the angle between the south-wall and west-wall sections of the pipe?

ANS. 120°.

1.3.6 Find the shortest distance of an observer at the point $(2, 1, 3)$ from a rocket in free flight with velocity $(1, 2, 3)$ m/s. The rocket was launched at time $t = 0$ from $(1, 1, 1)$. Lengths are in kilometers.

1.3.7 Prove the law of cosines from the triangle with corners at the point of **C** and **A** in Fig. 1.10 and the projection of vector **B** onto vector **A**.

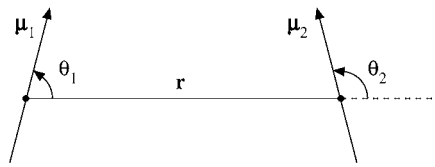

FIGURE 1.11 Two dipole moments.

1.4 VECTOR OR CROSS PRODUCT

A second form of vector multiplication employs the sine of the included angle instead of the cosine. For instance, the angular momentum of a body shown at the point of the distance vector in Fig. 1.12 is defined as

$$\begin{aligned}\text{angular momentum} &= \text{radius arm} \times \text{linear momentum}\\ &= \text{distance} \times \text{linear momentum} \times \sin\theta.\end{aligned}$$

For convenience in treating problems relating to quantities such as angular momentum, torque, and angular velocity, we define the vector product, or cross product, as

$$\boxed{\mathbf{C} = \mathbf{A} \times \mathbf{B}, \qquad \text{with } C = AB\sin\theta.} \tag{1.35}$$

Unlike the preceding case of the scalar product, **C** is now a vector, and we assign it a direction perpendicular to the plane of **A** and **B** such that **A**, **B**, and **C** form a right-handed system. With this choice of direction we have

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A}, \qquad \text{anticommutation.} \tag{1.36a}$$

From this definition of cross product we have

$$\hat{\mathbf{x}} \times \hat{\mathbf{x}} = \hat{\mathbf{y}} \times \hat{\mathbf{y}} = \hat{\mathbf{z}} \times \hat{\mathbf{z}} = \mathbf{0}, \tag{1.36b}$$

whereas

$$\begin{aligned}&\hat{\mathbf{x}} \times \hat{\mathbf{y}} = \hat{\mathbf{z}}, \qquad \hat{\mathbf{y}} \times \hat{\mathbf{z}} = \hat{\mathbf{x}}, \qquad \hat{\mathbf{z}} \times \hat{\mathbf{x}} = \hat{\mathbf{y}},\\ &\hat{\mathbf{y}} \times \hat{\mathbf{x}} = -\hat{\mathbf{z}}, \qquad \hat{\mathbf{z}} \times \hat{\mathbf{y}} = -\hat{\mathbf{x}}, \qquad \hat{\mathbf{x}} \times \hat{\mathbf{z}} = -\hat{\mathbf{y}}.\end{aligned} \tag{1.36c}$$

Among the examples of the cross product in mathematical physics are the relation between linear momentum **p** and angular momentum **L**, with **L** defined as

$$\mathbf{L} = \mathbf{r} \times \mathbf{p},$$

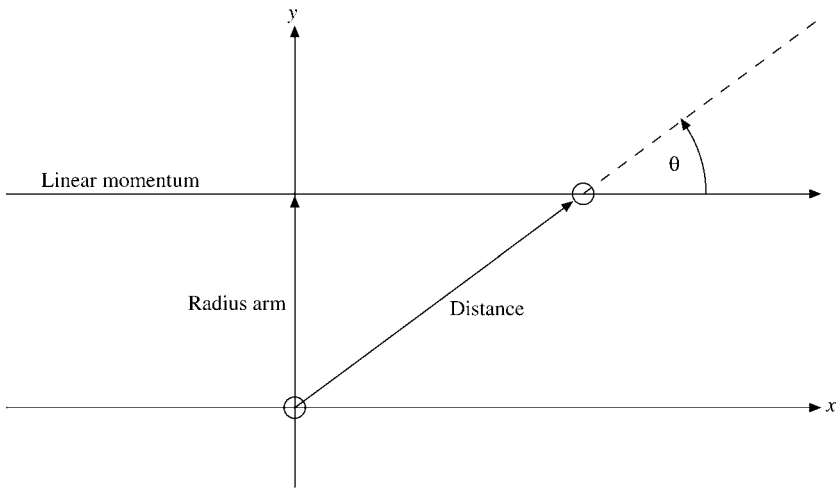

FIGURE 1.12 Angular momentum.

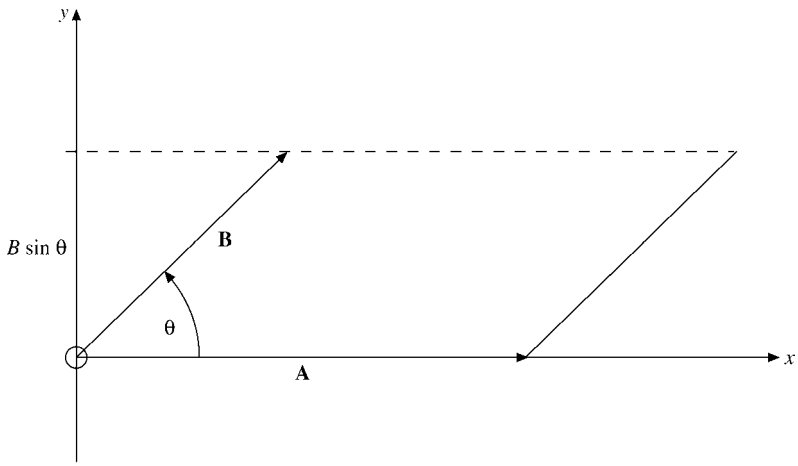

FIGURE 1.13 Parallelogram representation of the vector product.

and the relation between linear velocity **v** and angular velocity $\boldsymbol{\omega}$,

$$\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}.$$

Vectors **v** and **p** describe properties of the particle or physical system. However, the position vector **r** is determined by the choice of the origin of the coordinates. This means that $\boldsymbol{\omega}$ and **L** depend on the choice of the origin.

The familiar magnetic induction **B** is usually defined by the vector product force equation[8]

$$\mathbf{F}_M = q\mathbf{v} \times \mathbf{B} \text{ (mks units)}.$$

Here **v** is the velocity of the electric charge q and $\mathbf{F}_M$ is the resulting force on the moving charge.

The cross product has an important geometrical interpretation, which we shall use in subsequent sections. In the parallelogram defined by **A** and **B** (Fig. 1.13), $B\sin\theta$ is the height if A is taken as the length of the base. Then $|\mathbf{A} \times \mathbf{B}| = AB\sin\theta$ is the **area** of the parallelogram. As a vector, $\mathbf{A} \times \mathbf{B}$ is the area of the parallelogram defined by **A** and **B**, with the area vector normal to the plane of the parallelogram. This suggests that area (with its orientation in space) may be treated as a vector quantity.

An alternate definition of the vector product can be derived from the special case of the coordinate unit vectors in Eqs. (1.36c) in conjunction with the linearity of the cross product in both vector arguments, in analogy with Eqs. (1.23) for the dot product,

$$\mathbf{A} \times (\mathbf{B} + \mathbf{C}) = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C}, \tag{1.37a}$$

$$(\mathbf{A} + \mathbf{B}) \times \mathbf{C} = \mathbf{A} \times \mathbf{C} + \mathbf{B} \times \mathbf{C}, \tag{1.37b}$$

$$\mathbf{A} \times (y\mathbf{B}) = y\mathbf{A} \times \mathbf{B} = (y\mathbf{A}) \times \mathbf{B}, \tag{1.37c}$$

[8]The electric field **E** is assumed here to be zero.

where y is a number again. Using the decomposition of $\mathbf{A}$ and $\mathbf{B}$ into their Cartesian components according to Eq. (1.5), we find

$$\begin{aligned}\mathbf{A}\times\mathbf{B}\equiv\mathbf{C}&=(C_x,C_y,C_z)=(A_x\hat{\mathbf{x}}+A_y\hat{\mathbf{y}}+A_z\hat{\mathbf{z}})\times(B_x\hat{\mathbf{x}}+B_y\hat{\mathbf{y}}+B_z\hat{\mathbf{z}})\\&=(A_xB_y-A_yB_x)\hat{\mathbf{x}}\times\hat{\mathbf{y}}+(A_xB_z-A_zB_x)\hat{\mathbf{x}}\times\hat{\mathbf{z}}\\&\quad+(A_yB_z-A_zB_y)\hat{\mathbf{y}}\times\hat{\mathbf{z}}\end{aligned}$$

upon applying Eqs. (1.37a) and (1.37b) and substituting Eqs. (1.36a), (1.36b), and (1.36c) so that the Cartesian components of $\mathbf{A}\times\mathbf{B}$ become

$$C_x=A_yB_z-A_zB_y,\qquad C_y=A_zB_x-A_xB_z,\qquad C_z=A_xB_y-A_yB_x,\tag{1.38}$$

or

$$C_i=A_jB_k-A_kB_j,\qquad i,j,k \text{ all different},\tag{1.39}$$

and with cyclic permutation of the indices i, j, and k corresponding to x, y, and z, respectively. The vector product $\mathbf{C}$ may be mnemonically represented by a determinant,[9]

$$\mathbf{C}=\begin{vmatrix}\hat{\mathbf{x}}&\hat{\mathbf{y}}&\hat{\mathbf{z}}\\A_x&A_y&A_z\\B_x&B_y&B_z\end{vmatrix}$$

[9] See Section 3.1 for a brief summary of determinants.

Hence

$$C = AB\sin\theta. \tag{1.44}$$

The first step in Eq. (1.43) may be verified by expanding out in component form, using Eqs. (1.38) for $\mathbf{A}\times\mathbf{B}$ and Eq. (1.24) for the dot product. From Eqs. (1.41), (1.42), and (1.44) we see the equivalence of Eqs. (1.35) and (1.38), the two definitions of vector product.

There still remains the problem of verifying that $\mathbf{C} = \mathbf{A}\times\mathbf{B}$ is indeed a vector, that is, that it obeys Eq. (1.15), the vector transformation law. Starting in a rotated (primed system),

$$\begin{aligned} C_i' &= A_j' B_k' - A_k' B_j', \qquad i, j, \text{ and } k \text{ in cyclic order,} \\ &= \sum_l a_{jl} A_l \sum_m a_{km} B_m - \sum_l a_{kl} A_l \sum_m a_{jm} B_m \\ &= \sum_{l,m} (a_{jl}a_{km} - a_{kl}a_{jm}) A_l B_m. \end{aligned} \tag{1.45}$$

The combination of direction cosines in parentheses vanishes for $m = l$. We therefore have j and k taking on fixed values, dependent on the choice of i, and six combinations of l and m. If $i = 3$, then $j = 1, k = 2$ (cyclic order), and we have the following direction cosine combinations:[10]

$$\begin{aligned} a_{11}a_{22} - a_{21}a_{12} &= a_{33}, \\ a_{13}a_{21} - a_{23}a_{11} &= a_{32}, \\ a_{12}a_{23} - a_{22}a_{13} &= a_{31} \end{aligned} \tag{1.46}$$

and their negatives. Equations (1.46) are identities satisfied by the direction cosines. They may be verified with the use of determinants and matrices (see Exercise 3.3.3). Substituting back into Eq. (1.45),

$$\begin{aligned} C_3' &= a_{33}A_1B_2 + a_{32}A_3B_1 + a_{31}A_2B_3 - a_{33}A_2B_1 - a_{32}A_1B_3 - a_{31}A_3B_2 \\ &= a_{31}C_1 + a_{32}C_2 + a_{33}C_3 \\ &= \sum_n a_{3n}C_n. \end{aligned} \tag{1.47}$$

By permuting indices to pick up C_1' and C_2', we see that Eq. (1.15) is satisfied and $\mathbf{C}$ is indeed a vector. It should be mentioned here that this **vector nature** of the **cross product** is an accident associated with the **three-dimensional** nature of ordinary space.[11] It will be seen in Chapter 2 that the cross product may also be treated as a second-rank antisymmetric tensor.

[10] Equations (1.46) hold for rotations because they preserve volumes. For a more general orthogonal transformation, the r.h.s. of Eqs. (1.46) is multiplied by the determinant of the transformation matrix (see Chapter 3 for matrices and determinants).

[11] Specifically Eqs. (1.46) hold only for three-dimensional space. See D. Hestenes and G. Sobczyk, *Clifford Algebra to Geometric Calculus* (Dordrecht: Reidel, 1984) for a far-reaching generalization of the cross product.

If we define a vector as an ordered triplet of numbers (or functions), as in the latter part of Section 1.2, then there is no problem identifying the cross product as a vector. The cross-product operation maps the two triples **A** and **B** into a third triple, **C**, which by definition is a vector.

We now have two ways of multiplying vectors; a third form appears in Chapter 2. But what about division by a vector? It turns out that the ratio $\mathbf{B}/\mathbf{A}$ is not uniquely specified (Exercise 3.2.21) unless **A** and **B** are also required to be parallel. Hence division of one vector by another is not defined.

Exercises

1.4.1 Show that the medians of a triangle intersect in the center, which is $2/3$ of the median's length from each corner. Construct a numerical example and plot it.

1.4.2 Prove the law of cosines starting from $\mathbf{A}^2 = (\mathbf{B} - \mathbf{C})^2$.

1.4.3 Starting with $\mathbf{C} = \mathbf{A} + \mathbf{B}$, show that $\mathbf{C} \times \mathbf{C} = 0$ leads to

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A}.$$

1.4.4 Show that

(a) $(\mathbf{A} - \mathbf{B}) \cdot (\mathbf{A} + \mathbf{B}) = A^2 - B^2$,
(b) $(\mathbf{A} - \mathbf{B}) \times (\mathbf{A} + \mathbf{B}) = 2\mathbf{A} \times \mathbf{B}$.

The distributive laws needed here,

$$\mathbf{A} \cdot (\mathbf{B} + \mathbf{C}) = \mathbf{A} \cdot \mathbf{B} + \mathbf{A} \cdot \mathbf{C},$$

and

$$\mathbf{A} \times (\mathbf{B} + \mathbf{C}) = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C},$$

may easily be verified (if desired) by expansion in Cartesian components.

1.4.5 Given the three vectors,

$$\mathbf{P} = 3\hat{\mathbf{x}} + 2\hat{\mathbf{y}} - \hat{\mathbf{z}},$$
$$\mathbf{Q} = -6\hat{\mathbf{x}} - 4\hat{\mathbf{y}} + 2\hat{\mathbf{z}},$$
$$\mathbf{R} = \hat{\mathbf{x}} - 2\hat{\mathbf{y}} - \hat{\mathbf{z}},$$

find two that are perpendicular and two that are parallel or antiparallel.

1.4.6 If $\mathbf{P} = \hat{\mathbf{x}}P_x + \hat{\mathbf{y}}P_y$ and $\mathbf{Q} = \hat{\mathbf{x}}Q_x + \hat{\mathbf{y}}Q_y$ are any two nonparallel (also nonantiparallel) vectors in the xy-plane, show that $\mathbf{P} \times \mathbf{Q}$ is in the z-direction.

1.4.7 Prove that $(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{A} \times \mathbf{B}) = (AB)^2 - (\mathbf{A} \cdot \mathbf{B})^2$.

1.4.8 Using the vectors

$$\mathbf{P} = \hat{\mathbf{x}}\cos\theta + \hat{\mathbf{y}}\sin\theta,$$
$$\mathbf{Q} = \hat{\mathbf{x}}\cos\varphi - \hat{\mathbf{y}}\sin\varphi,$$
$$\mathbf{R} = \hat{\mathbf{x}}\cos\varphi + \hat{\mathbf{y}}\sin\varphi,$$

prove the familiar trigonometric identities

$$\sin(\theta+\varphi) = \sin\theta\cos\varphi + \cos\theta\sin\varphi,$$
$$\cos(\theta+\varphi) = \cos\theta\cos\varphi - \sin\theta\sin\varphi.$$

1.4.9 (a) Find a vector $\mathbf{A}$ that is perpendicular to

$$\mathbf{U} = 2\hat{\mathbf{x}} + \hat{\mathbf{y}} - \hat{\mathbf{z}},$$
$$\mathbf{V} = \hat{\mathbf{x}} - \hat{\mathbf{y}} + \hat{\mathbf{z}}.$$

(b) What is $\mathbf{A}$ if, in addition to this requirement, we demand that it have unit magnitude?

1.4.10 If four vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$ all lie in the same plane, show that

$$(\mathbf{a}\times\mathbf{b})\times(\mathbf{c}\times\mathbf{d}) = 0.$$

Hint. Consider the directions of the cross-product vectors.

1.4.11 The coordinates of the three vertices of a triangle are $(2, 1, 5)$, $(5, 2, 8)$, and $(4, 8, 2)$. Compute its area by vector methods, its center and medians. Lengths are in centimeters. *Hint*. See Exercise 1.4.1.

1.4.12 The vertices of parallelogram *ABCD* are $(1, 0, 0)$, $(2, -1, 0)$, $(0, -1, 1)$, and $(-1, 0, 1)$ in order. Calculate the vector areas of triangle *ABD* and of triangle *BCD*. Are the two vector areas equal?

ANS. $\text{Area}_{ABD} = -\frac{1}{2}(\hat{\mathbf{x}} + \hat{\mathbf{y}} + 2\hat{\mathbf{z}})$.

1.4.13 The origin and the three vectors $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ (all of which start at the origin) define a tetrahedron. Taking the outward direction as positive, calculate the total vector area of the four tetrahedral surfaces.
Note. In Section 1.11 this result is generalized to any closed surface.

1.4.14 Find the sides and angles of the spherical triangle *ABC* defined by the three vectors

$$\mathbf{A} = (1, 0, 0),$$
$$\mathbf{B} = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right),$$
$$\mathbf{C} = \left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right).$$

Each vector starts from the origin (Fig. 1.14).

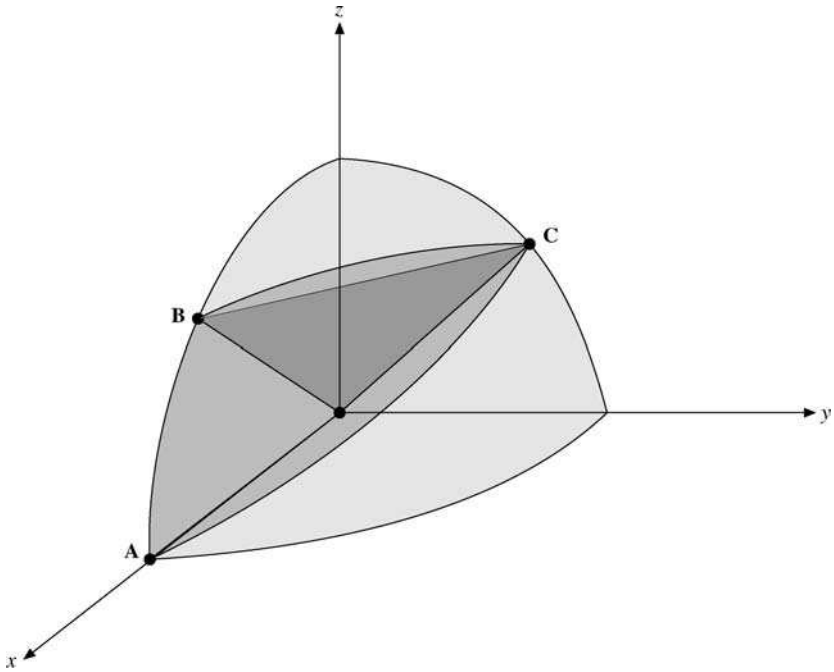

FIGURE 1.14 Spherical triangle.

1.4.15 Derive the law of sines (Fig. 1.15):

$$\frac{\sin\alpha}{|\mathbf{A}|} = \frac{\sin\beta}{|\mathbf{B}|} = \frac{\sin\gamma}{|\mathbf{C}|}.$$

1.4.16 The magnetic induction **B** is **defined** by the Lorentz force equation,

$$\mathbf{F} = q(\mathbf{v} \times \mathbf{B}).$$

Carrying out three experiments, we find that if

$$\mathbf{v} = \hat{\mathbf{x}}, \qquad \frac{\mathbf{F}}{q} = 2\hat{\mathbf{z}} - 4\hat{\mathbf{y}},$$

$$\mathbf{v} = \hat{\mathbf{y}}, \qquad \frac{\mathbf{F}}{q} = 4\hat{\mathbf{x}} - \hat{\mathbf{z}},$$

$$\mathbf{v} = \hat{\mathbf{z}}, \qquad \frac{\mathbf{F}}{q} = \hat{\mathbf{y}} - 2\hat{\mathbf{x}}.$$

From the results of these three separate experiments calculate the magnetic induction **B**.

1.4.17 Define a cross product of two vectors in two-dimensional space and give a geometrical interpretation of your construction.

1.4.18 Find the shortest distance between the paths of two rockets in free flight. Take the first rocket path to be $\mathbf{r} = \mathbf{r}_1 + t_1\mathbf{v}_1$ with launch at $\mathbf{r}_1 = (1, 1, 1)$ and velocity $\mathbf{v}_1 = (1, 2, 3)$

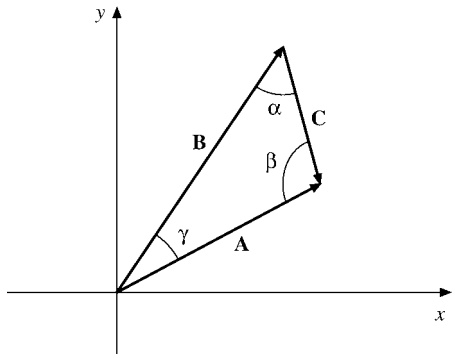

FIGURE 1.15 Law of sines.

and the second rocket path as $\mathbf{r} = \mathbf{r}_2 + t_2\mathbf{v}_2$ with $\mathbf{r}_2 = (5, 2, 1)$ and $\mathbf{v}_2 = (-1, -1, 1)$. Lengths are in kilometers, velocities in kilometers per hour.

1.5 TRIPLE SCALAR PRODUCT, TRIPLE VECTOR PRODUCT

Triple Scalar Product

Sections 1.3 and 1.4 cover the two types of multiplication of interest here. However, there are combinations of three vectors, $\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C})$ and $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$, that occur with sufficient frequency to deserve further attention. The combination

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C})$$

is known as the **triple scalar product**. $\mathbf{B} \times \mathbf{C}$ yields a vector that, dotted into $\mathbf{A}$, gives a scalar. We note that $(\mathbf{A} \cdot \mathbf{B}) \times \mathbf{C}$ represents a scalar crossed into a vector, an operation that is not defined. Hence, if we agree to exclude this undefined interpretation, the parentheses may be omitted and the triple scalar product written $\mathbf{A} \cdot \mathbf{B} \times \mathbf{C}$.

Using Eqs. (1.38) for the cross product and Eq. (1.24) for the dot product, we obtain

$$\begin{aligned}\mathbf{A} \cdot \mathbf{B} \times \mathbf{C} &= A_x(B_yC_z - B_zC_y) + A_y(B_zC_x - B_xC_z) + A_z(B_xC_y - B_yC_x) \\ &= \mathbf{B} \cdot \mathbf{C} \times \mathbf{A} = \mathbf{C} \cdot \mathbf{A} \times \mathbf{B} \\ &= -\mathbf{A} \cdot \mathbf{C} \times \mathbf{B} = -\mathbf{C} \cdot \mathbf{B} \times \mathbf{A} = -\mathbf{B} \cdot \mathbf{A} \times \mathbf{C}, \text{ and so on.}\end{aligned} \tag{1.48}$$

There is a high degree of symmetry in the component expansion. Every term contains the factors A_i, B_j, and C_k. If i, j, and k are in cyclic order (x, y, z), the sign is positive. If the order is anticyclic, the sign is negative. Further, the dot and the cross may be interchanged,

$$\mathbf{A} \cdot \mathbf{B} \times \mathbf{C} = \mathbf{A} \times \mathbf{B} \cdot \mathbf{C}. \tag{1.49}$$

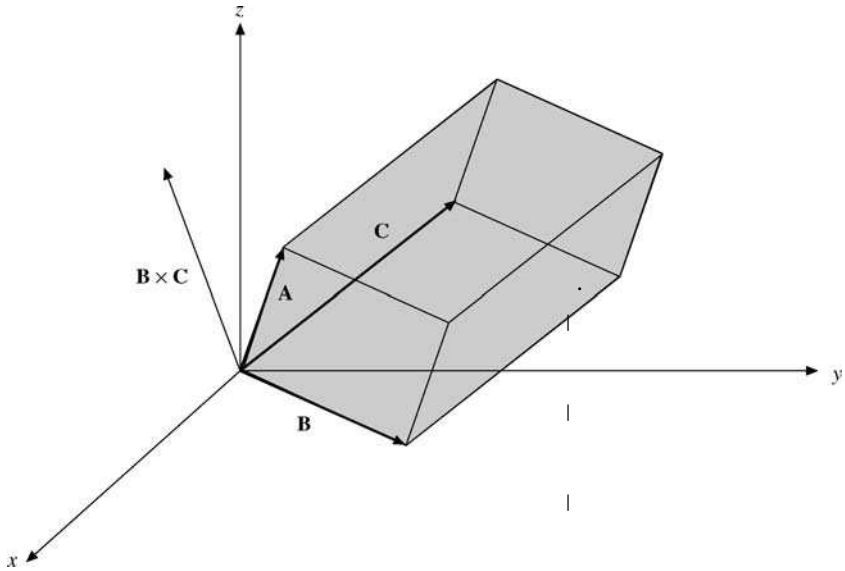

FIGURE 1.16 Parallelepiped representation of triple scalar product.

A convenient representation of the component expansion of Eq. (1.48) is provided by the determinant

$$\mathbf{A}\cdot\mathbf{B}\times\mathbf{C}=\begin{vmatrix} A_x & A_y & A_z \\ B_x & B_y & B_z \\ C_x & C_y & C_z \end{vmatrix} \tag{1.50}$$

The rules for interchanging rows and columns of a determinant[12] provide an immediate verification of the permutations listed in Eq. (1.48), whereas the symmetry of **A**, **B**, and **C** in the determinant form suggests the relation given in Eq. (1.49). The triple products encountered in Section 1.4, which showed that $\mathbf{A}\times\mathbf{B}$ was perpendicular to both **A** and **B**, were special cases of the general result (Eq. (1.48)).

The triple scalar product has a direct geometrical interpretation. The three vectors **A**, **B**, and **C** may be interpreted as defining a parallelepiped (Fig. 1.16):

$$\begin{aligned} |\mathbf{B}\times\mathbf{C}| &= BC\sin\theta \\ &= \text{area of parallelogram base.} \end{aligned} \tag{1.51}$$

The direction, of course, is normal to the base. Dotting **A** into this means multiplying the base area by the projection of **A** onto the normal, or base times height. Therefore

$$\mathbf{A}\cdot\mathbf{B}\times\mathbf{C}=\text{volume of parallelepiped defined by }\mathbf{A},\mathbf{B},\text{ and }\mathbf{C}.$$

The triple scalar product finds an interesting and important application in the construction of a reciprocal crystal lattice. Let **a**, **b**, and **c** (not necessarily mutually perpendicular)

[12] See Section 3.1 for a summary of the properties of determinants.

represent the vectors that define a crystal lattice. The displacement from one lattice point to another may then be written

$$\mathbf{r} = n_a\mathbf{a} + n_b\mathbf{b} + n_c\mathbf{c}, \tag{1.52}$$

with n_a, n_b, and n_c taking on integral values. With these vectors we may form

$$\mathbf{a}' = \frac{\mathbf{b}\times\mathbf{c}}{\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}}, \qquad \mathbf{b}' = \frac{\mathbf{c}\times\mathbf{a}}{\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}}, \qquad \mathbf{c}' = \frac{\mathbf{a}\times\mathbf{b}}{\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}}. \tag{1.53a}$$

We see that $\mathbf{a}'$ is perpendicular to the plane containing $\mathbf{b}$ and $\mathbf{c}$, and we can readily show that

$$\mathbf{a}'\cdot\mathbf{a} = \mathbf{b}'\cdot\mathbf{b} = \mathbf{c}'\cdot\mathbf{c} = 1, \tag{1.53b}$$

whereas

$$\mathbf{a}'\cdot\mathbf{b} = \mathbf{a}'\cdot\mathbf{c} = \mathbf{b}'\cdot\mathbf{a} = \mathbf{b}'\cdot\mathbf{c} = \mathbf{c}'\cdot\mathbf{a} = \mathbf{c}'\cdot\mathbf{b} = 0. \tag{1.53c}$$

It is from Eqs. (1.53b) and (1.53c) that the name **reciprocal lattice** is associated with the points $\mathbf{r}' = n_a'\mathbf{a}' + n_b'\mathbf{b}' + n_c'\mathbf{c}'$. The mathematical space in which this reciprocal lattice exists is sometimes called a *momentum space*, or *k* space (because of its use in describing the Fourier analysis of wavefunctions). Further details will be found in Chapters 14 and 15. This reciprocal lattice is useful in problems involving the scattering of waves from the various planes in a crystal. Further details may be found in R. B. Leighton's *Principles of Modern Physics*, pp. 440–448 [New York: McGraw-Hill (1959)].

Triple Vector Product

The second triple product of interest is $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})$, which is a vector. Here the parentheses must be retained, as may be seen from a special case $(\hat{\mathbf{x}}\times\hat{\mathbf{x}})\times\hat{\mathbf{y}} = 0$, while $\hat{\mathbf{x}}\times(\hat{\mathbf{x}}\times\hat{\mathbf{y}}) = \hat{\mathbf{x}}\times\hat{\mathbf{z}} = -\hat{\mathbf{y}}$.

Example 1.5.1 A Triple Vector Product

For the vectors

$$\mathbf{A} = \hat{\mathbf{x}} + 2\hat{\mathbf{y}} - \hat{\mathbf{z}} = (1, 2, -1), \quad \mathbf{B} = \hat{\mathbf{y}} + \hat{\mathbf{z}} = (0, 1, 1), \quad \mathbf{C} = \hat{\mathbf{x}} - \hat{\mathbf{y}} = (0, 1, 1),$$

$$\mathbf{B}\times\mathbf{C} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ 0 & 1 & 1 \\ 1 & -1 & 0 \end{vmatrix} = \hat{\mathbf{x}} + \hat{\mathbf{y}} - \hat{\mathbf{z}},$$

and

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) =$$

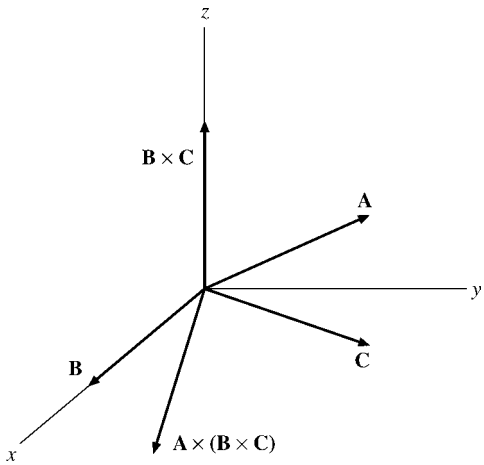

FIGURE 1.17 **B** and **C** are in the xy-plane. $\mathbf{B} \times \mathbf{C}$ is perpendicular to the xy-plane and is shown here along the z-axis. Then $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$ is perpendicular to the z-axis and therefore is back in the xy-plane.

to **A** and to $\mathbf{B} \times \mathbf{C}$. The plane defined by **B** and **C** is perpendicular to $\mathbf{B} \times \mathbf{C}$, and so the triple product lies in this plane (see Fig. 1.17):

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = u\mathbf{B} + v\mathbf{C}. \tag{1.54}$$

Taking the scalar product of Eq. (1.54) with **A** gives zero for the left-hand side, so $u\mathbf{A} \cdot \mathbf{B} + v\mathbf{A} \cdot \mathbf{C} = 0$. Hence $u = w\mathbf{A} \cdot \mathbf{C}$ and $v = -w\mathbf{A} \cdot \mathbf{B}$ for a suitable w. Substituting these values into Eq. (1.54) gives

$$\boxed{\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = w\big[\mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B})\big];} \tag{1.55}$$

we want to show that

$$w = 1$$

in Eq. (1.55), an important relation sometimes known as the **BAC–CAB** rule. Since Eq. (1.55) is linear in A, B, and C, w is independent of these magnitudes. That is, we only need to show that $w = 1$ for unit vectors $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, $\hat{\mathbf{C}}$. Let us denote $\hat{\mathbf{B}} \cdot \hat{\mathbf{C}} = \cos\alpha$, $\hat{\mathbf{C}} \cdot \hat{\mathbf{A}} = \cos\beta$, $\hat{\mathbf{A}} \cdot \hat{\mathbf{B}} = \cos\gamma$, and square Eq. (1.55) to obtain

$$\begin{aligned}
\big[\hat{\mathbf{A}} \times (\hat{\mathbf{B}} \times \hat{\mathbf{C}})\big]^2 &= \hat{\mathbf{A}}^2(\hat{\mathbf{B}} \times \hat{\mathbf{C}})^2 - \big[\hat{\mathbf{A}} \cdot (\hat{\mathbf{B}} \times \hat{\mathbf{C}})\big]^2 \\
&= 1 - \cos^2\alpha - \big[\hat{\mathbf{A}} \cdot (\hat{\mathbf{B}} \times \hat{\mathbf{C}})\big]^2 \\
&= w^2\big[(\hat{\mathbf{A}} \cdot \hat{\mathbf{C}})^2 + (\hat{\mathbf{A}} \cdot \hat{\mathbf{B}})^2 - 2(\hat{\mathbf{A}} \cdot \hat{\mathbf{B}})(\hat{\mathbf{A}} \cdot \hat{\mathbf{C}})(\hat{\mathbf{B}} \cdot \hat{\mathbf{C}})\big] \\
&= w^2\big(\cos^2\beta + \cos^2\gamma - 2\cos\alpha\cos\beta\cos\gamma\big),
\end{aligned} \tag{1.56}$$

using $(\hat{\mathbf{A}} \times \hat{\mathbf{B}})^2 = \hat{\mathbf{A}}^2\hat{\mathbf{B}}^2 - (\hat{\mathbf{A}} \cdot \hat{\mathbf{B}})^2$ repeatedly (see Eq. (1.43) for a proof). Consequently, the (squared) volume spanned by $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, $\hat{\mathbf{C}}$ that occurs in Eq. (1.56) can be written as

$$\left[\hat{\mathbf{A}} \cdot (\hat{\mathbf{B}} \times \hat{\mathbf{C}})\right]^2 = 1 - \cos^2\alpha - w^2\left(\cos^2\beta + \cos^2\gamma - 2\cos\alpha\cos\beta\cos\gamma\right).$$

Here $w^2 = 1$, since this volume is symmetric in α, β, γ. That is, $w = \pm 1$ and is independent of $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, $\hat{\mathbf{C}}$. Using again the special case $\hat{\mathbf{x}} \times (\hat{\mathbf{x}} \times \hat{\mathbf{y}}) = -\hat{\mathbf{y}}$ in Eq. (1.55) finally gives $w = 1$. (An alternate derivation using the Levi-Civita symbol ε_{ijk} of Chapter 2 is the topic of Exercise 2.9.8.)

It might be noted here that just as vectors are independent of the coordinates, so a vector equation is independent of the particular coordinate system. The coordinate system only determines the components. If the vector equation can be established in Cartesian coordinates, it is established and valid in any of the coordinate systems to be introduced in Chapter 2. Thus, Eq. (1.55) may be verified by a direct though not very elegant method of expanding into Cartesian components (see Exercise 1.5.2).

Exercises

1.5.1 One vertex of a glass parallelepiped is at the origin (Fig. 1.18). The three adjacent vertices are at $(3, 0, 0)$, $(0, 0, 2)$, and $(0, 3, 1)$. All lengths are in centimeters. Calculate the number of cubic centimeters of glass in the parallelepiped using the triple scalar product.

1.5.2 Verify the expansion of the triple vector product

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B})$$

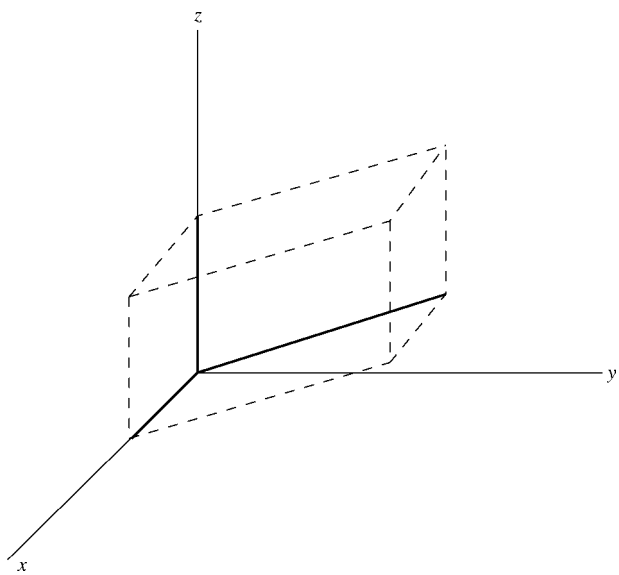

FIGURE 1.18 Parallelepiped: triple scalar product.

by direct expansion in Cartesian coordinates.

1.5.3 Show that the first step in Eq. (1.43), which is

$$(\mathbf{A}\times\mathbf{B})\cdot(\mathbf{A}\times\mathbf{B}) = A^2B^2 - (\mathbf{A}\cdot\mathbf{B})^2,$$

is consistent with the *BAC–CAB* rule for a triple vector product.

1.5.4 You are given the three vectors **A**, **B**, and **C**,

$$\mathbf{A} = \hat{\mathbf{x}} + \hat{\mathbf{y}},$$
$$\mathbf{B} = \hat{\mathbf{y}} + \hat{\mathbf{z}},$$
$$\mathbf{C} = \hat{\mathbf{x}} - \hat{\mathbf{z}}.$$

(a) Compute the triple scalar product, $\mathbf{A}\cdot\mathbf{B}\times\mathbf{C}$. Noting that $\mathbf{A} = \mathbf{B} + \mathbf{C}$, give a geometric interpretation of your result for the triple scalar product.
(b) Compute $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})$.

1.5.5 The orbital angular momentum **L** of a particle is given by $\mathbf{L} = \mathbf{r}\times\mathbf{p} = m\mathbf{r}\times\mathbf{v}$, where **p** is the linear momentum. With linear and angular velocity related by $\mathbf{v} = \boldsymbol{\omega}\times\mathbf{r}$, show that

$$\mathbf{L} = mr^2\big[\boldsymbol{\omega} - \hat{\mathbf{r}}(\hat{\mathbf{r}}\cdot\boldsymbol{\omega})\big].$$

Here $\hat{\mathbf{r}}$ is a unit vector in the **r**-direction. For $\mathbf{r}\cdot\boldsymbol{\omega} = 0$ this reduces to $\mathbf{L} = I\boldsymbol{\omega}$, with the moment of inertia I given by mr^2. In Section 3.5 this result is generalized to form an inertia tensor.

1.5.6 The kinetic energy of a single particle is given by $T = \frac{1}{2}mv^2$. For rotational motion this becomes $\frac{1}{2}m(\boldsymbol{\omega}\times\mathbf{r})^2$. Show that

$$T = \frac{1}{2}m\big[r^2\omega^2 - (\mathbf{r}\cdot\boldsymbol{\omega})^2\big].$$

For $\mathbf{r}\cdot\boldsymbol{\omega} = 0$ this reduces to $T = \frac{1}{2}I\omega^2$, with the moment of inertia I given by mr^2.

1.5.7 Show that[13]

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) + \mathbf{b}\times(\mathbf{c}\times\mathbf{a}) + \mathbf{c}\times(\mathbf{a}\times\mathbf{b}) = 0.$$

1.5.8 A vector **A** is decomposed into a radial vector $\mathbf{A}_r$ and a tangential vector $\mathbf{A}_t$. If $\hat{\mathbf{r}}$ is a unit vector in the radial direction, show that

(a) $\mathbf{A}_r = \hat{\mathbf{r}}(\mathbf{A}\cdot\hat{\mathbf{r}})$ and
(b) $\mathbf{A}_t = -\hat{\mathbf{r}}\times(\hat{\mathbf{r}}\times\mathbf{A})$.

1.5.9 Prove that a necessary and sufficient condition for the three (nonvanishing) vectors **A**, **B**, and **C** to be coplanar is the vanishing of the triple scalar product

$$\mathbf{A}\cdot\mathbf{B}\times\mathbf{C} = 0.$$

[13]This is Jacobi's identity for vector products; for commutators it is important in the context of Lie algebras (see Eq. (4.16) in Section 4.2).

1.5.10 Three vectors **A**, **B**, and **C** are given by

$$\mathbf{A} = 3\hat{\mathbf{x}} - 2\hat{\mathbf{y}} + 2\hat{\mathbf{z}},$$
$$\mathbf{B} = 6\hat{\mathbf{x}} + 4\hat{\mathbf{y}} - 2\hat{\mathbf{z}},$$
$$\mathbf{C} = -3\hat{\mathbf{x}} - 2\hat{\mathbf{y}} - 4\hat{\mathbf{z}}.$$

Compute the values of $\mathbf{A}\cdot\mathbf{B}\times\mathbf{C}$ and $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})$, $\mathbf{C}\times(\mathbf{A}\times\mathbf{B})$ and $\mathbf{B}\times(\mathbf{C}\times\mathbf{A})$.

1.5.11 Vector **D** is a linear combination of three noncoplanar (and nonorthogonal) vectors:

$$\mathbf{D} = a\mathbf{A} + b\mathbf{B} + c\mathbf{C}.$$

Show that the coefficients are given by a ratio of triple scalar products,

$$a = \frac{\mathbf{D}\cdot\mathbf{B}\times\mathbf{C}}{\mathbf{A}\cdot\mathbf{B}\times\mathbf{C}}, \qquad \text{and so on.}$$

1.5.12 Show that

$$(\mathbf{A}\times\mathbf{B})\cdot(\mathbf{C}\times\mathbf{D}) = (\mathbf{A}\cdot\mathbf{C})(\mathbf{B}\cdot\mathbf{D}) - (\mathbf{A}\cdot\mathbf{D})(\mathbf{B}\cdot\mathbf{C}).$$

1.5.13 Show that

$$(\mathbf{A}\times\mathbf{B})\times(\mathbf{C}\times\mathbf{D}) = (\mathbf{A}\cdot\mathbf{B}\times\mathbf{D})\mathbf{C} - (\mathbf{A}\cdot\mathbf{B}\times\mathbf{C})\mathbf{D}.$$

1.5.14 For a **spherical** triangle such as pictured in Fig. 1.14 show that

$$\frac{\sin A}{\sin\overline{BC}} = \frac{\sin B}{\sin\overline{CA}} = \frac{\sin C}{\sin\overline{AB}}.$$

Here $\sin A$ is the sine of the included angle at A, while $\overline{BC}$ is the side opposite (in radians).

1.5.15 Given

$$\mathbf{a}' = \frac{\mathbf{b}\times\mathbf{c}}{\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}}, \qquad \mathbf{b}' = \frac{\mathbf{c}\times\mathbf{a}}{\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}}, \qquad \mathbf{c}' = \frac{\mathbf{a}\times\mathbf{b}}{\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}},$$

and $\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}\neq 0$, show that

(a) $\mathbf{x}\cdot\mathbf{y}' = \delta_{xy}$, $(\mathbf{x},\mathbf{y} = \mathbf{a},\mathbf{b},\mathbf{c})$,
(b) $\mathbf{a}'\cdot\mathbf{b}'\times\mathbf{c}' = (\mathbf{a}\cdot\mathbf{b}\times\mathbf{c})^{-1}$,
(c) $\mathbf{a} = \dfrac{\mathbf{b}'\times\mathbf{c}'}{\mathbf{a}'\cdot\mathbf{b}'\times\mathbf{c}'}$.

1.5.16 If $\mathbf{x}\cdot\mathbf{y}' = \delta_{xy}$, $(\mathbf{x},\mathbf{y} = \mathbf{a},\mathbf{b},\mathbf{c})$, prove that

$$\mathbf{a}' = \frac{\mathbf{b}\times\mathbf{c}}{\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}}.$$

(This is the converse of Problem 1.5.15.)

1.5.17 Show that any vector **V** may be expressed in terms of the reciprocal vectors $\mathbf{a}'$, $\mathbf{b}'$, $\mathbf{c}'$ (of Problem 1.5.15) by

$$\mathbf{V} = (\mathbf{V}\cdot\mathbf{a})\mathbf{a}' + (\mathbf{V}\cdot\mathbf{b})\mathbf{b}' + (\mathbf{V}\cdot\mathbf{c})\mathbf{c}'.$$

1.5.18 An electric charge q_1 moving with velocity $\mathbf{v}_1$ produces a magnetic induction $\mathbf{B}$ given by

$$\mathbf{B} = \frac{\mu_0}{4\pi} q_1 \frac{\mathbf{v}_1 \times \hat{\mathbf{r}}}{r^2} \qquad \text{(mks units)},$$

where $\hat{\mathbf{r}}$ points from q_1 to the point at which $\mathbf{B}$ is measured (Biot and Savart law).

(a) Show that the magnetic force on a second charge q_2, velocity $\mathbf{v}_2$, is given by the triple vector product

$$\mathbf{F}_2 = \frac{\mu_0}{4\pi} \frac{q_1 q_2}{r^2} \mathbf{v}_2 \times (\mathbf{v}_1 \times \hat{\mathbf{r}}).$$

(b) Write out the corresponding magnetic force $\mathbf{F}_1$ that q_2 exerts on q_1. Define your unit radial vector. How do $\mathbf{F}_1$ and $\mathbf{F}_2$ compare?

(c) Calculate $\mathbf{F}_1$ and $\mathbf{F}_2$ for the case of q_1 and q_2 moving along parallel trajectories side by side.

ANS.

(b) $\mathbf{F}_1 = -\dfrac{\mu_0}{4\pi} \dfrac{q_1 q_2}{r^2} \mathbf{v}_1 \times (\mathbf{v}_2 \times \hat{\mathbf{r}})$.
In general, there is no simple relation between $\mathbf{F}_1$ and $\mathbf{F}_2$. Specifically, Newton's third law, $\mathbf{F}_1 = -\mathbf{F}_2$, does not hold.

(c) $\mathbf{F}_1 = \dfrac{\mu_0}{4\pi} \dfrac{q_1 q_2}{r^2} v^2 \hat{\mathbf{r}} = -\mathbf{F}_2$.
Mutual attraction.

1.6 GRADIENT, ∇

To provide a motivation for the vector nature of partial derivatives, we now introduce the **total variation of a function** $F(x, y)$,

$$\boxed{dF = \frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial y} dy.}$$

It consists of independent variations in the x- and y-directions. We write dF as a sum of two increments, one purely in the x- and the other in the y-direction,

$$\begin{aligned} dF(x, y) &\equiv F(x + dx, y + dy) - F(x, y) \\ &= \big[F(x + dx, y + dy) - F(x, y + dy)\big] + \big[F(x, y + dy) - F(x, y)\big] \\ &= \frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial y} dy, \end{aligned}$$

by adding and subtracting $F(x, y + dy)$. The mean value theorem (that is, continuity of F) tells us that here $\partial F/\partial x$, $\partial F/\partial y$ are evaluated at some point ξ, η between x and $x + dx$, y

and $y + dy$, respectively. As $dx \to 0$ and $dy \to 0$, $\xi \to x$ and $\eta \to y$. This result generalizes to three and higher dimensions. For example, for a function φ of three variables,

$$\begin{aligned} d\varphi(x, y, z) \equiv & \big[\varphi(x + dx, y + dy, z + dz) - \varphi(x, y + dy, z + dz)\big] \\ & + \big[\varphi(x, y + dy, z + dz) - \varphi(x, y, z + dz)\big] \\ & + \big[\varphi(x, y, z + dz) - \varphi(x, y, z)\big] \\ = & \frac{\partial \varphi}{\partial x} dx + \frac{\partial \varphi}{\partial y} dy + \frac{\partial \varphi}{\partial z} dz. \end{aligned} \tag{1.57}$$

Algebraically, $d\varphi$ in the total variation is a scalar product of the change in position $d\mathbf{r}$ and the **directional** change of φ. And now we are ready to recognize the three-dimensional partial derivative as a vector, which leads us to the concept of gradient.

Suppose that $\varphi(x, y, z)$ is a scalar point function, that is, a function whose value depends on the values of the coordinates (x, y, z). As a scalar, it must have the same value at a given fixed point in space, independent of the rotation of our coordinate system, or

$$\varphi'(x_1', x_2', x_3') = \varphi(x_1, x_2, x_3). \tag{1.58}$$

By differentiating with respect to x_i' we obtain

$$\frac{\partial \varphi'(x_1', x_2', x_3')}{\partial x_i'} = \frac{\partial \varphi(x_1, x_2, x_3)}{\partial x_i'} = \sum_j \frac{\partial \varphi}{\partial x_j} \frac{\partial x_j}{\partial x_i'} = \sum_j a_{ij} \frac{\partial \varphi}{\partial x_j} \tag{1.59}$$

by the rules of partial differentiation and Eqs. (1.16a) and (1.16b). But comparison with Eq. (1.17), the vector transformation law, now shows that we have **constructed** a vector with components $\partial \varphi / \partial x_j$. This vector we label the gradient of φ.

A convenient symbolism is

$$\nabla \varphi = \hat{\mathbf{x}} \frac{\partial \varphi}{\partial x} + \hat{\mathbf{y}} \frac{\partial \varphi}{\partial y} + \hat{\mathbf{z}} \frac{\partial \varphi}{\partial z} \tag{1.60}$$

or

$$\boxed{\nabla = \hat{\mathbf{x}} \frac{\partial}{\partial x} + \hat{\mathbf{y}} \frac{\partial}{\partial y} + \hat{\mathbf{z}} \frac{\partial}{\partial z}.} \tag{1.61}$$

$\nabla \varphi$ (or del φ) is our gradient of the scalar φ, whereas ∇ (del) itself is a vector differential operator (available to operate on or to differentiate a scalar φ). All the relationships for ∇ (del) can be derived from the hybrid nature of del in terms of both the partial derivatives and its vector nature.

The gradient of a scalar is extremely important in physics and engineering in expressing the relation between a force field and a potential field,

$$\text{force } \mathbf{F} = -\nabla(\text{potential } V), \tag{1.62}$$

which holds for both gravitational and electrostatic fields, among others. Note that the minus sign in Eq. (1.62) results in water flowing downhill rather than uphill! If a force can be described, as in Eq. (1.62), by a single function $V(\mathbf{r})$ everywhere, we call the scalar function V its **potential**. Because the force is the directional derivative of the potential, we can find the potential, if it exists, by integrating the force along a suitable path. Because the

total variation $dV = \nabla V \cdot d\mathbf{r} = -\mathbf{F} \cdot d\mathbf{r}$ is the work done against the force along the path $d\mathbf{r}$, we recognize the physical meaning of the potential (difference) as work and energy. Moreover, in a sum of path increments the intermediate points cancel,

$$\big[V(\mathbf{r} + d\mathbf{r}_1 + d\mathbf{r}_2) - V(\mathbf{r} + d\mathbf{r}_1)\big] + \big[V(\mathbf{r} + d\mathbf{r}_1) - V(\mathbf{r})\big] = V(\mathbf{r} + d\mathbf{r}_2 + d\mathbf{r}_1) - V(\mathbf{r}),$$

so the integrated work along some path from an initial point $\mathbf{r}_i$ to a final point $\mathbf{r}$ is given by the potential difference $V(\mathbf{r}) - V(\mathbf{r}_i)$ at the endpoints of the path. Therefore, such forces are especially simple and well behaved: They are called **conservative**. When there is loss of energy due to friction along the path or some other dissipation, the work will depend on the path, and such forces cannot be conservative: No potential exists. We discuss conservative forces in more detail in Section 1.13.

Example 1.6.1 THE GRADIENT OF A POTENTIAL $V(r)$

Let us calculate the gradient of $V(r) = V(\sqrt{x^2 + y^2 + z^2})$, so

$$\nabla V(r) = \hat{\mathbf{x}} \frac{\partial V(r)}{\partial x} + \hat{\mathbf{y}} \frac{\partial V(r)}{\partial y} + \hat{\mathbf{z}} \frac{\partial V(r)}{\partial z}.$$

Now, $V(r)$ depends on x through the dependence of r on x. Therefore[14]

$$\frac{\partial V(r)}{\partial x} = \frac{dV(r)}{dr} \cdot \frac{\partial r}{\partial x}.$$

From r as a function of x, y, z,

$$\frac{\partial r}{\partial x} = \frac{\partial (x^2 + y^2 + z^2)^{1/2}}{\partial x} = \frac{x}{(x^2 + y^2 + z^2)^{1/2}} = \frac{x}{r}.$$

Therefore

$$\frac{\partial V(r)}{\partial x} = \frac{dV(r)}{dr} \cdot \frac{x}{r}.$$

Permuting coordinates ($x \to y$, $y \to z$, $z \to x$) to obtain the y and z derivatives, we get

$$\boxed{\begin{aligned} \nabla V(r) &= (\hat{\mathbf{x}} x + \hat{\mathbf{y}} y + \hat{\mathbf{z}} z) \frac{1}{r} \frac{dV}{dr} \\ &= \frac{\mathbf{r}}{r} \frac{dV}{dr} = \hat{\mathbf{r}} \frac{dV}{dr}. \end{aligned}}$$

Here $\hat{\mathbf{r}}$ is a unit vector $(\mathbf{r}/r)$ in the **positive** radial direction. The gradient of a function of r is a vector in the (positive or negative) radial direction. In Section 2.5, $\hat{\mathbf{r}}$ is seen as one of the three orthonormal unit vectors of spherical polar coordinates and $\hat{\mathbf{r}}\partial/\partial r$ as the radial component of ∇. ■

[14] This is a special case of the **chain rule** of partial differentiation:

$$\frac{\partial V(r, \theta, \varphi)}{\partial x} = \frac{\partial V}{\partial r} \frac{\partial r}{\partial x} + \frac{\partial V}{\partial \theta} \frac{\partial \theta}{\partial x} + \frac{\partial V}{\partial \varphi} \frac{\partial \varphi}{\partial x},$$

where $\partial V/\partial\theta = \partial V/\partial\varphi = 0$, $\partial V/\partial r \to dV/dr$.

A Geometrical Interpretation

One immediate application of $\nabla\varphi$ is to dot it into an increment of length

$$d\mathbf{r} = \hat{\mathbf{x}}\,dx + \hat{\mathbf{y}}\,dy + \hat{\mathbf{z}}\,dz.$$

Thus we obtain

$$\nabla\varphi \cdot d\mathbf{r} = \frac{\partial\varphi}{\partial x}dx + \frac{\partial\varphi}{\partial y}dy + \frac{\partial\varphi}{\partial z}dz = d\varphi,$$

the change in the scalar function φ corresponding to a change in position $d\mathbf{r}$. Now consider P and Q to be two points on a surface $\varphi(x, y, z) = C$, a constant. These points are chosen so that Q is a distance $d\mathbf{r}$ from P. Then, moving from P to Q, the change in $\varphi(x, y, z) = C$ is given by

$$d\varphi = (\nabla\varphi) \cdot d\mathbf{r} = 0 \tag{1.63}$$

since we stay on the surface $\varphi(x, y, z) = C$. This shows that $\nabla\varphi$ is perpendicular to $d\mathbf{r}$. Since $d\mathbf{r}$ may have any direction from P **as long as it stays in the surface** of constant φ, point Q being restricted to the surface but having arbitrary direction, $\nabla\varphi$ is seen as normal to the surface $\varphi =$ constant (Fig. 1.19).

If we now permit $d\mathbf{r}$ to take us from one surface $\varphi = C_1$ to an adjacent surface $\varphi = C_2$ (Fig. 1.20),

$$d\varphi = C_1 - C_2 = \Delta C = (\nabla\varphi) \cdot d\mathbf{r}. \tag{1.64}$$

For a given $d\varphi$, $|d\mathbf{r}|$ is a minimum when it is chosen parallel to $\nabla\varphi$ ($\cos\theta = 1$); or, for a given $|d\mathbf{r}|$, the change in the scalar function φ is maximized by choosing $d\mathbf{r}$ parallel to

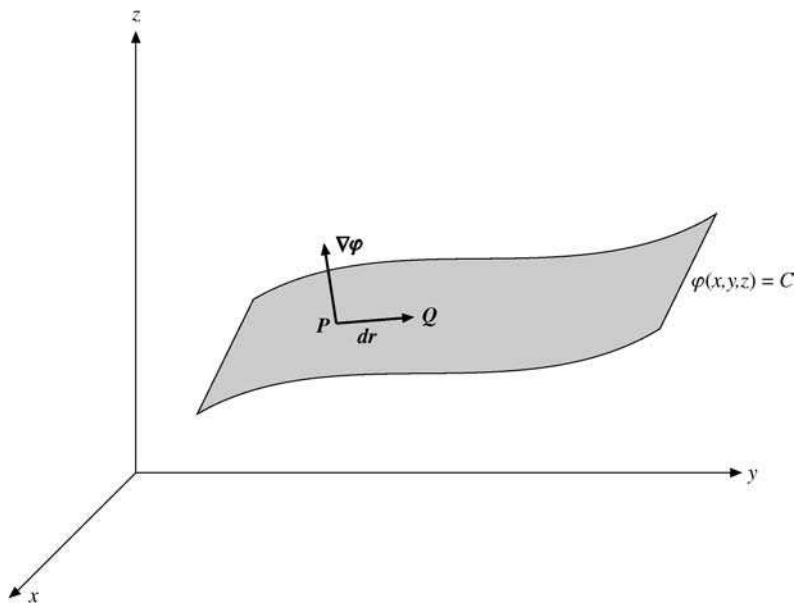

FIGURE 1.19 The length increment $d\mathbf{r}$ has to stay on the surface $\varphi = C$.

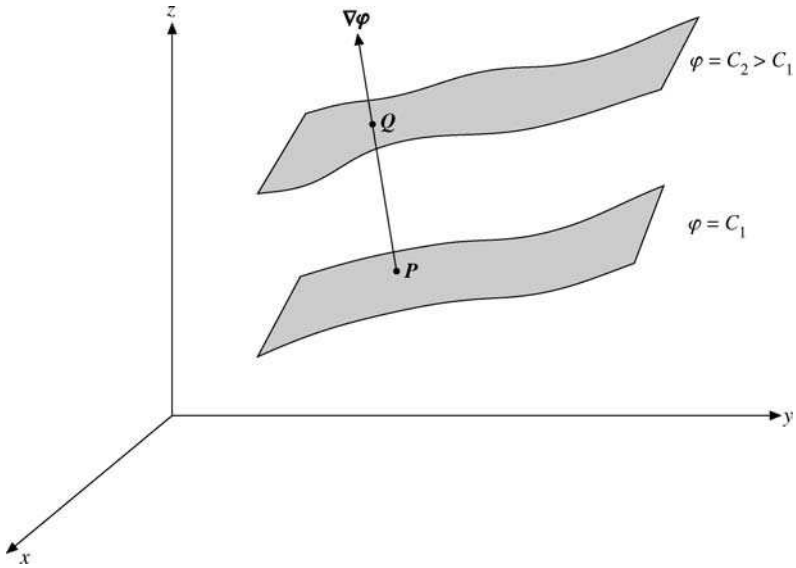

FIGURE 1.20 Gradient.

$\nabla\varphi$. **This identifies $\nabla\varphi$ as a vector having the direction of the maximum space rate of change of** φ, an identification that will be useful in Chapter 2 when we consider non-Cartesian coordinate systems. This identification of $\nabla\varphi$ may also be developed by using the calculus of variations subject to a constraint, Exercise 17.6.9.

Example 1.6.2 FORCE AS GRADIENT OF A POTENTIAL

As a specific example of the foregoing, and as an extension of Example 1.6.1, we consider the surfaces consisting of concentric spherical shells, Fig. 1.21. We have

$$\varphi(x, y, z) = \left(x^2 + y^2 + z^2\right)^{1/2} = r = C,$$

where r is the radius, equal to C, our constant. $\Delta C = \Delta\varphi = \Delta r$, the distance between two shells. From Example 1.6.1

$$\nabla\varphi(r) = \hat{\mathbf{r}}\frac{d\varphi(r)}{dr} = \hat{\mathbf{r}}.$$

The gradient is in the radial direction and is normal to the spherical surface $\varphi = C$. ■

Example 1.6.3 INTEGRATION BY PARTS OF GRADIENT

Let us prove the formula $\int \mathbf{A}(\mathbf{r}) \cdot \nabla f(\mathbf{r})\, d^3r = -\int f(\mathbf{r}) \nabla \cdot \mathbf{A}(\mathbf{r})\, d^3r$, where $\mathbf{A}$ or f or both vanish at infinity so that the integrated parts vanish. This condition is satisfied if, for example, $\mathbf{A}$ is the electromagnetic vector potential and f is a bound-state wave function $\psi(\mathbf{r})$.

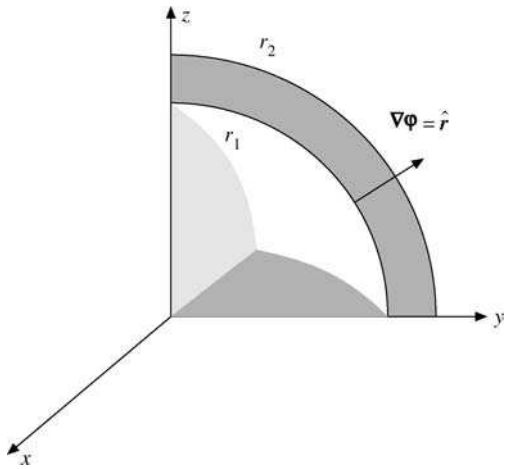

FIGURE 1.21 Gradient for $\varphi(x, y, z) = (x^2 + y^2 + z^2)^{1/2}$, spherical shells: $(x_2^2 + y_2^2 + z_2^2)^{1/2} = r_2 = C_2$, $(x_1^2 + y_1^2 + z_1^2)^{1/2} = r_1 = C_1$.

Writing the inner product in Cartesian coordinates, integrating each one-dimensional integral by parts, and dropping the integrated terms, we obtain

$$\begin{aligned}\int \mathbf{A}(\mathbf{r}) \cdot \nabla f(\mathbf{r})\, d^3r &= \iint \left[A_x f|_{x=-\infty}^{\infty} - \int f \frac{\partial A_x}{\partial x} dx \right] dy\, dz + \cdots \\ &= -\iiint f \frac{\partial A_x}{\partial x} dx\, dy\, dz - \iiint f \frac{\partial A_y}{\partial y} dy\, dx\, dz - \iiint f \frac{\partial A_z}{\partial z} dz\, dx\, dy \\ &= -\int f(\mathbf{r}) \nabla \cdot \mathbf{A}(\mathbf{r})\, d^3r.\end{aligned}$$

If $\mathbf{A} = e^{ikz}\hat{\mathbf{e}}$ describes an outgoing photon in the direction of the constant polarization unit vector $\hat{\mathbf{e}}$ and $f = \psi(\mathbf{r})$ is an exponentially decaying bound-state wave function, then

$$\int e^{ikz}\hat{\mathbf{e}} \cdot \nabla \psi(\mathbf{r})\, d^3r = -e_z \int \psi(\mathbf{r}) \frac{de^{ikz}}{dz} d^3r = -ike_z \int \psi(\mathbf{r}) e^{ikz}\, d^3r,$$

because only the z-component of the gradient contributes. ■

Exercises

1.6.1 If $S(x, y, z) = (x^2 + y^2 + z^2)^{-3/2}$, find

(a) ∇S at the point $(1, 2, 3)$;
(b) the magnitude of the gradient of S, $|\nabla S|$ at $(1, 2, 3)$; and
(c) the direction cosines of ∇S at $(1, 2, 3)$.

1.6.2 (a) Find a unit vector perpendicular to the surface

$$x^2 + y^2 + z^2 = 3$$

at the point $(1, 1, 1)$. Lengths are in centimeters.

(b) Derive the equation of the plane tangent to the surface at $(1, 1, 1)$.

ANS. (a) $(\hat{\mathbf{x}} + \hat{\mathbf{y}} + \hat{\mathbf{z}})/\sqrt{3}$, (b) $x + y + z = 3$.

1.6.3 Given a vector $\mathbf{r}_{12} = \hat{\mathbf{x}}(x_1 - x_2) + \hat{\mathbf{y}}(y_1 - y_2) + \hat{\mathbf{z}}(z_1 - z_2)$, show that $\nabla_1 r_{12}$ (gradient with respect to x_1, y_1, and z_1 of the magnitude r_{12}) is a unit vector in the direction of $\mathbf{r}_{12}$.

1.6.4 If a vector function $\mathbf{F}$ depends on both space coordinates (x, y, z) and time t, show that

$$d\mathbf{F} = (d\mathbf{r} \cdot \nabla)\mathbf{F} + \frac{\partial \mathbf{F}}{\partial t} dt.$$

1.6.5 Show that $\nabla(uv) = v\nabla u + u\nabla v$, where u and v are differentiable scalar functions of x, y, and z.

(a) Show that a necessary and sufficient condition that $u(x, y, z)$ and $v(x, y, z)$ are related by some function $f(u, v) = 0$ is that $(\nabla u) \times (\nabla v) = 0$.

(b) If $u = u(x, y)$ and $v = v(x, y)$, show that the condition $(\nabla u) \times (\nabla v) = 0$ leads to the two-dimensional Jacobian

$$J\left(\frac{u, v}{x, y}\right) = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} = 0.$$

The functions u and v are assumed differentiable.

1.7 DIVERGENCE, ∇

Differentiating a vector function is a simple extension of differentiating scalar quantities. Suppose $\mathbf{r}(t)$ describes the position of a satellite at some time t. Then, for differentiation with respect to time,

$$\frac{d\mathbf{r}(t)}{dt} = \lim_{\Delta \to 0} \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t)}{\Delta t} = \mathbf{v}, \text{ linear velocity.}$$

Graphically, we again have the slope of a curve, orbit, or trajectory, as shown in Fig. 1.22.

If we resolve $\mathbf{r}(t)$ into its Cartesian components, $d\mathbf{r}/dt$ always reduces directly to a vector sum of not more than three (for three-dimensional space) scalar derivatives. In other coordinate systems (Chapter 2) the situation is more complicated, for the unit vectors are no longer constant in direction. Differentiation with respect to the space coordinates is handled in the same way as differentiation with respect to time, as seen in the following paragraphs.

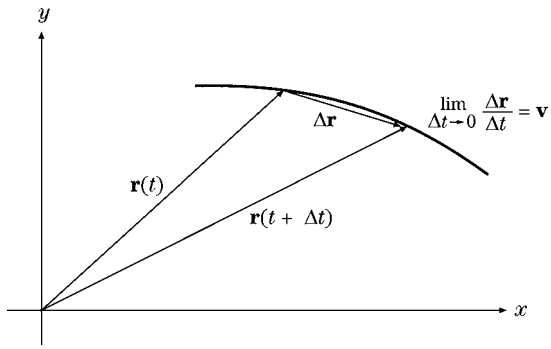

FIGURE 1.22 Differentiation of a vector.

In Section 1.6, ∇ was defined as a vector operator. Now, paying attention to both its vector and its differential properties, we let it operate on a vector. First, as a vector we dot it into a second vector to obtain

$$\boxed{\nabla \cdot \mathbf{V} = \frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y} + \frac{\partial V_z}{\partial z},} \tag{1.65a}$$

known as the divergence of $\mathbf{V}$. This is a scalar, as discussed in Section 1.3.

Example 1.7.1 DIVERGENCE OF COORDINATE VECTOR

Calculate $\nabla \cdot \mathbf{r}$:

$$\begin{aligned}\nabla \cdot \mathbf{r} &= \left(\hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{y}}\frac{\partial}{\partial y} + \hat{\mathbf{z}}\frac{\partial}{\partial z}\right) \cdot (\hat{\mathbf{x}}x + \hat{\mathbf{y}}y + \hat{\mathbf{z}}z) \\ &= \frac{\partial x}{\partial x} + \frac{\partial y}{\partial y} + \frac{\partial z}{\partial z},\end{aligned}$$

or $\nabla \cdot \mathbf{r} = 3$. ■

Example 1.7.2 DIVERGENCE OF CENTRAL FORCE FIELD

Generalizing Example 1.7.1,

$$\begin{aligned}\nabla \cdot \big(\mathbf{r} f(r)\big) &= \frac{\partial}{\partial x}\big[x\, f(r)\big] + \frac{\partial}{\partial y}\big[y\, f(r)\big] + \frac{\partial}{\partial z}\big[z\, f(r)\big] \\ &= 3f(r) + \frac{x^2}{r}\frac{df}{dr} + \frac{y^2}{r}\frac{df}{dr} + \frac{z^2}{r}\frac{df}{dr} \\ &= 3f(r) + r\frac{df}{dr}.\end{aligned}$$

The manipulation of the partial derivatives leading to the second equation in Example 1.7.2 is discussed in Example 1.6.1. In particular, if $f(r) = r^{n-1}$,

$$\begin{aligned}\nabla \cdot \left(\mathbf{r} r^{n-1}\right) &= \nabla \cdot \hat{\mathbf{r}} r^n \\ &= 3r^{n-1} + (n-1)r^{n-1} \\ &= (n+2)r^{n-1}. \end{aligned} \tag{1.65b}$$

This divergence vanishes for $n = -2$, except at $r = 0$, an important fact in Section 1.14. ■

Example 1.7.3 INTEGRATION BY PARTS OF DIVERGENCE

Let us prove the formula $\int f(\mathbf{r}) \nabla \cdot \mathbf{A}(\mathbf{r})\, d^3r = -\int \mathbf{A} \cdot \nabla f\, d^3r$, where $\mathbf{A}$ or f or both vanish at infinity.

To show this, we proceed, as in Example 1.6.3, by integration by parts after writing the inner product in Cartesian coordinates. Because the integrated terms are evaluated at infinity, where they vanish, we obtain

$$\begin{aligned}\int f(\mathbf{r}) \nabla \cdot \mathbf{A}(\mathbf{r})\, d^3r &= \int f \left(\frac{\partial A_x}{\partial x} dx\, dy\, dz + \frac{\partial A_y}{\partial y} dy\, dx\, dz + \frac{\partial A_z}{\partial z} dz\, dx\, dy \right) \\ &= -\int \left(A_x \frac{\partial f}{\partial x} dx\, dy\, dz + A_y \frac{\partial f}{\partial y} dy\, dx\, dz + A_z \frac{\partial f}{\partial z} dz\, dx\, dy \right) \\ &= -\int \mathbf{A} \cdot \nabla f\, d^3r. \end{aligned}$$

■

A Physical Interpretation

To develop a feeling for the physical significance of the divergence, consider $\nabla \cdot (\rho \mathbf{v})$ with $\mathbf{v}(x, y, z)$, the velocity of a compressible fluid, and $\rho(x, y, z)$, its density at point (x, y, z). If we consider a small volume $dx\, dy\, dz$ (Fig. 1.23) at $x = y = z = 0$, the fluid flowing into this volume per unit time (positive x-direction) through the face *EFGH* is (rate of flow in)$_{EFGH} = \rho v_x|_{x=0} = dy\, dz$. The components of the flow ρv_y and ρv_z tangential to this face contribute nothing to the flow through this face. The rate of flow out (still positive x-direction) through face *ABCD* is $\rho v_x|_{x=dx}\, dy\, dz$. To compare these flows and to find the net flow out, we expand this last result, like the total variation in Section 1.6.[15] This yields

$$\begin{aligned}(\text{rate of flow out})_{ABCD} &= \rho v_x|_{x=dx}\, dy\, dz \\ &= \left[\rho v_x + \frac{\partial}{\partial x}(\rho v_x)\, dx \right]_{x=0} dy\, dz. \end{aligned}$$

Here the derivative term is a first correction term, allowing for the possibility of nonuniform density or velocity or both.[16] The zero-order term $\rho v_x|_{x=0}$ (corresponding to uniform flow)

[15] Here we have the increment dx and we show a partial derivative with respect to x since ρv_x may also depend on y and z.

[16] Strictly speaking, ρv_x is averaged over face *EFGH* and the expression $\rho v_x + (\partial/\partial x)(\rho v_x)\, dx$ is similarly averaged over face *ABCD*. Using an arbitrarily small differential volume, we find that the averages reduce to the values employed here.

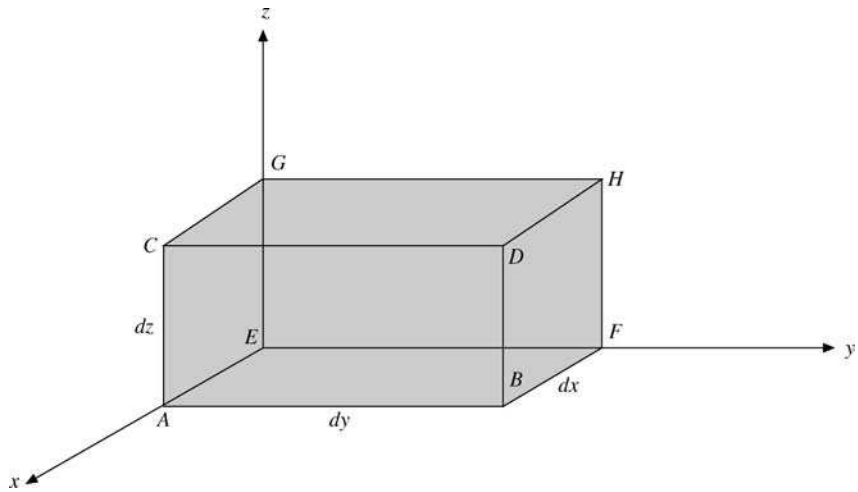

FIGURE 1.23 Differential rectangular parallelepiped (in first octant).

cancels out:

$$\text{Net rate of flow out}|_x = \frac{\partial}{\partial x}(\rho v_x)\,dx\,dy\,dz.$$

Equivalently, we can arrive at this result by

$$\lim_{\Delta x\to 0}\frac{\rho v_x(\Delta x,0,0)-\rho v_x(0,0,0)}{\Delta x}\equiv\left.\frac{\partial[\rho v_x(x,y,z)]}{\partial x}\right|_{0,0,0}.$$

Now, the x-axis is not entitled to any preferred treatment. The preceding result for the two faces perpendicular to the x-axis must hold for the two faces perpendicular to the y-axis, with x replaced by y and the corresponding changes for y and z: $y\to z$, $z\to x$. This is a cyclic permutation of the coordinates. A further cyclic permutation yields the result for the remaining two faces of our parallelepiped. Adding the net rate of flow out for all three pairs of surfaces of our volume element, we have

$$\begin{aligned}\text{net flow out (per unit time)} &= \left[\frac{\partial}{\partial x}(\rho v_x)+\frac{\partial}{\partial y}(\rho v_y)+\frac{\partial}{\partial z}(\rho v_z)\right]dx\,dy\,dz\\ &= \nabla\cdot(\rho\mathbf{v})\,dx\,dy\,dz. \end{aligned}\tag{1.66}$$

Therefore the net flow of our compressible fluid out of the volume element $dx\,dy\,dz$ per unit volume per unit time is $\nabla\cdot(\rho\mathbf{v})$. Hence the name **divergence**. A direct application is in the continuity equation

$$\frac{\partial\rho}{\partial t}+\nabla\cdot(\rho\mathbf{v})=0,\tag{1.67a}$$

which states that a net flow out of the volume results in a decreased density inside the volume. Note that in Eq. (1.67a), ρ is considered to be a possible function of time as well as of space: $\rho(x,y,z,t)$. The divergence appears in a wide variety of physical problems,

ranging from a probability current density in quantum mechanics to neutron leakage in a nuclear reactor.

The combination $\nabla \cdot (f\mathbf{V})$, in which f is a scalar function and $\mathbf{V}$ is a vector function, may be written

$$\begin{aligned}\nabla \cdot (f\mathbf{V}) &= \frac{\partial}{\partial x}(fV_x) + \frac{\partial}{\partial y}(fV_y) + \frac{\partial}{\partial z}(fV_z) \\ &= \frac{\partial f}{\partial x}V_x + f\frac{\partial V_x}{\partial x} + \frac{\partial f}{\partial y}V_y + f\frac{\partial V_y}{\partial y} + \frac{\partial f}{\partial z}V_z + f\frac{\partial V_z}{\partial z} \\ &= (\nabla f)\cdot\mathbf{V} + f\nabla\cdot\mathbf{V}, \end{aligned} \tag{1.67b}$$

which is just what we would expect for the derivative of a product. Notice that ∇ as a differential operator differentiates both f and $\mathbf{V}$; as a vector it is dotted into $\mathbf{V}$ (in each term).

If we have the special case of the divergence of a vector vanishing,

$$\nabla\cdot\mathbf{B} = 0, \tag{1.68}$$

the vector $\mathbf{B}$ is said to be **solenoidal**, the term coming from the example in which $\mathbf{B}$ is the magnetic induction and Eq. (1.68) appears as one of Maxwell's equations. When a vector is solenoidal, it may be written as the curl of another vector known as the vector potential. (In Section 1.13 we shall calculate such a vector potential.)

Exercises

1.7.1 For a particle moving in a circular orbit $\mathbf{r} = \hat{\mathbf{x}}r\cos\omega t + \hat{\mathbf{y}}r\sin\omega t$,

(a) evaluate $\mathbf{r}\times\dot{\mathbf{r}}$, with $\dot{\mathbf{r}} = \frac{d\mathbf{r}}{dt} = \mathbf{v}$.
(b) Show that $\ddot{\mathbf{r}} + \omega^2\mathbf{r} = 0$ with $\ddot{\mathbf{r}} = \frac{d\mathbf{v}}{dt}$.
The radius r and the angular velocity ω are constant.

ANS. (a) $\hat{\mathbf{z}}\omega r^2$.

1.7.2 Vector $\mathbf{A}$ satisfies the vector transformation law, Eq. (1.15). Show directly that its time derivative $d\mathbf{A}/dt$ also satisfies Eq. (1.15) and is therefore a vector.

1.7.3 Show, by differentiating components, that

(a) $\frac{d}{dt}(\mathbf{A}\cdot\mathbf{B}) = \frac{d\mathbf{A}}{dt}\cdot\mathbf{B} + \mathbf{A}\cdot\frac{d\mathbf{B}}{dt}$,
(b) $\frac{d}{dt}(\mathbf{A}\times\mathbf{B}) = \frac{d\mathbf{A}}{dt}\times\mathbf{B} + \mathbf{A}\times\frac{d\mathbf{B}}{dt}$,
just like the derivative of the product of two algebraic functions.

1.7.4 In Chapter 2 it will be seen that the unit vectors in non-Cartesian coordinate systems are usually functions of the coordinate variables, $\mathbf{e}_i = \mathbf{e}_i(q_1, q_2, q_3)$ but $|\mathbf{e}_i| = 1$. Show that either $\partial\mathbf{e}_i/\partial q_j = 0$ or $\partial\mathbf{e}_i/\partial q_j$ is orthogonal to $\mathbf{e}_i$.
Hint. $\partial\mathbf{e}_i^2/\partial q_j = 0$.

1.7.5 Prove $\nabla\cdot(\mathbf{a}\times\mathbf{b}) = \mathbf{b}\cdot(\nabla\times\mathbf{a}) - \mathbf{a}\cdot(\nabla\times\mathbf{b})$.
Hint. Treat as a triple scalar product.

1.7.6 The electrostatic field of a point charge q is

$$\mathbf{E} = \frac{q}{4\pi\varepsilon_0}\cdot\frac{\hat{\mathbf{r}}}{r^2}.$$

Calculate the divergence of $\mathbf{E}$. What happens at the origin?

1.8 CURL, $\nabla\times$

Another possible operation with the vector operator ∇ is to cross it into a vector. We obtain

$$\begin{aligned}\nabla\times\mathbf{V} &= \hat{\mathbf{x}}\left(\frac{\partial}{\partial y}V_z - \frac{\partial}{\partial z}V_y\right) + \hat{\mathbf{y}}\left(\frac{\partial}{\partial z}V_x - \frac{\partial}{\partial x}V_z\right) + \hat{\mathbf{z}}\left(\frac{\partial}{\partial x}V_y - \frac{\partial}{\partial y}V_x\right)\\ &= \begin{vmatrix}\hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}}\\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z}\\ V_x & V_y & V_z\end{vmatrix},\end{aligned} \tag{1.69}$$

which is called the **curl** of $\mathbf{V}$. In expanding this determinant we must consider the derivative nature of ∇. Specifically, $\mathbf{V}\times\nabla$ is defined only as an operator, another vector differential operator. It is certainly not equal, in general, to $-\nabla\times\mathbf{V}$.[17] In the case of Eq. (1.69) the determinant must be expanded **from the top down** so that we get the derivatives as shown in the middle portion of Eq. (1.69). If ∇ is crossed into the product of a scalar and a vector, we can show

$$\begin{aligned}\nabla\times(f\mathbf{V})|_x &= \left[\frac{\partial}{\partial y}(fV_z) - \frac{\partial}{\partial z}(fV_y)\right]\\ &= \left(f\frac{\partial V_z}{\partial y} + \frac{\partial f}{\partial y}V_z - f\frac{\partial V_y}{\partial z} - \frac{\partial f}{\partial z}V_y\right)\\ &= f\nabla\times\mathbf{V}|_x + (\nabla f)\times\mathbf{V}|_x.\end{aligned} \tag{1.70}$$

If we permute the coordinates $x\to y, y\to z, z\to x$ to pick up the y-component and then permute them a second time to pick up the z-component, then

$$\nabla\times(f\mathbf{V}) = f\nabla\times\mathbf{V} + (\nabla f)\times\mathbf{V}, \tag{1.71}$$

which is the vector product analog of Eq. (1.67b). Again, as a differential operator ∇ differentiates both f and $\mathbf{V}$. As a vector it is crossed into $\mathbf{V}$ (in each term).

[17] In this same spirit, if $\mathbf{A}$ is a differential operator, it is not necessarily true that $\mathbf{A}\times\mathbf{A} = 0$. Specifically, for the quantum mechanical angular momentum **operator** $\mathbf{L} = -i(\mathbf{r}\times\nabla)$, we find that $\mathbf{L}\times\mathbf{L} = i\mathbf{L}$. See Sections 4.3 and 4.4 for more details.

Example 1.8.1 VECTOR POTENTIAL OF A CONSTANT **B** FIELD

From electrodynamics we know that $\nabla \cdot \mathbf{B} = 0$, which has the general solution $\mathbf{B} = \nabla \times \mathbf{A}$, where $\mathbf{A}(\mathbf{r})$ is called the vector potential (of the magnetic induction), because $\nabla \cdot (\nabla \times \mathbf{A}) = (\nabla \times \nabla) \cdot \mathbf{A} \equiv 0$, as a triple scalar product with two identical vectors. This last identity will not change if we add the gradient of some scalar function to the vector potential, which, therefore, is not unique.

In our case, we want to show that a vector potential is $\mathbf{A} = \frac{1}{2}(\mathbf{B} \times \mathbf{r})$.

Using the *BAC–BAC* rule in conjunction with Example 1.7.1, we find that

$$2\nabla \times \mathbf{A} = \nabla \times (\mathbf{B} \times \mathbf{r}) = (\nabla \cdot \mathbf{r})\mathbf{B} - (\mathbf{B} \cdot \nabla)\mathbf{r} = 3\mathbf{B} - \mathbf{B} = 2\mathbf{B},$$

where we indicate by the ordering of the scalar product of the second term that the gradient still acts on the coordinate vector. ■

Example 1.8.2 CURL OF A CENTRAL FORCE FIELD

Calculate $\nabla \times (\mathbf{r} f(r))$.

By Eq. (1.71),

$$\nabla \times \big(\mathbf{r} f(r)\big) = f(r)\nabla \times \mathbf{r} + \big[\nabla f(r)\big] \times \mathbf{r}. \tag{1.72}$$

First,

$$\nabla \times \mathbf{r} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ x & y & z \end{vmatrix} = 0. \tag{1.73}$$

Second, using $\nabla f(r) = \hat{\mathbf{r}}(df/dr)$ (Example 1.6.1), we obtain

$$\nabla \times \mathbf{r} f(r) = \frac{df}{dr}\hat{\mathbf{r}} \times \mathbf{r} = 0. \tag{1.74}$$

This vector product vanishes, since $\mathbf{r} = \hat{\mathbf{r}}r$ and $\hat{\mathbf{r}} \times \hat{\mathbf{r}} = 0$. ■

To develop a better feeling for the physical significance of the curl, we consider the circulation of fluid around a differential loop in the xy-plane, Fig. 1.24.

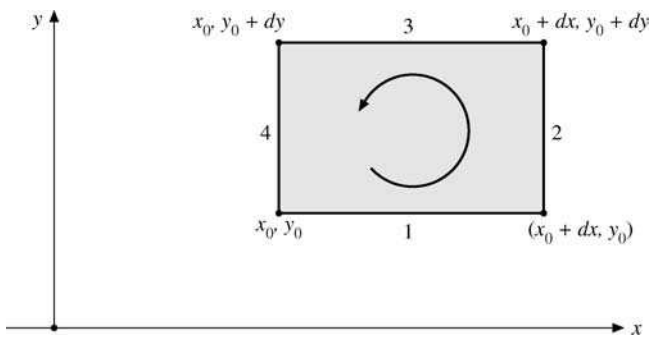

FIGURE 1.24 Circulation around a differential loop.

Although the circulation is technically given by a vector line integral $\int \mathbf{V}\cdot d\boldsymbol{\lambda}$ (Section 1.10), we can set up the equivalent scalar integrals here. Let us take the circulation to be

$$\text{circulation}_{1234} = \int_1 V_x(x,y)\,d\lambda_x + \int_2 V_y(x,y)\,d\lambda_y + \int_3 V_x(x,y)\,d\lambda_x + \int_4 V_y(x,y)\,d\lambda_y. \tag{1.75}$$

The numbers 1, 2, 3, and 4 refer to the numbered line segments in Fig. 1.24. In the first integral, $d\lambda_x = +dx$; but in the third integral, $d\lambda_x = -dx$ because the third line segment is traversed in the negative x-direction. Similarly, $d\lambda_y = +dy$ for the second integral, $-dy$ for the fourth. Next, the integrands are referred to the point (x_0, y_0) with a Taylor expansion[18] taking into account the displacement of line segment 3 from 1 and that of 2 from 4. For our differential line segments this leads to

$$\begin{aligned}\text{circulation}_{1234} &= V_x(x_0,y_0)\,dx + \left[V_y(x_0,y_0) + \frac{\partial V_y}{\partial x}dx\right]dy \\ &\quad + \left[V_x(x_0,y_0) + \frac{\partial V_x}{\partial y}dy\right](-dx) + V_y(x_0,y_0)(-dy) \\ &= \left(\frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y}\right)dx\,dy.\end{aligned} \tag{1.76}$$

Dividing by $dx\,dy$, we have

$$\text{circulation per unit area} = \nabla\times\mathbf{V}|_z. \tag{1.77}$$

The circulation[19] about our differential area in the xy-plane is given by the z-component of $\nabla\times\mathbf{V}$. In principle, the curl $\nabla\times\mathbf{V}$ at (x_0, y_0) could be determined by inserting a (differential) paddle wheel into the moving fluid at point (x_0, y_0). The rotation of the little paddle wheel would be a measure of the curl, and its axis would be along the direction of $\nabla\times\mathbf{V}$, which is perpendicular to the plane of circulation.

We shall use the result, Eq. (1.76), in Section 1.12 to derive Stokes' theorem. Whenever the curl of a vector $\mathbf{V}$ vanishes,

$$\nabla\times\mathbf{V} = 0, \tag{1.78}$$

$\mathbf{V}$ is labeled **irrotational**. The most important physical examples of irrotational vectors are the gravitational and electrostatic forces. In each case

$$\mathbf{V} = C\frac{\hat{\mathbf{r}}}{r^2} = C\frac{\mathbf{r}}{r^3}, \tag{1.79}$$

where C is a constant and $\hat{\mathbf{r}}$ is the unit vector in the outward radial direction. For the gravitational case we have $C = -Gm_1m_2$, given by Newton's law of universal gravitation. If $C = q_1q_2/4\pi\varepsilon_0$, we have Coulomb's law of electrostatics (mks units). The force $\mathbf{V}$

[18]Here, $V_y(x_0+dx, y_0) = V_y(x_0,y_0) + (\frac{\partial V_y}{\partial x})_{x_0y_0}\,dx + \cdots$. The higher-order terms will drop out in the limit as $dx\to 0$. A correction term for the variation of V_y with y is canceled by the corresponding term in the fourth integral.

[19]In fluid dynamics $\nabla\times\mathbf{V}$ is called the "vorticity."

given in Eq. (1.79) may be shown to be irrotational by direct expansion into Cartesian components, as we did in Example 1.8.1. Another approach is developed in Chapter 2, in which we express $\nabla\times$, the curl, in terms of spherical polar coordinates. In Section 1.13 we shall see that whenever a vector is irrotational, the vector may be written as the (negative) gradient of a scalar potential. In Section 1.16 we shall prove that a vector field may be resolved into an irrotational part and a solenoidal part (subject to conditions at infinity). In terms of the electromagnetic field this corresponds to the resolution into an irrotational electric field and a solenoidal magnetic field.

For waves in an elastic medium, if the displacement $\mathbf{u}$ is irrotational, $\nabla\times\mathbf{u}=0$, plane waves (or spherical waves at large distances) become longitudinal. If $\mathbf{u}$ is solenoidal, $\nabla\cdot\mathbf{u}=0$, then the waves become transverse. A seismic disturbance will produce a displacement that may be resolved into a solenoidal part and an irrotational part (compare Section 1.16). The irrotational part yields the longitudinal P (primary) earthquake waves. The solenoidal part gives rise to the slower transverse S (secondary) waves.

Using the gradient, divergence, and curl, and of course the *BAC–CAB* rule, we may construct or verify a large number of useful vector identities. For verification, complete expansion into Cartesian components is always a possibility. Sometimes if we use insight instead of routine shuffling of Cartesian components, the verification process can be shortened drastically.

Remember that ∇ is a vector operator, a hybrid creature satisfying two sets of rules:

1. vector rules, and
2. partial differentiation rules — including differentiation of a product.

Example 1.8.3 GRADIENT OF A DOT PRODUCT

Verify that

$$\nabla(\mathbf{A}\cdot\mathbf{B})=(\mathbf{B}\cdot\nabla)\mathbf{A}+(\mathbf{A}\cdot\nabla)\mathbf{B}+\mathbf{B}\times(\nabla\times\mathbf{A})+\mathbf{A}\times(\nabla\times\mathbf{B}). \tag{1.80}$$

This particular example hinges on the recognition that $\nabla(\mathbf{A}\cdot\mathbf{B})$ is the type of term that appears in the *BAC–CAB* expansion of a triple vector product, Eq. (1.55). For instance,

$$\mathbf{A}\times(\nabla\times\mathbf{B})=\nabla(\mathbf{A}\cdot\mathbf{B})-(\mathbf{A}\cdot\nabla)\mathbf{B},$$

with the ∇ differentiating only $\mathbf{B}$, not $\mathbf{A}$. From the commutativity of factors in a scalar product we may interchange $\mathbf{A}$ and $\mathbf{B}$ and write

$$\mathbf{B}\times(\nabla\times\mathbf{A})=\nabla(\mathbf{A}\cdot\mathbf{B})-(\mathbf{B}\cdot\nabla)\mathbf{A},$$

now with ∇ differentiating only $\mathbf{A}$, not $\mathbf{B}$. Adding these two equations, we obtain ∇ differentiating the product $\mathbf{A}\cdot\mathbf{B}$ and the identity, Eq. (1.80). This identity is used frequently in electromagnetic theory. Exercise 1.8.13 is a simple illustration. ■

Example 1.8.4 INTEGRATION BY PARTS OF CURL

Let us prove the formula $\int \mathbf{C}(\mathbf{r})\cdot(\nabla\times\mathbf{A}(\mathbf{r}))\,d^3r = \int \mathbf{A}(\mathbf{r})\cdot(\nabla\times\mathbf{C}(\mathbf{r}))\,d^3r$, where $\mathbf{A}$ or $\mathbf{C}$ or both vanish at infinity.

To show this, we proceed, as in Examples 1.6.3 and 1.7.3, by integration by parts after writing the inner product and the curl in Cartesian coordinates. Because the integrated terms vanish at infinity we obtain

$$\begin{aligned}
&\int \mathbf{C}(\mathbf{r})\cdot\big(\nabla\times\mathbf{A}(\mathbf{r})\big)\,d^3r\\
&\quad=\int\left[C_z\left(\frac{\partial A_y}{\partial x}-\frac{\partial A_x}{\partial y}\right)+C_x\left(\frac{\partial A_z}{\partial y}-\frac{\partial A_y}{\partial z}\right)+C_y\left(\frac{\partial A_x}{\partial z}-\frac{\partial A_z}{\partial x}\right)\right]d^3r\\
&\quad=\int\left[A_x\left(\frac{\partial C_z}{\partial y}-\frac{\partial C_y}{\partial z}\right)+A_y\left(\frac{\partial C_x}{\partial z}-\frac{\partial C_z}{\partial x}\right)+A_z\left(\frac{\partial C_y}{\partial x}-\frac{\partial C_x}{\partial y}\right)\right]d^3r\\
&\quad=\int \mathbf{A}(\mathbf{r})\cdot\big(\nabla\times\mathbf{C}(\mathbf{r})\big)\,d^3r,
\end{aligned}$$

just rearranging appropriately the terms after integration by parts. ■

Exercises

1.8.1 Show, by rotating the coordinates, that the components of the curl of a vector transform as a vector.
Hint. The direction cosine identities of Eq. (1.46) are available as needed.

1.8.2 Show that $\mathbf{u}\times\mathbf{v}$ is solenoidal if $\mathbf{u}$ and $\mathbf{v}$ are each irrotational.

1.8.3 If $\mathbf{A}$ is irrotational, show that $\mathbf{A}\times\mathbf{r}$ is solenoidal.

1.8.4 A rigid body is rotating with constant angular velocity $\boldsymbol{\omega}$. Show that the linear velocity $\mathbf{v}$ is solenoidal.

1.8.5 If a vector function $\mathbf{f}(x, y, z)$ is not irrotational but the product of f and a scalar function $g(x, y, z)$ is irrotational, show that then

$$\mathbf{f}\cdot\nabla\times\mathbf{f}=0.$$

1.8.6 If (a) $\mathbf{V}=\hat{\mathbf{x}}V_x(x, y)+\hat{\mathbf{y}}V_y(x, y)$ and (b) $\nabla\times\mathbf{V}\neq 0$, prove that $\nabla\times\mathbf{V}$ is perpendicular to $\mathbf{V}$.

1.8.7 Classically, orbital angular momentum is given by $\mathbf{L}=\mathbf{r}\times\mathbf{p}$, where $\mathbf{p}$ is the linear momentum. To go from classical mechanics to quantum mechanics, replace $\mathbf{p}$ by the operator $-i\nabla$ (Section 15.6). Show that the quantum mechanical angular momentum

operator has Cartesian components (in units of $\hbar$)

$$L_x = -i\left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}\right),$$
$$L_y = -i\left(z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z}\right),$$
$$L_z = -i\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right).$$

1.8.8 Using the angular momentum operators previously given, show that they satisfy commutation relations of the form

$$[L_x, L_y] \equiv L_x L_y - L_y L_x = i L_z$$

and hence

$$\mathbf{L} \times \mathbf{L} = i\mathbf{L}.$$

These commutation relations will be taken later as the defining relations of an angular momentum operator — Exercise 3.2.15 and the following one and Chapter 4.

1.8.9 With the commutator bracket notation $[L_x, L_y] = L_x L_y - L_y L_x$, the angular momentum vector $\mathbf{L}$ satisfies $[L_x, L_y] = iL_z$, etc., or $\mathbf{L} \times \mathbf{L} = i\mathbf{L}$.
If two other vectors $\mathbf{a}$ and $\mathbf{b}$ commute with each other and with $\mathbf{L}$, that is, $[\mathbf{a}, \mathbf{b}] = [\mathbf{a}, \mathbf{L}] = [\mathbf{b}, \mathbf{L}] = 0$, show that

$$[\mathbf{a} \cdot \mathbf{L}, \mathbf{b} \cdot \mathbf{L}] = i(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{L}.$$

1.8.10 For $\mathbf{A} = \hat{\mathbf{x}} A_x(x, y, z)$ and $\mathbf{B} = \hat{\mathbf{x}} B_x(x, y, z)$ evaluate each term in the vector identity

$$\nabla(\mathbf{A} \cdot \mathbf{B}) = (\mathbf{B} \cdot \nabla)\mathbf{A} + (\mathbf{A} \cdot \nabla)\mathbf{B} + \mathbf{B} \times (\nabla \times \mathbf{A}) + \mathbf{A} \times (\nabla \times \mathbf{B})$$

and verify that the identity is satisfied.

1.8.11 Verify the vector identity

$$\nabla \times (\mathbf{A} \times \mathbf{B}) = (\mathbf{B} \cdot \nabla)\mathbf{A} - (\mathbf{A} \cdot \nabla)\mathbf{B} - \mathbf{B}(\nabla \cdot \mathbf{A}) + \mathbf{A}(\nabla \cdot \mathbf{B}).$$

1.8.12 As an alternative to the vector identity of Example 1.8.3 show that

$$\nabla(\mathbf{A} \cdot \mathbf{B}) = (\mathbf{A} \times \nabla) \times \mathbf{B} + (\mathbf{B} \times \nabla) \times \mathbf{A} + \mathbf{A}(\nabla \cdot \mathbf{B}) + \mathbf{B}(\nabla \cdot \mathbf{A}).$$

1.8.13 Verify the identity

$$\mathbf{A} \times (\nabla \times \mathbf{A}) = \frac{1}{2}\nabla\left(A^2\right) - (\mathbf{A} \cdot \nabla)\mathbf{A}.$$

1.8.14 If $\mathbf{A}$ and $\mathbf{B}$ are constant vectors, show that

$$\nabla(\mathbf{A} \cdot \mathbf{B} \times \mathbf{r}) = \mathbf{A} \times \mathbf{B}.$$

1.8.15 A distribution of electric currents creates a constant magnetic moment $\mathbf{m} = \text{const}$. The force on $\mathbf{m}$ in an external magnetic induction $\mathbf{B}$ is given by

$$\mathbf{F} = \nabla \times (\mathbf{B} \times \mathbf{m}).$$

Show that

$$\mathbf{F} = (\mathbf{m} \cdot \nabla)\mathbf{B}.$$

Note. Assuming no time dependence of the fields, Maxwell's equations yield $\nabla \times \mathbf{B} = 0$. Also, $\nabla \cdot \mathbf{B} = 0$.

1.8.16 An electric dipole of moment $\mathbf{p}$ is located at the origin. The dipole creates an electric potential at $\mathbf{r}$ given by

$$\psi(\mathbf{r}) = \frac{\mathbf{p} \cdot \mathbf{r}}{4\pi\varepsilon_0 r^3}.$$

Find the electric field, $\mathbf{E} = -\nabla\psi$ at $\mathbf{r}$.

1.8.17 The vector potential $\mathbf{A}$ of a magnetic dipole, dipole moment $\mathbf{m}$, is given by $\mathbf{A}(\mathbf{r}) = (\mu_0/4\pi)(\mathbf{m} \times \mathbf{r}/r^3)$. Show that the magnetic induction $\mathbf{B} = \nabla \times \mathbf{A}$ is given by

$$\mathbf{B} = \frac{\mu_0}{4\pi} \frac{3\hat{\mathbf{r}}(\hat{\mathbf{r}} \cdot \mathbf{m}) - \mathbf{m}}{r^3}.$$

Note. The limiting process leading to point dipoles is discussed in Section 12.1 for electric dipoles, in Section 12.5 for magnetic dipoles.

1.8.18 The velocity of a two-dimensional flow of liquid is given by

$$\mathbf{V} = \hat{\mathbf{x}}u(x, y) - \hat{\mathbf{y}}v(x, y).$$

If the liquid is incompressible and the flow is irrotational, show that

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \text{and} \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

These are the Cauchy–Riemann conditions of Section 6.2.

1.8.19 The evaluation in this section of the four integrals for the circulation omitted Taylor series terms such as $\partial V_x/\partial x$, $\partial V_y/\partial y$ and all second derivatives. Show that $\partial V_x/\partial x$, $\partial V_y/\partial y$ cancel out when the four integrals are added and that the second derivative terms drop out in the limit as $dx \to 0$, $dy \to 0$.
Hint. Calculate the circulation per unit area and then take the limit $dx \to 0$, $dy \to 0$.

1.9 SUCCESSIVE APPLICATIONS OF ∇

We have now defined gradient, divergence, and curl to obtain vector, scalar, and vector quantities, respectively. Letting ∇ operate on each of these quantities, we obtain

(a) $\nabla \cdot \nabla\varphi$ (b) $\nabla \times \nabla\varphi$ (c) $\nabla\nabla \cdot \mathbf{V}$
(d) $\nabla \cdot \nabla \times \mathbf{V}$ (e) $\nabla \times (\nabla \times \mathbf{V})$

all five expressions involving second derivatives and all five appearing in the second-order differential equations of mathematical physics, particularly in electromagnetic theory.

The first expression, $\nabla \cdot \nabla\varphi$, the divergence of the gradient, is named the Laplacian of φ. We have

$$\begin{aligned}\nabla \cdot \nabla\varphi &= \left(\hat{\mathbf{x}}\frac{\partial}{\partial x} + \hat{\mathbf{y}}\frac{\partial}{\partial y} + \hat{\mathbf{z}}\frac{\partial}{\partial z}\right) \cdot \left(\hat{\mathbf{x}}\frac{\partial\varphi}{\partial x} + \hat{\mathbf{y}}\frac{\partial\varphi}{\partial y} + \hat{\mathbf{z}}\frac{\partial\varphi}{\partial z}\right) \\ &= \frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial^2\varphi}{\partial z^2}. \end{aligned} \tag{1.81a}$$

When φ is the electrostatic potential, we have

$$\nabla \cdot \nabla\varphi = 0 \tag{1.81b}$$

at points where the charge density vanishes, which is Laplace's equation of electrostatics. Often the combination $\nabla \cdot \nabla$ is written ∇^2, or Δ in the European literature.

Example 1.9.1 LAPLACIAN OF A POTENTIAL

Calculate $\nabla \cdot \nabla V(r)$.

Referring to Examples 1.6.1 and 1.7.2,

$$\nabla \cdot \nabla V(r) = \nabla \cdot \hat{\mathbf{r}}\frac{dV}{dr} = \frac{2}{r}\frac{dV}{dr} + \frac{d^2V}{dr^2},$$

replacing $f(r)$ in Example 1.7.2 by $1/r \cdot dV/dr$. If $V(r) = r^n$, this reduces to

$$\nabla \cdot \nabla r^n = n(n+1)r^{n-2}.$$

This vanishes for $n = 0$ [$V(r) =$ constant] and for $n = -1$; that is, $V(r) = 1/r$ is a solution of Laplace's equation, $\nabla^2 V(r) = 0$. This is for $r \neq 0$. At $r = 0$, a Dirac delta function is involved (see Eq. (1.169) and Section 9.7). ■

Expression (b) may be written

$$\nabla \times \nabla\varphi = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{\partial\varphi}{\partial x} & \frac{\partial\varphi}{\partial y} & \frac{\partial\varphi}{\partial z} \end{vmatrix}.$$

By expanding the determinant, we obtain

$$\begin{aligned}\nabla \times \nabla\varphi = \hat{\mathbf{x}}&\left(\frac{\partial^2\varphi}{\partial y\,\partial z} - \frac{\partial^2\varphi}{\partial z\,\partial y}\right) + \hat{\mathbf{y}}\left(\frac{\partial^2\varphi}{\partial z\,\partial x} - \frac{\partial^2\varphi}{\partial x\,\partial z}\right) \\ &+ \hat{\mathbf{z}}\left(\frac{\partial^2\varphi}{\partial x\,\partial y} - \frac{\partial^2\varphi}{\partial y\,\partial x}\right) = 0, \end{aligned} \tag{1.82}$$

assuming that the order of partial differentiation may be interchanged. This is true as long as these second partial derivatives of φ are continuous functions. Then, from Eq. (1.82), the curl of a gradient is identically zero. All gradients, therefore, are irrotational. Note that

the zero in Eq. (1.82) comes as a mathematical identity, independent of any physics. The zero in Eq. (1.81b) is a consequence of physics.

Expression (d) is a triple scalar product that may be written

$$\nabla \cdot \nabla \times \mathbf{V} = \begin{vmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ V_x & V_y & V_z \end{vmatrix}. \tag{1.83}$$

Again, assuming continuity so that the order of differentiation is immaterial, we obtain

$$\nabla \cdot \nabla \times \mathbf{V} = 0. \tag{1.84}$$

The divergence of a curl vanishes or all curls are solenoidal. In Section 1.16 we shall see that vectors may be resolved into solenoidal and irrotational parts by Helmholtz's theorem.

The two remaining expressions satisfy a relation

$$\nabla \times (\nabla \times \mathbf{V}) = \nabla\nabla \cdot \mathbf{V} - \nabla \cdot \nabla\mathbf{V}, \tag{1.85}$$

valid in Cartesian coordinates (but not in curved coordinates). This follows immediately from Eq. (1.55), the *BAC–CAB* rule, which we rewrite so that **C** appears at the extreme right of each term. The term $\nabla \cdot \nabla\mathbf{V}$ was not included in our list, but it may be **defined** by Eq. (1.85).

Example 1.9.2 ELECTROMAGNETIC WAVE EQUATION

One important application of this vector relation (Eq. (1.85)) is in the derivation of the electromagnetic wave equation. In vacuum Maxwell's equations become

$$\nabla \cdot \mathbf{B} = 0, \tag{1.86a}$$

$$\nabla \cdot \mathbf{E} = 0, \tag{1.86b}$$

$$\nabla \times \mathbf{B} = \varepsilon_0\mu_0 \frac{\partial \mathbf{E}}{\partial t}, \tag{1.86c}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \tag{1.86d}$$

Here **E** is the electric field, **B** is the magnetic induction, ε_0 is the electric permittivity, and μ_0 is the magnetic permeability (SI units), so $\varepsilon_0\mu_0 = 1/c^2$, c being the velocity of light. The relation has important consequences. Because ε_0, μ_0 can be measured in any frame, the velocity of light is the same in any frame.

Suppose we eliminate **B** from Eqs. (1.86c) and (1.86d). We may do this by taking the curl of both sides of Eq. (1.86d) and the time derivative of both sides of Eq. (1.86c). Since the space and time derivatives commute,

$$\frac{\partial}{\partial t}\nabla \times \mathbf{B} = \nabla \times \frac{\partial \mathbf{B}}{\partial t},$$

and we obtain

$$\nabla \times (\nabla \times \mathbf{E}) = -\varepsilon_0\mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}.$$

Application of Eqs. (1.85) and (1.86b) yields

$$\nabla \cdot \nabla \mathbf{E} = \varepsilon_0 \mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}, \tag{1.87}$$

the electromagnetic vector wave equation. Again, if $\mathbf{E}$ is expressed in Cartesian coordinates, Eq. (1.87) separates into three scalar wave equations, each involving the scalar Laplacian.

When external electric charge and current densities are kept as driving terms in Maxwell's equations, similar wave equations are valid for the electric potential and the vector potential. To show this, we solve Eq. (1.86a) by writing $\mathbf{B} = \nabla \times \mathbf{A}$ as a curl of the vector potential. This expression is substituted into Faraday's induction law in differential form, Eq. (1.86d), to yield $\nabla \times (\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}) = 0$. The vanishing curl implies that $\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}$ is a gradient and, therefore, can be written as $-\nabla\varphi$, where $\varphi(\mathbf{r}, t)$ is defined as the (nonstatic) electric potential. These results for the $\mathbf{B}$ and $\mathbf{E}$ fields,

$$\boxed{\mathbf{B} = \nabla \times \mathbf{A}, \qquad \mathbf{E} = -\nabla\varphi - \frac{\partial \mathbf{A}}{\partial t},} \tag{1.88}$$

solve the homogeneous Maxwell's equations.

We now show that the inhomogeneous Maxwell's equations,

$$\text{Gauss' law: } \boxed{\nabla \cdot \mathbf{E} = \rho/\varepsilon_0,} \qquad \text{Oersted's law: } \boxed{\nabla \times \mathbf{B} - \frac{1}{c^2}\frac{\partial \mathbf{E}}{\partial t} = \mu_0 \mathbf{J}} \tag{1.89}$$

in differential form lead to wave equations for the potentials φ and $\mathbf{A}$, provided that $\nabla \cdot \mathbf{A}$ is determined by the constraint $\frac{1}{c^2}\frac{\partial \varphi}{\partial t} + \nabla \cdot \mathbf{A} = 0$. This choice of fixing the divergence of the vector potential, called the **Lorentz gauge**, serves to uncouple the differential equations of both potentials. This gauge constraint is not a restriction; it has no physical effect.

Substituting our electric field solution into Gauss' law yields

$$\boxed{\frac{\rho}{\varepsilon_0} = \nabla \cdot \mathbf{E} = -\nabla^2\varphi - \frac{\partial}{\partial t}\nabla \cdot \mathbf{A} = -\nabla^2\varphi + \frac{1}{c^2}\frac{\partial^2 \varphi}{\partial t^2},} \tag{1.90}$$

the wave equation for the electric potential. In the last step we have used the Lorentz gauge to replace the divergence of the vector potential by the time derivative of the electric potential and thus decouple φ from $\mathbf{A}$.

Finally, we substitute $\mathbf{B} = \nabla \times \mathbf{A}$ into Oersted's law and use Eq. (1.85), which expands ∇^2 in terms of a longitudinal (the gradient term) and a transverse component (the curl term). This yields

$$\mu_0 \mathbf{J} + \frac{1}{c^2}\frac{\partial \mathbf{E}}{\partial t} = \nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} = \mu_0 \mathbf{J} - \frac{1}{c^2}\left(\nabla \frac{\partial \varphi}{\partial t} + \frac{\partial^2 \mathbf{A}}{\partial t^2}\right),$$

where we have used the electric field solution (Eq. (1.88)) in the last step. Now we see that the Lorentz gauge condition eliminates the gradient terms, so the wave equation

$$\boxed{\frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla^2 \mathbf{A} = \mu_0 \mathbf{J}} \tag{1.91}$$

for the vector potential remains.

Finally, looking back at Oersted's law, taking the divergence of Eq. (1.89), dropping $\nabla \cdot (\nabla \times \mathbf{B}) = 0$, and substituting Gauss' law for $\nabla \cdot \mathbf{E} = \rho/\epsilon_0$, we find $\mu_0 \nabla \cdot \mathbf{J} = -\frac{1}{\epsilon_0 c^2}\frac{\partial \rho}{\partial t}$, where $\epsilon_0 \mu_0 = 1/c^2$, that is, the continuity equation for the current density. This step justifies the inclusion of Maxwell's displacement current in the generalization of Oersted's law to nonstationary situations. ■

Exercises

1.9.1 Verify Eq. (1.85),

$$\nabla \times (\nabla \times \mathbf{V}) = \nabla\nabla \cdot \mathbf{V} - \nabla \cdot \nabla\mathbf{V},$$

by direct expansion in Cartesian coordinates.

1.9.2 Show that the identity

$$\nabla \times (\nabla \times \mathbf{V}) = \nabla\nabla \cdot \mathbf{V} - \nabla \cdot \nabla\mathbf{V}$$

follows from the *BAC–CAB* rule for a triple vector product. Justify any alteration of the order of factors in the *BAC* and *CAB* terms.

1.9.3 Prove that $\nabla \times (\varphi \nabla \varphi) = 0$.

1.9.4 You are given that the curl of **F** equals the curl of **G**. Show that **F** and **G** may differ by (a) a constant and (b) a gradient of a scalar function.

1.9.5 The Navier–Stokes equation of hydrodynamics contains a nonlinear term $(\mathbf{v} \cdot \nabla)\mathbf{v}$. Show that the curl of this term may be written as $-\nabla \times [\mathbf{v} \times (\nabla \times \mathbf{v})]$.

1.9.6 From the Navier–Stokes equation for the steady flow of an incompressible viscous fluid we have the term

$$\nabla \times \big[\mathbf{v} \times (\nabla \times \mathbf{v})\big],$$

where **v** is the fluid velocity. Show that this term vanishes for the special case

$$\mathbf{v} = \hat{\mathbf{x}} v(y, z).$$

1.9.7 Prove that $(\nabla u) \times (\nabla v)$ is solenoidal, where u and v are differentiable scalar functions.

1.9.8 φ is a scalar satisfying Laplace's equation, $\nabla^2 \varphi = 0$. Show that $\nabla\varphi$ is **both** solenoidal and irrotational.

1.9.9 With ψ a scalar (wave) function, show that

$$(\mathbf{r} \times \nabla) \cdot (\mathbf{r} \times \nabla)\psi = r^2 \nabla^2 \psi - r^2 \frac{\partial^2 \psi}{\partial r^2} - 2r \frac{\partial \psi}{\partial r}.$$

(This can actually be shown more easily in spherical polar coordinates, Section 2.5.)

1.9.10 In a (nonrotating) isolated mass such as a star, the condition for equilibrium is

$$\nabla P + \rho \nabla \varphi = 0.$$

Here P is the total pressure, ρ is the density, and φ is the gravitational potential. Show that at any given point the normals to the surfaces of constant pressure and constant gravitational potential are parallel.

1.9.11 In the Pauli theory of the electron, one encounters the expression

$$(\mathbf{p} - e\mathbf{A}) \times (\mathbf{p} - e\mathbf{A})\psi,$$

where ψ is a scalar (wave) function. $\mathbf{A}$ is the magnetic vector potential related to the magnetic induction $\mathbf{B}$ by $\mathbf{B} = \nabla \times \mathbf{A}$. Given that $\mathbf{p} = -i\nabla$, show that this expression reduces to $ie\mathbf{B}\psi$. Show that this leads to the orbital g-factor $g_L = 1$ upon writing the magnetic moment as $\boldsymbol{\mu} = g_L \mathbf{L}$ in units of Bohr magnetons and $\mathbf{L} = -i\mathbf{r} \times \nabla$. See also Exercise 1.13.7.

1.9.12 Show that any solution of the equation

$$\nabla \times (\nabla \times \mathbf{A}) - k^2 \mathbf{A} = 0$$

automatically satisfies the vector Helmholtz equation

$$\nabla^2 \mathbf{A} + k^2 \mathbf{A} = 0$$

and the solenoidal condition

$$\nabla \cdot \mathbf{A} = 0.$$

Hint. Let $\nabla\cdot$ operate on the first equation.

1.9.13 The theory of heat conduction leads to an equation

$$\nabla^2 \Psi = k|\nabla \Phi|^2,$$

where Φ is a potential satisfying Laplace's equation: $\nabla^2 \Phi = 0$. Show that a solution of this equation is

$$\Psi = \frac{1}{2} k \Phi^2.$$

1.10 VECTOR INTEGRATION

The next step after differentiating vectors is to integrate them. Let us start with line integrals and then proceed to surface and volume integrals. In each case the method of attack will be to reduce the vector integral to scalar integrals with which the reader is assumed familiar.

Line Integrals

Using an increment of length $d\mathbf{r} = \hat{\mathbf{x}}\,dx + \hat{\mathbf{y}}\,dy + \hat{\mathbf{z}}\,dz$, we may encounter the line integrals

$$\int_C \varphi\,d\mathbf{r}, \tag{1.92a}$$

$$\int_C \mathbf{V}\cdot d\mathbf{r}, \tag{1.92b}$$

$$\int_C \mathbf{V}\times d\mathbf{r}, \tag{1.92c}$$

in each of which the integral is over some contour C that may be open (with starting point and ending point separated) or closed (forming a loop). Because of its physical interpretation that follows, the second form, Eq. (1.92b) is by far the most important of the three.

With φ, a scalar, the first integral reduces immediately to

$$\int_C \varphi\,d\mathbf{r} = \hat{\mathbf{x}}\int_C \varphi(x,y,z)\,dx + \hat{\mathbf{y}}\int_C \varphi(x,y,z)\,dy + \hat{\mathbf{z}}\int_C \varphi(x,y,z)\,dz. \tag{1.93}$$

This separation has employed the relation

$$\int \hat{\mathbf{x}}\varphi\,dx = \hat{\mathbf{x}}\int \varphi\,dx, \tag{1.94}$$

which is permissible because the Cartesian unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ are constant in both magnitude and direction. Perhaps this relation is obvious here, but it will not be true in the non-Cartesian systems encountered in Chapter 2.

The three integrals on the right side of Eq. (1.93) are ordinary scalar integrals and, to avoid complications, we assume that they are Riemann integrals. Note, however, that the integral with respect to x cannot be evaluated unless y and z are known in terms of x and similarly for the integrals with respect to y and z. This simply means that the path of integration C must be specified. Unless the integrand has special properties so that the integral depends only on the value of the end points, the value will depend on the particular choice of contour C. For instance, if we choose the very special case $\varphi = 1$, Eq. (1.92a) is just the vector distance from the start of contour C to the endpoint, in this case independent of the choice of path connecting fixed endpoints. With $d\mathbf{r} = \hat{\mathbf{x}}\,dx + \hat{\mathbf{y}}\,dy + \hat{\mathbf{z}}\,dz$, the second and third forms also reduce to scalar integrals and, like Eq. (1.92a), are dependent, in general, on the choice of path. The form (Eq. (1.92b)) is exactly the same as that encountered when we calculate the work done by a force that varies along the path,

$$W = \int \mathbf{F}\cdot d\mathbf{r} = \int F_x(x,y,z)\,dx + \int F_y(x,y,z)\,dy + \int F_z(x,y,z)\,dz. \tag{1.95a}$$

In this expression $\mathbf{F}$ is the force exerted on a particle.

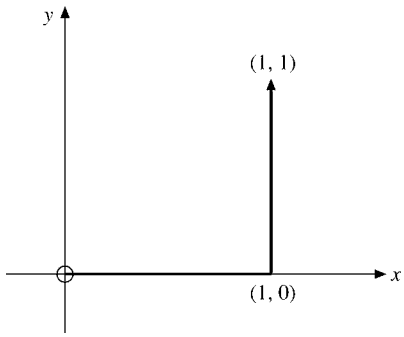

FIGURE 1.25 A path of integration.

Example 1.10.1 PATH-DEPENDENT WORK

The force exerted on a body is $\mathbf{F} = -\hat{\mathbf{x}}y + \hat{\mathbf{y}}x$. The problem is to calculate the work done going from the origin to the point $(1, 1)$:

$$W = \int_{0,0}^{1,1} \mathbf{F} \cdot d\mathbf{r} = \int_{0,0}^{1,1} (-y\,dx + x\,dy). \tag{1.95b}$$

Separating the two integrals, we obtain

$$W = -\int_0^1 y\,dx + \int_0^1 x\,dy. \tag{1.95c}$$

The first integral cannot be evaluated until we specify the values of y as x ranges from 0 to 1. Likewise, the second integral requires x as a function of y. Consider first the path shown in Fig. 1.25. Then

$$W = -\int_0^1 0\,dx + \int_0^1 1\,dy = 1, \tag{1.95d}$$

since $y = 0$ along the first segment of the path and $x = 1$ along the second. If we select the path $[x = 0, 0 \leqslant y \leqslant 1]$ and $[0 \leqslant x \leqslant 1, y = 1]$, then Eq. (1.95c) gives $W = -1$. For this force the work done depends on the choice of path. ■

Surface Integrals

Surface integrals appear in the same forms as line integrals, the element of area also being a vector, $d\boldsymbol{\sigma}$.[20] Often this area element is written $\mathbf{n}dA$, in which $\mathbf{n}$ is a unit (normal) vector to indicate the positive direction.[21] There are two conventions for choosing the positive direction. First, if the surface is a closed surface, we agree to take the outward normal as positive. Second, if the surface is an open surface, the positive normal depends on the direction in which the perimeter of the open surface is traversed. If the right-hand fingers

[20] Recall that in Section 1.4 the area (of a parallelogram) is represented by a cross-product **vector**.

[21] Although $\mathbf{n}$ always has unit length, its direction may well be a function of position.

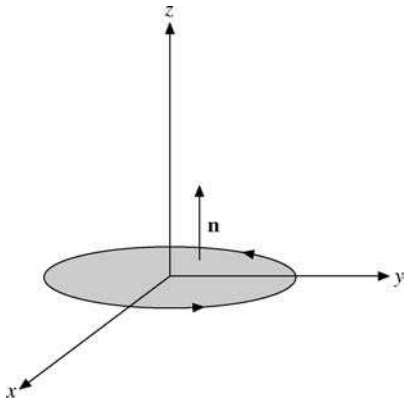

FIGURE 1.26 Right-hand rule for the positive normal.

are placed in the direction of travel around the perimeter, the positive normal is indicated by the thumb of the right hand. As an illustration, a circle in the xy-plane (Fig. 1.26) mapped out from x to y to $-x$ to $-y$ and back to x will have its positive normal parallel to the positive z-axis (for the right-handed coordinate system).

Analogous to the line integrals, Eqs. (1.92a) to (1.92c), surface integrals may appear in the forms

$$\int \varphi \, d\boldsymbol{\sigma}, \qquad \int \mathbf{V} \cdot d\boldsymbol{\sigma}, \qquad \int \mathbf{V} \times d\boldsymbol{\sigma}.$$

Again, the dot product is by far the most commonly encountered form. The surface integral $\int \mathbf{V} \cdot d\boldsymbol{\sigma}$ may be interpreted as a flow or flux through the given surface. This is really what we did in Section 1.7 to obtain the significance of the term divergence. This identification reappears in Section 1.11 as Gauss' theorem. Note that both physically and from the dot product the tangential components of the velocity contribute nothing to the flow through the surface.

Volume Integrals

Volume integrals are somewhat simpler, for the volume element $d\tau$ is a scalar quantity.[22] We have

$$\int_V \mathbf{V} \, d\tau = \hat{\mathbf{x}} \int_V V_x \, d\tau + \hat{\mathbf{y}} \int_V V_y \, d\tau + \hat{\mathbf{z}} \int_V V_z \, d\tau, \tag{1.96}$$

again reducing the vector integral to a vector sum of scalar integrals.

[22] Frequently the symbols d^3r and d^3x are used to denote a volume element in coordinate (xyz or $x_1x_2x_3$) space.

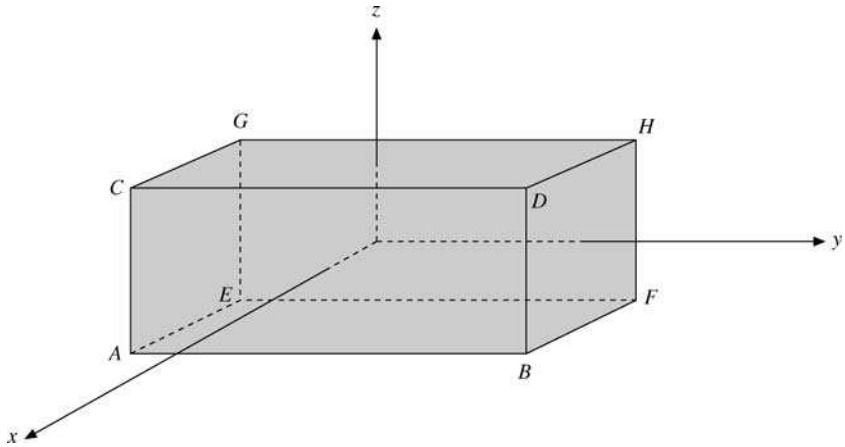

FIGURE 1.27 Differential rectangular parallelepiped (origin at center).

Integral Definitions of Gradient, Divergence, and Curl

One interesting and significant application of our surface and volume integrals is their use in developing alternate definitions of our differential relations. We find

$$\nabla\varphi = \lim_{\int d\tau\to 0} \frac{\int \varphi\, d\boldsymbol{\sigma}}{\int d\tau}, \tag{1.97}$$

$$\nabla\cdot\mathbf{V} = \lim_{\int d\tau\to 0} \frac{\int \mathbf{V}\cdot d\boldsymbol{\sigma}}{\int d\tau}, \tag{1.98}$$

$$\nabla\times\mathbf{V} = \lim_{\int d\tau\to 0} \frac{\int d\boldsymbol{\sigma}\times\mathbf{V}}{\int d\tau}. \tag{1.99}$$

In these three equations $\int d\tau$ is the volume of a small region of space and $d\boldsymbol{\sigma}$ is the vector area element of this volume. The identification of Eq. (1.98) as the divergence of $\mathbf{V}$ was carried out in Section 1.7. Here we show that Eq. (1.97) is consistent with our earlier definition of $\nabla\varphi$ (Eq. (1.60)). For simplicity we choose $d\tau$ to be the differential volume $dx\,dy\,dz$ (Fig. 1.27). This time we place the origin at the geometric center of our volume element. The area integral leads to six integrals, one for each of the six faces. Remembering that $d\boldsymbol{\sigma}$ is outward, $d\boldsymbol{\sigma}\cdot\hat{\mathbf{x}} = -|d\boldsymbol{\sigma}|$ for surface *EFHG*, and $+|d\boldsymbol{\sigma}|$ for surface *ABDC*, we have

$$\begin{aligned}\int \varphi\, d\boldsymbol{\sigma} = &-\hat{\mathbf{x}}\int_{EFHG}\left(\varphi - \frac{\partial\varphi}{\partial x}\frac{dx}{2}\right)dy\,dz + \hat{\mathbf{x}}\int_{ABDC}\left(\varphi + \frac{\partial\varphi}{\partial x}\frac{dx}{2}\right)dy\,dz\\ &-\hat{\mathbf{y}}\int_{AEGC}\left(\varphi - \frac{\partial\varphi}{\partial y}\frac{dy}{2}\right)dx\,dz + \hat{\mathbf{y}}\int_{BFHD}\left(\varphi + \frac{\partial\varphi}{\partial y}\frac{dy}{2}\right)dx\,dz\\ &-\hat{\mathbf{z}}\int_{ABFE}\left(\varphi - \frac{\partial\varphi}{\partial z}\frac{dz}{2}\right)dx\,dy + \hat{\mathbf{z}}\int_{CDHG}\left(\varphi + \frac{\partial\varphi}{\partial z}\frac{dz}{2}\right)dx\,dy.\end{aligned}$$

Using the total variations, we evaluate each integrand at the origin with a correction included to correct for the displacement ($\pm dx/2$, etc.) of the center of the face from the origin. Having chosen the total volume to be of differential size ($\int d\tau = dx\,dy\,dz$), we drop the integral signs on the right and obtain

$$\int \varphi\, d\boldsymbol{\sigma} = \left(\hat{\mathbf{x}}\frac{\partial\varphi}{\partial x} + \hat{\mathbf{y}}\frac{\partial\varphi}{\partial y} + \hat{\mathbf{z}}\frac{\partial\varphi}{\partial z}\right) dx\,dy\,dz. \tag{1.100}$$

Dividing by

$$\int d\tau = dx\,dy\,dz,$$

we verify Eq. (1.97).

This verification has been oversimplified in ignoring other correction terms beyond the first derivatives. These additional terms, which are introduced in Section 5.6 when the Taylor expansion is developed, vanish in the limit

$$\int d\tau \to 0 \ (dx \to 0, dy \to 0, dz \to 0).$$

This, of course, is the reason for specifying in Eqs. (1.97), (1.98), and (1.99) that this limit be taken. Verification of Eq. (1.99) follows these same lines exactly, using a differential volume $dx\,dy\,dz$.

Exercises

1.10.1 The force field acting on a two-dimensional linear oscillator may be described by

$$\mathbf{F} = -\hat{\mathbf{x}}kx - \hat{\mathbf{y}}ky.$$

Compare the work done moving against this force field when going from $(1, 1)$ to $(4, 4)$ by the following straight-line paths:

(a) $(1, 1) \to (4, 1) \to (4, 4)$
(b) $(1, 1) \to (1, 4) \to (4, 4)$
(c) $(1, 1) \to (4, 4)$ along $x = y$.
This means evaluating

$$-\int_{(1,1)}^{(4,4)} \mathbf{F}\cdot d\mathbf{r}$$

along each path.

1.10.2 Find the work done going around a unit circle in the xy-plane:

(a) counterclockwise from 0 to π,
(b) clockwise from 0 to $-\pi$, doing work **against** a force field given by

$$\mathbf{F} = \frac{-\hat{\mathbf{x}}y}{x^2+y^2} + \frac{\hat{\mathbf{y}}x}{x^2+y^2}.$$

Note that the work done depends on the path.

1.10.3 Calculate the work you do in going from point (1, 1) to point (3, 3). The force **you exert** is given by

$$\mathbf{F} = \hat{\mathbf{x}}(x - y) + \hat{\mathbf{y}}(x + y).$$

Specify clearly the path you choose. Note that this force field is nonconservative.

1.10.4 Evaluate $\oint \mathbf{r} \cdot d\mathbf{r}$.
Note. The symbol $\oint$ means that the path of integration is a closed loop.

1.10.5 Evaluate

$$\frac{1}{3}\int_s \mathbf{r} \cdot d\boldsymbol{\sigma}$$

over the unit cube defined by the point $(0, 0, 0)$ and the unit intercepts on the positive x-, y-, and z-axes. Note that (a) $\mathbf{r} \cdot d\boldsymbol{\sigma}$ is zero for three of the surfaces and (b) each of the three remaining surfaces contributes the same amount to the integral.

1.10.6 Show, by expansion of the surface integral, that

$$\lim_{\int d\tau \to 0} \frac{\int_s d\boldsymbol{\sigma} \times \mathbf{V}}{\int d\tau} = \nabla \times \mathbf{V}.$$

Hint. Choose the volume $\int d\tau$ to be a differential volume $dx\,dy\,dz$.

1.11 GAUSS' THEOREM

Here we derive a useful relation between a surface integral of a vector and the volume integral of the divergence of that vector. Let us assume that the vector $\mathbf{V}$ and its first derivatives are continuous over the simply connected region (that does not have any holes, such as a donut) of interest. Then Gauss' theorem states that

$$\boxed{\oiint_{\partial V} \mathbf{V} \cdot d\boldsymbol{\sigma} = \iiint_V \nabla \cdot \mathbf{V}\, d\tau.} \tag{1.101a}$$

In words, the surface integral of a vector over a closed surface equals the volume integral of the divergence of that vector integrated over the volume enclosed by the surface.

Imagine that volume V is subdivided into an arbitrarily large number of tiny (differential) parallelepipeds. For each parallelepiped

$$\sum_{\text{six surfaces}} \mathbf{V} \cdot d\boldsymbol{\sigma} = \nabla \cdot \mathbf{V}\, d\tau \tag{1.101b}$$

from the analysis of Section 1.7, Eq. (1.66), with $\rho\mathbf{v}$ replaced by $\mathbf{V}$. The summation is over the six faces of the parallelepiped. Summing over all parallelepipeds, we find that the $\mathbf{V} \cdot d\boldsymbol{\sigma}$ terms cancel (pairwise) for all **interior** faces; only the contributions of the **exterior** surfaces survive (Fig. 1.28). Analogous to the definition of a Riemann integral as the limit

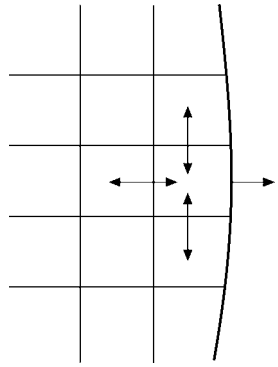

FIGURE 1.28 Exact cancellation of $d\boldsymbol{\sigma}$'s on **interior** surfaces. No cancellation on the **exterior** surface.

of a sum, we take the limit as the number of parallelepipeds approaches infinity ($\to \infty$) and the dimensions of each approach zero ($\to 0$):

$$\begin{array}{ccc} \sum\limits_{\text{exterior surfaces}} \mathbf{V}\cdot d\boldsymbol{\sigma} & = & \sum\limits_{\text{volumes}} \boldsymbol{\nabla}\cdot\mathbf{V}\, d\tau \\ \downarrow & & \downarrow \\ \int_S \mathbf{V}\cdot d\boldsymbol{\sigma} & = & \int_V \boldsymbol{\nabla}\cdot\mathbf{V}\, d\tau. \end{array}$$

The result is Eq. (1.101a), Gauss' theorem.

From a physical point of view Eq. (1.66) has established $\boldsymbol{\nabla}\cdot\mathbf{V}$ as the net outflow of fluid per unit volume. The volume integral then gives the total net outflow. But the surface integral $\int \mathbf{V}\cdot d\boldsymbol{\sigma}$ is just another way of expressing this same quantity, which is the equality, Gauss' theorem.

Green's Theorem

A frequently useful corollary of Gauss' theorem is a relation known as Green's theorem. If u and v are two scalar functions, we have the identities

$$\boldsymbol{\nabla}\cdot(u\,\boldsymbol{\nabla}v) = u\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}v + (\boldsymbol{\nabla}u)\cdot(\boldsymbol{\nabla}v), \tag{1.102}$$

$$\boldsymbol{\nabla}\cdot(v\,\boldsymbol{\nabla}u) = v\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}u + (\boldsymbol{\nabla}v)\cdot(\boldsymbol{\nabla}u). \tag{1.103}$$

Subtracting Eq. (1.103) from Eq. (1.102), integrating over a volume (u, v, and their derivatives, assumed continuous), and applying Eq. (1.101a) (Gauss' theorem), we obtain

$$\boxed{\iiint_V (u\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}v - v\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}u)\,d\tau = \oiint_{\partial V} (u\boldsymbol{\nabla}v - v\boldsymbol{\nabla}u)\cdot d\boldsymbol{\sigma}.} \tag{1.104}$$

This is Green's theorem. We use it for developing Green's functions in Chapter 9. An alternate form of Green's theorem, derived from Eq. (1.102) alone, is

$$\oiint_{\partial V} u\nabla v \cdot d\boldsymbol{\sigma} = \iiint_V u\nabla \cdot \nabla v \, d\tau + \iiint_V \nabla u \cdot \nabla v \, d\tau. \tag{1.105}$$

This is the form of Green's theorem used in Section 1.16.

Alternate Forms of Gauss' Theorem

Although Eq. (1.101a) involving the divergence is by far the most important form of Gauss' theorem, volume integrals involving the gradient and the curl may also appear. Suppose

$$\mathbf{V}(x, y, z) = V(x, y, z)\mathbf{a}, \tag{1.106}$$

in which $\mathbf{a}$ is a vector with constant magnitude and constant but arbitrary direction. (You pick the direction, but once you have chosen it, hold it fixed.) Equation (1.101a) becomes

$$\mathbf{a} \cdot \oiint_{\partial V} V \, d\boldsymbol{\sigma} = \iiint_V \nabla \cdot \mathbf{a} V \, d\tau = \mathbf{a} \cdot \iiint_V \nabla V \, d\tau \tag{1.107}$$

by Eq. (1.67b). This may be rewritten

$$\mathbf{a} \cdot \left[\oiint_{\partial V} V \, d\boldsymbol{\sigma} - \iiint_V \nabla V \, d\tau \right] = 0. \tag{1.108}$$

Since $|\mathbf{a}| \neq 0$ and its direction is arbitrary, meaning that the cosine of the included angle cannot **always** vanish, the terms in brackets must be zero.[23] The result is

$$\oiint_{\partial V} V \, d\boldsymbol{\sigma} = \iiint_V \nabla V \, d\tau. \tag{1.109}$$

In a similar manner, using $\mathbf{V} = \mathbf{a} \times \mathbf{P}$ in which $\mathbf{a}$ is a constant vector, we may show

$$\oiint_{\partial V} d\boldsymbol{\sigma} \times \mathbf{P} = \iiint_V \nabla \times \mathbf{P} \, d\tau. \tag{1.110}$$

These last two forms of Gauss' theorem are used in the vector form of Kirchoff diffraction theory. They may also be used to verify Eqs. (1.97) and (1.99). Gauss' theorem may also be extended to tensors (see Section 2.11).

Exercises

1.11.1 Using Gauss' theorem, prove that

$$\oiint_S d\boldsymbol{\sigma} = 0$$

if $S = \partial V$ is a closed surface.

[23] This exploitation of the **arbitrary** nature of a part of a problem is a valuable and widely used technique. The arbitrary vector is used again in Sections 1.12 and 1.13. Other examples appear in Section 1.14 (integrands equated) and in Section 2.8, quotient rule.

1.11.2 Show that

$$\frac{1}{3}\oiint_S \mathbf{r}\cdot d\boldsymbol{\sigma} = V,$$

where V is the volume enclosed by the closed surface $S = \partial V$.
Note. This is a generalization of Exercise 1.10.5.

1.11.3 If $\mathbf{B} = \nabla \times \mathbf{A}$, show that

$$\oiint_S \mathbf{B}\cdot d\boldsymbol{\sigma} = \mathbf{0}$$

for any closed surface S.

1.11.4 Over some volume V let ψ be a solution of Laplace's equation (with the derivatives appearing there continuous). Prove that the integral over any closed surface in V of the normal derivative of ψ ($\partial\psi/\partial n$, or $\nabla\psi\cdot\mathbf{n}$) will be zero.

1.11.5 In analogy to the integral definition of gradient, divergence, and curl of Section 1.10, show that

$$\nabla^2\varphi = \lim_{\int d\tau\to 0} \frac{\int \nabla\varphi\cdot d\boldsymbol{\sigma}}{\int d\tau}.$$

1.11.6 The electric displacement vector $\mathbf{D}$ satisfies the Maxwell equation $\nabla\cdot\mathbf{D} = \rho$, where ρ is the charge density (per unit volume). At the boundary between two media there is a surface charge density σ (per unit area). Show that a boundary condition for $\mathbf{D}$ is

$$(\mathbf{D}_2 - \mathbf{D}_1)\cdot\mathbf{n} = \sigma.$$

$\mathbf{n}$ is a unit vector normal to the surface and out of medium 1.
Hint. Consider a thin pillbox as shown in Fig. 1.29.

1.11.7 From Eq. (1.67b), with $\mathbf{V}$ the electric field $\mathbf{E}$ and f the electrostatic potential φ, show that, for integration over all space,

$$\int \rho\varphi\, d\tau = \varepsilon_0 \int E^2\, d\tau.$$

This corresponds to a three-dimensional integration by parts.
Hint. $\mathbf{E} = -\nabla\varphi$, $\nabla\cdot\mathbf{E} = \rho/\varepsilon_0$. You may assume that φ vanishes at large r at least as fast as r^{-1}.

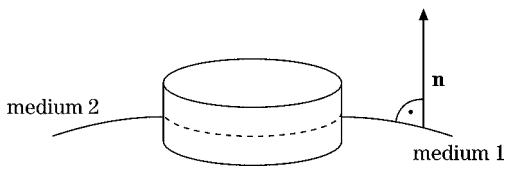

FIGURE 1.29 Pillbox.

1.11.8 A particular steady-state electric current distribution is localized in space. Choosing a bounding surface far enough out so that the current density $\mathbf{J}$ is zero everywhere on the surface, show that

$$\iiint \mathbf{J}\, d\tau = 0.$$

Hint. Take one component of $\mathbf{J}$ at a time. With $\nabla \cdot \mathbf{J} = 0$, show that $\mathbf{J}_i = \nabla \cdot (x_i \mathbf{J})$ and apply Gauss' theorem.

1.11.9 The creation of a **localized** system of steady electric currents (current density $\mathbf{J}$) and magnetic fields may be shown to require an amount of work

$$W = \frac{1}{2} \iiint \mathbf{H} \cdot \mathbf{B}\, d\tau.$$

Transform this into

$$W = \frac{1}{2} \iiint \mathbf{J} \cdot \mathbf{A}\, d\tau.$$

Here $\mathbf{A}$ is the magnetic vector potential: $\nabla \times \mathbf{A} = \mathbf{B}$.
Hint. In Maxwell's equations take the displacement current term $\partial \mathbf{D}/\partial t = 0$. If the fields and currents are localized, a bounding surface may be taken far enough out so that the integrals of the fields and currents over the surface yield zero.

1.11.10 Prove the generalization of Green's theorem:

$$\iiint_V (v\mathcal{L}u - u\mathcal{L}v)\, d\tau = \oiint_{\partial V} p(v\nabla u - u\nabla v) \cdot d\boldsymbol{\sigma}.$$

Here $\mathcal{L}$ is the self-adjoint operator (Section 10.1),

$$\mathcal{L} = \nabla \cdot \big[p(\mathbf{r})\nabla\big] + q(\mathbf{r})$$

and p, q, u, and v are functions of position, p and q having continuous first derivatives and u and v having continuous second derivatives.
Note. This generalized Green's theorem appears in Section 9.7.

1.12 STOKES' THEOREM

Gauss' theorem relates the volume integral of a derivative of a function to an integral of the function over the closed surface bounding the volume. Here we consider an analogous relation between the surface integral of a derivative of a function and the line integral of the function, the path of integration being the perimeter bounding the surface.

Let us take the surface and subdivide it into a network of arbitrarily small rectangles. In Section 1.8 we showed that the circulation about such a differential rectangle (in the xy-plane) is $\nabla \times \mathbf{V}|_z\, dx\, dy$. From Eq. (1.76) applied to **one** differential rectangle,

$$\sum_{\text{four sides}} \mathbf{V} \cdot d\boldsymbol{\lambda} = \nabla \times \mathbf{V} \cdot d\boldsymbol{\sigma}. \tag{1.111}$$

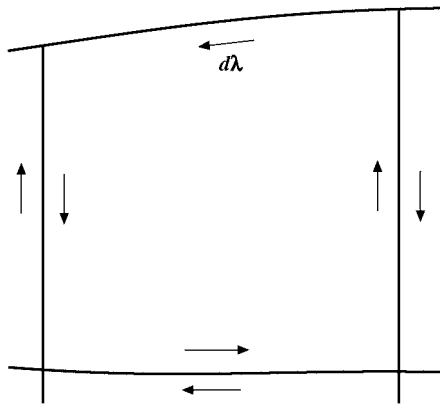

FIGURE 1.30 Exact cancellation on **interior** paths. No cancellation on the **exterior** path.

We sum over all the little rectangles, as in the definition of a Riemann integral. The surface contributions (right-hand side of Eq. (1.111)) are added together. The line integrals (left-hand side of Eq. (1.111)) of all **interior** line segments cancel identically. Only the line integral around the perimeter survives (Fig. 1.30). Taking the usual limit as the number of rectangles approaches infinity while $dx \to 0, dy \to 0$, we have

$$\sum_{\substack{\text{exterior line}\\ \text{segments}}} \mathbf{V}\cdot d\boldsymbol{\lambda} = \sum_{\text{rectangles}} \nabla\times\mathbf{V}\cdot d\boldsymbol{\sigma} \tag{1.112}$$

$$\boxed{\oint \mathbf{V}\cdot d\boldsymbol{\lambda} = \int_S \nabla\times\mathbf{V}\cdot d\boldsymbol{\sigma}.}$$

This is Stokes' theorem. The surface integral on the right is over the surface bounded by the perimeter or contour, for the line integral on the left. The direction of the vector representing the area is out of the paper plane toward the reader if the direction of traversal around the contour for the line integral is in the positive mathematical sense, as shown in Fig. 1.30.

This demonstration of Stokes' theorem is limited by the fact that we used a Maclaurin expansion of $\mathbf{V}(x, y, z)$ in establishing Eq. (1.76) in Section 1.8. Actually we need only demand that the curl of $\mathbf{V}(x, y, z)$ exist and that it be integrable over the surface. A proof of the Cauchy integral theorem analogous to the development of Stokes' theorem here but using these less restrictive conditions appears in Section 6.3.

Stokes' theorem obviously applies to an open surface. It is possible to consider a closed surface as a limiting case of an open surface, with the opening (and therefore the perimeter) shrinking to zero. This is the point of Exercise 1.12.7.

Alternate Forms of Stokes' Theorem

As with Gauss' theorem, other relations between surface and line integrals are possible. We find

$$\int_S d\boldsymbol{\sigma} \times \nabla\varphi = \oint_{\partial S} \varphi \, d\boldsymbol{\lambda} \tag{1.113}$$

and

$$\int_S (d\boldsymbol{\sigma} \times \nabla) \times \mathbf{P} = \oint_{\partial S} d\boldsymbol{\lambda} \times \mathbf{P}. \tag{1.114}$$

Equation (1.113) may readily be verified by the substitution $\mathbf{V} = \mathbf{a}\varphi$, in which $\mathbf{a}$ is a vector of constant magnitude and of constant direction, as in Section 1.11. Substituting into Stokes' theorem, Eq. (1.112),

$$\begin{aligned} \int_S (\nabla \times \mathbf{a}\varphi) \cdot d\boldsymbol{\sigma} &= -\int_S \mathbf{a} \times \nabla\varphi \cdot d\boldsymbol{\sigma} \\ &= -\mathbf{a} \cdot \int_S \nabla\varphi \times d\boldsymbol{\sigma}. \end{aligned} \tag{1.115}$$

For the line integral,

$$\oint_{\partial S} \mathbf{a}\varphi \cdot d\boldsymbol{\lambda} = \mathbf{a} \cdot \oint_{\partial S} \varphi \, d\boldsymbol{\lambda}, \tag{1.116}$$

and we obtain

$$\mathbf{a} \cdot \left(\oint_{\partial S} \varphi \, d\boldsymbol{\lambda} + \int_S \nabla\varphi \times d\boldsymbol{\sigma} \right) = 0. \tag{1.117}$$

Since the choice of direction of $\mathbf{a}$ is arbitrary, the expression in parentheses must vanish, thus verifying Eq. (1.113). Equation (1.114) may be derived similarly by using $\mathbf{V} = \mathbf{a} \times \mathbf{P}$, in which $\mathbf{a}$ is again a constant vector.

We can use Stokes' theorem to derive Oersted's and Faraday's laws from two of Maxwell's equations, and vice versa, thus recognizing that the former are an integrated form of the latter.

Example 1.12.1 OERSTED'S AND FARADAY'S LAWS

Consider the magnetic field generated by a long wire that carries a stationary current I. Starting from Maxwell's differential law $\nabla \times \mathbf{H} = \mathbf{J}$, Eq. (1.89) (with Maxwell's displacement current $\partial\mathbf{D}/\partial t = 0$ for a stationary current case by Ohm's law), we integrate over a closed area S perpendicular to and surrounding the wire and apply Stokes' theorem to get

$$I = \int_S \mathbf{J} \cdot d\boldsymbol{\sigma} = \int_S (\nabla \times \mathbf{H}) \cdot d\boldsymbol{\sigma} = \oint_{\partial S} \mathbf{H} \cdot d\mathbf{r},$$

which is Oersted's law. Here the line integral is along ∂S, the closed curve surrounding the cross-sectional area S.

Similarly, we can integrate Maxwell's equation for $\nabla \times \mathbf{E}$, Eq. (1.86d), to yield Faraday's induction law. Imagine moving a closed loop (∂S) of wire (of area S) across a magnetic induction field $\mathbf{B}$. We integrate Maxwell's equation and use Stokes' theorem, yielding

$$\int_{\partial S} \mathbf{E} \cdot d\mathbf{r} = \int_S (\nabla \times \mathbf{E}) \cdot d\boldsymbol{\sigma} = -\frac{d}{dt} \int_S \mathbf{B} \cdot d\boldsymbol{\sigma} = -\frac{d\Phi}{dt},$$

which is Faraday's law. The line integral on the left-hand side represents the voltage induced in the wire loop, while the right-hand side is the change with time of the magnetic flux Φ through the moving surface S of the wire. ■

Both Stokes' and Gauss' theorems are of tremendous importance in a wide variety of problems involving vector calculus. Some idea of their power and versatility may be obtained from the exercises of Sections 1.11 and 1.12 and the development of potential theory in Sections 1.13 and 1.14.

Exercises

1.12.1 Given a vector $\mathbf{t} = -\hat{\mathbf{x}}y + \hat{\mathbf{y}}x$, show, with the help of Stokes' theorem, that the integral around a continuous closed curve in the xy-plane

$$\frac{1}{2} \oint \mathbf{t} \cdot d\boldsymbol{\lambda} = \frac{1}{2} \oint (x\,dy - y\,dx) = A,$$

the area enclosed by the curve.

1.12.2 The calculation of the magnetic moment of a current loop leads to the line integral

$$\oint \mathbf{r} \times d\mathbf{r}.$$

(a) Integrate around the perimeter of a current loop (in the xy-plane) and show that the scalar magnitude of this line integral is twice the area of the enclosed surface.
(b) The perimeter of an ellipse is described by $\mathbf{r} = \hat{\mathbf{x}}a\cos\theta + \hat{\mathbf{y}}b\sin\theta$. From part (a) show that the area of the ellipse is πab.

1.12.3 Evaluate $\oint \mathbf{r} \times d\mathbf{r}$ by using the alternate form of Stokes' theorem given by Eq. (1.114):

$$\int_S (d\boldsymbol{\sigma} \times \nabla) \times \mathbf{P} = \oint d\boldsymbol{\lambda} \times \mathbf{P}.$$

Take the loop to be entirely in the xy-plane.

1.12.4 In steady state the magnetic field $\mathbf{H}$ satisfies the Maxwell equation $\nabla \times \mathbf{H} = \mathbf{J}$, where $\mathbf{J}$ is the current density (per square meter). At the boundary between two media there is a surface current density $\mathbf{K}$. Show that a boundary condition on $\mathbf{H}$ is

$$\mathbf{n} \times (\mathbf{H}_2 - \mathbf{H}_1) = \mathbf{K}.$$

$\mathbf{n}$ is a unit vector normal to the surface and out of medium 1.
Hint. Consider a narrow loop perpendicular to the interface as shown in Fig. 1.31.

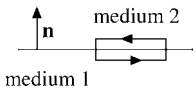

FIGURE 1.31
Integration path at the boundary of two media.

1.12.5 From Maxwell's equations, $\nabla \times \mathbf{H} = \mathbf{J}$, with $\mathbf{J}$ here the current density and $\mathbf{E} = 0$. Show from this that

$$\oint \mathbf{H} \cdot d\mathbf{r} = I,$$

where I is the net electric current enclosed by the loop integral. These are the differential and integral forms of Ampère's law of magnetism.

1.12.6 A magnetic induction $\mathbf{B}$ is generated by electric current in a ring of radius R. Show that the **magnitude** of the vector potential $\mathbf{A}$ ($\mathbf{B} = \nabla \times \mathbf{A}$) at the ring can be

$$|\mathbf{A}| = \frac{\varphi}{2\pi R},$$

where φ is the total magnetic flux passing through the ring.
Note. $\mathbf{A}$ is tangential to the ring and may be changed by adding the gradient of a scalar function.

1.12.7 Prove that

$$\int_S \nabla \times \mathbf{V} \cdot d\boldsymbol{\sigma} = 0,$$

if S is a closed surface.

1.12.8 Evaluate $\oint \mathbf{r} \cdot d\mathbf{r}$ (Exercise 1.10.4) by Stokes' theorem.

1.12.9 Prove that

$$\oint u \nabla v \cdot d\boldsymbol{\lambda} = -\oint v \nabla u \cdot d\boldsymbol{\lambda}.$$

1.12.10 Prove that

$$\oint u \nabla v \cdot d\boldsymbol{\lambda} = \int_S (\nabla u) \times (\nabla v) \cdot d\boldsymbol{\sigma}.$$

1.13 POTENTIAL THEORY

Scalar Potential

If a force over a given simply connected region of space S (which means that it has no holes) can be expressed as the negative gradient of a scalar function φ,

$$\mathbf{F} = -\nabla \varphi, \tag{1.118}$$

we call φ a scalar potential that describes the force by one function instead of three. A scalar potential is only determined up to an additive constant, which can be used to adjust its value at infinity (usually zero) or at some other point. The force $\mathbf{F}$ appearing as the negative gradient of a single-valued scalar potential is labeled a **conservative** force. We want to know when a scalar potential function exists. To answer this question we establish two other relations as equivalent to Eq. (1.118). These are

$$\nabla \times \mathbf{F} = 0 \tag{1.119}$$

and

$$\oint \mathbf{F} \cdot d\mathbf{r} = 0, \tag{1.120}$$

for every closed path in our simply connected region S. We proceed to show that each of these three equations implies the other two. Let us start with

$$\mathbf{F} = -\nabla\varphi. \tag{1.121}$$

Then

$$\nabla \times \mathbf{F} = -\nabla \times \nabla\varphi = 0 \tag{1.122}$$

by Eq. (1.82) or Eq. (1.118) implies Eq. (1.119). Turning to the line integral, we have

$$\oint \mathbf{F} \cdot d\mathbf{r} = -\oint \nabla\varphi \cdot d\mathbf{r} = -\oint d\varphi, \tag{1.123}$$

using Eq. (1.118). Now, $d\varphi$ integrates to give φ. Since we have specified a closed loop, the end points coincide and we get zero for every closed path in our region S for which Eq. (1.118) holds. It is important to note the restriction here that the potential be single-valued and that Eq. (1.118) hold for **all** points in S. This problem may arise in using a scalar magnetic potential, a perfectly valid procedure as long as no net current is encircled. As soon as we choose a path in space that encircles a net current, the scalar magnetic potential ceases to be single-valued and our analysis no longer applies.

Continuing this demonstration of equivalence, let us assume that Eq. (1.120) holds. If $\oint \mathbf{F} \cdot d\mathbf{r} = 0$ for all paths in S, we see that the value of the integral joining two distinct points A and B is independent of the path (Fig. 1.32). Our premise is that

$$\oint_{ACBDA} \mathbf{F} \cdot d\mathbf{r} = 0. \tag{1.124}$$

Therefore

$$\int_{ACB} \mathbf{F} \cdot d\mathbf{r} = -\int_{BDA} \mathbf{F} \cdot d\mathbf{r} = \int_{ADB} \mathbf{F} \cdot d\mathbf{r}, \tag{1.125}$$

reversing the sign by reversing the direction of integration. Physically, this means that the work done in going from A to B is independent of the path and that the work done in going around a closed path is zero. This is the reason for labeling such a force conservative: Energy is conserved.

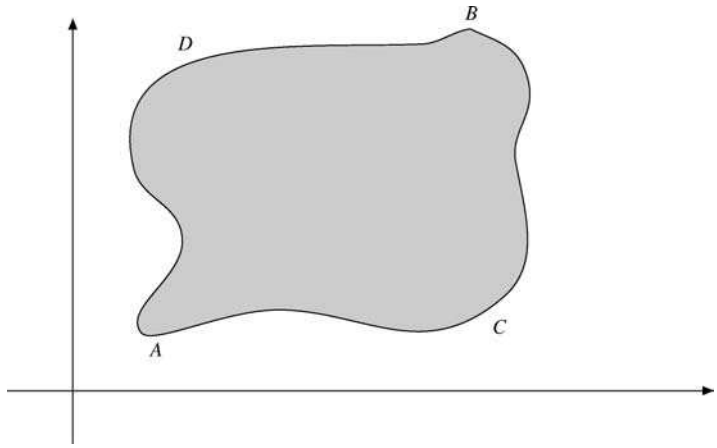

FIGURE 1.32 Possible paths for doing work.

With the result shown in Eq. (1.125), we have the work done dependent only on the endpoints A and B. That is,

$$\text{work done by force} = \int_A^B \mathbf{F} \cdot d\mathbf{r} = \varphi(A) - \varphi(B). \tag{1.126}$$

Equation (1.126) defines a scalar potential (strictly speaking, the difference in potential between points A and B) and provides a means of calculating the potential. If point B is taken as a variable, say, (x, y, z), then differentiation with respect to x, y, and z will recover Eq. (1.118).

The choice of sign on the right-hand side is arbitrary. The choice here is made to achieve agreement with Eq. (1.118) and to ensure that water will run downhill rather than uphill. For points A and B separated by a length $d\mathbf{r}$, Eq. (1.126) becomes

$$\mathbf{F} \cdot d\mathbf{r} = -d\varphi = -\nabla\varphi \cdot d\mathbf{r}. \tag{1.127}$$

This may be rewritten

$$(\mathbf{F} + \nabla\varphi) \cdot d\mathbf{r} = 0, \tag{1.128}$$

and since $d\mathbf{r}$ is arbitrary, Eq. (1.118) must follow. If

$$\oint \mathbf{F} \cdot d\mathbf{r} = 0, \tag{1.129}$$

we may obtain Eq. (1.119) by using Stokes' theorem (Eq. (1.112)):

$$\oint \mathbf{F} \cdot d\mathbf{r} = \int \nabla \times \mathbf{F} \cdot d\boldsymbol{\sigma}. \tag{1.130}$$

If we take the path of integration to be the perimeter of an arbitrary differential area $d\boldsymbol{\sigma}$, the integrand in the surface integral must vanish. Hence Eq. (1.120) implies Eq. (1.119).

Finally, if $\nabla \times \mathbf{F} = 0$, we need only reverse our statement of Stokes' theorem (Eq. (1.130)) to derive Eq. (1.120). Then, by Eqs. (1.126) to (1.128), the initial statement

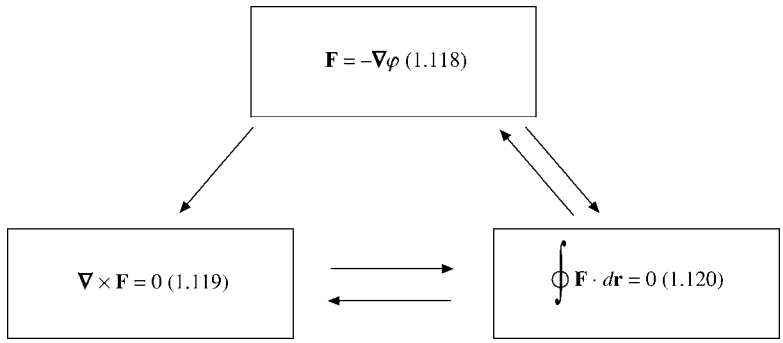

FIGURE 1.33 Equivalent formulations of a conservative force.

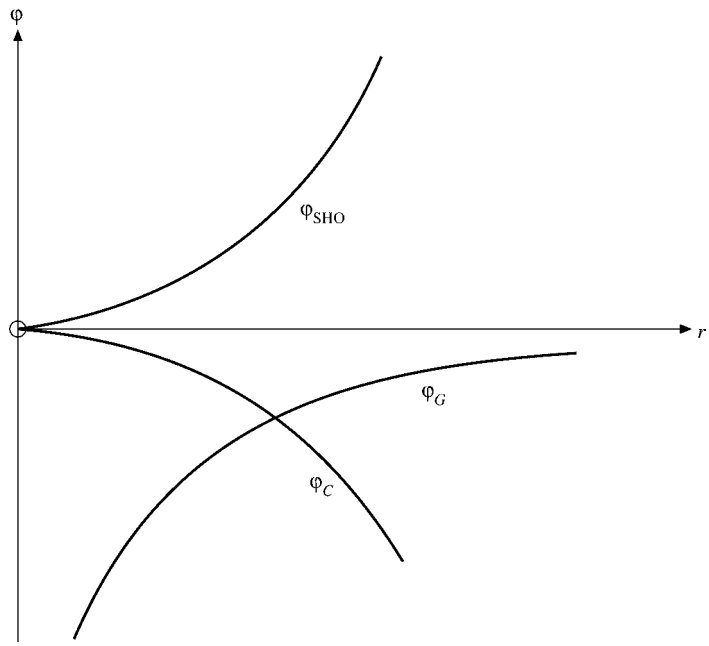

FIGURE 1.34 Potential energy versus distance (gravitational, centrifugal, and simple harmonic oscillator).

$\mathbf{F} = -\nabla\varphi$ is derived. The triple equivalence is demonstrated (Fig. 1.33). To summarize, a single-valued scalar potential function φ exists if and only if $\mathbf{F}$ is irrotational or the work done around every closed loop is zero. The gravitational and electrostatic force fields given by Eq. (1.79) are irrotational and therefore are conservative. Gravitational and electrostatic scalar potentials exist. Now, by calculating the work done (Eq. (1.126)), we proceed to determine three potentials (Fig. 1.34).

Example 1.13.1 GRAVITATIONAL POTENTIAL

Find the scalar potential for the gravitational force on a unit mass m_1,

$$\mathbf{F}_G = -\frac{Gm_1m_2\hat{\mathbf{r}}}{r^2} = -\frac{k\hat{\mathbf{r}}}{r^2}, \tag{1.131}$$

radially **inward**. By integrating Eq. (1.118) from infinity in to position $\mathbf{r}$, we obtain

$$\varphi_G(r) - \varphi_G(\infty) = -\int_\infty^{\mathbf{r}} \mathbf{F}_G \cdot d\mathbf{r} = +\int_{\mathbf{r}}^\infty \mathbf{F}_G \cdot d\mathbf{r}. \tag{1.132}$$

By use of $\mathbf{F}_G = -\mathbf{F}_{\text{applied}}$, a comparison with Eq. (1.95a) shows that the potential is the work done in bringing the unit mass in from infinity. (We can define only potential difference. Here we arbitrarily assign infinity to be a zero of potential.) The integral on the right-hand side of Eq. (1.132) is negative, meaning that $\varphi_G(r)$ is negative. Since $\mathbf{F}_G$ is radial, we obtain a contribution to φ only when $d\mathbf{r}$ is radial, or

$$\varphi_G(r) = -\int_r^\infty \frac{k\,dr}{r^2} = -\frac{k}{r} = -\frac{Gm_1m_2}{r}.$$

The final negative sign is a consequence of the attractive force of gravity. ■

Example 1.13.2 CENTRIFUGAL POTENTIAL

Calculate the scalar potential for the **centrifugal** force per unit mass, $\mathbf{F}_C = \omega^2 r\hat{\mathbf{r}}$, radially **outward**. Physically, you might feel this on a large horizontal spinning disk at an amusement park. Proceeding as in Example 1.13.1 but integrating from the origin outward and taking $\varphi_C(0) = 0$, we have

$$\varphi_C(r) = -\int_0^r \mathbf{F}_C \cdot d\mathbf{r} = -\frac{\omega^2 r^2}{2}.$$

If we reverse signs, taking $\mathbf{F}_{\text{SHO}} = -k\mathbf{r}$, we obtain $\varphi_{\text{SHO}} = \frac{1}{2}kr^2$, the simple harmonic oscillator potential.

The gravitational, centrifugal, and simple harmonic oscillator potentials are shown in Fig. 1.34. Clearly, the simple harmonic oscillator yields stability and describes a restoring force. The centrifugal potential describes an unstable situation. ■

Thermodynamics — Exact Differentials

In thermodynamics, which is sometimes called a search for exact differentials, we encounter equations of the form

$$df = P(x, y)\,dx + Q(x, y)\,dy. \tag{1.133a}$$

The usual problem is to determine whether $\int (P(x, y)\,dx + Q(x, y)\,dy)$ depends only on the endpoints, that is, whether df is indeed an exact differential. The necessary and sufficient condition is that

$$df = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy \tag{1.133b}$$

or that

$$\begin{aligned} P(x, y) &= \partial f/\partial x, \\ Q(x, y) &= \partial f/\partial y. \end{aligned} \tag{1.133c}$$

Equations (1.133c) depend on satisfying the relation

$$\frac{\partial P(x, y)}{\partial y} = \frac{\partial Q(x, y)}{\partial x}. \tag{1.133d}$$

This, however, is exactly analogous to Eq. (1.119), the requirement that **F** be irrotational. Indeed, the z-component of Eq. (1.119) yields

$$\frac{\partial F_x}{\partial y} = \frac{\partial F_y}{\partial x}, \tag{1.133e}$$

with

$$F_x = \frac{\partial f}{\partial x}, \qquad F_y = \frac{\partial f}{\partial y}.$$

Vector Potential

In some branches of physics, especially electrodynamics, it is convenient to introduce a vector potential **A** such that a (force) field **B** is given by

$$\mathbf{B} = \nabla \times \mathbf{A}. \tag{1.134}$$

Clearly, if Eq. (1.134) holds, $\nabla \cdot \mathbf{B} = 0$ by Eq. (1.84) and **B** is solenoidal. Here we want to develop a converse, to show that when **B** is solenoidal a vector potential **A** exists. We demonstrate the existence of **A** by actually calculating it. Suppose $\mathbf{B} = \hat{\mathbf{x}}b_1 + \hat{\mathbf{y}}b_2 + \hat{\mathbf{z}}b_3$ and our unknown $\mathbf{A} = \hat{\mathbf{x}}a_1 + \hat{\mathbf{y}}a_2 + \hat{\mathbf{z}}a_3$. By Eq. (1.134),

$$\frac{\partial a_3}{\partial y} - \frac{\partial a_2}{\partial z} = b_1, \tag{1.135a}$$

$$\frac{\partial a_1}{\partial z} - \frac{\partial a_3}{\partial x} = b_2, \tag{1.135b}$$

$$\frac{\partial a_2}{\partial x} - \frac{\partial a_1}{\partial y} = b_3. \tag{1.135c}$$

Let us assume that the coordinates have been chosen so that **A** is parallel to the yz-plane; that is, $a_1 = 0$.[24] Then

$$\begin{aligned} b_2 &= -\frac{\partial a_3}{\partial x} \\ b_3 &= \frac{\partial a_2}{\partial x}. \end{aligned} \tag{1.136}$$

[24]Clearly, this can be done at any one point. It is not at all obvious that this assumption will hold at all points; that is, **A** will be two-dimensional. The justification for the assumption is that it works; Eq. (1.141) satisfies Eq. (1.134).

Integrating, we obtain

$$\begin{aligned} a_2 &= \int_{x_0}^{x} b_3\,dx + f_2(y,z), \\ a_3 &= -\int_{x_0}^{x} b_2\,dx + f_3(y,z), \end{aligned} \tag{1.137}$$

where f_2 and f_3 are arbitrary functions of y and z but **not** functions of x. These two equations can be checked by differentiating and recovering Eq. (1.136). Equation (1.135a) becomes[25]

$$\begin{aligned} \frac{\partial a_3}{\partial y} - \frac{\partial a_2}{\partial z} &= -\int_{x_0}^{x}\left(\frac{\partial b_2}{\partial y} + \frac{\partial b_3}{\partial z}\right)dx + \frac{\partial f_3}{\partial y} - \frac{\partial f_2}{\partial z} \\ &= \int_{x_0}^{x} \frac{\partial b_1}{\partial x}\,dx + \frac{\partial f_3}{\partial y} - \frac{\partial f_2}{\partial z}, \end{aligned} \tag{1.138}$$

using $\nabla \cdot \mathbf{B} = 0$. Integrating with respect to x, we obtain

$$\frac{\partial a_3}{\partial y} - \frac{\partial a_2}{\partial z} = b_1(x,y,z) - b_1(x_0,y,z) + \frac{\partial f_3}{\partial y} - \frac{\partial f_2}{\partial z}. \tag{1.139}$$

Remembering that f_3 and f_2 are arbitrary functions of y and z, we choose

$$\begin{aligned} f_2 &= 0, \\ f_3 &= \int_{y_0}^{y} b_1(x_0,y,z)\,dy, \end{aligned} \tag{1.140}$$

so that the right-hand side of Eq. (1.139) reduces to $b_1(x,y,z)$, in agreement with Eq. (1.135a). With f_2 and f_3 given by Eq. (1.140), we can construct $\mathbf{A}$:

$$\mathbf{A} = \hat{\mathbf{y}}\int_{x_0}^{x} b_3(x,y,z)\,dx + \hat{\mathbf{z}}\left[\int_{y_0}^{y} b_1(x_0,y,z)\,dy - \int_{x_0}^{x} b_2(x,y,z)\,dx\right]. \tag{1.141}$$

However, this is not quite complete. We may add any constant since $\mathbf{B}$ is a derivative of $\mathbf{A}$. What is much more important, we may add any gradient of a scalar function $\nabla\varphi$ without affecting $\mathbf{B}$ at all. Finally, the functions f_2 and f_3 are not unique. Other choices could have been made. Instead of setting $a_1 = 0$ to get Eq. (1.136) any cyclic permutation of $1, 2, 3, x, y, z, x_0, y_0, z_0$ would also work.

Example 1.13.3 A Magnetic Vector Potential for a Constant Magnetic Field

To illustrate the construction of a magnetic vector potential, we take the special but still important case of a constant magnetic induction

$$\mathbf{B} = \hat{\mathbf{z}}B_z, \tag{1.142}$$

[25] Leibniz' formula in Exercise 9.6.13 is useful here.

in which B_z is a constant. Equations (1.135a to c) become

$$\begin{aligned}\frac{\partial a_3}{\partial y} - \frac{\partial a_2}{\partial z} &= 0,\\ \frac{\partial a_1}{\partial z} - \frac{\partial a_3}{\partial x} &= 0,\\ \frac{\partial a_2}{\partial x} - \frac{\partial a_1}{\partial y} &= B_z.\end{aligned} \tag{1.143}$$

If we assume that $a_1 = 0$, as before, then by Eq. (1.141)

$$\mathbf{A} = \hat{\mathbf{y}} \int^x B_z \, dx = \hat{\mathbf{y}} x B_z, \tag{1.144}$$

setting a constant of integration equal to zero. It can readily be seen that this $\mathbf{A}$ satisfies Eq. (1.134).

To show that the choice $a_1 = 0$ was not sacred or at least not required, let us try setting $a_3 = 0$. From Eq. (1.143)

$$\frac{\partial a_2}{\partial z} = 0, \tag{1.145a}$$

$$\frac{\partial a_1}{\partial z} = 0, \tag{1.145b}$$

$$\frac{\partial a_2}{\partial x} - \frac{\partial a_1}{\partial y} = B_z. \tag{1.145c}$$

We see a_1 and a_2 are independent of z, or

$$a_1 = a_1(x, y), \qquad a_2 = a_2(x, y). \tag{1.146}$$

Equation (1.145c) is satisfied if we take

$$a_2 = p \int^x B_z \, dx = p x B_z \tag{1.147}$$

and

$$a_1 = (p-1) \int^y B_z \, dy = (p-1) y B_z, \tag{1.148}$$

with p any constant. Then

$$\mathbf{A} = \hat{\mathbf{x}}(p-1) y B_z + \hat{\mathbf{y}} p x B_z. \tag{1.149}$$

Again, Eqs. (1.134), (1.142), and (1.149) are seen to be consistent. Comparison of Eqs. (1.144) and (1.149) shows immediately that $\mathbf{A}$ is not unique. The difference between Eqs. (1.144) and (1.149) and the appearance of the parameter p in Eq. (1.149) may be accounted for by rewriting Eq. (1.149) as

$$\begin{aligned}\mathbf{A} &= -\frac{1}{2}(\hat{\mathbf{x}} y - \hat{\mathbf{y}} x) B_z + \left(p - \frac{1}{2}\right)(\hat{\mathbf{x}} y + \hat{\mathbf{y}} x) B_z\\ &= -\frac{1}{2}(\hat{\mathbf{x}} y - \hat{\mathbf{y}} x) B_z + \left(p - \frac{1}{2}\right) B_z \nabla \varphi\end{aligned} \tag{1.150}$$

with

$$\varphi = xy. \tag{1.151}$$

■

The first term in **A** corresponds to the usual form

$$\mathbf{A} = \frac{1}{2}(\mathbf{B} \times \mathbf{r}) \tag{1.152}$$

for **B**, a constant.

Adding a gradient of a scalar function, Λ say, to the vector potential **A** does not affect **B**, by Eq. (1.82); this is known as a gauge transformation (see Exercises 1.13.9 and 4.6.4):

$$\mathbf{A} \to \mathbf{A}' = \mathbf{A} + \nabla\Lambda. \tag{1.153}$$

Suppose now that the wave function ψ_0 solves the Schrödinger equation of quantum mechanics without magnetic induction field **B**,

$$\left\{\frac{1}{2m}(-i\hbar\nabla)^2 + V - E\right\}\psi_0 = 0, \tag{1.154}$$

describing a particle with mass m and charge e. When **B** is switched on, the wave equation becomes

$$\boxed{\left\{\frac{1}{2m}(-i\hbar\nabla - e\mathbf{A})^2 + V - E\right\}\psi = 0.} \tag{1.155}$$

Its solution ψ picks up a phase factor that depends on the coordinates in general,

$$\boxed{\psi(\mathbf{r}) = \exp\left[\frac{ie}{\hbar}\int^{\mathbf{r}} \mathbf{A}(\mathbf{r}')\cdot d\mathbf{r}'\right]\psi_0(\mathbf{r}).} \tag{1.156}$$

From the relation

$$\begin{aligned}(-i\hbar\nabla - e\mathbf{A})\psi &= \exp\left[\frac{ie}{\hbar}\int \mathbf{A}\cdot d\mathbf{r}'\right]\left\{(-i\hbar\nabla - e\mathbf{A})\psi_0 - i\hbar\psi_0\frac{ie}{\hbar}\mathbf{A}\right\}\\ &= \exp\left[\frac{ie}{\hbar}\int \mathbf{A}\cdot d\mathbf{r}'\right](-i\hbar\nabla\psi_0),\end{aligned} \tag{1.157}$$

it is obvious that ψ solves Eq. (1.155) if ψ_0 solves Eq. (1.154). The **gauge covariant derivative** $\nabla - i(e/\hbar)\mathbf{A}$ describes the coupling of a charged particle with the magnetic field. It is often called **minimal substitution** and plays a central role in quantum electromagnetism, the first and simplest gauge theory in physics.

To summarize this discussion of the vector potential: *When a vector* **B** *is solenoidal, a vector potential* **A** *exists such that* $\mathbf{B} = \nabla \times \mathbf{A}$. **A** *is undetermined to within an additive gradient. This corresponds to the arbitrary zero of a potential, a constant of integration for the scalar potential*.

In many problems the magnetic vector potential **A** will be obtained from the current distribution that produces the magnetic induction **B**. This means solving Poisson's (vector) equation (see Exercise 1.14.4).

Exercises

1.13.1 If a force $\mathbf{F}$ is given by

$$\mathbf{F} = \left(x^2 + y^2 + z^2\right)^n (\hat{\mathbf{x}}x + \hat{\mathbf{y}}y + \hat{\mathbf{z}}z),$$

find

(a) $\nabla \cdot \mathbf{F}$.
(b) $\nabla \times \mathbf{F}$.
(c) A scalar potential $\varphi(x, y, z)$ so that $\mathbf{F} = -\nabla\varphi$.
(d) For what value of the exponent n does the scalar potential diverge at both the origin and infinity?

ANS. (a) $(2n+3)r^{2n}$, (b) 0,
(c) $-\frac{1}{2n+2}r^{2n+2}$, $n \neq -1$, (d) $n = -1$,
$\varphi = -\ln r$.

1.13.2 A sphere of radius a is uniformly charged (throughout its volume). Construct the electrostatic potential $\varphi(r)$ for $0 \leqslant r < \infty$.
Hint. In Section 1.14 it is shown that the Coulomb force on a test charge at $r = r_0$ depends only on the charge at distances less than r_0 and is independent of the charge at distances greater than r_0. Note that this applies to a **spherically symmetric** charge distribution.

1.13.3 The usual problem in classical mechanics is to calculate the motion of a particle given the potential. For a uniform density (ρ_0), nonrotating massive sphere, Gauss' law of Section 1.14 leads to a gravitational force on a unit mass m_0 at a point r_0 produced by the attraction of the mass at $r \leqslant r_0$. The mass at $r > r_0$ contributes nothing to the force.

(a) Show that $\mathbf{F}/m_0 = -(4\pi G\rho_0/3)\mathbf{r}$, $0 \leqslant r \leqslant a$, where a is the radius of the sphere.
(b) Find the corresponding gravitational potential, $0 \leqslant r \leqslant a$.
(c) Imagine a vertical hole running completely through the center of the Earth and out to the far side. Neglecting the rotation of the Earth and assuming a uniform density $\rho_0 = 5.5$ gm/cm^3, calculate the nature of the motion of a particle dropped into the hole. What is its period?

Note. $\mathbf{F} \propto \mathbf{r}$ is actually a very poor approximation. Because of varying density, the approximation $\mathbf{F} =$ constant along the outer half of a radial line and $\mathbf{F} \propto \mathbf{r}$ along the inner half is a much closer approximation.

1.13.4 The origin of the Cartesian coordinates is at the Earth's center. The moon is on the z-axis, a fixed distance R away (center-to-center distance). The tidal force exerted by the moon on a particle at the Earth's surface (point x, y, z) is given by

$$F_x = -GMm\frac{x}{R^3}, \qquad F_y = -GMm\frac{y}{R^3}, \qquad F_z = +2GMm\frac{z}{R^3}.$$

Find the potential that yields this tidal force.

ANS. $-\dfrac{GMm}{R^3}\left(z^2-\dfrac{1}{2}x^2-\dfrac{1}{2}y^2\right)$.

In terms of the Legendre polynomials of Chapter 12 this becomes

$-\dfrac{GMm}{R^3}r^2P_2(\cos\theta)$.

1.13.5 A long, straight wire carrying a current I produces a magnetic induction $\mathbf{B}$ with components

$$\mathbf{B}=\frac{\mu_0 I}{2\pi}\left(-\frac{y}{x^2+y^2},\frac{x}{x^2+y^2},0\right).$$

Find a magnetic vector potential $\mathbf{A}$.

ANS. $\mathbf{A}=-\hat{\mathbf{z}}(\mu_0 I/4\pi)\ln(x^2+y^2)$. (This solution is not unique.)

1.13.6 If

$$\mathbf{B}=\frac{\hat{\mathbf{r}}}{r^2}=\left(\frac{x}{r^3},\frac{y}{r^3},\frac{z}{r^3}\right),$$

find a vector $\mathbf{A}$ such that $\nabla\times\mathbf{A}=\mathbf{B}$. One possible solution is

$$\mathbf{A}=\frac{\hat{\mathbf{x}}yz}{r(x^2+y^2)}-\frac{\hat{\mathbf{y}}xz}{r(x^2+y^2)}.$$

1.13.7 Show that the pair of equations

$$\mathbf{A}=\frac{1}{2}(\mathbf{B}\times\mathbf{r}),\qquad \mathbf{B}=\nabla\times\mathbf{A}$$

is satisfied by any constant magnetic induction $\mathbf{B}$.

1.13.8 Vector $\mathbf{B}$ is formed by the product of two gradients

$$\mathbf{B}=(\nabla u)\times(\nabla v),$$

where u and v are scalar functions.

(a) Show that $\mathbf{B}$ is solenoidal.

(b) Show that

$$\mathbf{A}=\frac{1}{2}(u\,\nabla v-v\,\nabla u)$$

is a vector potential for $\mathbf{B}$, in that

$$\mathbf{B}=\nabla\times\mathbf{A}.$$

1.13.9 The magnetic induction $\mathbf{B}$ is related to the magnetic vector potential $\mathbf{A}$ by $\mathbf{B}=\nabla\times\mathbf{A}$. By Stokes' theorem

$$\int\mathbf{B}\cdot d\boldsymbol{\sigma}=\oint\mathbf{A}\cdot d\mathbf{r}.$$

Show that each side of this equation is invariant under the **gauge transformation**, $\mathbf{A} \to \mathbf{A} + \nabla\varphi$.
Note. Take the function φ to be single-valued. The complete gauge transformation is considered in Exercise 4.6.4.

1.13.10 With $\mathbf{E}$ the electric field and $\mathbf{A}$ the magnetic vector potential, show that $[\mathbf{E} + \partial\mathbf{A}/\partial t]$ is irrotational and that therefore we may write

$$E = -\nabla\varphi - \frac{\partial \mathbf{A}}{\partial t}.$$

1.13.11 The total force on a charge q moving with velocity $\mathbf{v}$ is

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}).$$

Using the scalar and vector potentials, show that

$$\mathbf{F} = q\left[-\nabla\varphi - \frac{d\mathbf{A}}{dt} + \nabla(\mathbf{A} \cdot \mathbf{v})\right].$$

Note that we now have a total time derivative of $\mathbf{A}$ in place of the partial derivative of Exercise 1.13.10.

1.14 Gauss' Law, Poisson's Equation

Gauss' Law

Consider a point electric charge q at the origin of our coordinate system. This produces an electric field $\mathbf{E}$ given by[26]

$$\mathbf{E} = \frac{q\hat{\mathbf{r}}}{4\pi\varepsilon_0 r^2}. \tag{1.158}$$

We now derive Gauss' law, which states that the surface integral in Fig. 1.35 is q/ε_0 if the closed surface $S = \partial V$ includes the origin (where q is located) and zero if the surface does not include the origin. The surface S is any closed surface; it need not be spherical.

Using Gauss' theorem, Eqs. (1.101a) and (1.101b) (and neglecting the $q/4\pi\varepsilon_0$), we obtain

$$\int_S \frac{\hat{\mathbf{r}} \cdot d\boldsymbol{\sigma}}{r^2} = \int_V \nabla \cdot \left(\frac{\hat{\mathbf{r}}}{r^2}\right) d\tau = 0 \tag{1.159}$$

by Example 1.7.2, provided the surface S does not include the origin, where the integrands are not defined. This proves the second part of Gauss' law.

The first part, in which the surface S must include the origin, may be handled by surrounding the origin with a small sphere $S' = \partial V'$ of radius δ (Fig. 1.36). So that there will be no question what is inside and what is outside, imagine the volume outside the outer surface S and the volume inside surface $S'(r < \delta)$ connected by a small hole. This

[26] The electric field $\mathbf{E}$ is defined as the force per unit charge on a small stationary test charge q_t: $\mathbf{E} = \mathbf{F}/q_t$. From Coulomb's law the force on q_t due to q is $\mathbf{F} = (qq_t/4\pi\varepsilon_0)(\hat{\mathbf{r}}/r^2)$. When we divide by q_t, Eq. (1.158) follows.

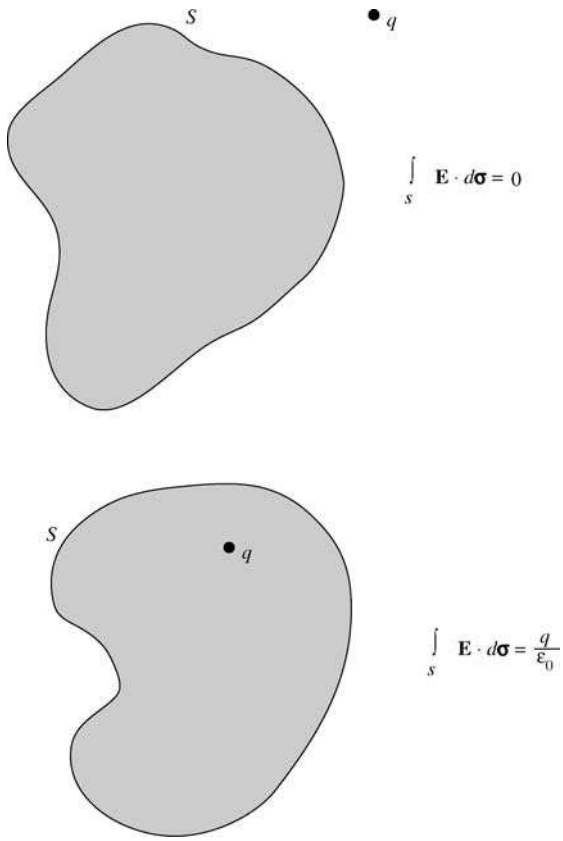

FIGURE 1.35 Gauss' law.

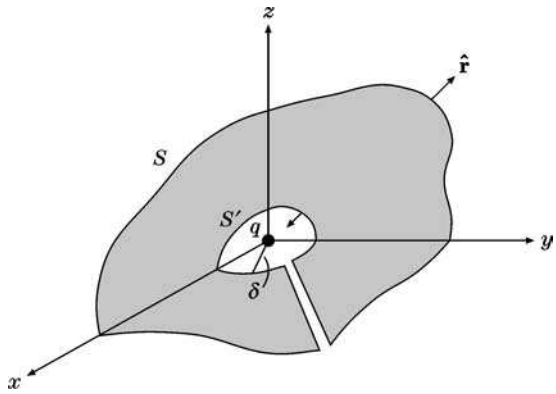

FIGURE 1.36 Exclusion of the origin.

joins surfaces S and S', combining them into one single simply connected closed surface. Because the radius of the imaginary hole may be made vanishingly small, there is no additional contribution to the surface integral. The inner surface is deliberately chosen to be

spherical so that we will be able to integrate over it. Gauss' theorem now applies to the volume between S and S' without any difficulty. We have

$$\int_S \frac{\hat{\mathbf{r}} \cdot d\boldsymbol{\sigma}}{r^2} + \int_{S'} \frac{\hat{\mathbf{r}} \cdot d\boldsymbol{\sigma}'}{\delta^2} = 0. \tag{1.160}$$

We may evaluate the second integral, for $d\boldsymbol{\sigma}' = -\hat{\mathbf{r}}\delta^2\, d\Omega$, in which $d\Omega$ is an element of solid angle. The minus sign appears because we agreed in Section 1.10 to have the positive normal $\hat{\mathbf{r}}'$ **outward** from the volume. In this case the outward $\hat{\mathbf{r}}'$ is in the negative radial direction, $\hat{\mathbf{r}}' = -\hat{\mathbf{r}}$. By integrating over all angles, we have

$$\int_{S'} \frac{\hat{\mathbf{r}} \cdot d\boldsymbol{\sigma}'}{\delta^2} = -\int_{S'} \frac{\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}\delta^2\, d\Omega}{\delta^2} = -4\pi, \tag{1.161}$$

independent of the radius δ. With the constants from Eq. (1.158), this results in

$$\int_S \mathbf{E} \cdot d\boldsymbol{\sigma} = \frac{q}{4\pi\varepsilon_0} 4\pi = \frac{q}{\varepsilon_0}, \tag{1.162}$$

completing the proof of Gauss' law. Notice that although the surface S may be spherical, it **need not** be spherical. Going just a bit further, we consider a distributed charge so that

$$q = \int_V \rho\, d\tau. \tag{1.163}$$

Equation (1.162) still applies, with q now interpreted as the total distributed charge enclosed by surface S:

$$\int_S \mathbf{E} \cdot d\boldsymbol{\sigma} = \int_V \frac{\rho}{\varepsilon_0} d\tau. \tag{1.164}$$

Using Gauss' theorem, we have

$$\int_V \boldsymbol{\nabla} \cdot \mathbf{E}\, d\tau = \int_V \frac{\rho}{\varepsilon_0} d\tau. \tag{1.165}$$

Since our volume is completely arbitrary, the integrands must be equal, or

$$\boldsymbol{\nabla} \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}, \tag{1.166}$$

one of Maxwell's equations. If we reverse the argument, Gauss' law follows immediately from Maxwell's equation.

Poisson's Equation

If we replace $\mathbf{E}$ by $-\boldsymbol{\nabla}\varphi$, Eq. (1.166) becomes

$$\boldsymbol{\nabla} \cdot \boldsymbol{\nabla}\varphi = -\frac{\rho}{\varepsilon_0}, \tag{1.167a}$$

which is Poisson's equation. For the condition $\rho = 0$ this reduces to an even more famous equation,

$$\nabla \cdot \nabla \varphi = 0, \tag{1.167b}$$

Laplace's equation. We encounter Laplace's equation frequently in discussing various coordinate systems (Chapter 2) and the special functions of mathematical physics that appear as its solutions. Poisson's equation will be invaluable in developing the theory of Green's functions (Section 9.7).

From direct comparison of the Coulomb electrostatic force law and Newton's law of universal gravitation,

$$\mathbf{F}_E = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r^2}\hat{\mathbf{r}}, \qquad \mathbf{F}_G = -G\frac{m_1 m_2}{r^2}\hat{\mathbf{r}}.$$

All of the potential theory of this section applies equally well to gravitational potentials. For example, the gravitational Poisson equation is

$$\nabla \cdot \nabla \varphi = +4\pi G\rho, \tag{1.168}$$

with ρ now a mass density.

Exercises

1.14.1 Develop Gauss' law for the two-dimensional case in which

$$\varphi = -q\frac{\ln \rho}{2\pi\varepsilon_0}, \qquad \mathbf{E} = -\nabla\varphi = q\frac{\hat{\boldsymbol{\rho}}}{2\pi\varepsilon_0\rho}.$$

Here q is the charge at the origin or the line charge per unit length if the two-dimensional system is a unit thickness slice of a three-dimensional (circular cylindrical) system. The variable ρ is measured radially outward from the line charge. $\hat{\boldsymbol{\rho}}$ is the corresponding unit vector (see Section 2.4).

1.14.2 (a) Show that Gauss' law follows from Maxwell's equation

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}.$$

Here ρ is the usual charge density.

(b) Assuming that the electric field of a point charge q is spherically symmetric, show that Gauss' law implies the Coulomb inverse square expression

$$\mathbf{E} = \frac{q\hat{\mathbf{r}}}{4\pi\varepsilon_0 r^2}.$$

1.14.3 Show that the value of the electrostatic potential φ at any point P is equal to the average of the potential over any spherical surface centered on P. There are no electric charges on or within the sphere.

Hint. Use Green's theorem, Eq. (1.104), with $u^{-1} = r$, the distance from P, and $v = \varphi$. Also note Eq. (1.170) in Section 1.15.

1.14.4 Using Maxwell's equations, show that for a system (steady current) the magnetic vector potential **A** satisfies a vector Poisson equation,

$$\nabla^2 \mathbf{A} = -\mu_0 \mathbf{J},$$

provided we require $\nabla \cdot \mathbf{A} = 0$.

1.15 DIRAC DELTA FUNCTION

From Example 1.6.1 and the development of Gauss' law in Section 1.14,

$$\int \nabla \cdot \nabla\left(\frac{1}{r}\right) d\tau = -\int \nabla \cdot \left(\frac{\hat{\mathbf{r}}}{r^2}\right) d\tau = \begin{cases} -4\pi \\ 0, \end{cases} \tag{1.169}$$

depending on whether or not the integration includes the origin $\mathbf{r} = 0$. This result may be conveniently expressed by introducing the Dirac delta function,

$$\boxed{\nabla^2\left(\frac{1}{r}\right) = -4\pi\delta(\mathbf{r}) \equiv -4\pi\delta(x)\delta(y)\delta(z).} \tag{1.170}$$

This Dirac delta function is **defined** by its assigned properties

$$\delta(x) = 0, \qquad x \neq 0 \tag{1.171a}$$

$$\boxed{f(0) = \int_{-\infty}^{\infty} f(x)\delta(x)\,dx,} \tag{1.171b}$$

where $f(x)$ is any well-behaved function and the integration includes the origin. As a special case of Eq. (1.171b),

$$\int_{-\infty}^{\infty} \delta(x)\,dx = 1. \tag{1.171c}$$

From Eq. (1.171b), $\delta(x)$ must be an infinitely high, infinitely thin spike at $x = 0$, as in the description of an impulsive force (Section 15.9) or the charge density for a point charge.[27] The problem is that **no such function exists**, in the usual sense of function. However, the crucial property in Eq. (1.171b) can be developed rigorously as the limit of a **sequence** of functions, a distribution. For example, the delta function may be approximated by the

[27] The delta function is frequently invoked to describe very short-range forces, such as nuclear forces. It also appears in the normalization of continuum wave functions of quantum mechanics. Compare Eq. (1.193c) for plane-wave eigenfunctions.

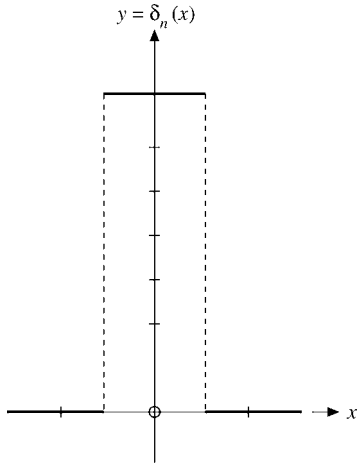

FIGURE 1.37 δ-Sequence function.

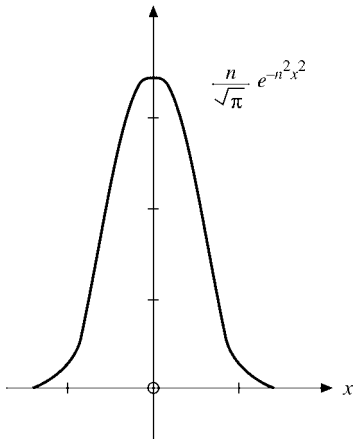

FIGURE 1.38 δ-Sequence function.

sequences of functions, Eqs. (1.172) to (1.175) and Figs. 1.37 to 1.40:

$$\delta_n(x) = \begin{cases} 0, & x < -\frac{1}{2n} \\ n, & -\frac{1}{2n} < x < \frac{1}{2n} \\ 0, & x > \frac{1}{2n} \end{cases} \tag{1.172}$$

$$\delta_n(x) = \frac{n}{\sqrt{\pi}} \exp\left(-n^2x^2\right) \tag{1.173}$$

$$\delta_n(x) = \frac{n}{\pi} \cdot \frac{1}{1+n^2x^2} \tag{1.174}$$

$$\delta_n(x) = \frac{\sin nx}{\pi x} = \frac{1}{2\pi} \int_{-n}^{n} e^{ixt}\, dt. \tag{1.175}$$

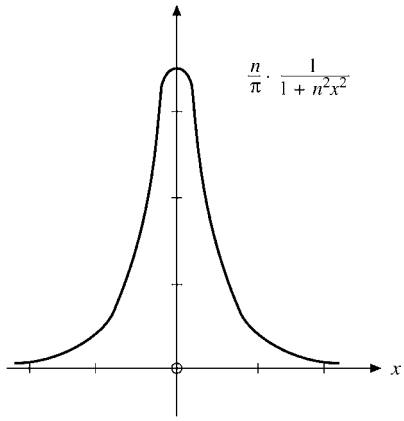

FIGURE 1.39 δ-Sequence function.

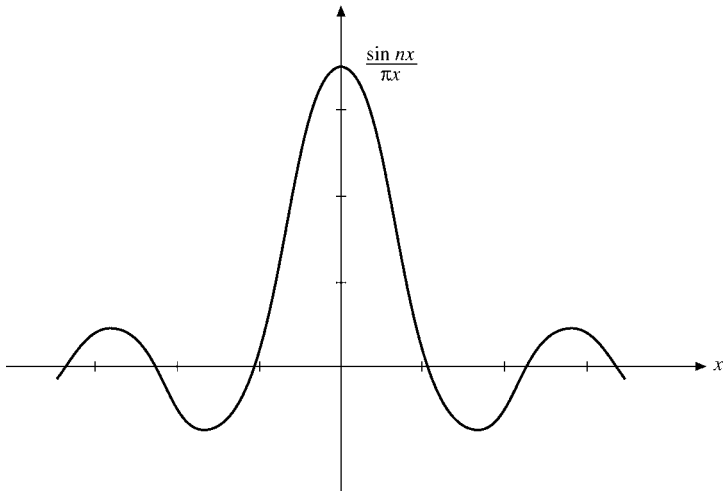

FIGURE 1.40 δ-Sequence function.

These approximations have varying degrees of usefulness. Equation (1.172) is useful in providing a simple derivation of the integral property, Eq. (1.171b). Equation (1.173) is convenient to differentiate. Its derivatives lead to the Hermite polynomials. Equation (1.175) is particularly useful in Fourier analysis and in its applications to quantum mechanics. In the theory of Fourier series, Eq. (1.175) often appears (modified) as the Dirichlet kernel:

$$\delta_n(x) = \frac{1}{2\pi} \frac{\sin[(n+\frac{1}{2})x]}{\sin(\frac{1}{2}x)}. \tag{1.176}$$

In using these approximations in Eq. (1.171b) and later, we assume that $f(x)$ is well behaved — it offers no problems at large x.

For most physical purposes such approximations are quite adequate. From a mathematical point of view the situation is still unsatisfactory: The limits

$$\lim_{n\to\infty} \delta_n(x)$$

do not exist.

A way out of this difficulty is provided by the theory of distributions. Recognizing that Eq. (1.171b) is the fundamental property, we focus our attention on it rather than on $\delta(x)$ itself. Equations (1.172) to (1.175) with $n = 1, 2, 3, \ldots$ may be interpreted as **sequences** of normalized functions:

$$\int_{-\infty}^{\infty} \delta_n(x)\,dx = 1. \tag{1.177}$$

The sequence of integrals has the limit

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} \delta_n(x) f(x)\,dx = f(0). \tag{1.178}$$

Note that Eq. (1.178) is the limit of a sequence of integrals. Again, the limit of $\delta_n(x)$, $n \to \infty$, does not exist. (The limits for all four forms of $\delta_n(x)$ diverge at $x = 0$.)

We may treat $\delta(x)$ consistently in the form

$$\int_{-\infty}^{\infty} \delta(x) f(x)\,dx = \lim_{n\to\infty} \int_{-\infty}^{\infty} \delta_n(x) f(x)\,dx. \tag{1.179}$$

$\delta(x)$ is labeled a distribution (not a function) defined by the sequences $\delta_n(x)$ as indicated in Eq. (1.179). We might emphasize that the integral on the left-hand side of Eq. (1.179) is not a Riemann integral.[28] It is a limit.

This distribution $\delta(x)$ is only one of an infinity of possible distributions, but it is the one we are interested in because of Eq. (1.171b).

From these sequences of functions we see that Dirac's delta function must be even in x, $\delta(-x) = \delta(x)$.

The integral property, Eq. (1.171b), is useful in cases where the argument of the delta function is a function $g(x)$ with simple zeros on the real axis, which leads to the rules

$$\delta(ax) = \frac{1}{a}\delta(x), \qquad a > 0, \tag{1.180}$$

$$\delta\big(g(x)\big) = \sum_{\substack{a,\\ g(a)=0,\\ g'(a)\neq 0}} \frac{\delta(x-a)}{|g'(a)|}. \tag{1.181a}$$

Equation (1.180) may be written

$$\int_{-\infty}^{\infty} f(x)\delta(ax)\,dx = \frac{1}{a}\int_{-\infty}^{\infty} f\left(\frac{y}{a}\right)\delta(y)\,dy = \frac{1}{a}f(0),$$

[28] It can be treated as a Stieltjes integral if desired. $\delta(x)\,dx$ is replaced by $du(x)$, where $u(x)$ is the Heaviside step function (compare Exercise 1.15.13).

applying Eq. (1.171b). Equation (1.180) may be written as $\delta(ax) = \frac{1}{|a|}\delta(x)$ for $a < 0$. To prove Eq. (1.181a) we decompose the integral

$$\int_{-\infty}^{\infty} f(x)\delta\big(g(x)\big)\,dx = \sum_a \int_{a-\varepsilon}^{a+\varepsilon} f(x)\delta\big((x-a)g'(a)\big)\,dx \tag{1.181b}$$

into a sum of integrals over small intervals containing the zeros of $g(x)$. In these intervals, $g(x) \approx g(a) + (x-a)g'(a) = (x-a)g'(a)$. Using Eq. (1.180) on the right-hand side of Eq. (1.181b) we obtain the integral of Eq. (1.181a).

Using integration by parts we can also **define the derivative** $\delta'(x)$ of the Dirac delta function by the relation

$$\int_{-\infty}^{\infty} f(x)\delta'(x-x')\,dx = -\int_{-\infty}^{\infty} f'(x)\delta(x-x')\,dx = -f'(x'). \tag{1.182}$$

We use $\delta(x)$ frequently and call it the Dirac delta function[29] — for historical reasons. Remember that it is not really a function. It is essentially a shorthand notation, defined implicitly as the limit of integrals in a sequence, $\delta_n(x)$, according to Eq. (1.179). It should be understood that our Dirac delta function has significance only as part of an integrand. In this spirit, the linear operator $\int dx\,\delta(x-x_0)$ operates on $f(x)$ and yields $f(x_0)$:

$$\mathcal{L}(x_0)f(x) \equiv \int_{-\infty}^{\infty} \delta(x-x_0)f(x)\,dx = f(x_0). \tag{1.183}$$

It may also be classified as a linear mapping or simply as a generalized function. Shifting our singularity to the point $x = x'$, we write the Dirac delta function as $\delta(x-x')$. Equation (1.171b) becomes

$$\int_{-\infty}^{\infty} f(x)\delta(x-x')\,dx = f(x'). \tag{1.184}$$

As a description of a singularity at $x = x'$, the Dirac delta function may be written as $\delta(x-x')$ or as $\delta(x'-x)$. Going to three dimensions and using spherical polar coordinates, we obtain

$$\int_0^{2\pi}\int_0^{\pi}\int_0^{\infty} \delta(\mathbf{r})r^2\,dr\sin\theta\,d\theta\,d\varphi = \iiint_{-\infty}^{\infty} \delta(x)\delta(y)\delta(z)\,dx\,dy\,dz = 1. \tag{1.185}$$

This corresponds to a singularity (or source) at the origin. Again, if our source is at $\mathbf{r} = \mathbf{r_1}$, Eq. (1.185) becomes

$$\iiint \delta(\mathbf{r}_2 - \mathbf{r}_1)r_2^2\,dr_2\sin\theta_2\,d\theta_2\,d\varphi_2 = 1. \tag{1.186}$$

[29] Dirac introduced the delta function to quantum mechanics. Actually, the delta function can be traced back to Kirchhoff, 1882. For further details see M. Jammer, *The Conceptual Development of Quantum Mechanics*. New York: McGraw–Hill (1966), p. 301.

Example 1.15.1 TOTAL CHARGE INSIDE A SPHERE

Consider the total electric flux $\oint \mathbf{E} \cdot d\boldsymbol{\sigma}$ out of a sphere of radius R around the origin surrounding n charges e_j, located at the points $\mathbf{r}_j$ with $r_j < R$, that is, inside the sphere. The electric field strength $\mathbf{E} = -\nabla\varphi(\mathbf{r})$, where the potential

$$\varphi = \sum_{j=1}^{n} \frac{e_j}{|\mathbf{r} - \mathbf{r}_j|} = \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3r'$$

is the sum of the Coulomb potentials generated by each charge and the total charge density is $\rho(\mathbf{r}) = \sum_j e_j \delta(\mathbf{r} - \mathbf{r}_j)$. The delta function is used here as an abbreviation of a pointlike density. Now we use Gauss' theorem for

$$\oint \mathbf{E} \cdot d\boldsymbol{\sigma} = -\oint \nabla\varphi \cdot d\boldsymbol{\sigma} = -\int \nabla^2\varphi \, d\tau = \int \frac{\rho(\mathbf{r})}{\varepsilon_0} d\tau = \frac{\sum_j e_j}{\varepsilon_0}$$

in conjunction with the differential form of Gauss's law, $\nabla \cdot \mathbf{E} = -\rho/\varepsilon_0$, and

$$\sum_j e_j \int \delta(\mathbf{r} - \mathbf{r}_j) \, d\tau = \sum_j e_j.$$

■

Example 1.15.2 PHASE SPACE

In the scattering theory of relativistic particles using Feynman diagrams, we encounter the following integral over energy of the scattered particle (we set the velocity of light $c = 1$):

$$\begin{aligned} \int d^4p \, \delta\big(p^2 - m^2\big) f(p) &\equiv \int d^3p \int dp_0 \, \delta\big(p_0^2 - \mathbf{p}^2 - m^2\big) f(p) \\ &= \int_{E>0} \frac{d^3p \, f(E, \mathbf{p})}{2\sqrt{m^2 + \mathbf{p}^2}} + \int_{E<0} \frac{d^3p \, f(E, \mathbf{p})}{2\sqrt{m^2 + \mathbf{p}^2}}, \end{aligned}$$

where we have used Eq. (1.181a) at the zeros $E = \pm\sqrt{m^2 + \mathbf{p}^2}$ of the argument of the delta function. The physical meaning of $\delta(p^2 - m^2)$ is that the particle of mass m and four-momentum $p^\mu = (p_0, \mathbf{p})$ is on its mass shell, because $p^2 = m^2$ is equivalent to $E = \pm\sqrt{m^2 + \mathbf{p}^2}$. Thus, the on-mass-shell volume element in momentum space is the Lorentz invariant $\frac{d^3p}{2E}$, in contrast to the nonrelativistic d^3p of momentum space. The fact that a negative energy occurs is a peculiarity of relativistic kinematics that is related to the antiparticle. ■

Delta Function Representation by Orthogonal Functions

Dirac's delta function[30] can be expanded in terms of any basis of real orthogonal functions $\{\varphi_n(x), n = 0, 1, 2, \ldots\}$. Such functions will occur in Chapter 10 as solutions of ordinary differential equations of the Sturm–Liouville form.

[30] This section is optional here. It is not needed until Chapter 10.

They satisfy the orthogonality relations

$$\int_a^b \varphi_m(x)\varphi_n(x)\,dx = \delta_{mn}, \tag{1.187}$$

where the interval (a, b) may be infinite at either end or both ends. [For convenience we assume that φ_n has been defined to include $(w(x))^{1/2}$ if the orthogonality relations contain an additional positive weight function $w(x)$.] We use the φ_n to expand the delta function as

$$\delta(x-t) = \sum_{n=0}^{\infty} a_n(t)\varphi_n(x), \tag{1.188}$$

where the coefficients a_n are functions of the variable t. Multiplying by $\varphi_m(x)$ and integrating over the orthogonality interval (Eq. (1.187)), we have

$$a_m(t) = \int_a^b \delta(x-t)\varphi_m(x)\,dx = \varphi_m(t) \tag{1.189}$$

or

$$\delta(x-t) = \sum_{n=0}^{\infty} \varphi_n(t)\varphi_n(x) = \delta(t-x). \tag{1.190}$$

This series is assuredly not uniformly convergent (see Chapter 5), but it may be used as part of an integrand in which the ensuing integration will make it convergent (compare Section 5.5).

Suppose we form the integral $\int F(t)\delta(t-x)\,dx$, where it is assumed that $F(t)$ can be expanded in a series of orthogonal functions $\varphi_p(t)$, a property called *completeness*. We then obtain

$$\begin{aligned}\int F(t)\delta(t-x)\,dt &= \int \sum_{p=0}^{\infty} a_p\varphi_p(t) \sum_{n=0}^{\infty} \varphi_n(x)\varphi_n(t)\,dt \\ &= \sum_{p=0}^{\infty} a_p\varphi_p(x) = F(x),\end{aligned} \tag{1.191}$$

the cross products $\int \varphi_p\varphi_n\,dt\,(n \neq p)$ vanishing by orthogonality (Eq. (1.187)). Referring back to the definition of the Dirac delta function, Eq. (1.171b), we see that our series representation, Eq. (1.190), satisfies the defining property of the Dirac delta function and therefore is a representation of it. This representation of the Dirac delta function is called **closure**. The assumption of completeness of a set of functions for expansion of $\delta(x-t)$ yields the closure relation. The converse, that closure implies completeness, is the topic of Exercise 1.15.16.

Integral Representations for the Delta Function

Integral transforms, such as the Fourier integral

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \exp(i\omega t)\, dt$$

of Chapter 15, lead to the corresponding integral representations of Dirac's delta function. For example, take

$$\delta_n(t-x) = \frac{\sin n(t-x)}{\pi(t-x)} = \frac{1}{2\pi} \int_{-n}^{n} \exp\big(i\omega(t-x)\big)\, d\omega, \tag{1.192}$$

using Eq. (1.175). We have

$$f(x) = \lim_{n\to\infty} \int_{-\infty}^{\infty} f(t)\delta_n(t-x)\, dt, \tag{1.193a}$$

where $\delta_n(t-x)$ is the sequence in Eq. (1.192) defining the distribution $\delta(t-x)$. Note that Eq. (1.193a) assumes that $f(t)$ is continuous at $t = x$. If we substitute Eq. (1.192) into Eq. (1.193a) we obtain

$$f(x) = \lim_{n\to\infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) \int_{-n}^{n} \exp\big(i\omega(t-x)\big)\, d\omega\, dt. \tag{1.193b}$$

Interchanging the order of integration and then taking the limit as $n \to \infty$, we have the Fourier integral theorem, Eq. (15.20).

With the understanding that it belongs under an integral sign, as in Eq. (1.193a), the identification

$$\boxed{\delta(t-x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\big(i\omega(t-x)\big)\, d\omega} \tag{1.193c}$$

provides a very useful integral representation of the delta function.

When the Laplace transform (see Sections 15.1 and 15.9)

$$L_\delta(s) = \int_0^{\infty} \exp(-st)\delta(t-t_0) = \exp(-st_0), \qquad t_0 > 0 \tag{1.194}$$

is inverted, we obtain the complex representation

$$\delta(t-t_0) = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \exp\big(s(t-t_0)\big)\, ds, \tag{1.195}$$

which is essentially equivalent to the previous Fourier representation of Dirac's delta function.

Exercises

1.15.1 Let

$$\delta_n(x) = \begin{cases} 0, & x < -\frac{1}{2n}, \\ n, & -\frac{1}{2n} < x < \frac{1}{2n}, \\ 0, & \frac{1}{2n} < x. \end{cases}$$

Show that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f(x)\delta_n(x)\,dx = f(0),$$

assuming that $f(x)$ is continuous at $x = 0$.

1.15.2 Verify that the sequence $\delta_n(x)$, based on the function

$$\delta_n(x) = \begin{cases} 0, & x < 0, \\ ne^{-nx}, & x > 0, \end{cases}$$

is a delta sequence (satisfying Eq. (1.178)). Note that the singularity is at $+0$, the positive side of the origin.
Hint. Replace the upper limit (∞) by c/n, where c is large but finite, and use the mean value theorem of integral calculus.

1.15.3 For

$$\delta_n(x) = \frac{n}{\pi} \cdot \frac{1}{1+n^2x^2},$$

(Eq. (1.174)), show that

$$\int_{-\infty}^{\infty} \delta_n(x)\,dx = 1.$$

1.15.4 Demonstrate that $\delta_n = \sin nx/\pi x$ is a delta distribution by showing that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f(x)\frac{\sin nx}{\pi x}dx = f(0).$$

Assume that $f(x)$ is continuous at $x = 0$ and vanishes as $x \to \pm\infty$.
Hint. Replace x by y/n and take $\lim n \to \infty$ **before** integrating.

1.15.5 Fejer's method of summing series is associated with the function

$$\delta_n(t) = \frac{1}{2\pi n}\left[\frac{\sin(nt/2)}{\sin(t/2)}\right]^2.$$

Show that $\delta_n(t)$ is a delta distribution, in the sense that

$$\lim_{n\to\infty} \frac{1}{2\pi n} \int_{-\infty}^{\infty} f(t)\left[\frac{\sin(nt/2)}{\sin(t/2)}\right]^2 dt = f(0).$$

1.15.6 Prove that

$$\delta\big[a(x-x_1)\big] = \frac{1}{a}\delta(x-x_1).$$

Note. If $\delta[a(x-x_1)]$ is considered even, relative to x_1, the relation holds for negative a and $1/a$ may be replaced by $1/|a|$.

1.15.7 Show that

$$\delta\big[(x-x_1)(x-x_2)\big] = \big[\delta(x-x_1)+\delta(x-x_2)\big]/|x_1-x_2|.$$

Hint. Try using Exercise 1.15.6.

1.15.8 Using the Gauss error curve delta sequence ($\delta_n = \frac{n}{\sqrt{\pi}}\, e^{-n^2x^2}$), show that

$$x\frac{d}{dx}\delta(x) = -\delta(x),$$

treating $\delta(x)$ and its derivative as in Eq. (1.179).

1.15.9 Show that

$$\int_{-\infty}^{\infty} \delta'(x)\, f(x)\, dx = -f'(0).$$

Here we assume that $f'(x)$ is continuous at $x = 0$.

1.15.10 Prove that

$$\delta\big(f(x)\big) = \left|\frac{df(x)}{dx}\right|^{-1}_{x=x_0} \delta(x-x_0),$$

where x_0 is chosen so that $f(x_0) = 0$.
Hint. Note that $\delta(f)\, df = \delta(x)\, dx$.

1.15.11 Show that in spherical polar coordinates $(r, \cos\theta, \varphi)$ the delta function $\delta(\mathbf{r}_1 - \mathbf{r}_2)$ becomes

$$\frac{1}{r_1^2}\delta(r_1-r_2)\delta(\cos\theta_1-\cos\theta_2)\delta(\varphi_1-\varphi_2).$$

Generalize this to the curvilinear coordinates (q_1, q_2, q_3) of Section 2.1 with scale factors h_1, h_2, and h_3.

1.15.12 A rigorous development of Fourier transforms[31] includes as a theorem the relations

$$\lim_{a\to\infty} \frac{2}{\pi}\int_{x_1}^{x_2} f(u+x)\frac{\sin ax}{x}dx$$

$$= \begin{cases} f(u+0)+f(u-0), & x_1 < 0 < x_2 \\ f(u+0), & x_1 = 0 < x_2 \\ f(u-0), & x_1 < 0 = x_2 \\ 0, & x_1 < x_2 < 0 \text{ or } 0 < x_1 < x_2. \end{cases}$$

Verify these results using the Dirac delta function.

[31] I. N. Sneddon, *Fourier Transforms*. New York: McGraw-Hill (1951).

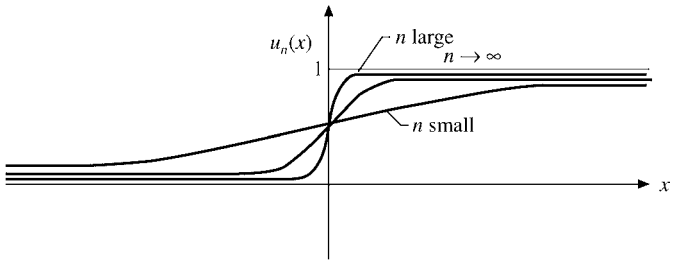

FIGURE 1.41 $\frac{1}{2}[1 + \tanh nx]$ and the Heaviside unit step function.

1.15.13 (a) If we define a sequence $\delta_n(x) = n/(2\cosh^2 nx)$, show that

$$\int_{-\infty}^{\infty} \delta_n(x)\,dx = 1, \qquad \text{independent of } n.$$

(b) Continuing this analysis, show that[32]

$$\int_{-\infty}^{x} \delta_n(x)\,dx = \frac{1}{2}[1 + \tanh nx] \equiv u_n(x),$$

$$\lim_{n\to\infty} u_n(x) = \begin{cases} 0, & x < 0, \\ 1, & x > 0. \end{cases}$$

This is the Heaviside unit step function (Fig. 1.41).

1.15.14 Show that the unit step function $u(x)$ may be represented by

$$u(x) = \frac{1}{2} + \frac{1}{2\pi i} P \int_{-\infty}^{\infty} e^{ixt} \frac{dt}{t},$$

where P means Cauchy principal value (Section 7.1).

1.15.15 As a variation of Eq. (1.175), take

$$\delta_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ixt - |t|/n}\,dt.$$

Show that this reduces to $(n/\pi)1/(1 + n^2x^2)$, Eq. (1.174), and that

$$\int_{-\infty}^{\infty} \delta_n(x)\,dx = 1.$$

Note. In terms of integral transforms, the initial equation here may be interpreted as either a Fourier exponential transform of $e^{-|t|/n}$ or a Laplace transform of e^{ixt}.

[32] Many other symbols are used for this function. This is the AMS-55 (see footnote 4 on p. 330 for the reference) notation: u for unit.

1.15.16 (a) The Dirac delta function representation given by Eq. (1.190),

$$\delta(x-t)=\sum_{n=0}^{\infty}\varphi_n(x)\varphi_n(t),$$

is often called the **closure relation**. For an orthonormal set of real functions, φ_n, show that closure implies completeness, that is, Eq. (1.191) follows from Eq. (1.190).

Hint. One can take

$$F(x)=\int F(t)\delta(x-t)\,dt.$$

(b) Following the hint of part (a) you encounter the integral $\int F(t)\varphi_n(t)\,dt$. How do you know that this integral is finite?

1.15.17 For the finite interval $(-\pi,\pi)$ write the Dirac delta function $\delta(x-t)$ as a series of sines and cosines: $\sin nx, \cos nx, n=0,1,2,\ldots$. Note that although these functions are orthogonal, they are not normalized to unity.

1.15.18 In the interval $(-\pi,\pi)$, $\delta_n(x)=\frac{n}{\sqrt{\pi}}\exp(-n^2x^2)$.

(a) Write $\delta_n(x)$ as a Fourier cosine series.
(b) Show that your Fourier series agrees with a Fourier expansion of $\delta(x)$ in the limit as $n\to\infty$.
(c) Confirm the delta function nature of your Fourier series by showing that for any $f(x)$ that is finite in the interval $[-\pi,\pi]$ and continuous at $x=0$,

$$\int_{-\pi}^{\pi} f(x)\big[\text{Fourier expansion of } \delta_\infty(x)\big]\,dx=f(0).$$

1.15.19 (a) Write $\delta_n(x)=\frac{n}{\sqrt{\pi}}\exp(-n^2x^2)$ in the interval $(-\infty,\infty)$ as a Fourier integral and compare the limit $n\to\infty$ with Eq. (1.193c).
(b) Write $\delta_n(x)=n\exp(-nx)$ as a Laplace transform and compare the limit $n\to\infty$ with Eq. (1.195).

Hint. See Eqs. (15.22) and (15.23) for (a) and Eq. (15.212) for (b).

1.15.20 (a) Show that the Dirac delta function $\delta(x-a)$, expanded in a Fourier sine series in the half-interval $(0,L)$, $(0<a<L)$, is given by

$$\delta(x-a)=\frac{2}{L}\sum_{n=1}^{\infty}\sin\left(\frac{n\pi a}{L}\right)\sin\left(\frac{n\pi x}{L}\right).$$

Note that this series actually describes

$$-\delta(x+a)+\delta(x-a)\qquad\text{in the interval } (-L,L).$$

(b) By integrating both sides of the preceding equation from 0 to x, show that the cosine expansion of the square wave

$$f(x)=\begin{cases}0, & 0\leqslant x<a\\ 1, & a<x<L,\end{cases}$$

is, for $0 \leqslant x < L$,

$$f(x) = \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin\left(\frac{n\pi a}{L}\right) - \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin\left(\frac{n\pi a}{L}\right)\cos\left(\frac{n\pi x}{L}\right).$$

(c) Verify that the term

$$\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin\left(\frac{n\pi a}{L}\right) \qquad \text{is} \qquad \langle f(x)\rangle \equiv \frac{1}{L}\int_0^L f(x)\,dx.$$

1.15.21 Verify the Fourier cosine expansion of the square wave, Exercise 1.15.20(b), by direct calculation of the Fourier coefficients.

1.15.22 We may define a sequence

$$\delta_n(x) = \begin{cases} n, & |x| < 1/2n, \\ 0, & |x| > 1/2n. \end{cases}$$

(This is Eq. (1.172).) Express $\delta_n(x)$ as a Fourier integral (via the Fourier integral theorem, inverse transform, etc.). Finally, show that we may write

$$\delta(x) = \lim_{n\to\infty}\delta_n(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ikx}\,dk.$$

1.15.23 Using the sequence

$$\delta_n(x) = \frac{n}{\sqrt{\pi}}\exp\left(-n^2x^2\right),$$

show that

$$\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ikx}\,dk.$$

Note. Remember that $\delta(x)$ is defined in terms of its behavior as part of an integrand—especially Eqs. (1.178) and (1.189).

1.15.24 Derive sine and cosine representations of $\delta(t-x)$ that are comparable to the exponential representation, Eq. (1.193c).

ANS. $\frac{2}{\pi}\int_0^\infty \sin\omega t \sin\omega x\,d\omega$, $\frac{2}{\pi}\int_0^\infty \cos\omega t\cos\omega x\,d\omega$.

1.16 HELMHOLTZ'S THEOREM

In Section 1.13 it was emphasized that the choice of a magnetic vector potential **A** was not unique. The divergence of **A** was still undetermined. In this section two theorems about the divergence and curl of a vector are developed. The first theorem is as follows:

A vector is uniquely specified by giving its divergence and its curl within a simply connected region (without holes) and its normal component over the boundary.

Note that the subregions, where the divergence and curl are defined (often in terms of Dirac delta functions), are part of our region and are not supposed to be removed here or in Helmholtz's theorem, which follows. Let us take

$$\begin{aligned} \nabla \cdot \mathbf{V}_1 &= s, \\ \nabla \times \mathbf{V}_1 &= \mathbf{c}, \end{aligned} \tag{1.196}$$

where s may be interpreted as a source (charge) density and $\mathbf{c}$ as a circulation (current) density. Assuming also that the normal component V_{1n} on the boundary is given, we want to show that $\mathbf{V}_1$ is unique. We do this by assuming the existence of a second vector, $\mathbf{V}_2$, which satisfies Eq. (1.196) and has the same normal component over the boundary, and then showing that $\mathbf{V}_1 - \mathbf{V}_2 = 0$. Let

$$\mathbf{W} = \mathbf{V}_1 - \mathbf{V}_2.$$

Then

$$\nabla \cdot \mathbf{W} = 0 \tag{1.197}$$

and

$$\nabla \times \mathbf{W} = 0. \tag{1.198}$$

Since $\mathbf{W}$ is irrotational we may write (by Section (1.13))

$$\mathbf{W} = -\nabla\varphi. \tag{1.199}$$

Substituting this into Eq. (1.197), we obtain

$$\nabla \cdot \nabla\varphi = 0, \tag{1.200}$$

Laplace's equation.

Now we draw upon Green's theorem in the form given in Eq. (1.105), letting u and v each equal φ. Since

$$W_n = V_{1n} - V_{2n} = 0 \tag{1.201}$$

on the boundary, Green's theorem reduces to

$$\int_V (\nabla\varphi) \cdot (\nabla\varphi)\, d\tau = \int_V \mathbf{W} \cdot \mathbf{W}\, d\tau = 0. \tag{1.202}$$

The quantity $\mathbf{W} \cdot \mathbf{W} = W^2$ is nonnegative and so we must have

$$\mathbf{W} = \mathbf{V}_1 - \mathbf{V}_2 = 0 \tag{1.203}$$

everywhere. Thus $\mathbf{V}_1$ is unique, proving the theorem.

For our magnetic vector potential $\mathbf{A}$ the relation $\mathbf{B} = \nabla \times \mathbf{A}$ specifies the curl of $\mathbf{A}$. Often for convenience we set $\nabla \cdot \mathbf{A} = 0$ (compare Exercise 1.14.4). Then (with boundary conditions) $\mathbf{A}$ is fixed.

This theorem may be written as a uniqueness theorem for solutions of Laplace's equation, Exercise 1.16.1. In this form, this uniqueness theorem is of great importance in solving electrostatic and other Laplace equation boundary value problems. If we can find a solution of Laplace's equation that satisfies the necessary boundary conditions, then our solution is the complete solution. Such boundary value problems are taken up in Sections 12.3 and 12.5.

Helmholtz's Theorem

The second theorem we shall prove is Helmholtz's theorem.

A vector $\mathbf{V}$ *satisfying Eq.* (1.196) *with both source and circulation densities vanishing at infinity may be written as the sum of two parts, one of which is irrotational, the other of which is solenoidal.*

Note that our region is simply connected, being all of space, for simplicity. Helmholtz's theorem will clearly be satisfied if we may write $\mathbf{V}$ as

$$\mathbf{V} = -\nabla\varphi + \nabla \times \mathbf{A}, \tag{1.204a}$$

$-\nabla\varphi$ being irrotational and $\nabla \times A$ being solenoidal. We proceed to justify Eq. (1.204a).

$\mathbf{V}$ is a known vector. We take the divergence and curl

$$\nabla \cdot \mathbf{V} = s(\mathbf{r}) \tag{1.204b}$$

$$\nabla \times \mathbf{V} = \mathbf{c}(\mathbf{r}) \tag{1.204c}$$

with $s(\mathbf{r})$ and $\mathbf{c}(\mathbf{r})$ now known functions of position. From these two functions we construct a scalar potential $\varphi(\mathbf{r}_1)$,

$$\varphi(\mathbf{r}_1) = \frac{1}{4\pi} \int \frac{s(\mathbf{r}_2)}{r_{12}} d\tau_2, \tag{1.205a}$$

and a vector potential $\mathbf{A}(\mathbf{r}_1)$,

$$\mathbf{A}(\mathbf{r}_1) = \frac{1}{4\pi} \int \frac{\mathbf{c}(\mathbf{r}_2)}{r_{12}} d\tau_2. \tag{1.205b}$$

If $s = 0$, then $\mathbf{V}$ is solenoidal and Eq. (1.205a) implies $\varphi = 0$. From Eq. (1.204a), $\mathbf{V} = \nabla \times \mathbf{A}$, with $\mathbf{A}$ as given in Eq. (1.141), which is consistent with Section 1.13. Further, if $\mathbf{c} = 0$, then $\mathbf{V}$ is irrotational and Eq. (1.205b) implies $\mathbf{A} = 0$, and Eq. (1.204a) implies $\mathbf{V} = -\nabla\varphi$, consistent with scalar potential theory of Section 1.13.

Here the argument $\mathbf{r}_1$ indicates (x_1, y_1, z_1), the field point; $\mathbf{r}_2$, the coordinates of the source point (x_2, y_2, z_2), whereas

$$r_{12} = \left[(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2\right]^{1/2}. \tag{1.206}$$

When a direction is associated with r_{12}, the positive direction is taken to be away from the source and toward the field point. Vectorially, $\mathbf{r}_{12} = \mathbf{r}_1 - \mathbf{r}_2$, as shown in Fig. 1.42. Of course, s and $\mathbf{c}$ must vanish sufficiently rapidly at large distance so that the integrals exist. The actual expansion and evaluation of integrals such as Eqs. (1.205a) and (1.205b) is treated in Section 12.1.

From the uniqueness theorem at the beginning of this section, $\mathbf{V}$ is uniquely specified by its divergence, s, and curl, $\mathbf{c}$ (and boundary conditions). Returning to Eq. (1.204a), we have

$$\nabla \cdot \mathbf{V} = -\nabla \cdot \nabla\varphi, \tag{1.207a}$$

the divergence of the curl vanishing, and

$$\nabla \times \mathbf{V} = \nabla \times (\nabla \times \mathbf{A}), \tag{1.207b}$$

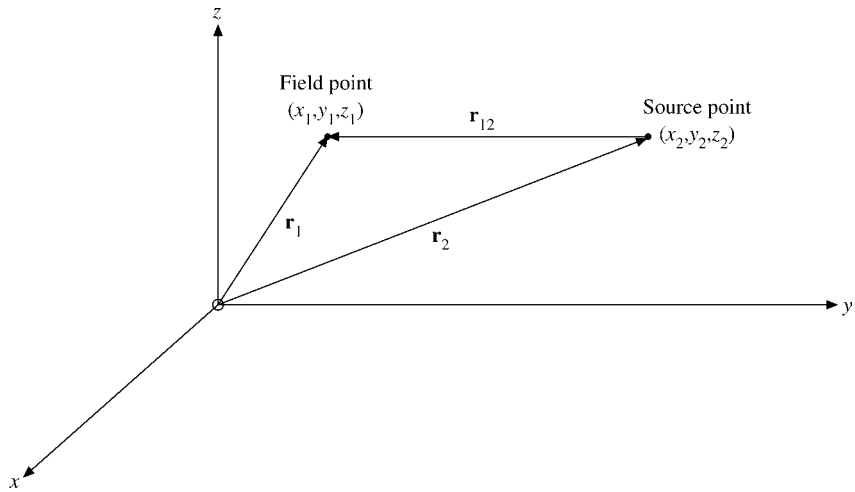

FIGURE 1.42 Source and field points.

the curl of the gradient vanishing. If we can show that

$$-\nabla \cdot \nabla\varphi(\mathbf{r}_1) = s(\mathbf{r}_1) \tag{1.207c}$$

and

$$\nabla \times \big(\nabla \times \mathbf{A}(\mathbf{r}_1)\big) = \mathbf{c}(\mathbf{r}_1), \tag{1.207d}$$

then $\mathbf{V}$ as given in Eq. (1.204a) will have the proper divergence and curl. Our description will be internally consistent and Eq. (1.204a) justified.[33]

First, we consider the divergence of $\mathbf{V}$:

$$\nabla \cdot \mathbf{V} = -\nabla \cdot \nabla\varphi = -\frac{1}{4\pi}\nabla \cdot \nabla \int \frac{s(\mathbf{r}_2)}{r_{12}} d\tau_2. \tag{1.208}$$

The Laplacian operator, $\nabla \cdot \nabla$, or ∇^2, operates on the field coordinates (x_1, y_1, z_1) and so commutes with the integration with respect to (x_2, y_2, z_2). We have

$$\nabla \cdot \mathbf{V} = -\frac{1}{4\pi} \int s(\mathbf{r}_2) \nabla_1^2 \left(\frac{1}{r_{12}}\right) d\tau_2. \tag{1.209}$$

We must make two minor modifications in Eq. (1.169) before applying it. First, our source is at r_2, not at the origin. This means that a nonzero result from Gauss' law appears if and only if the surface S includes the point $\mathbf{r} = \mathbf{r}_2$. To show this, we rewrite Eq. (1.170):

$$\nabla^2 \left(\frac{1}{r_{12}}\right) = -4\pi\delta(\mathbf{r}_1 - \mathbf{r}_2). \tag{1.210}$$

[33] Alternatively, we could solve Eq. (1.207c), Poisson's equation, and compare the solution with the constructed potential, Eq. (1.205a). The solution of Poisson's equation is developed in Section 9.7.

This shift of the source to $\mathbf{r}_2$ may be incorporated in the defining equation (1.171b) as

$$\delta(\mathbf{r}_1 - \mathbf{r}_2) = 0, \qquad \mathbf{r}_1 \neq \mathbf{r}_2, \tag{1.211a}$$

$$\int f(\mathbf{r}_1)\delta(\mathbf{r}_1 - \mathbf{r}_2)\, d\tau_1 = f(\mathbf{r}_2). \tag{1.211b}$$

Second, noting that differentiating r_{12}^{-1} twice with respect to x_2, y_2, z_2 is the same as differentiating **twice** with respect to x_1, y_1, z_1, we have

$$\begin{aligned}\nabla_1^2\left(\frac{1}{r_{12}}\right) = \nabla_2^2\left(\frac{1}{r_{12}}\right) &= -4\pi\delta(\mathbf{r}_1 - \mathbf{r}_2)\\ &= -4\pi\delta(\mathbf{r}_2 - \mathbf{r}_1).\end{aligned} \tag{1.212}$$

Rewriting Eq. (1.209) and using the Dirac delta function, Eq. (1.212), we may integrate to obtain

$$\begin{aligned}\nabla\cdot\mathbf{V} &= -\frac{1}{4\pi}\int s(\mathbf{r}_2)\nabla_2^2\left(\frac{1}{r_{12}}\right)d\tau_2\\ &= -\frac{1}{4\pi}\int s(\mathbf{r}_2)(-4\pi)\delta(\mathbf{r}_2 - \mathbf{r}_1)\, d\tau_2\\ &= s(\mathbf{r}_1).\end{aligned} \tag{1.213}$$

The final step follows from Eq. (1.211b), with the subscripts 1 and 2 exchanged. Our result, Eq. (1.213), shows that the assumed forms of $\mathbf{V}$ and of the scalar potential φ are in agreement with the given divergence (Eq. (1.204b)).

To complete the proof of Helmholtz's theorem, we need to show that our assumptions are consistent with Eq. (1.204c), that is, that the curl of $\mathbf{V}$ is equal to $\mathbf{c}(\mathbf{r}_1)$. From Eq. (1.204a),

$$\begin{aligned}\nabla\times\mathbf{V} &= \nabla\times(\nabla\times\mathbf{A})\\ &= \nabla\nabla\cdot\mathbf{A} - \nabla^2\mathbf{A}.\end{aligned} \tag{1.214}$$

The first term, $\nabla\nabla\cdot\mathbf{A}$, leads to

$$4\pi\nabla\nabla\cdot\mathbf{A} = \int \mathbf{c}(\mathbf{r}_2)\cdot\nabla_1\nabla_1\left(\frac{1}{r_{12}}\right)d\tau_2 \tag{1.215}$$

by Eq. (1.205b). Again replacing the second derivatives with respect to x_1, y_1, z_1 by second derivatives with respect to x_2, y_2, z_2, we integrate each component[34] of Eq. (1.215) by parts:

$$\begin{aligned}4\pi\nabla\nabla\cdot\mathbf{A}|_x &= \int \mathbf{c}(\mathbf{r}_2)\cdot\nabla_2\frac{\partial}{\partial x_2}\left(\frac{1}{r_{12}}\right)d\tau_2\\ &= \int \nabla_2\cdot\left[\mathbf{c}(\mathbf{r}_2)\frac{\partial}{\partial x_2}\left(\frac{1}{r_{12}}\right)\right]d\tau_2\\ &\quad - \int\left[\nabla_2\cdot\mathbf{c}(\mathbf{r}_2)\right]\frac{\partial}{\partial x_2}\left(\frac{1}{r_{12}}\right)d\tau_2.\end{aligned} \tag{1.216}$$

[34] This avoids creating the **tensor** $\mathbf{c}(\mathbf{r}_2)\nabla_2$.

The second integral vanishes because the circulation density $\mathbf{c}$ is solenoidal.[35] The first integral may be transformed to a surface integral by Gauss' theorem. If $\mathbf{c}$ is bounded in space or vanishes faster that $1/r$ for large r, so that the integral in Eq. (1.205b) exists, then by choosing a sufficiently large surface the first integral on the right-hand side of Eq. (1.216) also vanishes.

With $\nabla\nabla\cdot\mathbf{A}=0$, Eq. (1.214) now reduces to

$$\nabla\times\mathbf{V}=-\nabla^2\mathbf{A}=-\frac{1}{4\pi}\int\mathbf{c}(\mathbf{r}_2)\nabla_1^2\left(\frac{1}{r_{12}}\right)d\tau_2. \tag{1.217}$$

This is exactly like Eq. (1.209) except that the scalar $s(\mathbf{r}_2)$ is replaced by the vector circulation density $\mathbf{c}(\mathbf{r}_2)$. Introducing the Dirac delta function, as before, as a convenient way of carrying out the integration, we find that Eq. (1.217) reduces to Eq. (1.196). We see that our assumed forms of $\mathbf{V}$, given by Eq. (1.204a), and of the vector potential $\mathbf{A}$, given by Eq. (1.205b), are in agreement with Eq. (1.196) specifying the curl of $\mathbf{V}$.

This completes the proof of Helmholtz's theorem, showing that a vector may be resolved into irrotational and solenoidal parts. Applied to the electromagnetic field, we have resolved our field vector $\mathbf{V}$ into an irrotational electric field $\mathbf{E}$, derived from a scalar potential φ, and a solenoidal magnetic induction field $\mathbf{B}$, derived from a vector potential $\mathbf{A}$. The source density $s(\mathbf{r})$ may be interpreted as an electric charge density (divided by electric permittivity ε), whereas the circulation density $\mathbf{c}(\mathbf{r})$ becomes electric current density (times magnetic permeability μ).

Exercises

1.16.1 Implicit in this section is a proof that a function $\psi(\mathbf{r})$ is uniquely specified by requiring it to (1) satisfy Laplace's equation and (2) satisfy a complete set of boundary conditions. Develop this proof explicitly.

1.16.2 (a) Assuming that $\mathbf{P}$ is a solution of the vector Poisson equation, $\nabla_1^2\mathbf{P}(\mathbf{r}_1)=-\mathbf{V}(\mathbf{r}_1)$, develop an alternate proof of Helmholtz's theorem, showing that $\mathbf{V}$ may be written as

$$\mathbf{V}=-\nabla\varphi+\nabla\times\mathbf{A},$$

where

$$\mathbf{A}=\nabla\times\mathbf{P},$$

and

$$\varphi=\nabla\cdot\mathbf{P}.$$

(b) Solving the vector Poisson equation, we find

$$\mathbf{P}(\mathbf{r}_1)=\frac{1}{4\pi}\int_V\frac{\mathbf{V}(\mathbf{r}_2)}{r_{12}}d\tau_2.$$

Show that this solution substituted into φ and $\mathbf{A}$ of part (a) leads to the expressions given for φ and $\mathbf{A}$ in Section 1.16.

[35] Remember, $\mathbf{c}=\nabla\times\mathbf{V}$ is known.

Additional Readings

Borisenko, A. I., and I. E. Taropov, *Vector and Tensor Analysis with Applications*. Englewood Cliffs, NJ: Prentice-Hall (1968). Reprinted, Dover (1980).

Davis, H. F., and A. D. Snider, *Introduction to Vector Analysis*, 7th ed. Boston: Allyn & Bacon (1995).

Kellogg, O. D., *Foundations of Potential Theory*. New York: Dover (1953). Originally published (1929). The classic text on potential theory.

Lewis, P. E., and J. P. Ward, *Vector Analysis for Engineers and Scientists*. Reading, MA: Addison-Wesley (1989).

Marion, J. B., *Principles of Vector Analysis*. New York: Academic Press (1965). A moderately advanced presentation of vector analysis oriented toward tensor analysis. Rotations and other transformations are described with the appropriate matrices.

Spiegel, M. R., *Vector Analysis*. New York: McGraw-Hill (1989).

Tai, C.-T., *Generalized Vector and Dyadic Analysis*. Oxford: Oxford University Press (1996).

Wrede, R. C., *Introduction to Vector and Tensor Analysis*. New York: Wiley (1963). Reprinted, New York: Dover (1972). Fine historical introduction. Excellent discussion of differentiation of vectors and applications to mechanics.

This page intentionally left blank

CHAPTER 2

VECTOR ANALYSIS IN CURVED COORDINATES AND TENSORS

In Chapter 1 we restricted ourselves almost completely to rectangular or Cartesian coordinate systems. A Cartesian coordinate system offers the unique advantage that all three unit vectors, $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$, are constant in direction as well as in magnitude. We did introduce the radial distance r, but even this was treated as a function of x, y, and z. Unfortunately, not all physical problems are well adapted to a solution in Cartesian coordinates. For instance, if we have a central force problem, $\mathbf{F} = \hat{\mathbf{r}} F(r)$, such as gravitational or electrostatic force, Cartesian coordinates may be unusually inappropriate. Such a problem demands the use of a coordinate system in which the radial distance is taken to be one of the coordinates, that is, spherical polar coordinates.

The point is that the coordinate system should be chosen to fit the problem, to exploit any constraint or symmetry present in it. Then it is likely to be more readily soluble than if we had forced it into a Cartesian framework.

Naturally, there is a price that must be paid for the use of a non-Cartesian coordinate system. We have not yet written expressions for gradient, divergence, or curl in any of the non-Cartesian coordinate systems. Such expressions are developed in general form in Section 2.2. First, we develop a system of curvilinear coordinates, a general system that may be specialized to any of the particular systems of interest. We shall specialize to circular cylindrical coordinates in Section 2.4 and to spherical polar coordinates in Section 2.5.

2.1 ORTHOGONAL COORDINATES IN $\mathbb{R}^3$

In Cartesian coordinates we deal with three mutually perpendicular families of planes: $x =$ constant, $y =$ constant, and $z =$ constant. Imagine that we superimpose on this system

three other families of surfaces $q_i(x, y, z),\ i = 1, 2, 3$. The surfaces of any one family q_i need not be parallel to each other and they need not be planes. If this is difficult to visualize, the figure of a specific coordinate system, such as Fig. 2.3, may be helpful. The three new families of surfaces need not be mutually perpendicular, but for simplicity we impose this condition (Eq. (2.7)) because orthogonal coordinates are common in physical applications. This orthogonality has many advantages: Orthogonal coordinates are almost like Cartesian coordinates where infinitesimal areas and volumes are products of coordinate differentials.

In this section we develop the general formalism of orthogonal coordinates, derive from the geometry the coordinate differentials, and use them for line, area, and volume elements in multiple integrals and vector operators. We may describe any point (x, y, z) as the intersection of three planes in Cartesian coordinates or as the intersection of the three surfaces that form our new, curvilinear coordinates. Describing the curvilinear coordinate surfaces by $q_1 =$ constant, $q_2 =$ constant, $q_3 =$ constant, we may identify our point by (q_1, q_2, q_3) as well as by (x, y, z):

$$\begin{array}{cc} \text{General curvilinear coordinates} & \text{Circular cylindrical coordinates} \\ q_1, q_2, q_3 & \rho, \varphi, z \\ x = x(q_1, q_2, q_3) & -\infty < x = \rho\cos\varphi < \infty \\ y = y(q_1, q_2, q_3) & -\infty < y = \rho\sin\varphi < \infty \\ z = z(q_1, q_2, q_3) & -\infty < z = z < \infty \end{array} \tag{2.1}$$

specifying x, y, z in terms of q_1, q_2, q_3 and the inverse relations

$$\begin{array}{ll} q_1 = q_1(x, y, z) & 0 \leqslant \rho = \left(x^2 + y^2\right)^{1/2} < \infty \\ q_2 = q_2(x, y, z) & 0 \leqslant \varphi = \arctan(y/x) < 2\pi \\ q_3 = q_3(x, y, z) & -\infty < z = z < \infty. \end{array} \tag{2.2}$$

As a specific illustration of the general, abstract q_1, q_2, q_3, the transformation equations for circular cylindrical coordinates (Section 2.4) are included in Eqs. (2.1) and (2.2). With each family of surfaces $q_i =$ constant, we can associate a unit vector $\hat{\mathbf{q}}_i$ normal to the surface $q_i =$ constant and in the direction of increasing q_i. In general, these unit vectors will depend on the position in space. Then a vector $\mathbf{V}$ may be written

$$\boxed{\mathbf{V} = \hat{\mathbf{q}}_1 V_1 + \hat{\mathbf{q}}_2 V_2 + \hat{\mathbf{q}}_3 V_3,} \tag{2.3}$$

but the coordinate or position vector is different in general,

$$\mathbf{r} \neq \hat{\mathbf{q}}_1 q_1 + \hat{\mathbf{q}}_2 q_2 + \hat{\mathbf{q}}_3 q_3,$$

as the special cases $\mathbf{r} = r\hat{\mathbf{r}}$ for spherical polar coordinates and $\mathbf{r} = \rho\hat{\boldsymbol{\rho}} + z\hat{z}$ for cylindrical coordinates demonstrate. The $\hat{\mathbf{q}}_i$ are normalized to $\hat{\mathbf{q}}_i^2 = 1$ and form a right-handed coordinate system with volume $\hat{\mathbf{q}}_1 \cdot (\hat{\mathbf{q}}_2 \times \hat{\mathbf{q}}_3) > 0$.

Differentiation of x in Eqs. (2.1) leads to the total variation or differential

$$dx = \frac{\partial x}{\partial q_1} dq_1 + \frac{\partial x}{\partial q_2} dq_2 + \frac{\partial x}{\partial q_3} dq_3, \tag{2.4}$$

and similarly for differentiation of y and z. In vector notation $d\mathbf{r} = \sum_i \frac{\partial \mathbf{r}}{\partial q_i} dq_i$. From the Pythagorean theorem in Cartesian coordinates the square of the distance between two neighboring points is

$$ds^2 = dx^2 + dy^2 + dz^2.$$

Substituting $d\mathbf{r}$ shows that in our curvilinear coordinate space the square of the distance element can be written as a quadratic form in the differentials dq_i:

$$\begin{aligned} ds^2 = d\mathbf{r}\cdot d\mathbf{r} = d\mathbf{r}^2 &= \sum_{ij} \frac{\partial \mathbf{r}}{\partial q_i}\cdot\frac{\partial \mathbf{r}}{\partial q_j}\,dq_i\,dq_j \\ &= g_{11}\,dq_1^2 + g_{12}\,dq_1\,dq_2 + g_{13}\,dq_1\,dq_3 \\ &\quad + g_{21}\,dq_2\,dq_1 + g_{22}\,dq_2^2 + g_{23}\,dq_2\,dq_3 \\ &\quad + g_{31}\,dq_3\,dq_1 + g_{32}\,dq_3\,dq_2 + g_{33}\,dq_3^2 \\ &= \sum_{ij} g_{ij}\,dq_i\,dq_j, \end{aligned} \tag{2.5}$$

where nonzero mixed terms $dq_i\,dq_j$ with $i \neq j$ signal that these coordinates are not orthogonal, that is, that the tangential directions $\hat{\mathbf{q}}_i$ are not mutually orthogonal. Spaces for which Eq. (2.5) is a legitimate expression are called *metric* or *Riemannian*.

Writing Eq. (2.5) more explicitly, we see that

$$g_{ij}(q_1,q_2,q_3) = \frac{\partial x}{\partial q_i}\frac{\partial x}{\partial q_j} + \frac{\partial y}{\partial q_i}\frac{\partial y}{\partial q_j} + \frac{\partial z}{\partial q_i}\frac{\partial z}{\partial q_j} = \frac{\partial \mathbf{r}}{\partial q_i}\cdot\frac{\partial \mathbf{r}}{\partial q_j} \tag{2.6}$$

are scalar products of the **tangent vectors** $\frac{\partial \mathbf{r}}{\partial q_i}$ to the curves $\mathbf{r}$ for $q_j = \text{const.}$, $j \neq i$. These coefficient functions g_{ij}, which we now proceed to investigate, may be viewed as specifying the nature of the coordinate system (q_1, q_2, q_3). Collectively these coefficients are referred to as the **metric** and in Section 2.10 will be shown to form a second-rank symmetric tensor.[1] In general relativity the metric components are determined by the properties of matter; that is, the g_{ij} are solutions of Einstein's field equations with the energy–momentum tensor as driving term; this may be articulated as "geometry is merged with physics."

At usual we limit ourselves to orthogonal (mutually perpendicular surfaces) coordinate systems, which means (see Exercise 2.1.1)[2]

$$g_{ij} = 0, \qquad i \neq j, \tag{2.7}$$

and $\hat{\mathbf{q}}_i \cdot \hat{\mathbf{q}}_j = \delta_{ij}$. (Nonorthogonal coordinate systems are considered in some detail in Sections 2.10 and 2.11 in the framework of tensor analysis.) Now, to simplify the notation, we write $g_{ii} = h_i^2 > 0$, so

$$\boxed{ds^2 = (h_1\,dq_1)^2 + (h_2\,dq_2)^2 + (h_3\,dq_3)^2 = \sum_i (h_i\,dq_i)^2.} \tag{2.8}$$

[1] The tensor nature of the set of g_{ij}'s follows from the quotient rule (Section 2.8). Then the tensor transformation law yields Eq. (2.5).

[2] In relativistic cosmology the nondiagonal elements of the metric g_{ij} are usually set equal to zero as a consequence of physical assumptions such as no rotation, as for $d\varphi\,dt, d\theta\,dt$.

The specific orthogonal coordinate systems are described in subsequent sections by specifying these (positive) scale factors h_1, h_2, and h_3. Conversely, the scale factors may be conveniently identified by the relation

$$ds_i = h_i\,dq_i, \qquad \frac{\partial \mathbf{r}}{\partial q_i} = h_i \hat{\mathbf{q}}_i \tag{2.9}$$

for any given dq_i, holding all other q constant. Here, ds_i is a differential length along the direction $\hat{\mathbf{q}}_i$. Note that the three curvilinear coordinates q_1, q_2, q_3 need not be lengths. The scale factors h_i may depend on q and they may have dimensions. The **product** $h_i\,dq_i$ must have a dimension of length. The differential distance vector $d\mathbf{r}$ may be written

$$d\mathbf{r} = h_1\,dq_1\,\hat{\mathbf{q}}_1 + h_2\,dq_2\,\hat{\mathbf{q}}_2 + h_3\,dq_3\,\hat{\mathbf{q}}_3 = \sum_i h_i\,dq_i\,\hat{\mathbf{q}}_i.$$

Using this curvilinear component form, we find that a line integral becomes

$$\boxed{\int \mathbf{V}\cdot d\mathbf{r} = \sum_i \int V_i h_i\,dq_i.}$$

From Eqs. (2.9) we may immediately develop the area and volume elements

$$d\sigma_{ij} = ds_i\,ds_j = h_i h_j\,dq_i\,dq_j \tag{2.10}$$

and

$$d\tau = ds_1\,ds_2\,ds_3 = h_1h_2h_3\,dq_1\,dq_2\,dq_3. \tag{2.11}$$

The expressions in Eqs. (2.10) and (2.11) agree, of course, with the results of using the transformation equations, Eq. (2.1), and Jacobians (described shortly; see also Exercise 2.1.5).

From Eq. (2.10) an area element may be expanded:

$$\begin{aligned} d\boldsymbol{\sigma} &= ds_2\,ds_3\,\hat{\mathbf{q}}_1 + ds_3\,ds_1\,\hat{\mathbf{q}}_2 + ds_1\,ds_2\,\hat{\mathbf{q}}_3 \\ &= h_2h_3\,dq_2\,dq_3\,\hat{\mathbf{q}}_1 + h_3h_1\,dq_3\,dq_1\,\hat{\mathbf{q}}_2 \\ &\quad + h_1h_2\,dq_1\,dq_2\,\hat{\mathbf{q}}_3. \end{aligned}$$

A surface integral becomes

$$\boxed{\begin{aligned} \int \mathbf{V}\cdot d\boldsymbol{\sigma} &= \int V_1h_2h_3\,dq_2\,dq_3 + \int V_2h_3h_1\,dq_3\,dq_1 \\ &\quad + \int V_3h_1h_2\,dq_1\,dq_2. \end{aligned}}$$

(Examples of such line and surface integrals appear in Sections 2.4 and 2.5.)

In anticipation of the new forms of equations for vector **calculus** that appear in the next section, let us emphasize that vector **algebra** is the same in orthogonal curvilinear coordinates as in Cartesian coordinates. Specifically, for the dot product,

$$\boxed{\begin{aligned}\mathbf{A}\cdot\mathbf{B} &= \sum_{ik} A_i\hat{\mathbf{q}}_i\cdot\hat{\mathbf{q}}_k B_k = \sum_{ik} A_i B_k\delta_{ik}\\ &= \sum_i A_i B_i = A_1B_1 + A_2B_2 + A_3B_3,\end{aligned}} \tag{2.12}$$

where the subscripts indicate curvilinear components. For the cross product,

$$\boxed{\mathbf{A}\times\mathbf{B} = \begin{vmatrix}\hat{\mathbf{q}}_1 & \hat{\mathbf{q}}_2 & \hat{\mathbf{q}}_3\\ A_1 & A_2 & A_3\\ B_1 & B_2 & B_3\end{vmatrix}} \tag{2.13}$$

as in Eq. (1.40).

Previously, we specialized to locally rectangular coordinates that are adapted to special symmetries. Let us now briefly look at the more general case, where the coordinates are not necessarily orthogonal. Surface and volume elements are part of multiple integrals, which are common in physical applications, such as center of mass determinations and moments of inertia. Typically, we choose coordinates according to the symmetry of the particular problem. In Chapter 1 we used Gauss' theorem to transform a volume integral into a surface integral and Stokes' theorem to transform a surface integral into a line integral. For orthogonal coordinates, the surface and volume elements are simply products of the line elements $h_i\,dq_i$ (see Eqs. (2.10) and (2.11)). For the general case, we use the geometric meaning of $\partial\mathbf{r}/\partial q_i$ in Eq. (2.5) as tangent vectors. We start with the Cartesian surface element $dx\,dy$, which becomes an infinitesimal rectangle in the new coordinates q_1, q_2 formed by the two incremental vectors

$$\begin{aligned}d\mathbf{r_1} &= \mathbf{r}(q_1+dq_1,q_2)-\mathbf{r}(q_1,q_2) = \frac{\partial\mathbf{r}}{\partial q_1}dq_1,\\ d\mathbf{r_2} &= \mathbf{r}(q_1,q_2+dq_2)-\mathbf{r}(q_1,q_2) = \frac{\partial\mathbf{r}}{\partial q_2}dq_2,\end{aligned} \tag{2.14}$$

whose area is the z-component of their cross product, or

$$\begin{aligned}dx\,dy = d\mathbf{r_1}\times d\mathbf{r_2}\big|_z &= \left[\frac{\partial x}{\partial q_1}\frac{\partial y}{\partial q_2} - \frac{\partial x}{\partial q_2}\frac{\partial y}{\partial q_1}\right]dq_1\,dq_2\\ &= \begin{vmatrix}\frac{\partial x}{\partial q_1} & \frac{\partial x}{\partial q_2}\\ \frac{\partial y}{\partial q_1} & \frac{\partial y}{\partial q_2}\end{vmatrix}dq_1\,dq_2.\end{aligned} \tag{2.15}$$

The transformation coefficient in determinant form is called the **Jacobian**.

Similarly, the volume element $dx\,dy\,dz$ becomes the triple scalar product of the three infinitesimal displacement vectors $d\mathbf{r_i} = dq_i\frac{\partial\mathbf{r}}{\partial q_i}$ along the q_i directions $\hat{\mathbf{q}}_\mathbf{i}$, which, according

to Section 1.5, takes on the form

$$dx\,dy\,dz = \begin{vmatrix} \frac{\partial x}{\partial q_1} & \frac{\partial x}{\partial q_2} & \frac{\partial x}{\partial q_3} \\ \frac{\partial y}{\partial q_1} & \frac{\partial y}{\partial q_2} & \frac{\partial y}{\partial q_3} \\ \frac{\partial z}{\partial q_1} & \frac{\partial z}{\partial q_2} & \frac{\partial z}{\partial q_3} \end{vmatrix} dq_1\,dq_2\,dq_3. \tag{2.16}$$

Here the determinant is also called the Jacobian, and so on in higher dimensions.

For orthogonal coordinates the Jacobians simplify to products of the orthogonal vectors in Eq. (2.9). It follows that they are just products of the h_i; for example, the volume Jacobian becomes

$$h_1 h_2 h_3 (\hat{\mathbf{q}}_1 \times \hat{\mathbf{q}}_2) \cdot \hat{\mathbf{q}}_3 = h_1 h_2 h_3,$$

by expanding the determinant along the third line. Hence the volume element becomes

$$d\rho\,d\varphi = \rho\,d\rho\,d\varphi.$$

Similarly, in spherical coordinates (see Section 2.5) we get, from $x = r\sin\theta\cos\varphi$, $y = r\sin\theta\sin\varphi$, $z = r\cos\theta$, the Jacobian

$$= r^2\left(\cos^2\theta\sin\theta + \sin^3\theta\right) = r^2\sin\theta$$

and so on, $r^2\,dr\sin\theta\,d\theta\,d\varphi$. The volume integral can be written as

$$\int f(x, y, z)\,dx\,dy\,dz = \int f\big(x(r,\theta,\varphi), y(r,\theta,\varphi), z(r,\theta,\varphi)\big)\,r^2\,dr\sin\theta\,d\theta\,d\varphi. \quad \blacksquare$$

Example 2.1.1 JACOBIANS FOR POLAR COORDINATES

Let us illustrate the transformation of the Cartesian two-dimensional volume element $dx\,dy$ to polar coordinates ρ, φ, with $x = \rho\cos\varphi$, $y = \rho\sin\varphi$. (See also Section 2.4.) Here,

$$J = \begin{vmatrix} \frac{\partial x}{\partial \rho} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial \rho} & \frac{\partial y}{\partial \varphi} \end{vmatrix} = \begin{vmatrix} \cos\varphi & -\rho\sin\varphi \\ \sin\varphi & \rho\cos\varphi \end{vmatrix} = \rho\cos^2\varphi + \rho\sin^2\varphi = \rho,$$

$$dx\,dy = \begin{vmatrix} \frac{\partial x}{\partial \rho} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial \rho} & \frac{\partial y}{\partial \varphi} \end{vmatrix} d\rho\,d\varphi = \rho\,d\rho\,d\varphi.$$

In summary, we have developed the general formalism for vector analysis in orthogonal curvilinear coordinates in $\mathbb{R}^3$. For most applications, locally orthogonal coordinates can be chosen for which surface and volume elements in multiple integrals are products of line elements. For the general nonorthogonal case, Jacobian determinants apply.

$$\begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \varphi} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \varphi} \end{vmatrix} = \begin{vmatrix} \sin\theta\cos\varphi & r\cos\theta\cos\varphi & -r\sin\theta\sin\varphi \\ \sin\theta\sin\varphi & r\cos\theta\sin\varphi & r\sin\theta\cos\varphi \\ \cos\theta & -r\sin\theta & 0 \end{vmatrix}$$

$$= \cos\theta\left(r^2\sin\theta\cos\theta\right) + r\sin\theta\left(r\sin^2\theta\right) = r^2\sin\theta.$$

Exercises

2.1.1 Show that limiting our attention to orthogonal coordinate systems implies that $g_{ij} = 0$ for $i \neq j$ (Eq. (2.7)).
Hint. Construct a triangle with sides ds_1, ds_2, and ds_2. Equation (2.9) must hold regardless of whether $g_{ij} = 0$. Then compare ds^2 from Eq. (2.5) with a calculation using the law of cosines. Show that $\cos\theta_{12} = g_{12}/\sqrt{g_{11}g_{22}}$.

2.1.2 In the spherical polar coordinate system, $q_1 = r$, $q_2 = \theta$, $q_3 = \varphi$. The transformation equations corresponding to Eq. (2.1) are

$$x = r\sin\theta\cos\varphi, \qquad y = r\sin\theta\sin\varphi, \qquad z = r\cos\theta.$$

(a) Calculate the spherical polar coordinate scale factors: h_r, h_θ, and h_φ.
(b) Check your calculated scale factors by the relation $ds_i = h_i\, dq_i$.

2.1.3 The u-, v-, z-coordinate system frequently used in electrostatics and in hydrodynamics is defined by

$$xy = u, \qquad x^2 - y^2 = v, \qquad z = z.$$

This u-, v-, z-system is orthogonal.

(a) In words, describe briefly the nature of each of the three families of coordinate surfaces.
(b) Sketch the system in the xy-plane showing the intersections of surfaces of constant u and surfaces of constant v with the xy-plane.
(c) Indicate the directions of the unit vector $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ in all four quadrants.
(d) Finally, is this u-, v-, z-system right-handed ($\hat{\mathbf{u}} \times \hat{\mathbf{v}} = +\hat{\mathbf{z}}$) or left-handed ($\hat{\mathbf{u}} \times \hat{\mathbf{v}} = -\hat{\mathbf{z}}$)?

2.1.4 The elliptic cylindrical coordinate system consists of three families of surfaces:

$$1)\ \frac{x^2}{a^2\cosh^2 u} + \frac{y^2}{a^2\sinh^2 u} = 1; \qquad 2)\ \frac{x^2}{a^2\cos^2 v} - \frac{y^2}{a^2\sin^2 v} = 1; \qquad 3)\ z = z.$$

Sketch the coordinate surfaces $u =$ constant and $v =$ constant as they intersect the first quadrant of the xy-plane. Show the unit vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$. The range of u is $0 \leqslant u < \infty$. The range of v is $0 \leqslant v \leqslant 2\pi$.

2.1.5 A two-dimensional orthogonal system is described by the coordinates q_1 and q_2. Show that the Jacobian

$$J\left(\frac{x, y}{q_1, q_2}\right) \equiv \frac{\partial(x, y)}{\partial(q_1, q_2)} \equiv \frac{\partial x}{\partial q_1}\frac{\partial y}{\partial q_2} - \frac{\partial x}{\partial q_2}\frac{\partial y}{\partial q_1} = h_1 h_2$$

is in agreement with Eq. (2.10).
Hint. It's easier to work with the square of each side of this equation.

2.1.6 In Minkowski space we define $x_1 = x$, $x_2 = y$, $x_3 = z$, and $x_0 = ct$. This is done so that the metric interval becomes $ds^2 = dx_0^2 - dx_1^2 - dx_2^2 - dx_3^2$ (with $c =$ velocity of light). Show that the metric in Minkowski space is

$$(g_{ij}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

We use Minkowski space in Sections 4.5 and 4.6 for describing Lorentz transformations.

2.2 DIFFERENTIAL VECTOR OPERATORS

We return to our restriction to orthogonal coordinate systems.

Gradient

The starting point for developing the gradient, divergence, and curl operators in curvilinear coordinates is the geometric interpretation of the gradient as the vector having the magnitude and direction of the maximum space rate of change (compare Section 1.6). From this interpretation the component of $\nabla\psi(q_1, q_2, q_3)$ in the direction normal to the family of surfaces $q_1 =$ constant is given by[3]

$$\hat{\mathbf{q}}_1 \cdot \nabla\psi = \nabla\psi|_1 = \frac{\partial\psi}{\partial s_1} = \frac{1}{h_1}\frac{\partial\psi}{\partial q_1}, \tag{2.17}$$

since this is the rate of change of ψ for varying q_1, holding q_2 and q_3 fixed. The quantity ds_1 is a differential length in the direction of increasing q_1 (compare Eqs. (2.9)). In Section 2.1 we introduced a unit vector $\hat{\mathbf{q}}_1$ to indicate this direction. By repeating Eq. (2.17) for q_2 and again for q_3 and adding vectorially, we see that the gradient becomes

$$\begin{aligned} \nabla\psi(q_1, q_2, q_3) &= \hat{\mathbf{q}}_1\frac{\partial\psi}{\partial s_1} + \hat{\mathbf{q}}_2\frac{\partial\psi}{\partial s_2} + \hat{\mathbf{q}}_3\frac{\partial\psi}{\partial s_3} \\ &= \hat{\mathbf{q}}_1\frac{1}{h_1}\frac{\partial\psi}{\partial q_1} + \hat{\mathbf{q}}_2\frac{1}{h_2}\frac{\partial\psi}{\partial q_2} + \hat{\mathbf{q}}_3\frac{1}{h_3}\frac{\partial\psi}{\partial q_3} \\ &= \sum_i \hat{\mathbf{q}}_i\frac{1}{h_i}\frac{\partial\psi}{\partial q_i}. \end{aligned} \tag{2.18}$$

Exercise 2.2.4 offers a mathematical alternative independent of this physical interpretation of the gradient. The total variation of a function,

$$d\psi = \nabla\psi \cdot d\mathbf{r} = \sum_i \frac{1}{h_i}\frac{\partial\psi}{\partial q_i} ds_i = \sum_i \frac{\partial\psi}{\partial q_i} dq_i$$

is consistent with Eq. (2.18), of course.

[3] Here the use of φ to label a function is avoided because it is conventional to use this symbol to denote an azimuthal coordinate.

Divergence

The divergence operator may be obtained from the second definition (Eq. (1.98)) of Chapter 1 or equivalently from Gauss' theorem, Section 1.11. Let us use Eq. (1.98),

$$\nabla \cdot \mathbf{V}(q_1, q_2, q_3) = \lim_{\int d\tau \to 0} \frac{\int \mathbf{V} \cdot d\boldsymbol{\sigma}}{\int d\tau}, \tag{2.19}$$

with a differential volume $h_1 h_2 h_3\, dq_1\, dq_2\, dq_3$ (Fig. 2.1). Note that the positive directions have been chosen so that $(\hat{\mathbf{q}}_1, \hat{\mathbf{q}}_2, \hat{\mathbf{q}}_3)$ form a right-handed set, $\hat{\mathbf{q}}_1 \times \hat{\mathbf{q}}_2 = \hat{\mathbf{q}}_3$.

The difference of area integrals for the two faces $q_1 =$ constant is given by

$$\left[V_1 h_2 h_3 + \frac{\partial}{\partial q_1}(V_1 h_2 h_3)\, dq_1\right] dq_2\, dq_3 - V_1 h_2 h_3\, dq_2\, dq_3$$
$$= \frac{\partial}{\partial q_1}(V_1 h_2 h_3)\, dq_1\, dq_2\, dq_3, \tag{2.20}$$

exactly as in Sections 1.7 and 1.10.[4] Here, $V_i = \mathbf{V} \cdot \hat{\mathbf{q}}_i$ is the projection of $\mathbf{V}$ onto the $\hat{\mathbf{q}}_i$-direction. Adding in the similar results for the other two pairs of surfaces, we obtain

$$\int \mathbf{V}(q_1, q_2, q_3) \cdot d\boldsymbol{\sigma}$$
$$= \left[\frac{\partial}{\partial q_1}(V_1 h_2 h_3) + \frac{\partial}{\partial q_2}(V_2 h_3 h_1) + \frac{\partial}{\partial q_3}(V_3 h_1 h_2)\right] dq_1\, dq_2\, dq_3.$$

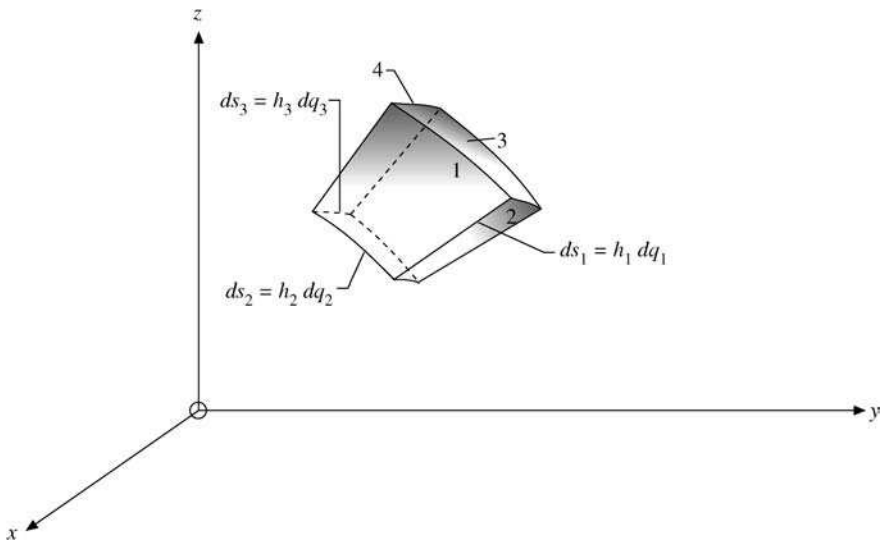

FIGURE 2.1 Curvilinear volume element.

[4] Since we take the limit $dq_1, dq_2, dq_3 \to 0$, the second- and higher-order derivatives will drop out.

Now, using Eq. (2.19), division by our differential volume yields

$$\boxed{\nabla \cdot \mathbf{V}(q_1, q_2, q_3) = \frac{1}{h_1 h_2 h_3}\left[\frac{\partial}{\partial q_1}(V_1 h_2 h_3) + \frac{\partial}{\partial q_2}(V_2 h_3 h_1) + \frac{\partial}{\partial q_3}(V_3 h_1 h_2)\right].} \tag{2.21}$$

We may obtain the Laplacian by combining Eqs. (2.18) and (2.21), using $\mathbf{V} = \nabla\psi(q_1, q_2, q_3)$. This leads to

$$\boxed{\begin{aligned}&\nabla \cdot \nabla\psi(q_1, q_2, q_3)\\ &\quad = \frac{1}{h_1 h_2 h_3}\left[\frac{\partial}{\partial q_1}\left(\frac{h_2 h_3}{h_1}\frac{\partial\psi}{\partial q_1}\right) + \frac{\partial}{\partial q_2}\left(\frac{h_3 h_1}{h_2}\frac{\partial\psi}{\partial q_2}\right) + \frac{\partial}{\partial q_3}\left(\frac{h_1 h_2}{h_3}\frac{\partial\psi}{\partial q_3}\right)\right].\end{aligned}} \tag{2.22}$$

Curl

Finally, to develop $\nabla \times \mathbf{V}$, let us apply Stokes' theorem (Section 1.12) and, as with the divergence, take the limit as the surface area becomes vanishingly small. Working on one component at a time, we consider a differential surface element in the curvilinear surface $q_1 =$ constant. From

$$\int_s \nabla \times \mathbf{V} \cdot d\boldsymbol{\sigma} = \hat{\mathbf{q}}_1 \cdot (\nabla \times \mathbf{V}) h_2 h_3\, dq_2\, dq_3 \tag{2.23}$$

(mean value theorem of integral calculus), Stokes' theorem yields

$$\hat{\mathbf{q}}_1 \cdot (\nabla \times \mathbf{V}) h_2 h_3\, dq_2\, dq_3 = \oint \mathbf{V} \cdot d\mathbf{r}, \tag{2.24}$$

with the line integral lying in the surface $q_1 =$ constant. Following the loop (1, 2, 3, 4) of Fig. 2.2,

$$\begin{aligned}\oint \mathbf{V}(q_1, q_2, q_3) \cdot d\mathbf{r} &= V_2 h_2\, dq_2 + \left[V_3 h_3 + \frac{\partial}{\partial q_2}(V_3 h_3)\, dq_2\right] dq_3\\ &\quad - \left[V_2 h_2 + \frac{\partial}{\partial q_3}(V_2 h_2) dq_3\right] dq_2 - V_3 h_3\, dq_3\\ &= \left[\frac{\partial}{\partial q_2}(h_3 V_3) - \frac{\partial}{\partial q_3}(h_2 V_2)\right] dq_2\, dq_3.\end{aligned} \tag{2.25}$$

We pick up a positive sign when going in the positive direction on parts 1 and 2 and a negative sign on parts 3 and 4 because here we are going in the negative direction. (Higher-order terms in Maclaurin or Taylor expansions have been omitted. They will vanish in the limit as the surface becomes vanishingly small ($dq_2 \to 0$, $dq_3 \to 0$).)

From Eq. (2.24),

$$\nabla \times \mathbf{V}|_1 = \frac{1}{h_2 h_3}\left[\frac{\partial}{\partial q_2}(h_3 V_3) - \frac{\partial}{\partial q_3}(h_2 V_2)\right]. \tag{2.26}$$

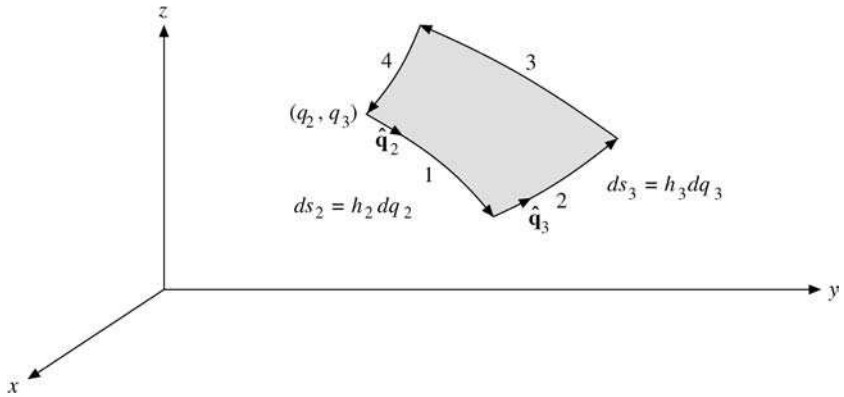

FIGURE 2.2 Curvilinear surface element with $q_1 =$ constant.

The remaining two components of $\nabla \times \mathbf{V}$ may be picked up by cyclic permutation of the indices. As in Chapter 1, it is often convenient to write the curl in determinant form:

$$\nabla \times \mathbf{V} = \frac{1}{h_1 h_2 h_3} \begin{vmatrix} \hat{\mathbf{q}}_1 h_1 & \hat{\mathbf{q}}_2 h_2 & \hat{\mathbf{q}}_3 h_3 \\ \dfrac{\partial}{\partial q_1} & \dfrac{\partial}{\partial q_2} & \dfrac{\partial}{\partial q_3} \\ h_1 V_1 & h_2 V_2 & h_3 V_3 \end{vmatrix}. \tag{2.27}$$

Remember that, because of the presence of the differential operators, this determinant must be expanded from the top down. Note that this equation is **not** identical with the form for the cross product of two vectors, Eq. (2.13). ∇ is not an ordinary vector; it is a vector **operator**.

Our geometric interpretation of the gradient and the use of Gauss' and Stokes' theorems (or integral definitions of divergence and curl) have enabled us to obtain these quantities **without having to differentiate the unit vectors $\hat{\mathbf{q}}_i$**. There exist alternate ways to determine grad, div, and curl based on direct differentiation of the $\hat{\mathbf{q}}_i$. One approach resolves the $\hat{\mathbf{q}}_i$ of a specific coordinate system into its Cartesian components (Exercises 2.4.1 and 2.5.1) and differentiates this Cartesian form (Exercises 2.4.3 and 2.5.2). The point here is that the derivatives of the Cartesian $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ vanish since $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ are constant in direction as well as in magnitude. A second approach [L. J. Kijewski, *Am. J. Phys.* **33**: 816 (1965)] assumes the equality of $\partial^2\mathbf{r}/\partial q_i\,\partial q_j$ and $\partial^2\mathbf{r}/\partial q_j\,\partial q_i$ and develops the derivatives of $\hat{\mathbf{q}}_i$ in a general curvilinear form. Exercises 2.2.3 and 2.2.4 are based on this method.

Exercises

2.2.1 Develop arguments to show that dot and cross products (not involving ∇) in orthogonal curvilinear coordinates in $\mathbb{R}^3$ proceed, as in Cartesian coordinates, **with no involvement of scale factors**.

2.2.2 With $\hat{\mathbf{q}}_1$ a unit vector in the direction of increasing q_1, show that

(a) $$\nabla \cdot \hat{\mathbf{q}}_1 = \frac{1}{h_1 h_2 h_3} \frac{\partial (h_2 h_3)}{\partial q_1}$$

(b) $$\nabla \times \hat{\mathbf{q}}_1 = \frac{1}{h_1} \left[\hat{\mathbf{q}}_2 \frac{1}{h_3} \frac{\partial h_1}{\partial q_3} - \hat{\mathbf{q}}_3 \frac{1}{h_2} \frac{\partial h_1}{\partial q_2} \right].$$

Note that even though $\hat{\mathbf{q}}_1$ is a unit vector, its divergence and curl **do not necessarily vanish**.

2.2.3 Show that the orthogonal unit vectors $\hat{\mathbf{q}}_j$ may be defined by

$$\hat{\mathbf{q}}_i = \frac{1}{h_i} \frac{\partial \mathbf{r}}{\partial q_i}. \tag{a}$$

In particular, show that $\hat{\mathbf{q}}_i \cdot \hat{\mathbf{q}}_i = 1$ leads to an expression for h_i in agreement with Eqs. (2.9).
Equation (a) may be taken as a starting point for deriving

$$\frac{\partial \hat{\mathbf{q}}_i}{\partial q_j} = \hat{\mathbf{q}}_j \frac{1}{h_i} \frac{\partial h_j}{\partial q_i}, \qquad i \neq j$$

and

$$\frac{\partial \hat{\mathbf{q}}_i}{\partial q_i} = -\sum_{j \neq i} \hat{\mathbf{q}}_j \frac{1}{h_j} \frac{\partial h_i}{\partial q_j}.$$

2.2.4 Derive

$$\nabla \psi = \hat{\mathbf{q}}_1 \frac{1}{h_1} \frac{\partial \psi}{\partial q_1} + \hat{\mathbf{q}}_2 \frac{1}{h_2} \frac{\partial \psi}{\partial q_2} + \hat{\mathbf{q}}_3 \frac{1}{h_3} \frac{\partial \psi}{\partial q_3}$$

by direct application of Eq. (1.97),

$$\nabla \psi = \lim_{\int d\tau \to 0} \frac{\int \psi \, d\boldsymbol{\sigma}}{\int d\tau}.$$

Hint. Evaluation of the surface integral will lead to terms like $(h_1 h_2 h_3)^{-1} (\partial / \partial q_1) \times (\hat{\mathbf{q}}_1 h_2 h_3)$. The results listed in Exercise 2.2.3 will be helpful. Cancellation of unwanted terms occurs when the contributions of all three pairs of surfaces are added together.

2.3 SPECIAL COORDINATE SYSTEMS: INTRODUCTION

There are at least 11 coordinate systems in which the three-dimensional Helmholtz equation can be separated into three ordinary differential equations. Some of these coordinate systems have achieved prominence in the historical development of quantum mechanics. Other systems, such as bipolar coordinates, satisfy special needs. Partly because the needs are rather infrequent but mostly because the development of computers and efficient programming techniques reduce the need for these coordinate systems, the discussion in this chapter is limited to (1) Cartesian coordinates, (2) spherical polar coordinates, and (3) circular cylindrical coordinates. Specifications and details of the other coordinate systems will be found in the first two editions of this work and in Additional Readings at the end of this chapter (Morse and Feshbach, Margenau and Murphy).

2.4 CIRCULAR CYLINDER COORDINATES

In the circular cylindrical coordinate system the three curvilinear coordinates (q_1, q_2, q_3) are relabeled (ρ, φ, z). We are using ρ for the perpendicular distance from the z-axis and saving r for the distance from the origin. The limits on ρ, φ and z are

$$0 \leqslant \rho < \infty, \qquad 0 \leqslant \varphi \leqslant 2\pi, \qquad \text{and} \qquad -\infty < z < \infty.$$

For $\rho = 0$, φ is not well defined. The coordinate surfaces, shown in Fig. 2.3, are:

1. Right circular cylinders having the z-axis as a common axis,
$$\rho = \left(x^2 + y^2\right)^{1/2} = \text{constant}.$$
2. Half-planes through the z-axis,
$$\varphi = \tan^{-1}\left(\frac{y}{x}\right) = \text{constant}.$$
3. Planes parallel to the xy-plane, as in the Cartesian system,
$$z = \text{constant}.$$

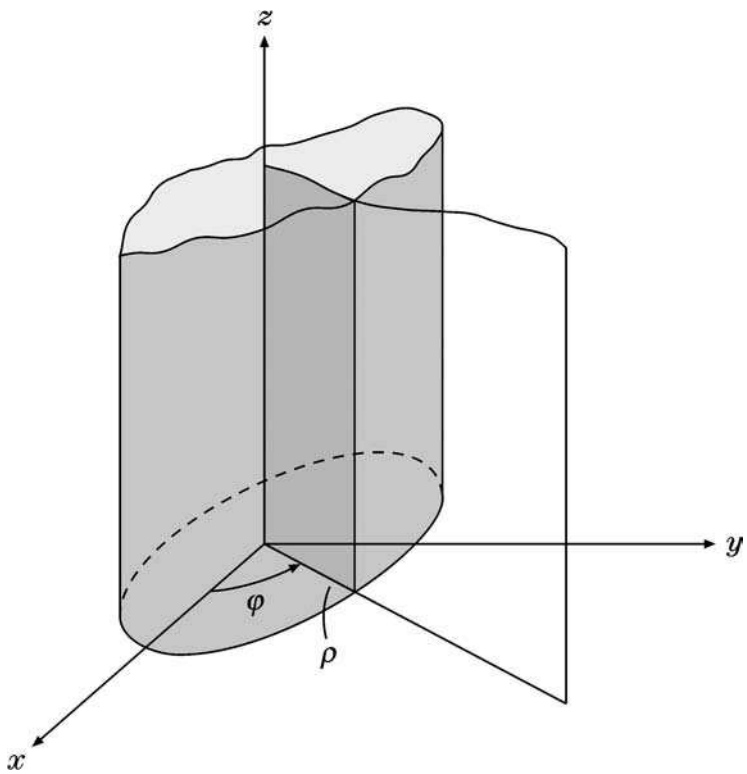

FIGURE 2.3 Circular cylinder coordinates.

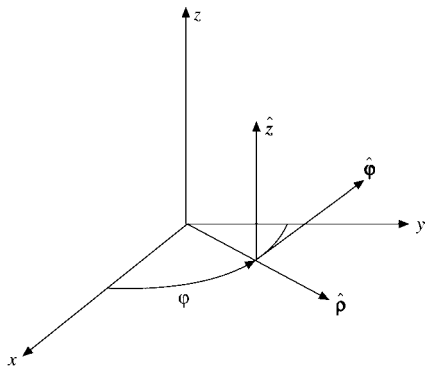

FIGURE 2.4 Circular cylindrical coordinate unit vectors.

Inverting the preceding equations for ρ and φ (or going directly to Fig. 2.3), we obtain the transformation relations

$$x = \rho\cos\varphi, \qquad y = \rho\sin\varphi, \qquad z = z. \tag{2.28}$$

The z-axis remains unchanged. This is essentially a two-dimensional curvilinear system with a Cartesian z-axis added on to form a three-dimensional system.

According to Eq. (2.5) or from the length elements ds_i, the scale factors are

$$h_1 = h_\rho = 1, \qquad h_2 = h_\varphi = \rho, \qquad h_3 = h_z = 1. \tag{2.29}$$

The unit vectors $\hat{\mathbf{q}}_1, \hat{\mathbf{q}}_2, \hat{\mathbf{q}}_3$ are relabeled $(\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\varphi}}, \hat{\mathbf{z}})$, as in Fig. 2.4. The unit vector $\hat{\boldsymbol{\rho}}$ is normal to the cylindrical surface, pointing in the direction of increasing radius ρ. The unit vector $\hat{\boldsymbol{\varphi}}$ is tangential to the cylindrical surface, perpendicular to the half plane $\varphi =$ constant and pointing in the direction of increasing azimuth angle φ. The third unit vector, $\hat{\mathbf{z}}$, is the usual Cartesian unit vector. They are mutually orthogonal,

$$\hat{\boldsymbol{\rho}} \cdot \hat{\boldsymbol{\varphi}} = \hat{\boldsymbol{\varphi}} \cdot \hat{\mathbf{z}} = \hat{\mathbf{z}} \cdot \hat{\boldsymbol{\rho}} = 0,$$

and the coordinate vector and a general vector $\mathbf{V}$ are expressed as

$$\mathbf{r} = \hat{\boldsymbol{\rho}}\rho + \hat{\mathbf{z}}z, \qquad \mathbf{V} = \hat{\boldsymbol{\rho}}V_\rho + \hat{\boldsymbol{\varphi}}V_\varphi + \hat{\mathbf{z}}V_z.$$

A differential displacement $d\mathbf{r}$ may be written

$$\begin{aligned} d\mathbf{r} &= \hat{\boldsymbol{\rho}}\,ds_\rho + \hat{\boldsymbol{\varphi}}\,ds_\varphi + \hat{\mathbf{z}}\,dz \\ &= \hat{\boldsymbol{\rho}}\,d\rho + \hat{\boldsymbol{\varphi}}\rho\,d\varphi + \hat{\mathbf{z}}\,dz. \end{aligned} \tag{2.30}$$

Example 2.4.1 AREA LAW FOR PLANETARY MOTION

First we derive Kepler's law in cylindrical coordinates, saying that the radius vector sweeps out equal areas in equal time, from angular momentum conservation.

We consider the sun at the origin as a source of the **central** gravitational force $\mathbf{F} = f(r)\hat{\mathbf{r}}$. Then the orbital angular momentum $\mathbf{L} = m\mathbf{r} \times \mathbf{v}$ of a planet of mass m and velocity $\mathbf{v}$ is conserved, because the torque

$$\frac{d\mathbf{L}}{dt} = m\frac{d\mathbf{r}}{dt} \times \frac{d\mathbf{r}}{dt} + \mathbf{r} \times m\frac{d\mathbf{v}}{dt} = \mathbf{r} \times \mathbf{F} = \frac{f(r)}{r}\mathbf{r} \times \mathbf{r} = 0.$$

Hence $\mathbf{L} = \text{const}$. Now we can choose the z-axis to lie along the direction of the orbital angular momentum vector, $\mathbf{L} = L\hat{\mathbf{z}}$, and work in cylindrical coordinates $\mathbf{r} = (\rho, \varphi, z) = \rho\hat{\boldsymbol{\rho}}$ with $z = 0$. The planet moves in the xy-plane because $\mathbf{r}$ and $\mathbf{v}$ are perpendicular to $\mathbf{L}$. Thus, we expand its velocity as follows:

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \dot{\rho}\hat{\boldsymbol{\rho}} + \rho\frac{d\hat{\boldsymbol{\rho}}}{dt}.$$

From

$$\hat{\boldsymbol{\rho}} = (\cos\varphi, \sin\varphi), \qquad \frac{\partial\hat{\boldsymbol{\rho}}}{d\varphi} = (-\sin\varphi, \cos\varphi) = \hat{\boldsymbol{\varphi}},$$

we find that $\frac{d\hat{\boldsymbol{\rho}}}{dt} = \frac{d\hat{\boldsymbol{\rho}}}{d\varphi}\frac{d\varphi}{dt} = \dot{\varphi}\hat{\boldsymbol{\varphi}}$ using the chain rule, so $\mathbf{v} = \dot{\rho}\hat{\rho} + \rho\frac{d\hat{\rho}}{dt} = \dot{\rho}\hat{\boldsymbol{\rho}} + \rho\dot{\varphi}\hat{\boldsymbol{\varphi}}$. When we substitute the expansions of $\hat{\boldsymbol{\rho}}$ and $\mathbf{v}$ in polar coordinates, we obtain

$$\mathbf{L} = m\boldsymbol{\rho} \times \mathbf{v} = m\rho(\rho\dot{\varphi})(\hat{\boldsymbol{\rho}} \times \hat{\boldsymbol{\varphi}}) = m\rho^2\dot{\varphi}\hat{\mathbf{z}} = \text{constant}.$$

The triangular area swept by the radius vector $\boldsymbol{\rho}$ in the time dt (**area law**), when integrated over one revolution, is given by

$$A = \frac{1}{2}\int \rho(\rho\, d\varphi) = \frac{1}{2}\int \rho^2\dot{\varphi}\, dt = \frac{L}{2m}\int dt = \frac{L\tau}{2m}, \tag{2.31}$$

if we substitute $m\rho^2\dot{\varphi} = L = \text{const}$. Here τ is the period, that is, the time for one revolution of the planet in its orbit.

Kepler's first law says that the orbit is an ellipse. Now we derive the orbit equation $\rho(\varphi)$ of the ellipse in polar coordinates, where in Fig. 2.5 the sun is at one focus, which is the origin of our cylindrical coordinates. From the geometrical construction of the ellipse we know that $\rho' + \rho = 2a$, where a is the major half-axis; we shall show that this is equivalent to the conventional form of the ellipse equation. The distance between both foci is $0 < 2a\epsilon < 2a$, where $0 < \epsilon < 1$ is called the eccentricity of the ellipse. For a circle $\epsilon = 0$ because both foci coincide with the center. There is an angle, as shown in Fig. 2.5, where the distances $\rho' = \rho = a$ are equal, and Pythagoras' theorem applied to this right triangle

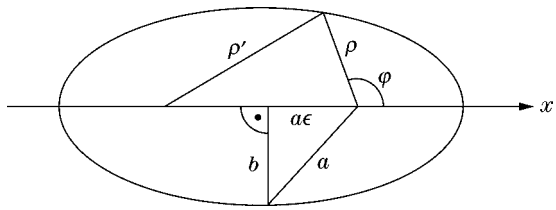

FIGURE 2.5 Ellipse in polar coordinates.

gives $b^2 + a^2\epsilon^2 = a^2$. As a result, $\sqrt{1-\epsilon^2} = b/a$ is the ratio of the minor half-axis (b) to the major half-axis, a.

Now consider the triangle with the sides labeled by ρ', ρ, $2a\epsilon$ in Fig. 2.5 and angle opposite ρ' equal to $\pi - \varphi$. Then, applying the law of cosines, gives

$$\rho'^2 = \rho^2 + 4a^2\epsilon^2 + 4\rho a\epsilon\cos\varphi.$$

Now substituting $\rho' = 2a - \rho$, canceling ρ^2 on both sides and dividing by $4a$ yields

$$\rho(1 + \epsilon\cos\varphi) = a\left(1 - \epsilon^2\right) \equiv p, \tag{2.32}$$

the **Kepler orbit equation in polar coordinates**.

Alternatively, we revert to Cartesian coordinates to find, from Eq. (2.32) with $x = \rho\cos\varphi$, that

$$\rho^2 = x^2 + y^2 = (p - x\epsilon)^2 = p^2 + x^2\epsilon^2 - 2px\epsilon,$$

so the familiar ellipse equation in Cartesian coordinates,

$$\left(1-\epsilon^2\right)\left(x + \frac{p\epsilon}{1-\epsilon^2}\right)^2 + y^2 = p^2 + \frac{p^2\epsilon^2}{1-\epsilon^2} = \frac{p^2}{1-\epsilon^2},$$

obtains. If we compare this result with the standard form of the ellipse,

$$\frac{(x-x_0)^2}{a^2} + \frac{y^2}{b^2} = 1,$$

we confirm that

$$b = \frac{p}{\sqrt{1-\epsilon^2}} = a\sqrt{1-\epsilon^2}, \qquad a = \frac{p}{1-\epsilon^2},$$

and that the distance x_0 between the center and focus is $a\epsilon$, as shown in Fig. 2.5. ■

The differential operations involving $\boldsymbol{\nabla}$ follow from Eqs. (2.18), (2.21), (2.22), and (2.27):

$$\boldsymbol{\nabla}\psi(\rho,\varphi,z) = \hat{\boldsymbol{\rho}}\frac{\partial\psi}{\partial\rho} + \hat{\boldsymbol{\varphi}}\frac{1}{\rho}\frac{\partial\psi}{\partial\varphi} + \hat{\mathbf{z}}\frac{\partial\psi}{\partial z}, \tag{2.33}$$

$$\boldsymbol{\nabla}\cdot\mathbf{V} = \frac{1}{\rho}\frac{\partial}{\partial\rho}(\rho V_\rho) + \frac{1}{\rho}\frac{\partial V_\varphi}{\partial\varphi} + \frac{\partial V_z}{\partial z}, \tag{2.34}$$

$$\boldsymbol{\nabla}^2\psi = \frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial\psi}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2\psi}{\partial\varphi^2} + \frac{\partial^2\psi}{\partial z^2}, \tag{2.35}$$

$$\boldsymbol{\nabla}\times\mathbf{V} = \frac{1}{\rho}\begin{vmatrix} \hat{\boldsymbol{\rho}} & \rho\hat{\boldsymbol{\varphi}} & \hat{\mathbf{z}} \\ \dfrac{\partial}{\partial\rho} & \dfrac{\partial}{\partial\varphi} & \dfrac{\partial}{\partial z} \\ V_\rho & \rho V_\varphi & V_z \end{vmatrix}. \tag{2.36}$$

Finally, for problems such as circular wave guides and cylindrical cavity resonators the vector Laplacian $\nabla^2\mathbf{V}$ resolved in circular cylindrical coordinates is

$$\begin{aligned}\nabla^2\mathbf{V}|_\rho &= \nabla^2 V_\rho - \frac{1}{\rho^2}V_\rho - \frac{2}{\rho^2}\frac{\partial V_\varphi}{\partial\varphi},\\ \nabla^2\mathbf{V}|_\varphi &= \nabla^2 V_\varphi - \frac{1}{\rho^2}V_\varphi + \frac{2}{\rho^2}\frac{\partial V_\rho}{\partial\varphi},\\ \nabla^2\mathbf{V}|_z &= \nabla^2 V_z,\end{aligned} \tag{2.37}$$

which follow from Eq. (1.85). The basic reason for this particular form of the z-component is that the z-axis is a Cartesian axis; that is,

$$\begin{aligned}\nabla^2(\hat{\boldsymbol{\rho}}V_\rho + \hat{\boldsymbol{\varphi}}V_\varphi + \hat{\mathbf{z}}V_z) &= \nabla^2(\hat{\boldsymbol{\rho}}V_\rho + \hat{\boldsymbol{\varphi}}V_\varphi) + \hat{\mathbf{z}}\nabla^2 V_z\\ &= \hat{\boldsymbol{\rho}}f(V_\rho, V_\varphi) + \hat{\boldsymbol{\varphi}}g(V_\rho, V_\varphi) + \hat{\mathbf{z}}\nabla^2 V_z.\end{aligned}$$

Finally, the operator ∇^2 operating on the $\hat{\boldsymbol{\rho}}$, $\hat{\boldsymbol{\varphi}}$ unit vectors stays in the $\hat{\boldsymbol{\rho}}\hat{\boldsymbol{\varphi}}$-plane.

Example 2.4.2 A Navier–Stokes Term

The Navier–Stokes equations of hydrodynamics contain a nonlinear term

$$\nabla\times\big[\mathbf{v}\times(\nabla\times\mathbf{v})\big],$$

where $\mathbf{v}$ is the fluid velocity. For fluid flowing through a cylindrical pipe in the z-direction,

$$\mathbf{v} = \hat{\mathbf{z}}v(\rho).$$

From Eq. (2.36),

$$\nabla\times\mathbf{v} = \frac{1}{\rho}\begin{vmatrix}\hat{\boldsymbol{\rho}} & \rho\hat{\boldsymbol{\varphi}} & \hat{\mathbf{z}}\\ \dfrac{\partial}{\partial\rho} & \dfrac{\partial}{\partial\varphi} & \dfrac{\partial}{\partial z}\\ 0 & 0 & v(\rho)\end{vmatrix} = -\hat{\boldsymbol{\varphi}}\frac{\partial v}{\partial\rho}$$

$$\mathbf{v}\times(\nabla\times\mathbf{v}) = \begin{vmatrix}\hat{\boldsymbol{\rho}} & \hat{\boldsymbol{\varphi}} & \hat{\mathbf{z}}\\ 0 & 0 & v\\ 0 & -\dfrac{\partial v}{\partial\rho} & 0\end{vmatrix} = \hat{\boldsymbol{\rho}}v(\rho)\frac{\partial v}{\partial\rho}.$$

Finally,

$$\nabla\times\big(\mathbf{v}\times(\nabla\times\mathbf{v})\big) = \frac{1}{\rho}\begin{vmatrix}\hat{\boldsymbol{\rho}} & \rho\hat{\boldsymbol{\varphi}} & \hat{\mathbf{z}}\\ \dfrac{\partial}{\partial\rho} & \dfrac{\partial}{\partial\varphi} & \dfrac{\partial}{\partial z}\\ v\dfrac{\partial v}{\partial\rho} & 0 & 0\end{vmatrix} = 0,$$

so, for this particular case, the nonlinear term vanishes. ■

Exercises

2.4.1 Resolve the circular cylindrical unit vectors into their Cartesian components (Fig. 2.6).

ANS. $\hat{\boldsymbol{\rho}} = \hat{\mathbf{x}}\cos\varphi + \hat{\mathbf{y}}\sin\varphi,$
$\hat{\boldsymbol{\varphi}} = -\hat{\mathbf{x}}\sin\varphi + \hat{\mathbf{y}}\cos\varphi,$
$\hat{\mathbf{z}} = \hat{\mathbf{z}}.$

2.4.2 Resolve the Cartesian unit vectors into their circular cylindrical components (Fig. 2.6).

ANS. $\hat{\mathbf{x}} = \hat{\boldsymbol{\rho}}\cos\varphi - \hat{\boldsymbol{\varphi}}\sin\varphi,$
$\hat{\mathbf{y}} = \hat{\boldsymbol{\rho}}\sin\varphi + \hat{\boldsymbol{\varphi}}\cos\varphi,$
$\hat{\mathbf{z}} = \hat{\mathbf{z}}.$

2.4.3 From the results of Exercise 2.4.1 show that

$$\frac{\partial\hat{\boldsymbol{\rho}}}{\partial\varphi} = \hat{\boldsymbol{\varphi}}, \qquad \frac{\partial\hat{\boldsymbol{\varphi}}}{\partial\varphi} = -\hat{\boldsymbol{\rho}}$$

and that all other first derivatives of the circular cylindrical unit vectors with respect to the circular cylindrical coordinates vanish.

2.4.4 Compare $\nabla\cdot\mathbf{V}$ (Eq. (2.34)) with the gradient operator

$$\nabla = \hat{\boldsymbol{\rho}}\frac{\partial}{\partial\rho} + \hat{\boldsymbol{\varphi}}\frac{1}{\rho}\frac{\partial}{\partial\varphi} + \hat{\mathbf{z}}\frac{\partial}{\partial z}$$

(Eq. (2.33)) dotted into $\mathbf{V}$. Note that the differential operators of ∇ differentiate **both** the unit vectors and the components of $\mathbf{V}$.
Hint. $\hat{\boldsymbol{\varphi}}(1/\rho)(\partial/\partial\varphi)\cdot\hat{\boldsymbol{\rho}}V_\rho$ becomes $\hat{\boldsymbol{\varphi}}\cdot\frac{1}{\rho}\frac{\partial}{\partial\varphi}(\hat{\boldsymbol{\rho}}V_\rho)$ and does **not** vanish.

2.4.5 (a) Show that $\mathbf{r} = \hat{\boldsymbol{\rho}}\rho + \hat{\mathbf{z}}z$.

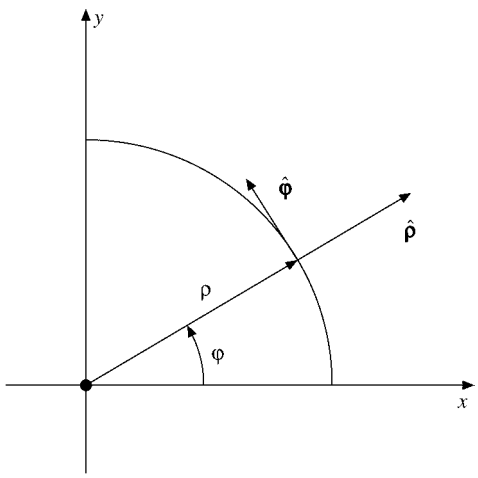

FIGURE 2.6 Plane polar coordinates.

(b) Working entirely in circular cylindrical coordinates, show that

$$\nabla \cdot \mathbf{r} = 3 \qquad \text{and} \qquad \nabla \times \mathbf{r} = 0.$$

2.4.6 (a) Show that the parity operation (reflection through the origin) on a point (ρ, φ, z) relative to **fixed** x-, y-, z-axes consists of the transformation

$$\rho \to \rho, \qquad \varphi \to \varphi \pm \pi, \qquad z \to -z.$$

(b) Show that $\hat{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{\varphi}}$ have odd parity (reversal of direction) and that $\hat{\mathbf{z}}$ has even parity.

Note. The Cartesian unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ remain constant.

2.4.7 A rigid body is rotating about a fixed axis with a constant angular velocity $\boldsymbol{\omega}$. Take $\boldsymbol{\omega}$ to lie along the z-axis. Express the position vector $\mathbf{r}$ in circular cylindrical coordinates and using circular cylindrical coordinates,
(a) calculate $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$, (b) calculate $\nabla \times \mathbf{v}$.

ANS. (a) $\mathbf{v} = \hat{\boldsymbol{\varphi}}\omega\rho$,
(b) $\nabla \times \mathbf{v} = 2\boldsymbol{\omega}$.

2.4.8 Find the circular cylindrical components of the velocity and acceleration of a movingparticle,

$$\begin{aligned} v_\rho &= \dot{\rho}, & a_\rho &= \ddot{\rho} - \rho\dot{\varphi}^2, \\ v_\varphi &= \rho\dot{\varphi}, & a_\varphi &= \rho\ddot{\varphi} + 2\dot{\rho}\dot{\varphi}, \\ v_z &= \dot{z}, & a_z &= \ddot{z}. \end{aligned}$$

Hint.

$$\begin{aligned} \mathbf{r}(t) &= \hat{\boldsymbol{\rho}}(t)\rho(t) + \hat{\mathbf{z}}z(t) \\ &= \big[\hat{\mathbf{x}}\cos\varphi(t) + \hat{\mathbf{y}}\sin\varphi(t)\big]\rho(t) + \hat{\mathbf{z}}z(t). \end{aligned}$$

Note. $\dot{\rho} = d\rho/dt$, $\ddot{\rho} = d^2\rho/dt^2$, and so on.

2.4.9 Solve Laplace's equation, $\nabla^2\psi = 0$, in cylindrical coordinates for $\psi = \psi(\rho)$.

ANS. $\psi = k \ln \dfrac{\rho}{\rho_0}$.

2.4.10 In right circular cylindrical coordinates a particular vector function is given by

$$\mathbf{V}(\rho, \varphi) = \hat{\boldsymbol{\rho}}V_\rho(\rho, \varphi) + \hat{\boldsymbol{\varphi}}V_\varphi(\rho, \varphi).$$

Show that $\nabla \times \mathbf{V}$ has only a z-component. Note that this result will hold for any vector confined to a surface $q_3 =$ constant as long as the products h_1V_1 and h_2V_2 are each independent of q_3.

2.4.11 For the flow of an incompressible viscous fluid the Navier–Stokes equations lead to

$$-\nabla \times \big(\mathbf{v} \times (\nabla \times \mathbf{v})\big) = \frac{\eta}{\rho_0}\nabla^2(\nabla \times \mathbf{v}).$$

Here η is the viscosity and ρ_0 is the density of the fluid. For axial flow in a cylindrical pipe we take the velocity $\mathbf{v}$ to be

$$\mathbf{v} = \hat{\mathbf{z}}v(\rho).$$

From Example 2.4.2,

$$\nabla \times \big(\mathbf{v} \times (\nabla \times \mathbf{v})\big) = 0$$

for this choice of **v**.
Show that

$$\nabla^2(\nabla \times \mathbf{v}) = 0$$

leads to the differential equation

$$\frac{1}{\rho}\frac{d}{d\rho}\left(\rho\frac{d^2v}{d\rho^2}\right) - \frac{1}{\rho^2}\frac{dv}{d\rho} = 0$$

and that this is satisfied by

$$v = v_0 + a_2\rho^2.$$

2.4.12 A conducting wire along the z-axis carries a current I. The resulting magnetic vector potential is given by

$$\mathbf{A} = \hat{\mathbf{z}}\frac{\mu I}{2\pi}\ln\left(\frac{1}{\rho}\right).$$

Show that the magnetic induction **B** is given by

$$\mathbf{B} = \hat{\boldsymbol{\varphi}}\frac{\mu I}{2\pi\rho}.$$

2.4.13 A force is described by

$$\mathbf{F} = -\hat{\mathbf{x}}\frac{y}{x^2+y^2} + \hat{\mathbf{y}}\frac{x}{x^2+y^2}.$$

(a) Express **F** in circular cylindrical coordinates.

Operating entirely in circular cylindrical coordinates for (b) and (c),

(b) calculate the curl of **F** and
(c) calculate the work done by **F** in travers the unit circle once counterclockwise.
(d) How do you reconcile the results of (b) and (c)?

2.4.14 A transverse electromagnetic wave (TEM) in a coaxial waveguide has an electric field $\mathbf{E} = \mathbf{E}(\rho, \varphi)e^{i(kz-\omega t)}$ and a magnetic induction field of $\mathbf{B} = \mathbf{B}(\rho, \varphi)e^{i(kz-\omega t)}$. Since the wave is transverse, neither **E** nor **B** has a z component. The two fields satisfy the **vector** Laplacian equation

$$\nabla^2\mathbf{E}(\rho, \varphi) = 0$$

$$\nabla^2\mathbf{B}(\rho, \varphi) = 0.$$

(a) Show that $\mathbf{E} = \hat{\boldsymbol{\rho}}E_0(a/\rho)e^{i(kz-\omega t)}$ and $\mathbf{B} = \hat{\boldsymbol{\varphi}}B_0(a/\rho)e^{i(kz-\omega t)}$ are solutions. Here a is the radius of the inner conductor and E_0 and B_0 are constant amplitudes.

(b) Assuming a vacuum inside the waveguide, verify that Maxwell's equations are satisfied with

$$B_0/E_0 = k/\omega = \mu_0\varepsilon_0(\omega/k) = 1/c.$$

2.4.15 A calculation of the magnetohydrodynamic pinch effect involves the evaluation of $(\mathbf{B}\cdot\nabla)\mathbf{B}$. If the magnetic induction $\mathbf{B}$ is taken to be $\mathbf{B} = \hat{\boldsymbol{\varphi}} B_\varphi(\rho)$, show that

$$(\mathbf{B}\cdot\nabla)\mathbf{B} = -\hat{\boldsymbol{\rho}} B_\varphi^2/\rho.$$

2.4.16 The linear velocity of particles in a rigid body rotating with angular velocity ω is given by

$$\mathbf{v} = \hat{\boldsymbol{\varphi}}\rho\omega.$$

Integrate $\oint \mathbf{v}\cdot d\boldsymbol{\lambda}$ around a circle in the xy-plane and verify that

$$\frac{\oint \mathbf{v}\cdot d\boldsymbol{\lambda}}{\text{area}} = \nabla\times\mathbf{v}|_z.$$

2.4.17 A proton of mass m, charge $+e$, and (asymptotic) momentum $p = mv$ is incident on a nucleus of charge $+Ze$ at an impact parameter b. Determine the proton's distance of closest approach.

2.5 SPHERICAL POLAR COORDINATES

Relabeling (q_1, q_2, q_3) as (r, θ, φ), we see that the spherical polar coordinate system consists of the following:

1. Concentric spheres centered at the origin,

$$r = \left(x^2 + y^2 + z^2\right)^{1/2} = \text{constant}.$$

2. Right circular cones centered on the z-(polar) axis, vertices at the origin,

$$\theta = \arccos\frac{z}{(x^2 + y^2 + z^2)^{1/2}} = \text{constant}.$$

3. Half-planes through the z-(polar) axis,

$$\varphi = \arctan\frac{y}{x} = \text{constant}.$$

By our arbitrary choice of definitions of θ, the polar angle, and φ, the azimuth angle, the z-axis is singled out for special treatment. The transformation equations corresponding to Eq. (2.1) are

$$x = r\sin\theta\cos\varphi, \qquad y = r\sin\theta\sin\varphi, \qquad z = r\cos\theta, \tag{2.38}$$

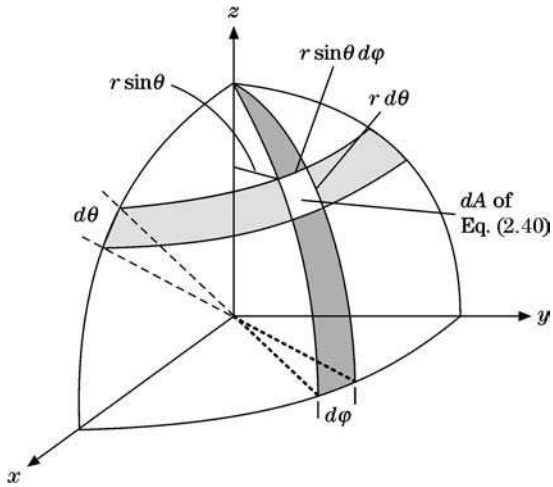

FIGURE 2.7 Spherical polar coordinate area elements.

measuring θ from the positive z-axis and φ in the xy-plane from the positive x-axis. The ranges of values are $0 \leqslant r < \infty$, $0 \leqslant \theta \leqslant \pi$, and $0 \leqslant \varphi \leqslant 2\pi$. At $r = 0$, θ and φ are undefined. From differentiation of Eq. (2.38),

$$\begin{aligned} h_1 &= h_r = 1, \\ h_2 &= h_\theta = r, \\ h_3 &= h_\varphi = r\sin\theta. \end{aligned} \tag{2.39}$$

This gives a line element

$$d\mathbf{r} = \hat{\mathbf{r}}\,dr + \hat{\boldsymbol{\theta}} r\,d\theta + \hat{\boldsymbol{\varphi}} r\sin\theta\,d\varphi,$$

so

$$ds^2 = d\mathbf{r}\cdot d\mathbf{r} = dr^2 + r^2 d\theta^2 + r^2\sin^2\theta\,d\varphi^2,$$

the coordinates being obviously orthogonal. In this spherical coordinate system the area element (for $r =$ constant) is

$$dA = d\sigma_{\theta\varphi} = r^2\sin\theta\,d\theta\,d\varphi, \tag{2.40}$$

the light, unshaded area in Fig. 2.7. Integrating over the azimuth φ, we find that the area element becomes a ring of width $d\theta$,

$$dA_\theta = 2\pi r^2\sin\theta\,d\theta. \tag{2.41}$$

This form will appear repeatedly in problems in spherical polar coordinates with azimuthal symmetry, such as the scattering of an unpolarized beam of particles. By definition of solid radians, or steradians, an element of solid angle $d\Omega$ is given by

$$d\Omega = \frac{dA}{r^2} = \sin\theta\,d\theta\,d\varphi. \tag{2.42}$$

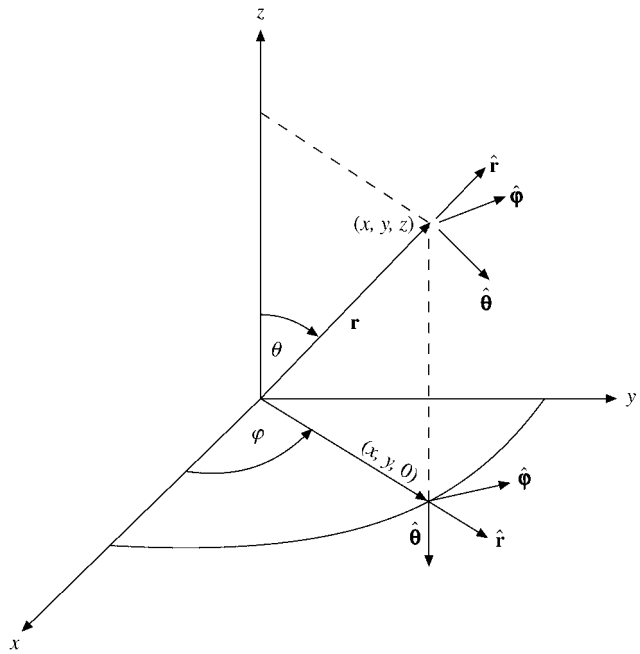

FIGURE 2.8 Spherical polar coordinates.

Integrating over the entire spherical surface, we obtain

$$\int d\Omega = 4\pi.$$

From Eq. (2.11) the volume element is

$$d\tau = r^2\,dr\sin\theta\,d\theta\,d\varphi = r^2\,dr\,d\Omega. \tag{2.43}$$

The spherical polar coordinate unit vectors are shown in Fig. 2.8.

It must be emphasized that **the unit vectors $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\varphi}}$ vary in direction as the angles θ and φ vary.** Specifically, the θ and φ derivatives of these spherical polar coordinate unit vectors do not vanish (Exercise 2.5.2). When differentiating vectors in spherical polar (or in any non-Cartesian system), this variation of the unit vectors with position must not be neglected. In terms of the fixed-direction Cartesian unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ (cp. Eq. (2.38)),

$$\begin{aligned}
\hat{\mathbf{r}} &= \hat{\mathbf{x}}\sin\theta\cos\varphi + \hat{\mathbf{y}}\sin\theta\sin\varphi + \hat{\mathbf{z}}\cos\theta,\\
\hat{\boldsymbol{\theta}} &= \hat{\mathbf{x}}\cos\theta\cos\varphi + \hat{\mathbf{y}}\cos\theta\sin\varphi - \hat{\mathbf{z}}\sin\theta = \frac{\partial\hat{\mathbf{r}}}{\partial\theta},\\
\hat{\boldsymbol{\varphi}} &= -\hat{\mathbf{x}}\sin\varphi + \hat{\mathbf{y}}\cos\varphi = \frac{1}{\sin\theta}\frac{\partial\hat{\mathbf{r}}}{\partial\varphi},
\end{aligned} \tag{2.44}$$

which follow from

$$0 = \frac{\partial\hat{\mathbf{r}}^2}{\partial\theta} = 2\hat{\mathbf{r}}\cdot\frac{\partial\hat{\mathbf{r}}}{\partial\theta}, \qquad 0 = \frac{\partial\hat{\mathbf{r}}^2}{\partial\varphi} = 2\hat{\mathbf{r}}\cdot\frac{\partial\hat{\mathbf{r}}}{\partial\varphi}.$$

Note that Exercise 2.5.5 gives the inverse transformation and that a given vector can now be expressed in a number of different (but equivalent) ways. For instance, the position vector $\mathbf{r}$ may be written

$$\begin{aligned}\mathbf{r} &= \hat{\mathbf{r}}r = \hat{\mathbf{r}}\left(x^2+y^2+z^2\right)^{1/2}\\ &= \hat{\mathbf{x}}x+\hat{\mathbf{y}}y+\hat{\mathbf{z}}z\\ &= \hat{\mathbf{x}}r\sin\theta\cos\varphi+\hat{\mathbf{y}}r\sin\theta\sin\varphi+\hat{\mathbf{z}}r\cos\theta. \end{aligned}\tag{2.45}$$

Select the form that is most useful for your particular problem.

From Section 2.2, relabeling the curvilinear coordinate unit vectors $\hat{\mathbf{q}}_1$, $\hat{\mathbf{q}}_2$, and $\hat{\mathbf{q}}_3$ as $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\varphi}}$ gives

$$\nabla\psi = \hat{\mathbf{r}}\frac{\partial\psi}{\partial r}+\hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial\psi}{\partial\theta}+\hat{\boldsymbol{\varphi}}\frac{1}{r\sin\theta}\frac{\partial\psi}{\partial\varphi}, \tag{2.46}$$

$$\nabla\cdot\mathbf{V} = \frac{1}{r^2\sin\theta}\left[\sin\theta\frac{\partial}{\partial r}(r^2V_r)+r\frac{\partial}{\partial\theta}(\sin\theta V_\theta)+r\frac{\partial V_\varphi}{\partial\varphi}\right], \tag{2.47}$$

$$\nabla\cdot\nabla\psi = \frac{1}{r^2\sin\theta}\left[\sin\theta\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)+\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right)+\frac{1}{\sin\theta}\frac{\partial^2\psi}{\partial\varphi^2}\right], \tag{2.48}$$

$$\nabla\times\mathbf{V} = \frac{1}{r^2\sin\theta}\begin{vmatrix}\hat{\mathbf{r}} & r\hat{\boldsymbol{\theta}} & r\sin\theta\hat{\boldsymbol{\varphi}}\\ \dfrac{\partial}{\partial r} & \dfrac{\partial}{\partial\theta} & \dfrac{\partial}{\partial\varphi}\\ V_r & rV_\theta & r\sin\theta V_\varphi\end{vmatrix}. \tag{2.49}$$

Occasionally, the vector Laplacian $\nabla^2\mathbf{V}$ is needed in spherical polar coordinates. It is best obtained by using the vector identity (Eq. (1.85)) of Chapter 1. For reference

$$\begin{aligned}\nabla^2\mathbf{V}|_r &= \left(-\frac{2}{r^2}+\frac{2}{r}\frac{\partial}{\partial r}+\frac{\partial^2}{\partial r^2}+\frac{\cos\theta}{r^2\sin\theta}\frac{\partial}{\partial\theta}+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}+\frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}\right)V_r\\ &\quad+\left(-\frac{2}{r^2}\frac{\partial}{\partial\theta}-\frac{2\cos\theta}{r^2\sin\theta}\right)V_\theta+\left(-\frac{2}{r^2\sin\theta}\frac{\partial}{\partial\varphi}\right)V_\varphi\\ &= \nabla^2V_r-\frac{2}{r^2}V_r-\frac{2}{r^2}\frac{\partial V_\theta}{\partial\theta}-\frac{2\cos\theta}{r^2\sin\theta}V_\theta-\frac{2}{r^2\sin\theta}\frac{\partial V_\varphi}{\partial\varphi}, \end{aligned}\tag{2.50}$$

$$\nabla^2\mathbf{V}|_\theta = \nabla^2V_\theta-\frac{1}{r^2\sin^2\theta}V_\theta+\frac{2}{r^2}\frac{\partial V_r}{\partial\theta}-\frac{2\cos\theta}{r^2\sin^2\theta}\frac{\partial V_\varphi}{\partial\varphi}, \tag{2.51}$$

$$\nabla^2\mathbf{V}|_\varphi = \nabla^2V_\varphi-\frac{1}{r^2\sin^2\theta}V_\varphi+\frac{2}{r^2\sin\theta}\frac{\partial V_r}{\partial\varphi}+\frac{2\cos\theta}{r^2\sin^2\theta}\frac{\partial V_\theta}{\partial\varphi}. \tag{2.52}$$

These expressions for the components of $\nabla^2\mathbf{V}$ are undeniably messy, but sometimes they are needed.

Example 2.5.1 ∇, $\nabla\cdot$, $\nabla\times$ FOR A CENTRAL FORCE

Using Eqs. (2.46) to (2.49), we can reproduce by inspection some of the results derived in Chapter 1 by laborious application of Cartesian coordinates.

From Eq. (2.46),

$$\begin{aligned}\nabla f(r) &= \hat{\mathbf{r}}\frac{df}{dr},\\ \nabla r^n &= \hat{\mathbf{r}} n r^{n-1}.\end{aligned} \tag{2.53}$$

For the Coulomb potential $V = Ze/(4\pi\varepsilon_0 r)$, the electric field is $\mathbf{E} = -\nabla V = \frac{Ze}{4\pi\varepsilon_0 r^2}\hat{\mathbf{r}}$.

From Eq. (2.47),

$$\begin{aligned}\nabla\cdot\hat{\mathbf{r}} f(r) &= \frac{2}{r}f(r) + \frac{df}{dr},\\ \nabla\cdot\hat{\mathbf{r}} r^n &= (n+2)r^{n-1}.\end{aligned} \tag{2.54}$$

For $r > 0$ the charge density of the electric field of the Coulomb potential is $\rho = \nabla\cdot\mathbf{E} = \frac{Ze}{4\pi\varepsilon_0}\nabla\cdot\frac{\hat{\mathbf{r}}}{r^2} = 0$ because $n = -2$.

From Eq. (2.48),

$$\nabla^2 f(r) = \frac{2}{r}\frac{df}{dr} + \frac{d^2 f}{dr^2}, \tag{2.55}$$

$$\nabla^2 r^n = n(n+1)r^{n-2}, \tag{2.56}$$

in contrast to the ordinary radial second derivative of r^n involving $n-1$ instead of $n+1$.

Finally, from Eq. (2.49),

$$\nabla\times\hat{\mathbf{r}} f(r) = 0. \tag{2.57}$$

■

Example 2.5.2 MAGNETIC VECTOR POTENTIAL

The computation of the magnetic vector potential of a single current loop in the xy-plane uses Oersted's law, $\nabla\times\mathbf{H} = \mathbf{J}$, in conjunction with $\mu_0\mathbf{H} = \mathbf{B} = \nabla\times\mathbf{A}$ (see Examples 1.9.2 and 1.12.1), and involves the evaluation of

$$\mu_0\mathbf{J} = \nabla\times\big[\nabla\times\hat{\boldsymbol{\varphi}}A_\varphi(r,\theta)\big].$$

In spherical polar coordinates this reduces to

$$\begin{aligned}\mu_0\mathbf{J} &= \nabla\times\frac{1}{r^2\sin\theta}\begin{vmatrix}\hat{\mathbf{r}} & r\hat{\boldsymbol{\theta}} & r\sin\theta\hat{\boldsymbol{\varphi}}\\ \dfrac{\partial}{\partial r} & \dfrac{\partial}{\partial\theta} & \dfrac{\partial}{\partial\varphi}\\ 0 & 0 & r\sin\theta A_\varphi(r,\theta)\end{vmatrix}\\ &= \nabla\times\frac{1}{r^2\sin\theta}\left[\hat{\mathbf{r}}\frac{\partial}{\partial\theta}(r\sin\theta A_\varphi) - r\hat{\boldsymbol{\theta}}\frac{\partial}{\partial r}(r\sin\theta A_\varphi)\right].\end{aligned}$$

Taking the curl a second time, we obtain

$$\mu_0\mathbf{J} = \frac{1}{r^2\sin\theta}\begin{vmatrix} \hat{\mathbf{r}} & r\hat{\boldsymbol{\theta}} & r\sin\theta\hat{\boldsymbol{\varphi}} \\ \dfrac{\partial}{\partial r} & \dfrac{\partial}{\partial\theta} & \dfrac{\partial}{\partial\varphi} \\ \dfrac{1}{r^2\sin\theta}\dfrac{\partial}{\partial\theta}(r\sin\theta A_\varphi) & -\dfrac{1}{r\sin\theta}\dfrac{\partial}{\partial r}(r\sin\theta A_\varphi) & 0 \end{vmatrix}.$$

By expanding the determinant along the top row, we have

$$\begin{aligned}\mu_0\mathbf{J} &= -\hat{\boldsymbol{\varphi}}\left\{\frac{1}{r}\frac{\partial^2}{\partial r^2}(rA_\varphi) + \frac{1}{r^2}\frac{\partial}{\partial\theta}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta A_\varphi)\right]\right\} \\ &= -\hat{\boldsymbol{\varphi}}\left[\nabla^2 A_\varphi(r,\theta) - \frac{1}{r^2\sin^2\theta}A_\varphi(r,\theta)\right]. \end{aligned} \tag{2.58}$$

■

Exercises

2.5.1 Express the spherical polar unit vectors in Cartesian unit vectors.

ANS. $\hat{\mathbf{r}} = \hat{\mathbf{x}}\sin\theta\cos\varphi + \hat{\mathbf{y}}\sin\theta\sin\varphi + \hat{\mathbf{z}}\cos\theta,$
$\hat{\boldsymbol{\theta}} = \hat{\mathbf{x}}\cos\theta\cos\varphi + \hat{\mathbf{y}}\cos\theta\sin\varphi - \hat{\mathbf{z}}\sin\theta,$
$\hat{\boldsymbol{\varphi}} = -\hat{\mathbf{x}}\sin\varphi + \hat{\mathbf{y}}\cos\varphi.$

2.5.2 (a) From the results of Exercise 2.5.1, calculate the partial derivatives of $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\varphi}}$ with respect to r, θ, and φ.

(b) With ∇ given by

$$\hat{\mathbf{r}}\frac{\partial}{\partial r} + \hat{\boldsymbol{\theta}}\frac{1}{r}\frac{\partial}{\partial\theta} + \hat{\boldsymbol{\varphi}}\frac{1}{r\sin\theta}\frac{\partial}{\partial\varphi}$$

(greatest space rate of change), use the results of part (a) to calculate $\nabla\cdot\nabla\psi$. This is an alternate derivation of the Laplacian.

Note. The derivatives of the left-hand ∇ operate on the unit vectors of the right-hand ∇ **before** the unit vectors are dotted together.

2.5.3 A rigid body is rotating about a fixed axis with a constant angular velocity $\boldsymbol{\omega}$. Take $\boldsymbol{\omega}$ to be along the z-axis. Using spherical polar coordinates,

(a) Calculate

$$\mathbf{v} = \boldsymbol{\omega}\times\mathbf{r}.$$

(b) Calculate

$$\nabla\times\mathbf{v}.$$

ANS. (a) $\mathbf{v} = \hat{\boldsymbol{\varphi}}\omega r\sin\theta,$
(b) $\nabla\times\mathbf{v} = 2\boldsymbol{\omega}.$

2.5.4 The coordinate system (x, y, z) is rotated through an angle Φ counterclockwise about an axis defined by the unit vector $\mathbf{n}$ into system (x', y', z'). In terms of the new coordinates the radius vector becomes

$$\mathbf{r}' = \mathbf{r}\cos\Phi + \mathbf{r}\times\mathbf{n}\sin\Phi + \mathbf{n}(\mathbf{n}\cdot\mathbf{r})(1-\cos\Phi).$$

(a) Derive this expression from geometric considerations.
(b) Show that it reduces as expected for $\mathbf{n} = \hat{\mathbf{z}}$. The answer, in matrix form, appears in Eq. (3.90).
(c) Verify that $r'^2 = r^2$.

2.5.5 Resolve the Cartesian unit vectors into their spherical polar components:

$$\hat{\mathbf{x}} = \hat{\mathbf{r}}\sin\theta\cos\varphi + \hat{\boldsymbol{\theta}}\cos\theta\cos\varphi - \hat{\boldsymbol{\varphi}}\sin\varphi,$$
$$\hat{\mathbf{y}} = \hat{\mathbf{r}}\sin\theta\sin\varphi + \hat{\boldsymbol{\theta}}\cos\theta\sin\varphi + \hat{\boldsymbol{\varphi}}\cos\varphi,$$
$$\hat{\mathbf{z}} = \hat{\mathbf{r}}\cos\theta - \hat{\boldsymbol{\theta}}\sin\theta.$$

2.5.6 The direction of one vector is given by the angles θ_1 and φ_1. For a second vector the corresponding angles are θ_2 and φ_2. Show that the cosine of the included angle γ is given by

$$\cos\gamma = \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2\cos(\varphi_1 - \varphi_2).$$

See Fig. 12.15.

2.5.7 A certain vector $\mathbf{V}$ has no radial component. Its curl has no tangential components. What does this imply about the radial dependence of the tangential components of $\mathbf{V}$?

2.5.8 Modern physics lays great stress on the property of parity — whether a quantity remains invariant or changes sign under an inversion of the coordinate system. In Cartesian coordinates this means $x \to -x$, $y \to -y$, and $z \to -z$.

(a) Show that the inversion (reflection through the origin) of a point (r, θ, φ) relative to **fixed** x-, y-, z-axes consists of the transformation

$$r \to r, \qquad \theta \to \pi - \theta, \qquad \varphi \to \varphi \pm \pi.$$

(b) Show that $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\varphi}}$ have odd parity (reversal of direction) and that $\hat{\boldsymbol{\theta}}$ has even parity.

2.5.9 With $\mathbf{A}$ any vector,

$$\mathbf{A}\cdot\nabla\mathbf{r} = \mathbf{A}.$$

(a) Verify this result in Cartesian coordinates.
(b) Verify this result using spherical polar coordinates. (Equation (2.46) provides ∇.)

2.5.10 Find the spherical coordinate components of the velocity and acceleration of a moving particle:

$$
\begin{aligned}
v_r &= \dot{r},\\
v_\theta &= r\dot{\theta},\\
v_\varphi &= r\sin\theta\dot{\varphi},\\
a_r &= \ddot{r} - r\dot{\theta}^2 - r\sin^2\theta\dot{\varphi}^2,\\
a_\theta &= r\ddot{\theta} + 2\dot{r}\dot{\theta} - r\sin\theta\cos\theta\dot{\varphi}^2,\\
a_\varphi &= r\sin\theta\ddot{\varphi} + 2\dot{r}\sin\theta\dot{\varphi} + 2r\cos\theta\dot{\theta}\dot{\varphi}.
\end{aligned}
$$

Hint.

$$
\begin{aligned}
\mathbf{r}(t) &= \hat{\mathbf{r}}(t)r(t)\\
&= \left[\hat{\mathbf{x}}\sin\theta(t)\cos\varphi(t) + \hat{\mathbf{y}}\sin\theta(t)\sin\varphi(t) + \hat{\mathbf{z}}\cos\theta(t)\right]r(t).
\end{aligned}
$$

Note. Using the Lagrangian techniques of Section 17.3, we may obtain these results somewhat more elegantly. The dot in $\dot{r}, \dot{\theta}, \dot{\varphi}$ means time derivative, $\dot{r} = dr/dt, \dot{\theta} = d\theta/dt, \dot{\varphi} = d\varphi/dt$. The notation was originated by Newton.

2.5.11 A particle m moves in response to a central force according to Newton's second law,

$$m\ddot{\mathbf{r}} = \hat{\mathbf{r}}f(r).$$

Show that $\mathbf{r}\times\dot{\mathbf{r}} = \mathbf{c}$, a constant, and that the geometric interpretation of this leads to Kepler's second law.

2.5.12 Express $\partial/\partial x$, $\partial/\partial y$, $\partial/\partial z$ in spherical polar coordinates.

$$
\begin{aligned}
\text{ANS. } \frac{\partial}{\partial x} &= \sin\theta\cos\varphi\frac{\partial}{\partial r} + \cos\theta\cos\varphi\frac{1}{r}\frac{\partial}{\partial\theta} - \frac{\sin\varphi}{r\sin\theta}\frac{\partial}{\partial\varphi},\\
\frac{\partial}{\partial y} &= \sin\theta\sin\varphi\frac{\partial}{\partial r} + \cos\theta\sin\varphi\frac{1}{r}\frac{\partial}{\partial\theta} + \frac{\cos\varphi}{r\sin\theta}\frac{\partial}{\partial\varphi},\\
\frac{\partial}{\partial z} &= \cos\theta\frac{\partial}{\partial r} - \sin\theta\frac{1}{r}\frac{\partial}{\partial\theta}.
\end{aligned}
$$

Hint. Equate ∇_{xyz} and $\nabla_{r\theta\varphi}$.

2.5.13 From Exercise 2.5.12 show that

$$-i\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right) = -i\frac{\partial}{\partial\varphi}.$$

This is the quantum mechanical operator corresponding to the z-component of orbital angular momentum.

2.5.14 With the quantum mechanical orbital angular momentum operator defined as $\mathbf{L} = -i(\mathbf{r}\times\nabla)$, show that

(a) $L_x + iL_y = e^{i\varphi}\left(\dfrac{\partial}{\partial\theta} + i\cot\theta\dfrac{\partial}{\partial\varphi}\right)$,

(b) $L_x - iL_y = -e^{-i\varphi}\left(\dfrac{\partial}{\partial\theta} - i\cot\theta\dfrac{\partial}{\partial\varphi}\right)$.

(These are the raising and lowering operators of Section 4.3.)

2.5.15 Verify that $\mathbf{L}\times\mathbf{L} = i\mathbf{L}$ in spherical polar coordinates. $\mathbf{L} = -i(\mathbf{r}\times\boldsymbol{\nabla})$, the quantum mechanical orbital angular momentum operator.
Hint. Use spherical polar coordinates for $\mathbf{L}$ but Cartesian components for the cross product.

2.5.16 (a) From Eq. (2.46) show that

$$\mathbf{L} = -i(\mathbf{r}\times\boldsymbol{\nabla}) = i\left(\hat{\boldsymbol{\theta}}\frac{1}{\sin\theta}\frac{\partial}{\partial\varphi} - \hat{\boldsymbol{\varphi}}\frac{\partial}{\partial\theta}\right).$$

(b) Resolving $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\varphi}}$ into Cartesian components, determine L_x, L_y, and L_z in terms of θ, φ, and their derivatives.

(c) From $L^2 = L_x^2 + L_y^2 + L_z^2$ show that

$$\mathbf{L}^2 = -\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) - \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}$$
$$= -r^2\boldsymbol{\nabla}^2 + \frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right).$$

This latter identity is useful in relating orbital angular momentum and Legendre's differential equation, Exercise 9.3.8.

2.5.17 With $\mathbf{L} = -i\mathbf{r}\times\boldsymbol{\nabla}$, verify the operator identities

(a) $\boldsymbol{\nabla} = \hat{\mathbf{r}}\dfrac{\partial}{\partial r} - i\dfrac{\mathbf{r}\times\mathbf{L}}{r^2}$,

(b) $\mathbf{r}\boldsymbol{\nabla}^2 - \boldsymbol{\nabla}\left(1 + r\dfrac{\partial}{\partial r}\right) = i\boldsymbol{\nabla}\times\mathbf{L}$.

2.5.18 Show that the following three forms (spherical coordinates) of $\boldsymbol{\nabla}^2\psi(r)$ are equivalent:

(a) $\dfrac{1}{r^2}\dfrac{d}{dr}\left[r^2\dfrac{d\psi(r)}{dr}\right]$; (b) $\dfrac{1}{r}\dfrac{d^2}{dr^2}\big[r\psi(r)\big]$; (c) $\dfrac{d^2\psi(r)}{dr^2} + \dfrac{2}{r}\dfrac{d\psi(r)}{dr}$.

The second form is particularly convenient in establishing a correspondence between spherical polar and Cartesian descriptions of a problem.

2.5.19 One model of the solar corona assumes that the steady-state equation of heat flow,

$$\boldsymbol{\nabla}\cdot(k\boldsymbol{\nabla}T) = 0,$$

is satisfied. Here, k, the thermal conductivity, is proportional to $T^{5/2}$. Assuming that the temperature T is proportional to r^n, show that the heat flow equation is satisfied by $T = T_0(r_0/r)^{2/7}$.

2.5.20 A certain force field is given by

$$\mathbf{F} = \hat{\mathbf{r}}\frac{2P\cos\theta}{r^3} + \hat{\boldsymbol{\theta}}\frac{P}{r^3}\sin\theta, \qquad r \geqslant P/2$$

(in spherical polar coordinates).

(a) Examine $\nabla \times \mathbf{F}$ to see if a potential exists.
(b) Calculate $\oint \mathbf{F} \cdot d\boldsymbol{\lambda}$ for a unit circle in the plane $\theta = \pi/2$. What does this indicate about the force being conservative or nonconservative?
(c) If you believe that $\mathbf{F}$ may be described by $\mathbf{F} = -\nabla\psi$, find ψ. Otherwise simply state that no acceptable potential exists.

2.5.21 (a) Show that $\mathbf{A} = -\hat{\boldsymbol{\varphi}}\cot\theta/r$ is a solution of $\nabla \times \mathbf{A} = \hat{\mathbf{r}}/r^2$.
(b) Show that this spherical polar coordinate solution agrees with the solution given for Exercise 1.13.6:

$$\mathbf{A} = \hat{\mathbf{x}}\frac{yz}{r(x^2+y^2)} - \hat{\mathbf{y}}\frac{xz}{r(x^2+y^2)}.$$

Note that the solution diverges for $\theta = 0, \pi$ corresponding to $x, y = 0$.
(c) Finally, show that $\mathbf{A} = -\hat{\boldsymbol{\theta}}\varphi\sin\theta/r$ is a solution. Note that although this solution does not diverge ($r \neq 0$), it is no longer single-valued for all possible azimuth angles.

2.5.22 A magnetic vector potential is given by

$$\mathbf{A} = \frac{\mu_0}{4\pi}\frac{\mathbf{m}\times\mathbf{r}}{r^3}.$$

Show that this leads to the magnetic induction $\mathbf{B}$ of a point magnetic dipole with dipole moment $\mathbf{m}$.

ANS. for $\mathbf{m} = \hat{\mathbf{z}}m$,

$$\nabla \times \mathbf{A} = \hat{\mathbf{r}}\frac{\mu_0}{4\pi}\frac{2m\cos\theta}{r^3} + \hat{\boldsymbol{\theta}}\frac{\mu_0}{4\pi}\frac{m\sin\theta}{r^3}.$$

Compare Eqs. (12.133) and (12.134)

2.5.23 At large distances from its source, electric dipole radiation has fields

$$\mathbf{E} = a_E\sin\theta\frac{e^{i(kr-\omega t)}}{r}\hat{\boldsymbol{\theta}}, \qquad \mathbf{B} = a_B\sin\theta\frac{e^{i(kr-\omega t)}}{r}\hat{\boldsymbol{\varphi}}.$$

Show that Maxwell's equations

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \text{and} \qquad \nabla \times \mathbf{B} = \varepsilon_0\mu_0\frac{\partial \mathbf{E}}{\partial t}$$

are satisfied, if we take

$$\frac{a_E}{a_B} = \frac{\omega}{k} = c = (\varepsilon_0\mu_0)^{-1/2}.$$

Hint. Since r is large, terms of order r^{-2} may be dropped.

2.5.24 The magnetic vector potential for a uniformly charged rotating spherical shell is

$$\mathbf{A} = \begin{cases} \hat{\boldsymbol{\varphi}} \dfrac{\mu_0 a^4 \sigma \omega}{3} \cdot \dfrac{\sin\theta}{r^2}, & r > a \\ \hat{\boldsymbol{\varphi}} \dfrac{\mu_0 a \sigma \omega}{3} \cdot r\cos\theta, & r < a. \end{cases}$$

(a = radius of spherical shell, σ = surface charge density, and ω = angular velocity.) Find the magnetic induction $\mathbf{B} = \nabla \times \mathbf{A}$.

$$\text{ANS. } B_r(r,\theta) = \frac{2\mu_0 a^4 \sigma \omega}{3} \cdot \frac{\cos\theta}{r^3}, \qquad r > a,$$

$$B_\theta(r,\theta) = \frac{\mu_0 a^4 \sigma \omega}{3} \cdot \frac{\sin\theta}{r^3}, \qquad r > a,$$

$$\mathbf{B} = \hat{\mathbf{z}} \frac{2\mu_0 a \sigma \omega}{3}, \qquad r < a.$$

2.5.25 (a) Explain why ∇^2 in plane polar coordinates follows from ∇^2 in circular cylindrical coordinates with z = constant.

(b) Explain why taking ∇^2 in spherical polar coordinates and restricting θ to $\pi/2$ does **not** lead to the plane polar form of ∇.

Note.

$$\nabla^2(\rho,\varphi) = \frac{\partial^2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial}{\partial\rho} + \frac{1}{\rho^2}\frac{\partial^2}{\partial\varphi^2}.$$

2.6 TENSOR ANALYSIS

Introduction, Definitions

Tensors are important in many areas of physics, including general relativity and electrodynamics. Scalars and vectors are special cases of tensors. In Chapter 1, a quantity that did not change under rotations of the coordinate system in three-dimensional space, an invariant, was labeled a scalar. A **scalar** is specified by one real number and is a **tensor of rank 0**. A quantity whose components transformed under rotations like those of the distance of a point from a chosen origin (Eq. (1.9), Section 1.2) was called a vector. The transformation of the components of the vector under a rotation of the coordinates preserves the vector as a geometric entity (such as an arrow in space), independent of the orientation of the reference frame. In three-dimensional space, a **vector** is specified by $3 = 3^1$ real numbers, for example, its Cartesian components, and is a **tensor of rank 1**. A **tensor of rank** n has 3^n components that transform in a definite way.[5] This transformation philosophy is of central importance for tensor analysis and conforms with the mathematician's concept of vector and vector (or linear) space and the physicist's notion that physical observables must not depend on the choice of coordinate frames. There is a physical basis for such a philosophy: We describe the physical world by mathematics, but any physical predictions we make

[5] In N-dimensional space a tensor of rank n has N^n components.

must be independent of our mathematical conventions, such as a coordinate system with its arbitrary origin and orientation of its axes.

There is a possible ambiguity in the transformation law of a vector

$$A_i' = \sum_j a_{ij} A_j, \tag{2.59}$$

in which a_{ij} is the cosine of the angle between the x_i'-axis and the x_j-axis.

If we start with a differential distance vector $d\mathbf{r}$, then, taking dx_i' to be a function of the unprimed variables,

$$dx_i' = \sum_j \frac{\partial x_i'}{\partial x_j}\, dx_j \tag{2.60}$$

by partial differentiation. If we set

$$a_{ij} = \frac{\partial x_i'}{\partial x_j}, \tag{2.61}$$

Eqs. (2.59) and (2.60) are consistent. Any set of quantities A^j transforming according to

$$A'^i = \sum_j \frac{\partial x_i'}{\partial x_j} A^j \tag{2.62a}$$

is defined as a **contravariant** vector, whose indices we write as **superscript**; this includes the Cartesian coordinate vector $x^i = x_i$ from now on.

However, we have already encountered a slightly different type of vector transformation. The gradient of a scalar $\nabla\varphi$, defined by

$$\nabla\varphi = \hat{\mathbf{x}}\frac{\partial\varphi}{\partial x^1} + \hat{\mathbf{y}}\frac{\partial\varphi}{\partial x^2} + \hat{\mathbf{z}}\frac{\partial\varphi}{\partial x^3} \tag{2.63}$$

(using x^1, x^2, x^3 for x, y, z), transforms as

$$\frac{\partial\varphi'}{\partial x'^i} = \sum_j \frac{\partial\varphi}{\partial x^j}\frac{\partial x^j}{\partial x'^i}, \tag{2.64}$$

using $\varphi = \varphi(x, y, z) = \varphi(x', y', z') = \varphi'$, φ defined as a scalar quantity. Notice that this differs from Eq. (2.62) in that we have $\partial x^j/\partial x'^i$ instead of $\partial x'^i/\partial x^j$. Equation (2.64) is taken as the definition of a **covariant** vector, with the gradient as the prototype. The covariant analog of Eq. (2.62a) is

$$A_i' = \sum_j \frac{\partial x^j}{\partial x'^i}\, A_j. \tag{2.62b}$$

Only in Cartesian coordinates is

$$\frac{\partial x^j}{\partial x'^i} = \frac{\partial x'^i}{\partial x^j} = a_{ij} \tag{2.65}$$

so that there no difference between contravariant and covariant transformations. In other systems, Eq. (2.65) in general does not apply, and the distinction between contravariant and covariant is real and must be observed. This is of prime importance in the curved Riemannian space of general relativity.

In the remainder of this section the components of any **contravariant** vector are denoted by a **superscript**, A^i, whereas a **subscript** is used for the components of a **covariant** vector A_i.[6]

Definition of Tensors of Rank 2

Now we proceed to define **contravariant, mixed, and covariant tensors of rank 2** by the following equations for their components under coordinate transformations:

$$A'^{ij} = \sum_{kl} \frac{\partial x'^i}{\partial x^k} \frac{\partial x'^j}{\partial x^l} A^{kl},$$

$$B'^i{}_j = \sum_{kl} \frac{\partial x'^i}{\partial x^k} \frac{\partial x^l}{\partial x'^j} B^k{}_l, \tag{2.66}$$

$$C'_{ij} = \sum_{kl} \frac{\partial x^k}{\partial x'^i} \frac{\partial x^l}{\partial x'^j} C_{kl}.$$

Clearly, the rank goes as the number of partial derivatives (or direction cosines) in the definition: 0 for a scalar, 1 for a vector, 2 for a second-rank tensor, and so on. Each index (subscript or superscript) ranges over the number of dimensions of the space. The number of indices (equal to the rank of tensor) is independent of the dimensions of the space. We see that A^{kl} is contravariant with respect to both indices, C_{kl} is covariant with respect to both indices, and $B^k{}_l$ transforms contravariantly with respect to the first index k but covariantly with respect to the second index l. Once again, if we are using Cartesian coordinates, all three forms of the tensors of second rank contravariant, mixed, and covariant are — the same.

As with the components of a vector, the transformation laws for the components of a tensor, Eq. (2.66), yield entities (and properties) that are independent of the choice of reference frame. This is what makes tensor analysis important in physics. The independence of reference frame (invariance) is ideal for expressing and investigating universal physical laws.

The second-rank tensor $\mathbf{A}$ (components A^{kl}) may be conveniently represented by writing out its components in a square array (3×3 if we are in three-dimensional space):

$$\mathbf{A} = \begin{pmatrix} A^{11} & A^{12} & A^{13} \\ A^{21} & A^{22} & A^{23} \\ A^{31} & A^{32} & A^{33} \end{pmatrix}. \tag{2.67}$$

This does not mean that any square array of numbers or functions forms a tensor. The essential condition is that the components transform according to Eq. (2.66).

[6] This means that the coordinates (x, y, z) are written (x^1, x^2, x^3) since $\mathbf{r}$ transforms as a contravariant vector. The ambiguity of x^2 representing both x squared and y is the price we pay.

In the context of matrix analysis the preceding transformation equations become (for Cartesian coordinates) an orthogonal similarity transformation; see Section 3.3. A geometrical interpretation of a second-rank tensor (the inertia tensor) is developed in Section 3.5.

In summary, tensors are systems of components organized by one or more indices that transform according to specific rules under a set of transformations. The number of indices is called the rank of the tensor. If the transformations are coordinate rotations in three-dimensional space, then tensor analysis amounts to what we did in the sections on curvilinear coordinates and in Cartesian coordinates in Chapter 1. *In four dimensions of Minkowski space–time, the transformations are Lorentz transformations, and tensors of rank* 1 *are called four-vectors.*

Addition and Subtraction of Tensors

The addition and subtraction of tensors is defined in terms of the individual elements, just as for vectors. If

$$\mathbf{A} + \mathbf{B} = \mathbf{C}, \tag{2.68}$$

then

$$A^{ij} + B^{ij} = C^{ij}.$$

Of course, **A** and **B** must be tensors of the same rank and both expressed in a space of the same number of dimensions.

Summation Convention

In tensor analysis it is customary to adopt a summation convention to put Eq. (2.66) and subsequent tensor equations in a more compact form. As long as we are distinguishing between contravariance and covariance, let us agree that when an index appears on one side of an equation, once as a superscript and once as a subscript (except for the coordinates where both are subscripts), we automatically sum over that index. Then we may write the second expression in Eq. (2.66) as

$$B'^{i}{}_{j} = \frac{\partial x'^{i}}{\partial x^{k}} \frac{\partial x^{l}}{\partial x'^{j}} B^{k}{}_{l}, \tag{2.69}$$

with the summation of the right-hand side over k and l implied. This is Einstein's summation convention.[7] The index i is superscript because it is associated with the contravariant x'^{i}; likewise j is subscript because it is related to the covariant gradient.

To illustrate the use of the summation convention and some of the techniques of tensor analysis, let us show that the now-familiar Kronecker delta, δ_{kl}, is really a mixed tensor

[7] In this context $\partial x'^{i}/\partial x^{k}$ might better be written as a^{i}_{k} and $\partial x^{l}/\partial x'^{j}$ as b^{l}_{j}.

of rank 2, $\delta^k{}_l$.[8] The question is: Does $\delta^k{}_l$ transform according to Eq. (2.66)? This is our criterion for calling it a tensor. We have, using the summation convention,

$$\delta^k{}_l \frac{\partial x'^i}{\partial x^k} \frac{\partial x^l}{\partial x'^j} = \frac{\partial x'^i}{\partial x^k} \frac{\partial x^k}{\partial x'^j} \tag{2.70}$$

by definition of the Kronecker delta. Now,

$$\frac{\partial x'^i}{\partial x^k} \frac{\partial x^k}{\partial x'^j} = \frac{\partial x'^i}{\partial x'^j} \tag{2.71}$$

by direct partial differentiation of the right-hand side (chain rule). However, x'^i and x'^j are independent coordinates, and therefore the variation of one with respect to the other must be zero if they are different, unity if they coincide; that is,

$$\frac{\partial x'^i}{\partial x'^j} = \delta'^i{}_j. \tag{2.72}$$

Hence

$$\delta'^i{}_j = \frac{\partial x'^i}{\partial x^k} \frac{\partial x^l}{\partial x'^j} \delta^k{}_l,$$

showing that the $\delta^k{}_l$ are indeed the components of a mixed second-rank tensor. Notice that this result is independent of the number of dimensions of our space. The reason for the upper index i and lower index j is the same as in Eq. (2.69).

The Kronecker delta has one further interesting property. It has the same components in all of our rotated coordinate systems and is therefore called **isotropic**. In Section 2.9 we shall meet a third-rank isotropic tensor and three fourth-rank isotropic tensors. No isotropic first-rank tensor (vector) exists.

Symmetry–Antisymmetry

The order in which the indices appear in our description of a tensor is important. In general, A^{mn} is independent of A^{nm}, but there are some cases of special interest. If, for all m and n,

$$A^{mn} = A^{nm}, \tag{2.73}$$

we call the tensor **symmetric**. If, on the other hand,

$$A^{mn} = -A^{nm}, \tag{2.74}$$

the tensor is **antisymmetric**. Clearly, every (second-rank) tensor can be resolved into symmetric and antisymmetric parts by the identity

$$A^{mn} = \tfrac{1}{2}\big(A^{mn} + A^{nm}\big) + \tfrac{1}{2}\big(A^{mn} - A^{nm}\big), \tag{2.75}$$

the first term on the right being a symmetric tensor, the second, an antisymmetric tensor. A similar resolution of functions into symmetric and antisymmetric parts is of extreme importance to quantum mechanics.

[8] It is common practice to refer to a tensor $\mathbf{A}$ by specifying a typical component, A_{ij}. As long as the reader refrains from writing nonsense such as $\mathbf{A} = A_{ij}$, no harm is done.

Spinors

It was once thought that the system of scalars, vectors, tensors (second-rank), and so on formed a complete mathematical system, one that is adequate for describing a physics independent of the choice of reference frame. But the universe and mathematical physics are not that simple. In the realm of elementary particles, for example, spin zero particles[9] (π mesons, α particles) may be described with scalars, spin 1 particles (deuterons) by vectors, and spin 2 particles (gravitons) by tensors. This listing omits the most common particles: electrons, protons, and neutrons, all with spin $\frac{1}{2}$. These particles are properly described by **spinors**. A spinor is not a scalar, vector, or tensor. A brief introduction to spinors in the context of group theory ($J = 1/2$) appears in Section 4.3.

Exercises

2.6.1 Show that if all the components of any tensor of any rank vanish in one particular coordinate system, they vanish in all coordinate systems.
Note. This point takes on special importance in the four-dimensional curved space of general relativity. If a quantity, expressed as a tensor, exists in one coordinate system, it exists in all coordinate systems and is not just a consequence of a **choice** of a coordinate system (as are centrifugal and Coriolis forces in Newtonian mechanics).

2.6.2 The components of tensor **A** are equal to the corresponding components of tensor **B** in one particular coordinate system, denoted by the superscript 0; that is,

$$A^0_{ij} = B^0_{ij}.$$

Show that tensor **A** is equal to tensor **B**, $A_{ij} = B_{ij}$, in all coordinate systems.

2.6.3 The last three components of a four-dimensional vector vanish in each of two reference frames. If the second reference frame is not merely a rotation of the first about the x_0 axis, that is, if at least one of the coefficients a_{i0} $(i = 1, 2, 3) \neq 0$, show that the zeroth component vanishes in all reference frames. Translated into relativistic mechanics this means that if momentum is conserved in two Lorentz frames, then energy is conserved in all Lorentz frames.

2.6.4 From an analysis of the behavior of a general second-rank tensor under 90° and 180° rotations about the coordinate axes, show that an isotropic second-rank tensor in three-dimensional space must be a multiple of δ_{ij}.

2.6.5 The four-dimensional fourth-rank Riemann–Christoffel curvature tensor of general relativity, R_{iklm}, satisfies the symmetry relations

$$R_{iklm} = -R_{ikml} = -R_{kilm}.$$

With the indices running from 0 to 3, show that the number of independent components is reduced from 256 to 36 and that the condition

$$R_{iklm} = R_{lmik}$$

[9]The particle spin is intrinsic angular momentum (in units of $\hbar$). It is distinct from classical, orbital angular momentum due to motion.

further reduces the number of independent components to 21. Finally, if the components satisfy an identity $R_{iklm} + R_{ilmk} + R_{imkl} = 0$, show that the number of independent components is reduced to 20.
Note. The final three-term identity furnishes new information only if all four indices are different. Then it reduces the number of independent components by one-third.

2.6.6 T_{iklm} is antisymmetric with respect to all pairs of indices. How many independent components has it (in three-dimensional space)?

2.7 CONTRACTION, DIRECT PRODUCT

Contraction

When dealing with vectors, we formed a scalar product (Section 1.3) by summing products of corresponding components:

$$\mathbf{A} \cdot \mathbf{B} = A_i B_i \qquad (\textbf{summation convention}). \tag{2.76}$$

The generalization of this expression in tensor analysis is a process known as contraction. Two indices, one covariant and the other contravariant, are set equal to each other, and then (as implied by the summation convention) we sum over this repeated index. For example, let us contract the second-rank mixed tensor $B'^{i}{}_{j}$,

$$B'^{i}{}_{i} = \frac{\partial x'^{i}}{\partial x^{k}} \frac{\partial x^{l}}{\partial x'^{i}} B^{k}{}_{l} = \frac{\partial x^{l}}{\partial x^{k}} B^{k}{}_{l} \tag{2.77}$$

using Eq. (2.71), and then by Eq. (2.72)

$$B'^{i}{}_{i} = \delta^{l}{}_{k} B^{k}{}_{l} = B^{k}{}_{k}. \tag{2.78}$$

Our contracted second-rank mixed tensor is invariant and therefore a scalar.[10] This is exactly what we obtained in Section 1.3 for the dot product of two vectors and in Section 1.7 for the divergence of a vector. In general, the operation of contraction reduces the rank of a tensor by 2. An example of the use of contraction appears in Chapter 4.

Direct Product

The components of a covariant vector (first-rank tensor) a_i and those of a contravariant vector (first-rank tensor) b^j may be multiplied component by component to give the general term $a_i b^j$. This, by Eq. (2.66) is actually a second-rank tensor, for

$$a'_i b'^{j} = \frac{\partial x^{k}}{\partial x'^{i}} a_k \frac{\partial x'^{j}}{\partial x^{l}} b^{l} = \frac{\partial x^{k}}{\partial x'^{i}} \frac{\partial x'^{j}}{\partial x^{l}} \left(a_k b^{l}\right). \tag{2.79}$$

Contracting, we obtain

$$a'_i b'^{i} = a_k b^{k}, \tag{2.80}$$

[10] In matrix analysis this scalar is the **trace** of the matrix, Section 3.2.

as in Eqs. (2.77) and (2.78), to give the regular scalar product.

The operation of adjoining two vectors a_i and b^j as in the last paragraph is known as forming the **direct product**. For the case of two vectors, the direct product is a tensor of second rank. In this sense we may attach meaning to $\nabla\mathbf{E}$, which was not defined within the framework of vector analysis. In general, the direct product of two tensors is a tensor of rank equal to the sum of the two initial ranks; that is,

$$A^i{}_j B^{kl} = C^i{}_j{}^{kl}, \tag{2.81a}$$

where $C^i{}_j{}^{kl}$ is a tensor of fourth rank. From Eqs. (2.66),

$$C'^i{}_j{}^{kl} = \frac{\partial x'^i}{\partial x^m}\frac{\partial x^n}{\partial x'^j}\frac{\partial x'^k}{\partial x^p}\frac{\partial x'^l}{\partial x^q} C^m{}_n{}^{pq}. \tag{2.81b}$$

The direct product is a technique for creating new, higher-rank tensors. Exercise 2.7.1 is a form of the direct product in which the first factor is ∇. Applications appear in Section 4.6.

When $\mathbf{T}$ is an nth-rank Cartesian tensor, $(\partial/\partial x^i)T_{jkl}\ldots$, a component of $\nabla\mathbf{T}$, is a **Cartesian** tensor of rank $n+1$ (Exercise 2.7.1). However, $(\partial/\partial x^i)T_{jkl}\ldots$ is not a tensor in more general spaces. In non-Cartesian systems $\partial/\partial x'^i$ will act on the partial derivatives $\partial x^p/\partial x'^q$ and destroy the simple tensor transformation relation (see Eq. (2.129)).

So far the distinction between a covariant transformation and a contravariant transformation has been maintained because it does exist in non-Euclidean space and because it is of great importance in general relativity. In Sections 2.10 and 2.11 we shall develop differential relations for general tensors. Often, however, because of the simplification achieved, we restrict ourselves to Cartesian tensors. As noted in Section 2.6, the distinction between contravariance and covariance disappears.

Exercises

2.7.1 If $T_{\ldots i}$ is a tensor of rank n, show that $\partial T_{\ldots i}/\partial x^j$ is a tensor of rank $n+1$ (Cartesian coordinates).
Note. In non-Cartesian coordinate systems the coefficients a_{ij} are, in general, functions of the coordinates, and the simple derivative of a tensor of rank n is not a tensor except in the special case of $n=0$. In this case the derivative does yield a covariant vector (tensor of rank 1) by Eq. (2.64).

2.7.2 If $T_{ijk\ldots}$ is a tensor of rank n, show that $\sum_j \partial T_{ijk\ldots}/\partial x^j$ is a tensor of rank $n-1$ (Cartesian coordinates).

2.7.3 The operator

$$\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}$$

may be written as

$$\sum_{i=1}^{4} \frac{\partial^2}{\partial x_i^2},$$

using $x_4 = ict$. This is the four-dimensional Laplacian, sometimes called the d'Alembertian and denoted by $\Box^2$. Show that it is a **scalar** operator, that is, is invariant under Lorentz transformations.

2.8 QUOTIENT RULE

If A_i and B_j are vectors, as seen in Section 2.7, we can easily show that $A_i B_j$ is a second-rank tensor. Here we are concerned with a variety of inverse relations. Consider such equations as

$$K_i A_i = B \tag{2.82a}$$

$$K_{ij} A_j = B_i \tag{2.82b}$$

$$K_{ij} A_{jk} = B_{ik} \tag{2.82c}$$

$$K_{ijkl} A_{ij} = B_{kl} \tag{2.82d}$$

$$K_{ij} A_k = B_{ijk}. \tag{2.82e}$$

Inline with our restriction to Cartesian systems, we write all indices as subscripts and, unless specified otherwise, sum repeated indices.

In each of these expressions **A** and **B** are known tensors of rank indicated by the number of indices and **A** is arbitrary. In each case K is an unknown quantity. We wish to establish the transformation properties of K. The quotient rule asserts that if the equation of interest holds in all (rotated) Cartesian coordinate systems, K is a tensor of the indicated rank. The importance in physical theory is that the quotient rule can establish the tensor nature of quantities. Exercise 2.8.1 is a simple illustration of this. The quotient rule (Eq. (2.82b)) shows that the inertia matrix appearing in the angular momentum equation $\mathbf{L} = I\boldsymbol{\omega}$, Section 3.5, is a tensor.

In proving the quotient rule, we consider Eq. (2.82b) as a typical case. In our primed coordinate system

$$K'_{ij} A'_j = B'_i = a_{ik} B_k, \tag{2.83}$$

using the vector transformation properties of **B**. Since the equation holds in all rotated Cartesian coordinate systems,

$$a_{ik} B_k = a_{ik} (K_{kl} A_l). \tag{2.84}$$

Now, transforming **A** back into the primed coordinate system[11] (compare Eq. (2.62)), we have

$$K'_{ij} A'_j = a_{ik} K_{kl} a_{jl} A'_j. \tag{2.85}$$

Rearranging, we obtain

$$(K'_{ij} - a_{ik} a_{jl} K_{kl}) A'_j = 0. \tag{2.86}$$

[11] Note the order of the indices of the direction cosine a_{jl} in this **inverse** transformation. We have

$$A_l = \sum_j \frac{\partial x_l}{\partial x'_j} A'_j = \sum_j a_{jl} A'_j.$$

This must hold for each value of the index i and for every primed coordinate system. Since the A'_j is arbitrary,[12] we conclude

$$K'_{ij} = a_{ik}a_{jl}K_{kl}, \tag{2.87}$$

which is our definition of second-rank tensor.

The other equations may be treated similarly, giving rise to other forms of the quotient rule. One minor pitfall should be noted: The quotient rule does not necessarily apply if B is zero. The transformation properties of zero are indeterminate.

Example 2.8.1 EQUATIONS OF MOTION AND FIELD EQUATIONS

In classical mechanics, Newton's equations of motion $m\dot{\mathbf{v}} = \mathbf{F}$ tell us on the basis of the quotient rule that, if the mass is a scalar and the force a vector, then the acceleration $\mathbf{a} \equiv \dot{\mathbf{v}}$ is a vector. In other words, the vector character of the force as the driving term imposes its vector character on the acceleration, provided the scale factor m is scalar.

The wave equation of electrodynamics $\partial^2 A^\mu = J^\mu$ involves the four-dimensional version of the Laplacian $\partial^2 = \frac{\partial^2}{c^2\partial t^2} - \nabla^2$, a Lorentz scalar, and the external four-vector current J^μ as its driving term. From the quotient rule, we infer that the vector potential A^μ is a four-vector as well. If the driving current is a four-vector, the vector potential must be of rank 1 by the quotient rule. ■

The quotient rule is a substitute for the illegal division of tensors.

Exercises

2.8.1 The double summation $K_{ij}A_iB_j$ is invariant for any two vectors A_i and B_j. Prove that K_{ij} is a second-rank tensor.
Note. In the form ds^2 (invariant) $= g_{ij}\,dx^i\,dx^j$, this result shows that the matrix g_{ij} is a tensor.

2.8.2 The equation $K_{ij}A_{jk} = B_{ik}$ holds for all orientations of the coordinate system. If $\mathbf{A}$ and $\mathbf{B}$ are arbitrary second-rank tensors, show that $\mathbf{K}$ is a second-rank tensor also.

2.8.3 The exponential in a plane wave is $\exp[i(\mathbf{k}\cdot\mathbf{r} - \omega t)]$. We recognize $x^\mu = (ct, x_1, x_2, x_3)$ as a prototype vector in Minkowski space. If $\mathbf{k}\cdot\mathbf{r} - \omega t$ is a scalar under Lorentz transformations (Section 4.5), show that $k^\mu = (\omega/c, k_1, k_2, k_3)$ is a vector in Minkowski space.
Note. Multiplication by $\hbar$ yields $(E/c, \mathbf{p})$ as a vector in Minkowski space.

2.9 PSEUDOTENSORS, DUAL TENSORS

So far our coordinate transformations have been restricted to pure passive rotations. We now consider the effect of reflections or inversions.

[12] We might, for instance, take $A'_1 = 1$ and $A'_m = 0$ for $m \neq 1$. Then the equation $K'_{i1} = a_{ik}a_{1l}K_{kl}$ follows immediately. The rest of Eq. (2.87) comes from other special choices of the arbitrary A'_j.

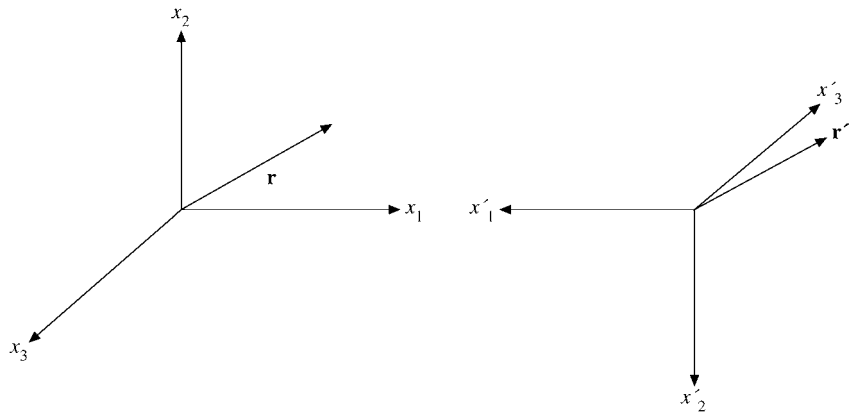

FIGURE 2.9 Inversion of Cartesian coordinates — polar vector.

If we have transformation coefficients $a_{ij} = -\delta_{ij}$, then by Eq. (2.60)

$$x^i = -x'^i, \tag{2.88}$$

which is an inversion or parity transformation. Note that this transformation changes our initial right-handed coordinate system into a left-handed coordinate system.[13] Our prototype vector $\mathbf{r}$ with components (x^1, x^2, x^3) transforms to

$$\mathbf{r}' = \left(x'^1, x'^2, x'^3\right) = \left(-x^1, -x^2, -x^3\right).$$

This new vector $\mathbf{r}'$ has negative components, relative to the new transformed set of axes. As shown in Fig. 2.9, reversing the directions of the coordinate axes and changing the signs of the components gives $\mathbf{r}' = \mathbf{r}$. The vector (an arrow in space) stays exactly as it was before the transformation was carried out. The position vector $\mathbf{r}$ and all other vectors whose components behave this way (reversing sign with a reversal of the coordinate axes) are called **polar vectors** and have odd parity.

A fundamental difference appears when we encounter a vector defined as the cross product of two polar vectors. Let $\mathbf{C} = \mathbf{A} \times \mathbf{B}$, where both $\mathbf{A}$ and $\mathbf{B}$ are polar vectors. From Eq. (1.33), the components of $\mathbf{C}$ are given by

$$C^1 = A^2 B^3 - A^3 B^2 \tag{2.89}$$

and so on. Now, when the coordinate axes are inverted, $A^i \to -A'^i$, $B_j \to -B'_j$, but from its definition $C^k \to +C'^k$; that is, our cross-product vector, vector $\mathbf{C}$, does **not** behave like a polar vector under inversion. To distinguish, we label it a pseudovector or axial vector (see Fig. 2.10) that has even parity. The term **axial vector** is frequently used because these cross products often arise from a description of rotation.

[13] This is an inversion of the coordinate system or coordinate axes, objects in the physical world remaining fixed.

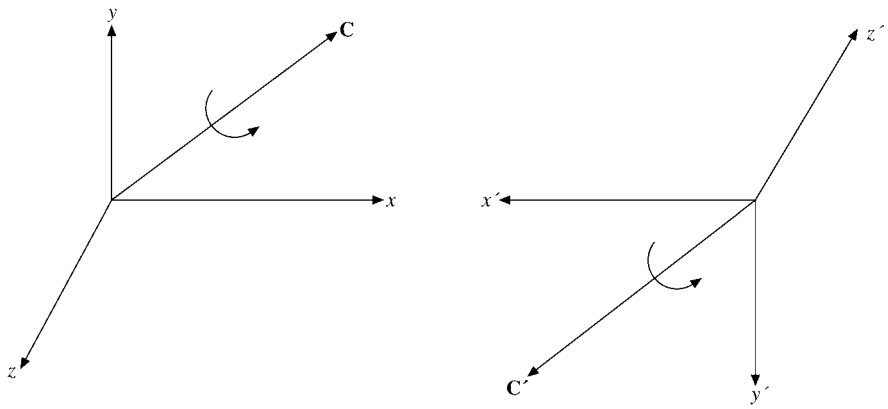

FIGURE 2.10 Inversion of Cartesian coordinates — axial vector.

Examples are

angular velocity,	$\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$,
orbital angular momentum,	$\mathbf{L} = \mathbf{r} \times \mathbf{p}$,
torque, force $= \mathbf{F}$,	$\mathbf{N} = \mathbf{r} \times \mathbf{F}$,
magnetic induction field $\mathbf{B}$,	$\dfrac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E}$.

In $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$, the axial vector is the angular velocity $\boldsymbol{\omega}$, and $\mathbf{r}$ and $\mathbf{v} = d\mathbf{r}/dt$ are polar vectors. Clearly, axial vectors occur frequently in physics, although this fact is usually not pointed out. In a right-handed coordinate system an axial vector $\mathbf{C}$ has a sense of rotation associated with it given by a right-hand rule (compare Section 1.4). In the inverted left-handed system the sense of rotation is a left-handed rotation. This is indicated by the curved arrows in Fig. 2.10.

The distinction between polar and axial vectors may also be illustrated by a reflection. A polar vector reflects in a mirror like a real physical arrow, Fig. 2.11a. In Figs. 2.9 and 2.10 the coordinates are inverted; the physical world remains fixed. Here the coordinate axes remain fixed; the world is reflected — as in a mirror in the xz-plane. Specifically, in this representation we keep the axes fixed and associate a change of sign with the component of the vector. For a mirror in the xz-plane, $P_y \to -P_y$. We have

$$\mathbf{P} = (P_x, P_y, P_z)$$
$$\mathbf{P}' = (P_x, -P_y, P_z) \qquad \text{polar vector.}$$

An axial vector such as a magnetic field $\mathbf{H}$ or a magnetic moment $\boldsymbol{\mu}$ (= current × area of current loop) behaves quite differently under reflection. Consider the magnetic field $\mathbf{H}$ and magnetic moment $\boldsymbol{\mu}$ to be produced by an electric charge moving in a circular path (Exercise 5.8.4 and Example 12.5.3). Reflection reverses the sense of rotation of the charge.

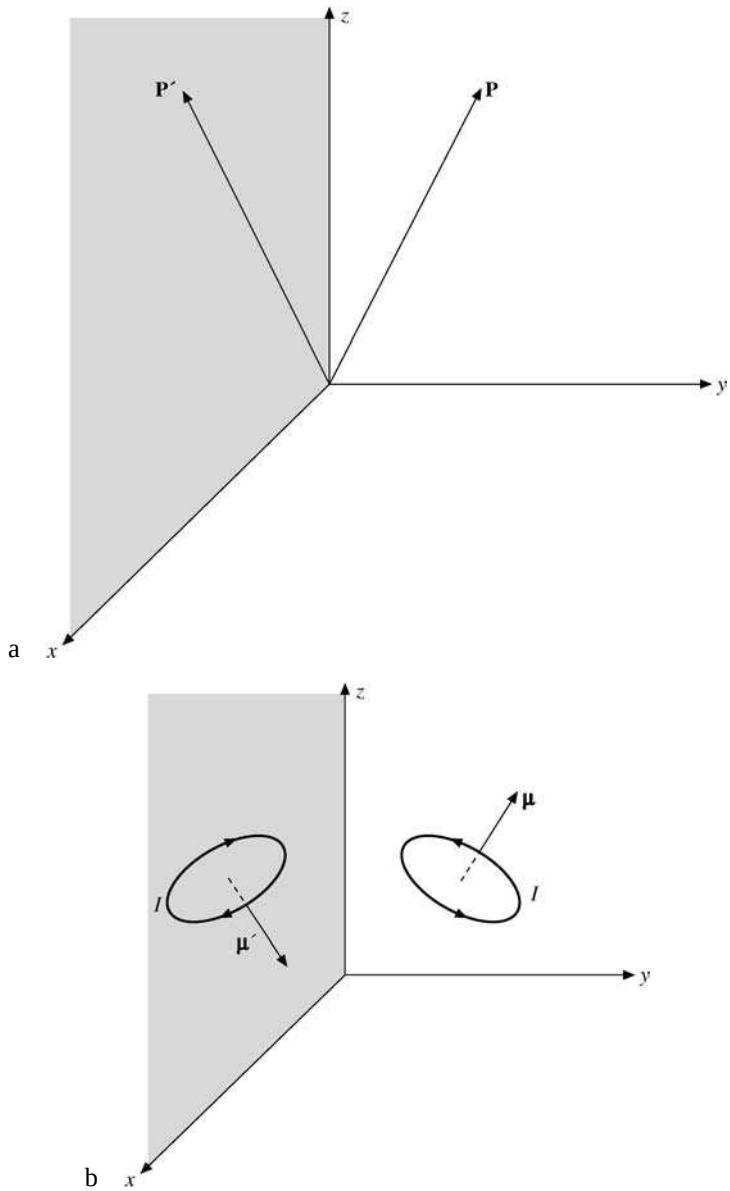

FIGURE 2.11 (a) Mirror in xz-plane; (b) mirror in xz-plane.

The two current loops and the resulting magnetic moments are shown in Fig. 2.11b. We have

$$\boldsymbol{\mu} = (\mu_x, \mu_y, \mu_z)$$

$$\boldsymbol{\mu}' = (-\mu_x, \mu_y, -\mu_z) \quad \text{reflected axial vector.}$$

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1, \tag{2.92}$$

and again the Jacobian $J = -1$. On the other hand, for all pure rotations, the Jacobian J is always $+1$. Rotation matrices discussed further in Section 3.3.

In Chapter 1 the triple scalar product $S = \mathbf{A} \times \mathbf{B} \cdot \mathbf{C}$ was shown to be a scalar (under rotations). Now by considering the parity transformation given by Eq. (2.88), we see that S is actually a pseudoscalar: This behavior was foreshadowed by the geometrical analogy of a volume. If all three parameters of the volume—length, depth, and height—change from positive distances to negative distances, the product of the three will be negative.

Usually, pseudoscalars, pseudovectors, and pseudotensors will transform as

$$S' = |J| S, \tag{2.90}$$

where J is the determinant[15] of the array of coefficients a_{mn}, the Jacobian of the parity transformation. In our inversion the Jacobian is

$$J = \begin{vmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{vmatrix}$$

Levi-Civita Symbol

For future use it is convenient to introduce the three-dimensional Levi-Civita symbol ε_{ijk}, defined by

$$\begin{aligned} \varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} &= 1, \\ \varepsilon_{132} = \varepsilon_{213} = \varepsilon_{321} &= -1, \\ \text{all other } \varepsilon_{ijk} &= 0. \end{aligned} \tag{2.93}$$

Note that ε_{ijk} is antisymmetric with respect to all pairs of indices. Suppose now that we have a third-rank pseudotensor δ_{ijk}, which in one particular coordinate system is equal to ε_{ijk}. Then

$$\delta'_{ijk} = |a| a_{ip} a_{jq} a_{kr} \varepsilon_{pqr} \tag{2.94}$$

[14] The big exception to this is in beta decay, weak interactions. Here the universe distinguishes between right- and left-handed systems, and we add polar and axial vector interactions.

[15] Determinants are described in Section 3.1.

by definition of pseudotensor. Now,

$$a_{1p}a_{2q}a_{3r}\varepsilon_{pqr} = |a| \tag{2.95}$$

by direct expansion of the determinant, showing that $\delta'_{123} = |a|^2 = 1 = \varepsilon_{123}$. Considering the other possibilities one by one, we find

$$\delta'_{ijk} = \varepsilon_{ijk} \tag{2.96}$$

for rotations and reflections. Hence ε_{ijk} is a pseudotensor.[16,17] Furthermore, it is seen to be an isotropic pseudotensor with the same components in all rotated Cartesian coordinate systems.

Dual Tensors

With any **antisymmetric** second-rank tensor $\mathbf{C}$ (in three-dimensional space) we may associate a dual pseudovector C_i defined by

$$C_i = \frac{1}{2}\varepsilon_{ijk}C^{jk}. \tag{2.97}$$

Here the antisymmetric $\mathbf{C}$ may be written

$$\mathbf{C} = \begin{pmatrix} 0 & C^{12} & -C^{31} \\ -C^{12} & 0 & C^{23} \\ C^{31} & -C^{23} & 0 \end{pmatrix}. \tag{2.98}$$

We know that C_i must transform as a vector under rotations from the double contraction of the fifth-rank (pseudo) tensor $\varepsilon_{ijk}C_{mn}$ but that it is really a pseudovector from the pseudo nature of ε_{ijk}. Specifically, the components of $\mathbf{C}$ are given by

$$(C_1, C_2, C_3) = \left(C^{23}, C^{31}, C^{12}\right). \tag{2.99}$$

Notice the cyclic order of the indices that comes from the cyclic order of the components of ε_{ijk}. Eq. (2.99) means that our three-dimensional vector product may literally be taken to be either a pseudovector or an antisymmetric second-rank tensor, depending on how we choose to write it out.

If we take three (polar) vectors **A**, **B**, and **C**, we may define the direct product

$$V^{ijk} = A^i B^j C^k. \tag{2.100}$$

By an extension of the analysis of Section 2.6, V^{ijk} is a tensor of third rank. The dual quantity

$$V = \frac{1}{3!}\varepsilon_{ijk}V^{ijk} \tag{2.101}$$

[16] The usefulness of ε_{pqr} extends far beyond this section. For instance, the matrices M_k of Exercise 3.2.16 are derived from $(M_r)_{pq} = -i\varepsilon_{pqr}$. Much of elementary vector analysis can be written in a very compact form by using ε_{ijk} and the identity of Exercise 2.9.4 See A. A. Evett, Permutation symbol approach to elementary vector analysis. *Am. J. Phys.* **34**: 503 (1966).

[17] The numerical value of ε_{pqr} is given by the triple scalar product of coordinate unit vectors:

$$\hat{\mathbf{x}}_p \cdot \hat{\mathbf{x}}_q \times \hat{\mathbf{x}}_r.$$

From this point of view each element of ε_{pqr} is a pseudoscalar, but the ε_{pqr} collectively form a third-rank pseudotensor.

We introduce the Levi-Civita symbol ε_{ijkl}, the four-dimensional analog of ε_{ijk}. This quantity ε_{ijkl} is defined as totally antisymmetric in all four indices. If $(ijkl)$ is an even permutation[18] of (0, 1, 2, 3), then ε_{ijkl} is defined as $+1$; if it is an odd permutation, then ε_{ijkl} is -1, and 0 if any two indices are equal. The Levi-Civita ε_{ijkl} may be proved a pseudotensor of rank 4 by analysis similar to that used for establishing the tensor nature of ε_{ijk}. Introducing the direct product of four vectors as fourth-rank tensor with components

$$H^{ijkl} = A^i B^j C^k D^l, \tag{2.103}$$

is clearly a pseudoscalar. By expansion it is seen that

$$V = \begin{vmatrix} A^1 & B^1 & C^1 \\ A^2 & B^2 & C^2 \\ A^3 & B^3 & C^3 \end{vmatrix}$$

built from the polar vectors **A**, **B**, **C**, and **D**, we may define the dual quantity

$$H = \frac{1}{4!}\varepsilon_{ijkl}H^{ijkl}, \tag{2.104}$$

a pseudoscalar due to the quadruple contraction with the pseudotensor ε_{ijkl}. Now we let **A**, **B**, **C**, and **D** be infinitesimal displacements along the four coordinate axes (Minkowski space),

$$\begin{aligned} \mathbf{A} &= \left(dx^0, 0, 0, 0\right) \\ \mathbf{B} &= \left(0, dx^1, 0, 0\right), \qquad \text{and so on,} \end{aligned} \tag{2.105}$$

and

$$H = dx^0\, dx^1\, dx^2\, dx^3. \tag{2.106}$$

The four-dimensional volume element is now identified as a pseudoscalar. We use this result in Section 4.6. This result could have been expected from the results of the special theory of relativity. The Lorentz–Fitzgerald contraction of $dx^1\, dx^2\, dx^3$ just balances the time dilation of dx^0.

We slipped into this four-dimensional space as a simple mathematical extension of the three-dimensional space and, indeed, we could just as easily have discussed 5-, 6-, or N-dimensional space. This is typical of the power of the component analysis. Physically, this four-dimensional space may be taken as Minkowski space,

$$\left(x^0, x^1, x^2, x^3\right) = (ct, x, y, z), \tag{2.107}$$

where t is time. This is the merger of space and time achieved in special relativity. The transformations that describe the rotations in four-dimensional space are the Lorentz transformations of special relativity. We encounter these Lorentz transformations in Section 4.6.

[18] A permutation is odd if it involves an odd number of interchanges of adjacent indices, such as (0 1 2 3) → (0 2 1 3). Even permutations arise from an even number of transpositions of adjacent indices. (Actually the word *adjacent* is unnecessary.) $\varepsilon_{0123} = +1$.

Irreducible Tensors

For some applications, particularly in the quantum theory of angular momentum, our Cartesian tensors are not particularly convenient. In mathematical language our general second-rank tensor A_{ij} is reducible, which means that it can be decomposed into parts of lower tensor rank. In fact, we have already done this. From Eq. (2.78),

$$A = A^i{}_i \tag{2.108}$$

is a scalar quantity, the trace of A_{ij}.[19]

The antisymmetric portion,

$$B_{ij} = \tfrac{1}{2}(A_{ij} - A_{ji}), \tag{2.109}$$

has just been shown to be equivalent to a (pseudo) vector, or

$$B_{ij} = C_k \qquad \text{cyclic permutation of } i, j, k. \tag{2.110}$$

By subtracting the scalar A and the vector C_k from our original tensor, we have an irreducible, symmetric, zero-trace second-rank tensor, S_{ij}, in which

$$S_{ij} = \tfrac{1}{2}(A_{ij} + A_{ji}) - \tfrac{1}{3}A\delta_{ij}, \tag{2.111}$$

with five independent components. Then, finally, our original Cartesian tensor may be written

$$A_{ij} = \tfrac{1}{3}A\delta_{ij} + C_k + S_{ij}. \tag{2.112}$$

The three quantities A, C_k, and S_{ij} form spherical tensors of rank 0, 1, and 2, respectively, transforming like the spherical harmonics Y_L^M (Chapter 12) for $L = 0$, 1, and 2. Further details of such spherical tensors and their uses will be found in Chapter 4 and the books by Rose and Edmonds cited there.

A specific example of the preceding reduction is furnished by the symmetric electric quadrupole tensor

$$Q_{ij} = \int \left(3x_i x_j - r^2\delta_{ij}\right)\rho(x_1, x_2, x_3)\, d^3x.$$

The $-r^2\delta_{ij}$ term represents a subtraction of the scalar trace (the three $i = j$ terms). The resulting Q_{ij} has zero trace.

Exercises

2.9.1 An antisymmetric square array is given by

$$\begin{pmatrix} 0 & C_3 & -C_2 \\ -C_3 & 0 & C_1 \\ C_2 & -C_1 & 0 \end{pmatrix}$$

[19] An alternate approach, using matrices, is given in Section 3.3 (see Exercise 3.3.9).

where (C_1, C_2, C_3) form a pseudovector. Assuming that the relation

$$C_i = \frac{1}{2!}\varepsilon_{ijk}C^{jk}$$

holds in all coordinate systems, prove that C^{jk} is a tensor. (This is another form of the quotient theorem.)

2.9.2 Show that the vector product is unique to three-dimensional space; that is, only in three dimensions can we establish a one-to-one correspondence between the components of an antisymmetric tensor (second-rank) and the components of a vector.

2.9.3 Show that in $\mathbb{R}^3$

(a) $\delta_{ii} = 3$,
(b) $\delta_{ij}\varepsilon_{ijk} = 0$,
(c) $\varepsilon_{ipq}\varepsilon_{jpq} = 2\delta_{ij}$,
(d) $\varepsilon_{ijk}\varepsilon_{ijk} = 6$.

2.9.4 Show that in $\mathbb{R}^3$

$$\varepsilon_{ijk}\varepsilon_{pqk} = \delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}.$$

2.9.5 (a) Express the components of a cross-product vector $\mathbf{C}$, $\mathbf{C} = \mathbf{A} \times \mathbf{B}$, in terms of ε_{ijk} and the components of $\mathbf{A}$ and $\mathbf{B}$.
(b) Use the antisymmetry of ε_{ijk} to show that $\mathbf{A} \cdot \mathbf{A} \times \mathbf{B} = 0$.

ANS. (a) $C_i = \varepsilon_{ijk}A_jB_k$.

2.9.6 (a) Show that the inertia tensor (matrix) may be written

$$I_{ij} = m(x_i x_j \delta_{ij} - x_i x_j)$$

for a particle of mass m at (x_1, x_2, x_3).
(b) Show that

$$I_{ij} = -M_{il}M_{lj} = -m\varepsilon_{ilk}x_k\varepsilon_{ljm}x_m,$$

where $M_{il} = m^{1/2}\varepsilon_{ilk}x_k$. This is the contraction of two second-rank tensors and is identical with the matrix product of Section 3.2.

2.9.7 Write $\nabla \cdot \nabla \times \mathbf{A}$ and $\nabla \times \nabla\varphi$ in tensor (index) notation in $\mathbb{R}^3$ so that it becomes obvious that each expression vanishes.

ANS. $\nabla \cdot \nabla \times \mathbf{A} = \varepsilon_{ijk}\dfrac{\partial}{\partial x^i}\dfrac{\partial}{\partial x^j}A^k$,

$(\nabla \times \nabla\varphi)_i = \varepsilon_{ijk}\dfrac{\partial}{\partial x^j}\dfrac{\partial}{\partial x^k}\varphi$.

2.9.8 Expressing cross products in terms of Levi-Civita symbols (ε_{ijk}), derive the *BAC–CAB* rule, Eq. (1.55).
Hint. The relation of Exercise 2.9.4 is helpful.

2.9.9 Verify that each of the following fourth-rank tensors is isotropic, that is, that it has the same form independent of any rotation of the coordinate systems.

(a) $A_{ijkl} = \delta_{ij}\delta_{kl}$,
(b) $B_{ijkl} = \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}$,
(c) $C_{ijkl} = \delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}$.

2.9.10 Show that the two-index Levi-Civita symbol ε_{ij} is a second-rank pseudotensor (in two-dimensional space). Does this contradict the uniqueness of δ_{ij} (Exercise 2.6.4)?

2.9.11 Represent ε_{ij} by a 2×2 matrix, and using the 2×2 rotation matrix of Section 3.3 show that ε_{ij} is invariant under orthogonal similarity transformations.

2.9.12 Given $A_k = \frac{1}{2}\varepsilon_{ijk}B^{ij}$ with $B^{ij} = -B^{ji}$, antisymmetric, show that

$$B^{mn} = \varepsilon^{mnk}A_k.$$

2.9.13 Show that the vector identity

$$(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) = (\mathbf{A} \cdot \mathbf{C})(\mathbf{B} \cdot \mathbf{D}) - (\mathbf{A} \cdot \mathbf{D})(\mathbf{B} \cdot \mathbf{C})$$

(Exercise 1.5.12) follows directly from the description of a cross product with ε_{ijk} and the identity of Exercise 2.9.4.

2.9.14 Generalize the cross product of two vectors to n-dimensional space for $n = 4, 5, \ldots$. Check the consistency of your construction and discuss concrete examples. See Exercise 1.4.17 for the case $n = 2$.

2.10 General Tensors

The distinction between contravariant and covariant transformations was established in Section 2.6. Then, for convenience, we restricted our attention to Cartesian coordinates (in which the distinction disappears). Now in these two concluding sections we return to non-Cartesian coordinates and resurrect the contravariant and covariant dependence. As in Section 2.6, a superscript will be used for an index denoting contravariant and a subscript for an index denoting covariant dependence. The metric tensor of Section 2.1 will be used to relate contravariant and covariant indices.

The emphasis in this section is on differentiation, culminating in the construction of the **covariant derivative**. We saw in Section 2.7 that the derivative of a vector yields a second-rank tensor — in Cartesian coordinates. In non-Cartesian coordinate systems, it is the covariant derivative of a vector rather than the ordinary derivative that yields a second-rank tensor by differentiation of a vector.

Metric Tensor

Let us start with the transformation of vectors from one set of coordinates (q^1, q^2, q^3) to another $\mathbf{r} = (x^1, x^2, x^3)$. The new coordinates are (in general **nonlinear**) functions

$x^i(q^1, q^2, q^3)$ of the old, such as spherical polar coordinates (r, θ, ϕ). But their **differentials** obey the **linear** transformation law

$$dx^i = \frac{\partial x^i}{\partial q^j} dq^j, \tag{2.113a}$$

or

$$d\mathbf{r} = \boldsymbol{\varepsilon}_j dq^j \tag{2.113b}$$

in vector notation. For convenience we take the basis vectors $\boldsymbol{\varepsilon}_1 = (\frac{\partial x^1}{\partial q^1}, \frac{\partial x^1}{\partial q^2}, \frac{\partial x^1}{\partial q^3})$, $\boldsymbol{\varepsilon}_2$, and $\boldsymbol{\varepsilon}_3$ to form a right-handed set. These vectors are not necessarily orthogonal. Also, a limitation to three-dimensional space will be required only for the discussions of cross products and curls. Otherwise these $\boldsymbol{\varepsilon}_i$ may be in N-dimensional space, including the four-dimensional space–time of special and general relativity. The basis vectors $\boldsymbol{\varepsilon}_i$ may be expressed by

$$\boldsymbol{\varepsilon}_i = \frac{\partial \mathbf{r}}{\partial q^i}, \tag{2.114}$$

as in Exercise 2.2.3. Note, however, that the $\boldsymbol{\varepsilon}_i$ here do **not** necessarily have unit magnitude. From Exercise 2.2.3, the unit vectors are

$$\mathbf{e}_i = \frac{1}{h_i} \frac{\partial \mathbf{r}}{\partial q_i} \qquad \text{(no summation)},$$

and therefore

$$\boldsymbol{\varepsilon}_i = h_i \mathbf{e}_i \qquad \text{(no summation)}. \tag{2.115}$$

The $\boldsymbol{\varepsilon}_i$ are related to the unit vectors $\mathbf{e}_i$ by the scale factors h_i of Section 2.2. The $\mathbf{e}_i$ have no dimensions; the $\boldsymbol{\varepsilon}_i$ have the dimensions of h_i. In spherical polar coordinates, as a specific example,

$$\boldsymbol{\varepsilon}_r = \mathbf{e}_r = \hat{\mathbf{r}}, \qquad \boldsymbol{\varepsilon}_\theta = r\mathbf{e}_\theta = r\hat{\boldsymbol{\theta}}, \qquad \boldsymbol{\varepsilon}_\varphi = r\sin\theta \mathbf{e}_\varphi = r\sin\theta \hat{\boldsymbol{\varphi}}. \tag{2.116}$$

In Euclidean spaces, or in Minkowski space of special relativity, the partial derivatives in Eq. (2.113) are constants that define the new coordinates in terms of the old ones. We used them to define the transformation laws of vectors in Eq. (2.59) and (2.62) and tensors in Eq. (2.66). Generalizing, we define a **contravariant** vector V^i under **general** coordinate transformations if its components transform according to

$$V'^i = \frac{\partial x^i}{\partial q^j} V^j, \tag{2.117a}$$

or

$$\mathbf{V}' = V^j \boldsymbol{\varepsilon}_j \tag{2.117b}$$

in vector notation. For **covariant** vectors we inspect the transformation of the gradient operator

$$\frac{\partial}{\partial x^i} = \frac{\partial q^j}{\partial x^i} \frac{\partial}{\partial q^j} \tag{2.118}$$

using the chain rule. From

$$\frac{\partial x^i}{\partial q^j}\frac{\partial q^j}{\partial x^k} = \delta^i{}_k \tag{2.119}$$

it is clear that Eq. (2.118) is related to the **inverse** transformation of Eq. (2.113),

$$dq^j = \frac{\partial q^j}{\partial x^i} dx^i. \tag{2.120}$$

Hence we define a **covariant** vector V_i if

$$V_i' = \frac{\partial q^j}{\partial x^i} V_j \tag{2.121a}$$

holds or, in vector notation,

$$\mathbf{V}' = V_j \boldsymbol{\varepsilon}^j, \tag{2.121b}$$

where $\boldsymbol{\varepsilon}^j$ are the contravariant vectors $g^{ji}\boldsymbol{\varepsilon}_i = \boldsymbol{\varepsilon}^j$.

Second-rank tensors are defined as in Eq. (2.66),

$$A'^{ij} = \frac{\partial x^i}{\partial q^k}\frac{\partial x^j}{\partial q^l} A^{kl}, \tag{2.122}$$

and tensors of higher rank similarly.

As in Section 2.1, we construct the square of a differential displacement

$$(ds)^2 = d\mathbf{r}\cdot d\mathbf{r} = \left(\boldsymbol{\varepsilon}_i\, dq^i\right)^2 = \boldsymbol{\varepsilon}_i \cdot \boldsymbol{\varepsilon}_j\, dq^i\, dq^j. \tag{2.123}$$

Comparing this with $(ds)^2$ of Section 2.1, Eq. (2.5), we identify $\boldsymbol{\varepsilon}_i \cdot \boldsymbol{\varepsilon}_j$ as the covariant metric tensor

$$\boldsymbol{\varepsilon}_i \cdot \boldsymbol{\varepsilon}_j = g_{ij}. \tag{2.124}$$

Clearly, g_{ij} is symmetric. The tensor nature of g_{ij} follows from the quotient rule, Exercise 2.8.1. We take the relation

$$g^{ik} g_{kj} = \delta^i{}_j \tag{2.125}$$

to define the corresponding contravariant tensor g^{ik}. Contravariant g^{ik} enters as the inverse[20] of covariant g_{kj}. We use this contravariant g^{ik} to raise indices, converting a covariant index into a contravariant index, as shown subsequently. Likewise the covariant g_{kj} will be used to lower indices. The choice of g^{ik} and g_{kj} for this raising–lowering operation is arbitrary. Any second-rank tensor (and its inverse) would do. Specifically, we have

$$\begin{aligned} g^{ij}\boldsymbol{\varepsilon}_j &= \boldsymbol{\varepsilon}^i && \text{relating covariant and contravariant basis vectors,} \\ g^{ij}F_j &= F^i && \text{relating covariant and contravariant vector components.} \end{aligned} \tag{2.126}$$

[20] If the tensor g_{kj} is written as a matrix, the tensor g^{ik} is given by the inverse matrix.

Then

$$\begin{aligned} g_{ij}\boldsymbol{\varepsilon}^j &= \boldsymbol{\varepsilon}_i \qquad \text{as the corresponding index} \\ g_{ij}F^j &= F_i \qquad \text{lowering relations.} \end{aligned} \tag{2.127}$$

It should be emphasized again that the $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\varepsilon}^j$ do **not** have unit magnitude. This may be seen in Eqs. (2.116) and in the metric tensor g_{ij} for spherical polar coordinates and its inverse g^{ij}:

$$(g_{ij}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin^2\theta \end{pmatrix} \qquad \left(g^{ij}\right) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \dfrac{1}{r^2} & 0 \\ 0 & 0 & \dfrac{1}{r^2\sin^2\theta} \end{pmatrix}.$$

Christoffel Symbols

Let us form the differential of a scalar ψ,

$$d\psi = \frac{\partial\psi}{\partial q^i}\,dq^i. \tag{2.128}$$

Since the dq^i are the components of a contravariant vector, the partial derivatives $\partial\psi/\partial q^i$ must form a covariant vector — by the quotient rule. The gradient of a scalar becomes

$$\nabla\psi = \frac{\partial\psi}{\partial q^i}\boldsymbol{\varepsilon}^i. \tag{2.129}$$

Note that $\partial\psi/\partial q^i$ are not the gradient components of Section 2.2 — because $\boldsymbol{\varepsilon}^i \neq \mathbf{e}_i$ of Section 2.2.

Moving on to the derivatives of a vector, we find that the situation is much more complicated because the basis vectors $\boldsymbol{\varepsilon}_i$ are in general not constant. Remember, we are no longer restricting ourselves to Cartesian coordinates and the nice, convenient $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$! Direct differentiation of Eq. (2.117a) yields

$$\frac{\partial V'^k}{\partial q^j} = \frac{\partial x^k}{\partial q^i}\frac{\partial V^i}{\partial q^j} + \frac{\partial^2 x^k}{\partial q^j\partial q^i}V^i, \tag{2.130a}$$

or, in vector notation,

$$\frac{\partial\mathbf{V}'}{\partial q^j} = \frac{\partial V^i}{\partial q^j}\boldsymbol{\varepsilon}_i + V^i\frac{\partial\boldsymbol{\varepsilon}_i}{\partial q^j}. \tag{2.130b}$$

The right side of Eq. (2.130a) differs from the transformation law for a second-rank mixed tensor by the second term, which contains second derivatives of the coordinates x^k. The latter are nonzero for nonlinear coordinate transformations.

Now, $\partial\boldsymbol{\varepsilon}_i/\partial q^j$ will be some linear combination of the $\boldsymbol{\varepsilon}_k$, with the coefficient depending on the indices i and j from the partial derivative and index k from the base vector. We write

$$\frac{\partial\boldsymbol{\varepsilon}_i}{\partial q^j} = \Gamma^k_{ij}\boldsymbol{\varepsilon}_k. \tag{2.131a}$$

Multiplying by $\boldsymbol{\varepsilon}^m$ and using $\boldsymbol{\varepsilon}^m \cdot \boldsymbol{\varepsilon}_k = \delta_k^m$ from Exercise 2.10.2, we have

$$\Gamma_{ij}^m = \boldsymbol{\varepsilon}^m \cdot \frac{\partial \boldsymbol{\varepsilon}_i}{\partial q^j}. \tag{2.131b}$$

The Γ_{ij}^k is a Christoffel symbol of the **second kind**. It is also called a *coefficient of connection*. These Γ_{ij}^k are **not** third-rank tensors and the $\partial V^i/\partial q^j$ of Eq. (2.130a) are not second-rank tensors. Equations (2.131) should be compared with the results quoted in Exercise 2.2.3 (remembering that in general $\boldsymbol{\varepsilon}_i \neq \mathbf{e}_i$). In Cartesian coordinates, $\Gamma_{ij}^k = 0$ for all values of the indices i, j, and k. These Christoffel three-index symbols may be computed by the techniques of Section 2.2. This is the topic of Exercise 2.10.8. Equation (2.138) offers an easier method. Using Eq. (2.114), we obtain

$$\frac{\partial \boldsymbol{\varepsilon}_i}{\partial q^j} = \frac{\partial^2 \mathbf{r}}{\partial q^j \, \partial q^i} = \frac{\partial \boldsymbol{\varepsilon}_j}{\partial q^i} = \Gamma_{ji}^k \boldsymbol{\varepsilon}_k. \tag{2.132}$$

Hence these Christoffel symbols are symmetric in the two lower indices:

$$\Gamma_{ij}^k = \Gamma_{ji}^k. \tag{2.133}$$

Christoffel Symbols as Derivatives of the Metric Tensor

It is often convenient to have an explicit expression for the Christoffel symbols in terms of derivatives of the metric tensor. As an initial step, we define the Christoffel symbol of the **first kind** $[ij, k]$ by

$$[ij, k] \equiv g_{mk} \Gamma_{ij}^m, \tag{2.134}$$

from which the symmetry $[ij, k] = [ji, k]$ follows. Again, this $[ij, k]$ is not a third-rank tensor. From Eq. (2.131b),

$$\begin{aligned} [ij, k] &= g_{mk} \boldsymbol{\varepsilon}^m \cdot \frac{\partial \boldsymbol{\varepsilon}_i}{\partial q^j} \\ &= \boldsymbol{\varepsilon}_k \cdot \frac{\partial \boldsymbol{\varepsilon}_i}{\partial q^j}. \end{aligned} \tag{2.135}$$

Now we differentiate $g_{ij} = \boldsymbol{\varepsilon}_i \cdot \boldsymbol{\varepsilon}_j$, Eq. (2.124):

$$\begin{aligned} \frac{\partial g_{ij}}{\partial q^k} &= \frac{\partial \boldsymbol{\varepsilon}_i}{\partial q^k} \cdot \boldsymbol{\varepsilon}_j + \boldsymbol{\varepsilon}_i \cdot \frac{\partial \boldsymbol{\varepsilon}_j}{\partial q^k} \\ &= [ik, j] + [jk, i] \end{aligned} \tag{2.136}$$

by Eq. (2.135). Then

$$[ij, k] = \frac{1}{2} \left\{ \frac{\partial g_{ik}}{\partial q^j} + \frac{\partial g_{jk}}{\partial q^i} - \frac{\partial g_{ij}}{\partial q^k} \right\}, \tag{2.137}$$

and

$$\begin{aligned} \Gamma_{ij}^s &= g^{ks} [ij, k] \\ &= \frac{1}{2} g^{ks} \left\{ \frac{\partial g_{ik}}{\partial q^j} + \frac{\partial g_{jk}}{\partial q^i} - \frac{\partial g_{ij}}{\partial q^k} \right\}. \end{aligned} \tag{2.138}$$

These Christoffel symbols are applied in the next section.

Covariant Derivative

With the Christoffel symbols, Eq. (2.130b) may be rewritten

$$\frac{\partial \mathbf{V}'}{\partial q^j} = \frac{\partial V^i}{\partial q^j}\boldsymbol{\varepsilon}_i + V^i \Gamma^k_{ij}\boldsymbol{\varepsilon}_k. \tag{2.139}$$

Now, i and k in the last term are dummy indices. Interchanging i and k (in this one term), we have

$$\frac{\partial \mathbf{V}'}{\partial q^j} = \left(\frac{\partial V^i}{\partial q^j} + V^k \Gamma^i_{kj}\right)\boldsymbol{\varepsilon}_i. \tag{2.140}$$

The quantity in parenthesis is labeled a **covariant derivative**, $V^i_{;j}$. We have

$$V^i_{;j} \equiv \frac{\partial V^i}{\partial q^j} + V^k \Gamma^i_{kj}. \tag{2.141}$$

The $;j$ subscript indicates differentiation with respect to q^j. The differential $d\mathbf{V}'$ becomes

$$d\mathbf{V}' = \frac{\partial \mathbf{V}'}{\partial q^j}\, dq^j = [V^i_{;j}\, dq^j]\boldsymbol{\varepsilon}_i. \tag{2.142}$$

A comparison with Eq. (2.113) or (2.122) shows that the quantity in square brackets is the ith contravariant component of a vector. Since dq^j is the jth contravariant component of a vector (again, Eq. (2.113)), $V^i_{;j}$ must be the ijth component of a (mixed) second-rank tensor (quotient rule). The covariant derivatives of the contravariant components of a vector form a mixed second-rank tensor, $V^i_{;j}$.

Since the Christoffel symbols vanish in Cartesian coordinates, the covariant derivative and the ordinary partial derivative coincide:

$$\frac{\partial V^i}{\partial q^j} = V^i_{;j} \qquad \text{(Cartesian coordinates).} \tag{2.143}$$

The covariant derivative of a covariant vector V_i is given by (Exercise 2.10.9)

$$V_{i;j} = \frac{\partial V_i}{\partial q^j} - V_k \Gamma^k_{ij}. \tag{2.144}$$

Like $V^i_{;j}$, $V_{i;j}$ is a second-rank tensor.

The physical importance of the covariant derivative is that "A consistent replacement of regular partial derivatives by covariant derivatives carries the laws of physics (in component form) from flat space–time into the curved (Riemannian) space–time of general relativity. Indeed, this substitution may be taken as a mathematical statement of Einstein's principle of equivalence."[21]

[21] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*. San Francisco: W. H. Freeman (1973), p. 387.

Geodesics, Parallel Transport

The covariant derivative of vectors, tensors, and the Christoffel symbols may also be approached from geodesics. A geodesic in Euclidean space is a straight line. In general, it is the curve of shortest length between two points and the curve along which a freely falling particle moves. The ellipses of planets are geodesics around the sun, and the moon is in free fall around the Earth on a geodesic. Since we can throw a particle in any direction, a geodesic can have any direction through a given point. Hence the geodesic equation can be obtained from Fermat's variational principle of optics (see Chapter 17 for Euler's equation),

$$\delta \int ds = 0, \tag{2.145}$$

where ds^2 is the metric, Eq. (2.123), of our space. Using the variation of ds^2,

$$2\,ds\,\delta\,ds = dq^i\,dq^j\,\delta\,g_{ij} + g_{ij}\,dq^i\,\delta\,dq^j + g_{ij}\,dq^j\,\delta\,dq^i \tag{2.146}$$

in Eq. (2.145) yields

$$\frac{1}{2}\int\left[\frac{dq^i}{ds}\frac{dq^j}{ds}\delta g_{ij} + g_{ij}\frac{dq^i}{ds}\frac{d}{ds}\delta\,dq^j + g_{ij}\frac{dq^j}{ds}\frac{d}{ds}\delta\,dq^i\right]ds = 0, \tag{2.147}$$

where ds measures the length on the geodesic. Expressing the variations

$$\delta g_{ij} = \frac{\partial g_{ij}}{\partial q^k}\delta\,dq^k \equiv (\partial_k g_{ij})\delta\,dq^k$$

in terms of the **independent variations** $\delta\,dq^k$, shifting their derivatives in the other two terms of Eq. (2.147) upon integrating by parts, and renaming dummy summation indices, we obtain

$$\frac{1}{2}\int\left[\frac{dq^i}{ds}\frac{dq^j}{ds}\partial_k g_{ij} - \frac{d}{ds}\left(g_{ik}\frac{dq^i}{ds} + g_{kj}\frac{dq^j}{ds}\right)\right]\delta\,dq^k\,ds = 0. \tag{2.148}$$

The integrand of Eq. (2.148), set equal to zero, is the geodesic equation. It is the Euler equation of our variational problem. Upon expanding

$$\frac{dg_{ik}}{ds} = (\partial_j g_{ik})\frac{dq^j}{ds}, \qquad \frac{dg_{kj}}{ds} = (\partial_i g_{kj})\frac{dq^i}{ds} \tag{2.149}$$

along the geodesic we find

$$\frac{1}{2}\frac{dq^i}{ds}\frac{dq^j}{ds}(\partial_k g_{ij} - \partial_j g_{ik} - \partial_i g_{kj}) - g_{ik}\frac{d^2q^i}{ds^2} = 0. \tag{2.150}$$

Multiplying Eq. (2.150) with g^{kl} and using Eq. (2.125), we find the **geodesic** equation

$$\frac{d^2q^l}{ds^2} + \frac{dq^i}{ds}\frac{dq^j}{ds}\frac{1}{2}g^{kl}(\partial_i g_{kj} + \partial_j g_{ik} - \partial_k g_{ij}) = 0, \tag{2.151}$$

where the coefficient of the velocities is the Christoffel symbol Γ^l_{ij} of Eq. (2.138).

Geodesics are curves that are independent of the choice of coordinates. They can be drawn through any point in space in various directions. Since the length ds measured along

the geodesic is a scalar, the velocities dq^i/ds (of a freely falling particle along the geodesic, for example) form a contravariant vector. Hence $V_k\,dq^k/ds$ is a well-defined scalar on any geodesic, which we can differentiate in order to define the covariant derivative of any covariant vector V_k. Using Eq. (2.151) we obtain from the scalar

$$\begin{aligned}\frac{d}{ds}\left(V_k\frac{dq^k}{ds}\right) &= \frac{dV_k}{ds}\frac{dq^k}{ds} + V_k\frac{d^2q^k}{ds^2}\\ &= \frac{\partial V_k}{\partial q^i}\frac{dq^i}{ds}\frac{dq^k}{ds} - V_k\Gamma^k_{ij}\frac{dq^i}{ds}\frac{dq^j}{ds}\\ &= \frac{dq^i}{ds}\frac{dq^k}{ds}\left(\frac{\partial V_k}{\partial q^i} - \Gamma^l_{ik}V_l\right).\end{aligned} \tag{2.152}$$

When the quotient theorem is applied to Eq. (2.152) it tells us that

$$V_{k;i} = \frac{\partial V_k}{\partial q^i} - \Gamma^l_{ik}V_l \tag{2.153}$$

is a covariant tensor that defines the covariant derivative of V_k, consistent with Eq. (2.144). Similarly, higher-order tensors may be derived.

The second term in Eq. (2.153) defines the **parallel transport or displacement**,

$$\delta V_k = \Gamma^l_{ki}V_l\delta q^i, \tag{2.154}$$

of the covariant vector V_k from the point with coordinates q^i to $q^i + \delta q^i$. The parallel transport, δU^k, of a contravariant vector U^k may be found from the invariance of the scalar product U^kV_k under parallel transport,

$$\delta(U^kV_k) = \delta U^kV_k + U^k\delta V_k = 0, \tag{2.155}$$

in conjunction with the quotient theorem.

In summary, when we shift a vector to a neighboring point, parallel transport prevents it from sticking out of our space. This can be clearly seen on the surface of a sphere in spherical geometry, where a tangent vector is supposed to remain a tangent upon translating it along some path on the sphere. This explains why the covariant derivative of a vector or tensor is naturally defined by translating it along a geodesic in the desired direction.

Exercises

2.10.1 Equations (2.115) and (2.116) use the scale factor h_i, citing Exercise 2.2.3. In Section 2.2 we had restricted ourselves to orthogonal coordinate systems, yet Eq. (2.115) holds for nonorthogonal systems. Justify the use of Eq. (2.115) for nonorthogonal systems.

2.10.2 (a) Show that $\boldsymbol{\varepsilon}^i \cdot \boldsymbol{\varepsilon}_j = \delta^i_j$.
(b) From the result of part (a) show that

$$F^i = \mathbf{F}\cdot\boldsymbol{\varepsilon}^i \qquad \text{and} \qquad F_i = \mathbf{F}\cdot\boldsymbol{\varepsilon}_i.$$

2.10.3 For the special case of three-dimensional space ($\boldsymbol{\varepsilon}_1$, $\boldsymbol{\varepsilon}_2$, $\boldsymbol{\varepsilon}_3$ defining a right-handed coordinate system, not necessarily orthogonal), show that

$$\boldsymbol{\varepsilon}^i = \frac{\boldsymbol{\varepsilon}_j \times \boldsymbol{\varepsilon}_k}{\boldsymbol{\varepsilon}_j \times \boldsymbol{\varepsilon}_k \cdot \boldsymbol{\varepsilon}_i}, \qquad i, j, k = 1,\ 2, 3 \text{ and cyclic permutations.}$$

Note. These contravariant basis vectors $\boldsymbol{\varepsilon}^i$ define the reciprocal lattice space of Section 1.5.

2.10.4 Prove that the contravariant metric tensor is given by

$$g^{ij} = \boldsymbol{\varepsilon}^i \cdot \boldsymbol{\varepsilon}^j.$$

2.10.5 If the covariant vectors $\boldsymbol{\varepsilon}_i$ are orthogonal, show that

(a) g_{ij} is diagonal,
(b) $g^{ii} = 1/g_{ii}$ (no summation),
(c) $|\boldsymbol{\varepsilon}^i| = 1/|\boldsymbol{\varepsilon}_i|$.

2.10.6 Derive the covariant and contravariant metric tensors for circular cylindrical coordinates.

2.10.7 Transform the right-hand side of Eq. (2.129),

$$\nabla\psi = \frac{\partial\psi}{\partial q^i}\boldsymbol{\varepsilon}^i,$$

into the $\mathbf{e}_i$ basis, and verify that this expression agrees with the gradient developed in Section 2.2 (for orthogonal coordinates).

2.10.8 Evaluate $\partial\boldsymbol{\varepsilon}_i/\partial q^j$ for spherical polar coordinates, and from these results calculate Γ^k_{ij} for spherical polar coordinates.
Note. Exercise 2.5.2 offers a way of calculating the needed partial derivatives. Remember,

$$\boldsymbol{\varepsilon}_1 = \hat{\mathbf{r}} \qquad \text{but} \qquad \boldsymbol{\varepsilon}_2 = r\hat{\boldsymbol{\theta}} \qquad \text{and} \qquad \boldsymbol{\varepsilon}_3 = r\sin\theta\hat{\boldsymbol{\varphi}}.$$

2.10.9 Show that the covariant derivative of a covariant vector is given by

$$V_{i;j} \equiv \frac{\partial V_i}{\partial q^j} - V_k\Gamma^k_{ij}.$$

Hint. Differentiate

$$\boldsymbol{\varepsilon}^i \cdot \boldsymbol{\varepsilon}_j = \delta^i_j.$$

2.10.10 Verify that $V_{i;j} = g_{ik}V^k_{;j}$ by showing that

$$\frac{\partial V_i}{\partial q^j} - V_s\Gamma^s_{ij} = g_{ik}\left\{\frac{\partial V^k}{\partial q^j} + V^m\Gamma^k_{mj}\right\}.$$

2.10.11 From the circular cylindrical metric tensor g_{ij}, calculate the Γ^k_{ij} for circular cylindrical coordinates.
Note. There are only three nonvanishing Γ.

2.10.12 Using the Γ^k_{ij} from Exercise 2.10.11, write out the covariant derivatives $V^i_{;j}$ of a vector $\mathbf{V}$ in circular cylindrical coordinates.

2.10.13 A triclinic crystal is described using an oblique coordinate system. The three covariant base vectors are

$$\boldsymbol{\varepsilon}_1 = 1.5\hat{\mathbf{x}},$$
$$\boldsymbol{\varepsilon}_2 = 0.4\hat{\mathbf{x}} + 1.6\hat{\mathbf{y}},$$
$$\boldsymbol{\varepsilon}_3 = 0.2\hat{\mathbf{x}} + 0.3\hat{\mathbf{y}} + 1.0\hat{\mathbf{z}}.$$

(a) Calculate the elements of the covariant metric tensor g_{ij}.
(b) Calculate the Christoffel three-index symbols, Γ^k_{ij}. (This is a "by inspection" calculation.)
(c) From the cross-product form of Exercise 2.10.3 calculate the contravariant base vector $\boldsymbol{\varepsilon}^3$.
(d) Using the explicit forms $\boldsymbol{\varepsilon}^3$ and $\boldsymbol{\varepsilon}_i$, verify that $\boldsymbol{\varepsilon}^3 \cdot \boldsymbol{\varepsilon}_i = \delta^3{}_i$.

Note. If it were needed, the contravariant metric tensor could be determined by finding the inverse of g_{ij} or by finding the $\boldsymbol{\varepsilon}^i$ and using $g^{ij} = \boldsymbol{\varepsilon}^i \cdot \boldsymbol{\varepsilon}^j$.

2.10.14 Verify that

$$[ij, k] = \frac{1}{2}\left\{\frac{\partial g_{ik}}{\partial q^j} + \frac{\partial g_{jk}}{\partial q^i} - \frac{\partial g_{ij}}{\partial q^k}\right\}.$$

Hint. Substitute Eq. (2.135) into the right-hand side and show that an identity results.

2.10.15 Show that for the metric tensor $g_{ij;k} = 0$, $g^{ij}{}_{;k} = 0$.

2.10.16 Show that parallel displacement $\delta\, dq^i = d^2q^i$ along a geodesic. Construct a geodesic by parallel displacement of $\delta\, dq^i$.

2.10.17 Construct the covariant derivative of a vector V^i by parallel transport starting from the limiting procedure

$$\lim_{dq^j \to 0} \frac{V^i(q^j + dq^j) - V^i(q^j)}{dq^j}.$$

2.11 TENSOR DERIVATIVE OPERATORS

In this section the covariant differentiation of Section 2.10 is applied to rederive the vector differential operations of Section 2.2 in general tensor form.

Divergence

Replacing the partial derivative by the covariant derivative, we take the divergence to be

$$\nabla \cdot \mathbf{V} = V^i_{;i} = \frac{\partial V^i}{\partial q^i} + V^k \Gamma^i_{ik}. \tag{2.156}$$

Expressing Γ^i_{ik} by Eq. (2.138), we have

$$\Gamma^i_{ik} = \frac{1}{2} g^{im} \left\{ \frac{\partial g_{im}}{\partial q^k} + \frac{\partial g_{km}}{\partial q^i} - \frac{\partial g_{ik}}{\partial q^m} \right\}. \tag{2.157}$$

When contracted with g^{im} the last two terms in the curly bracket cancel, since

$$g^{im} \frac{\partial g_{km}}{\partial q^i} = g^{mi} \frac{\partial g_{ki}}{\partial q^m} = g^{im} \frac{\partial g_{ik}}{\partial q^m}. \tag{2.158}$$

Then

$$\Gamma^i_{ik} = \frac{1}{2} g^{im} \frac{\partial g_{im}}{\partial q^k}. \tag{2.159}$$

From the theory of determinants, Section 3.1,

$$\frac{\partial g}{\partial q^k} = g g^{im} \frac{\partial g_{im}}{\partial q^k}, \tag{2.160}$$

where g is the determinant of the metric, $g = \det(g_{ij})$. Substituting this result into Eq. (2.158), we obtain

$$\Gamma^i_{ik} = \frac{1}{2g} \frac{\partial g}{\partial q^k} = \frac{1}{g^{1/2}} \frac{\partial g^{1/2}}{\partial q^k}. \tag{2.161}$$

This yields

$$\nabla \cdot \mathbf{V} = V^i_{;i} = \frac{1}{g^{1/2}} \frac{\partial}{\partial q^k} \left(g^{1/2} V^k \right). \tag{2.162}$$

To compare this result with Eq. (2.21), note that $h_1 h_2 h_3 = g^{1/2}$ and V^i (contravariant coefficient of $\boldsymbol{\varepsilon}_i$) $= V_i / h_i$ (no summation), where V_i is Section 2.2 coefficient of $\mathbf{e}_i$.

Laplacian

In Section 2.2, replacement of the vector $\mathbf{V}$ in $\nabla \cdot \mathbf{V}$ by $\nabla \psi$ led to the Laplacian $\nabla \cdot \nabla \psi$. Here we have a contravariant V^i. Using the metric tensor to create a contravariant $\nabla \psi$, we make the substitution

$$V^i \to g^{ik} \frac{\partial \psi}{\partial q^k}.$$

Then the Laplacian $\nabla \cdot \nabla \psi$ becomes

$$\boxed{\nabla \cdot \nabla \psi = \frac{1}{g^{1/2}} \frac{\partial}{\partial q^i} \left(g^{1/2} g^{ik} \frac{\partial \psi}{\partial q^k} \right).} \tag{2.163}$$

For the **orthogonal** systems of Section 2.2 the metric tensor is diagonal and the contravariant g^{ii} (no summation) becomes

$$g^{ii} = (h_i)^{-2}.$$

Equation (2.163) reduces to

$$\nabla \cdot \nabla \psi = \frac{1}{h_1 h_2 h_3} \frac{\partial}{\partial q^i} \left(\frac{h_1 h_2 h_3}{h_i^2} \frac{\partial \psi}{\partial q^i} \right),$$

in agreement with Eq. (2.22).

Curl

The difference of derivatives that appears in the curl (Eq. (2.27)) will be written

$$\frac{\partial V_i}{\partial q^j} - \frac{\partial V_j}{\partial q^i}.$$

Again, remember that the components V_i here are coefficients of the contravariant (nonunit) base vectors $\boldsymbol{\varepsilon}^i$. The V_i of Section 2.2 are coefficients of unit vectors $\mathbf{e}_i$. Adding and subtracting, we obtain

$$\begin{aligned} \frac{\partial V_i}{\partial q^j} - \frac{\partial V_j}{\partial q^i} &= \frac{\partial V_i}{\partial q^j} - V_k \Gamma_{ij}^k - \frac{\partial V_j}{\partial q^i} + V_k \Gamma_{ji}^k \\ &= V_{i;j} - V_{j;i} \end{aligned} \tag{2.164}$$

using the symmetry of the Christoffel symbols. The characteristic difference of derivatives of the curl becomes a difference of covariant derivatives and therefore is a second-rank tensor (covariant in both indices). As emphasized in Section 2.9, the special vector form of the curl exists only in three-dimensional space.

From Eq. (2.138) it is clear that all the Christoffel three index symbols vanish in Minkowski space and in the real space–time of special relativity with

$$g_{\lambda\mu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Here

$$x_0 = ct, \qquad x_1 = x, \qquad x_2 = y, \qquad \text{and} \qquad x_3 = z.$$

This completes the development of the differential operators in general tensor form. (The gradient was given in Section 2.10.) In addition to the fields of elasticity and electromagnetism, these differentials find application in mechanics (Lagrangian mechanics, Hamiltonian mechanics, and the Euler equations for rotation of rigid body); fluid mechanics; and perhaps most important of all, the curved space–time of modern theories of gravity.

Exercises

2.11.1 Verify Eq. (2.160),

$$\frac{\partial g}{\partial q^k} = g g^{im} \frac{\partial g_{im}}{\partial q^k},$$

for the specific case of spherical polar coordinates.

2.11.2 Starting with the divergence in tensor notation, Eq. (2.162), develop the divergence of a vector in spherical polar coordinates, Eq. (2.47).

2.11.3 The covariant vector A_i is the gradient of a scalar. Show that the difference of covariant derivatives $A_{i;j} - A_{j;i}$ vanishes.

Additional Readings

Dirac, P. A. M., *General Theory of Relativity*. Princeton, NJ: Princeton University Press (1996).

Hartle, J. B., *Gravity*, San Francisco: Addison-Wesley (2003). This text uses a minimum of tensor analysis.

Jeffreys, H., *Cartesian Tensors*. Cambridge: Cambridge University Press (1952). This is an excellent discussion of Cartesian tensors and their application to a wide variety of fields of classical physics.

Lawden, D. F., *An Introduction to Tensor Calculus, Relativity and Cosmology*, 3rd ed. New York: Wiley (1982).

Margenau, H., and G. M. Murphy, *The Mathematics of Physics and Chemistry*, 2nd ed. Princeton, NJ: Van Nostrand (1956). Chapter 5 covers curvilinear coordinates and 13 specific coordinate systems.

Misner, C. W., K. S. Thorne, and J. A. Wheeler, *Gravitation*. San Francisco: W. H. Freeman (1973), p. 387.

Moller, C., *The Theory of Relativity*. Oxford: Oxford University Press (1955). Reprinted (1972). Most texts on general relativity include a discussion of tensor analysis. Chapter 4 develops tensor calculus, including the topic of dual tensors. The extension to non-Cartesian systems, as required by general relativity, is presented in Chapter 9.

Morse, P. M., and H. Feshbach, *Methods of Theoretical Physics*. New York: McGraw-Hill (1953). Chapter 5 includes a description of several different coordinate systems. Note that Morse and Feshbach are not above using left-handed coordinate systems even for Cartesian coordinates. Elsewhere in this excellent (and difficult) book there are many examples of the use of the various coordinate systems in solving physical problems. Eleven additional fascinating but seldom-encountered orthogonal coordinate systems are discussed in the second (1970) edition of *Mathematical Methods for Physicists*.

Ohanian, H. C., and R. Ruffini, *Gravitation and Spacetime*, 2nd ed. New York: Norton & Co. (1994). A well-written introduction to Riemannian geometry.

Sokolnikoff, I. S., *Tensor Analysis — Theory and Applications*, 2nd ed. New York: Wiley (1964). Particularly useful for its extension of tensor analysis to non-Euclidean geometries.

Weinberg, S., *Gravitation and Cosmology. Principles and Applications of the General Theory of Relativity*. New York: Wiley (1972). This book and the one by Misner, Thorne, and Wheeler are the two leading texts on general relativity and cosmology (with tensors in non-Cartesian space).

Young, E. C., *Vector and Tensor Analysis*, 2nd ed. New York: Marcel Dekker (1993).

This page intentionally left blank

CHAPTER 3

DETERMINANTS AND MATRICES

3.1 DETERMINANTS

We begin the study of matrices by solving linear equations that will lead us to determinants and matrices. The concept of *determinant* and the notation were introduced by the renowned German mathematician and philosopher Gottfried Wilhelm von Leibniz.

Homogeneous Linear Equations

One of the major applications of determinants is in the establishment of a condition for the existence of a nontrivial solution for a set of linear homogeneous algebraic equations. Suppose we have three unknowns x_1, x_2, x_3 (or n equations with n unknowns):

$$\begin{aligned} a_1x_1 + a_2x_2 + a_3x_3 &= 0, \\ b_1x_1 + b_2x_2 + b_3x_3 &= 0, \\ c_1x_1 + c_2x_2 + c_3x_3 &= 0. \end{aligned} \tag{3.1}$$

The problem is to determine under what conditions there is any solution, apart from the trivial one $x_1 = 0,\ x_2 = 0,\ x_3 = 0$. If we use vector notation $\mathbf{x} = (x_1, x_2, x_3)$ for the solution and three rows $\mathbf{a} = (a_1, a_2, a_3),\ \mathbf{b} = (b_1, b_2, b_3),\ \mathbf{c} = (c_1, c_2, c_3)$ of coefficients, then the three equations, Eqs. (3.1), become

$$\mathbf{a} \cdot \mathbf{x} = 0, \qquad \mathbf{b} \cdot \mathbf{x} = 0, \qquad \mathbf{c} \cdot \mathbf{x} = 0. \tag{3.2}$$

These three vector equations have the **geometrical** interpretation that **x** is orthogonal to **a**, **b**, and **c**. If the volume spanned by **a**, **b**, **c** given by the determinant (or triple scalar

product, see Eq. (1.50) of Section 1.5)

$$D_3 = (\mathbf{a}\times\mathbf{b})\cdot\mathbf{c} = \det(\mathbf{a},\mathbf{b},\mathbf{c}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \tag{3.3}$$

is not zero, then there is only the trivial solution $\mathbf{x} = 0$.

Conversely, if the aforementioned determinant of coefficients vanishes, then one of the row vectors is a linear combination of the other two. Let us assume that $\mathbf{c}$ lies in the plane spanned by $\mathbf{a}$ and $\mathbf{b}$, that is, that the third equation is a linear combination of the first two and not independent. Then $\mathbf{x}$ is orthogonal to that plane so that $\mathbf{x} \sim \mathbf{a}\times\mathbf{b}$. Since homogeneous equations can be multiplied by arbitrary numbers, only ratios of the x_i are relevant, for which we then obtain ratios of 2×2 determinants

$$\begin{aligned} \frac{x_1}{x_3} &= \frac{a_2b_3 - a_3b_2}{a_1b_2 - a_2b_1} \\ \frac{x_2}{x_3} &= -\frac{a_1b_3 - a_3b_1}{a_1b_2 - a_2b_1} \end{aligned} \tag{3.4}$$

from the components of the cross product $\mathbf{a}\times\mathbf{b}$, provided $x_3 \sim a_1b_2 - a_2b_1 \neq 0$. This is **Cramer's rule** for three homogeneous linear equations.

Inhomogeneous Linear Equations

The simplest case of two equations with two unknowns,

$$a_1x_1 + a_2x_2 = a_3, \qquad b_1x_1 + b_2x_2 = b_3, \tag{3.5}$$

can be reduced to the previous case by imbedding it in three-dimensional space with a solution vector $\mathbf{x} = (x_1, x_2, -1)$ and row vectors $\mathbf{a} = (a_1, a_2, a_3)$, $\mathbf{b} = (b_1, b_2, b_3)$. As before, Eqs. (3.5) in vector notation, $\mathbf{a}\cdot\mathbf{x} = 0$ and $\mathbf{b}\cdot\mathbf{x} = 0$, imply that $\mathbf{x} \sim \mathbf{a}\times\mathbf{b}$, so the analog of Eqs. (3.4) holds. For this to apply, though, the third component of $\mathbf{a}\times\mathbf{b}$ must not be zero, that is, $a_1b_2 - a_2b_1 \neq 0$, because the third component of $\mathbf{x}$ is $-1 \neq 0$. This yields the x_i as

$$\boxed{\begin{aligned} x_2 &= \frac{a_1b_3 - a_3b_1}{a_1b_2 - a_2b_1} = \frac{\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}}{\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}}, \\ x_1 &= \frac{a_3b_2 - b_3a_2}{a_1b_2 - a_2b_1} = \frac{\begin{vmatrix} a_3 & a_2 \\ b_3 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}} \end{aligned}} \tag{3.6a}$$

$$\frac{}{a_1b_2 - a_2b_1} = \frac{\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}}{\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}} \tag{3.6b}$$

The determinant in the numerator of x_i (x_2) is obtained from the determinant of the coefficients $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$ by replacing the first (second) column vector by the vector $\binom{a_3}{b_3}$ of the inhomogeneous side of Eq. (3.5). This is **Cramer's rule** for a set of two inhomogeneous linear equations with two unknowns.

tation for Eq. (3.9). The number of terms in the sum (Eq. (3.8)) is 24 for a fourth-order determinant, $n!$ for an nth-order determinant. Because of the appearance of the negative signs in Eq. (3.9) (and possibly in the individual elements as well), there may be considerable cancellation. It is quite possible that a determinant of large elements will have a very small value.

Several useful properties of the nth-order determinants follow from Eq. (3.8). Again, to be specific, Eq. (3.9) for third-order determinants is used to illustrate these properties.

These solutions of linear equations in terms of determinants can be generalized to n dimensions. The determinant is a square array

$$D_n = \begin{vmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \\ c_1 & c_2 & \cdots & c_n \\ \cdot & \cdot & \cdots & \cdot \end{vmatrix}$$

Laplacian Development by Minors

Equation (3.9) may be written

$$\begin{aligned} D_3 &= a_1(b_2c_3 - b_3c_2) - a_2(b_1c_3 - b_3c_1) + a_3(b_1c_2 - b_2c_1) \\ &= a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} \end{aligned}$$

[1] In a linear sequence $abcd\cdots$, any single, simple transposition of adjacent elements yields an **odd** permutation of the original sequence: $abcd \to bacd$. Two such transpositions yield an even permutation. In general, an odd number of such interchanges of adjacent elements results in an odd permutation; an even number of such transpositions yields an even permutation.

(This determinant D (Eq. (3.12)) is formed from one of the Dirac matrices appearing in Dirac's relativistic electron theory in Section 3.4.)

Antisymmetry

The determinant changes sign if any two rows are interchanged or if any two columns are interchanged. This follows from the even–odd character of the Levi-Civita ε in Eq. (3.8) or explicitly from the form of Eqs. (3.9) and (3.10).[2]

This property was used in Section 2.9 to develop a totally antisymmetric linear combination. It is also frequently used in quantum mechanics in the construction of a many-particle wave function that, in accordance with the Pauli exclusion principle, will be antisymmetric under the interchange of any two identical spin $\frac{1}{2}$ particles (electrons, protons, neutrons, etc.).

[2]The sign reversal is reasonably obvious for the interchange of two adjacent rows (or columns), this clearly being an odd permutation. Show that the interchange of **any** two rows is still an odd permutation.

- As a special case of antisymmetry, any determinant with two rows equal or two columns equal equals zero.
- If each element in a row or each element in a column is zero, the determinant is equal to zero.
- If each element in a row or each element in a column is multiplied by a constant, the determinant is multiplied by that constant.
- The value of a determinant is unchanged if a multiple of one row is added (column by column) to another row or if a multiple of one column is added (row by row) to another column.[3]

We have

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

[3]This derives from the geometric meaning of the determinant as the volume of the parallelepiped spanned by its column vectors. Pulling it to the side without changing its height leaves the volume unchanged.

vanishes, the inhomogeneous set of equations has no solution — unless the numerators also vanish. In this case solutions may exist but they are not unique (see Exercise 3.1.3 for a specific example).

For numerical work, this determinant solution, Eq. (3.18), is exceedingly unwieldy. The determinant may involve large numbers with alternate signs, and in the subtraction of two large numbers the relative error may soar to a point that makes the result worthless. Also, although the determinant method is illustrated here with three equations and three unknowns, we might easily have 200 equations with 200 unknowns, which, involving up to 200! terms in each determinant, pose a challenge even to high-speed computers. There must be a better way.

If our linear equations are **inhomogeneous**, that is, as in Eqs. (3.5), if the zeros on the right-hand side of Eqs. (3.10) are replaced by a_4, b_4, and c_4, respectively, then from Eq. (3.17) we obtain instead

$$x_1 = \frac{\begin{vmatrix} a_4 & a_2 & a_3 \\ b_4 & b_2 & b_3 \\ c_4 & c_2 & c_3 \end{vmatrix}}{\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}}$$

In fact, there are better ways. One of the best is a straightforward process often called **Gauss elimination**. To illustrate this technique, consider the following set of equations.

Example 3.1.1 GAUSS ELIMINATION

Solve

$$\begin{aligned} 3x + 2y + z &= 11 \\ 2x + 3y + z &= 13 \\ x + y + 4z &= 12. \end{aligned} \tag{3.19}$$

The determinant of the inhomogeneous linear equations (3.19) is 18, so a solution exists.

For convenience and for the optimum numerical accuracy, the equations are rearranged so that the largest coefficients run along the main diagonal (upper left to lower right). This has already been done in the preceding set.

The Gauss technique is to use the first equation to eliminate the first unknown, x, from the remaining equations. Then the (new) second equation is used to eliminate y from the last equation. In general, we work down through the set of equations, and then, with one unknown determined, we work back up to solve for each of the other unknowns in succession.

Dividing each row by its initial coefficient, we see that Eqs. (3.19) become

$$x + {}^{2}$$

$$\begin{aligned} \tfrac{}{3}y + \tfrac{1}{3}z &= \tfrac{11}{3} \\ x + \tfrac{3}{2}y + \tfrac{1}{2}z &= \tfrac{13}{2} \\ x + y + 4z &= 12. \end{aligned} \tag{3.20}$$

Now, using the first equation, we eliminate x from the second and third equations:

$$\begin{aligned} x + \tfrac{2}{3}y + \tfrac{1}{3}z &= \tfrac{11}{3} \\ \tfrac{5}{6}y + \tfrac{1}{6}z &= \tfrac{17}{6} \\ \tfrac{1}{3}y + \tfrac{11}{3}z &= \tfrac{25}{3} \end{aligned} \tag{3.21}$$

and

$$\begin{aligned} x + \tfrac{2}{3}y + \tfrac{1}{3}z &= \tfrac{11}{3} \\ y + \tfrac{1}{5}z &= \tfrac{17}{5} \\ y + 11z &= 25. \end{aligned} \tag{3.22}$$

Repeating the technique, we use the new second equation to eliminate y from the third equation:

$$\begin{aligned} x + \tfrac{2}{3}y + \tfrac{1}{3}z &= \tfrac{11}{3} \\ y + \tfrac{1}{5}z &= \tfrac{17}{5} \\ 54z &= 108, \end{aligned} \tag{3.23}$$

or

$$z = 2.$$

Finally, working back up, we get

$$y + \tfrac{1}{5} \times 2 = \tfrac{17}{5},$$

or

$$y = 3.$$

Then with z and y determined,

$$x + \tfrac{2}{3} \times 3 + \tfrac{1}{3} \times 2 = \tfrac{11}{3},$$

and

$$x = 1.$$

The technique may not seem so elegant as Eq. (3.18), but it is well adapted to computers and is far faster than the time spent with determinants.

This Gauss technique may be used to convert a determinant into triangular form:

$$D = \begin{vmatrix} a_1 & b_1 & c_1 \\ 0 & b_2 & c_2 \\ 0 & 0 & c_3 \end{vmatrix}$$

for a third-order determinant whose elements are not to be confused with those in Eq. (3.3). In this form $D = a_1 b_2 c_3$. For an nth-order determinant the evaluation of the triangular form requires only $n - 1$ multiplications, compared with the $n!$ required for the general case.

A variation of this progressive elimination is known as Gauss–Jordan elimination. We start as with the preceding Gauss elimination, but each new equation considered is used to eliminate a variable from **all** the other equations, not just those below it. If we had used this Gauss–Jordan elimination, Eq. (3.23) would become

$$\begin{aligned} x + \tfrac{1}{5}z &= \tfrac{7}{5} \\ y + \tfrac{1}{5}z &= \tfrac{17}{5} \\ z &= 2, \end{aligned} \tag{3.24}$$

using the second equation of Eqs. (3.22) to eliminate y from both the first and third equations. Then the third equation of Eqs. (3.24) is used to eliminate z from the first and second, giving

$$\begin{aligned} x &= 1 \\ y &= 3 \\ z &= 2. \end{aligned} \tag{3.25}$$

We return to this Gauss–Jordan technique in Section 3.2 for inverting matrices.

Another technique suitable for computer use is the Gauss–Seidel iteration technique. Each technique has its advantages and disadvantages. The Gauss and Gauss–Jordan methods may have accuracy problems for large determinants. This is also a problem for matrix inversion (Section 3.2). The Gauss–Seidel method, as an iterative method, may have convergence problems. The IBM Scientific Subroutine Package (SSP) uses Gauss and Gauss–Jordan techniques. The Gauss–Seidel iterative method and the Gauss and Gauss–Jordan elimination methods are discussed in considerable detail by Ralston and Wilf and also by Pennington.[4] Computer codes in FORTRAN and other programming languages and extensive literature for the Gauss–Jordan elimination and others are also given by Press *et al.*[5] ■

Linear Dependence of Vectors

Two nonzero two-dimensional vectors

$$\mathbf{a}_1 = \begin{pmatrix} a_{11} \\ a_{12} \end{pmatrix} \neq 0, \qquad \mathbf{a}_2 = \begin{pmatrix} a_{21} \\ a_{22} \end{pmatrix} \neq 0$$

are defined to be **linearly dependent** if two numbers x_1, x_2 can be found that are not both zero so that the linear relation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = 0$ holds. They are **linearly independent** if $x_1 = 0 = x_2$ is the only solution of this linear relation. Writing it in Cartesian components, we obtain two homogeneous linear equations

$$a_{11}x_1 + a_{21}x_2 = 0, \qquad a_{12}x_1 + a_{22}x_2 = 0$$

[4] A. Ralston and H. Wilf, eds., *Mathematical Methods for Digital Computers*. New York: Wiley (1960); R. H. Pennington, *Introductory Computer Methods and Numerical Analysis*. New York: Macmillan (1970).

[5] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*, 2nd ed. Cambridge, UK: Cambridge University Press (1992), Chapter 2.

from which we extract the following criterion for linear independence of two vectors using Cramer's rule. **If $\mathbf{a}_1$, $\mathbf{a}_2$ span a nonzero area**, that is, **their determinant** $\begin{vmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{vmatrix} \neq 0$, **then the set of homogeneous linear equations has only the solution** $x_1 = 0 = x_2$. **If the determinant is zero**, **then there is a nontrivial solution** x_1, x_2, and **our vectors are linearly dependent**. In particular, the unit vectors in the x- and y-directions are linearly independent, the linear relation $x_1\hat{x}_1 + x_2\hat{x}_2 = \binom{x_1}{x_2} = \binom{0}{0}$ having only the trivial solution $x_1 = 0 = x_2$.

Three or more vectors in two-dimensional space are always linearly dependent. Thus, the maximum number of linearly independent vectors in two-dimensional space is 2. For example, given $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, the linear relation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 = 0$ always has nontrivial solutions. If one of the vectors is zero, linear dependence is obvious because the coefficient of the zero vector may be chosen to be nonzero and that of the others as zero. So we assume all of them as nonzero. If $\mathbf{a}_1$ and $\mathbf{a}_2$ are linearly independent, we write the linear relation

$$a_{11}x_1 + a_{21}x_2 = -a_{31}x_3, \qquad a_{12}x_1 + a_{22}x_2 = -a_{32}x_3,$$

as a set of two inhomogeneous linear equations and apply Cramer's rule. Since the determinant is nonzero, we can find a nontrivial solution x_1, x_2 for any nonzero x_3. This argument goes through for any pair of linearly independent vectors. If all pairs are linearly dependent, any of these linear relations is a linear relation among the three vectors, and we are finished. If there are more than three vectors, we pick any three of them and apply the foregoing reasoning and put the coefficients of the other vectors, $x_j = 0$, in the linear relation.

- Mutually orthogonal vectors are linearly independent.

Assume a linear relation $\sum_i c_i\mathbf{v}_i = 0$. Dotting $\mathbf{v}_j$ into this using $\mathbf{v}_j \cdot \mathbf{v}_i = 0$ for $j \neq i$, we obtain $c_j\mathbf{v}_j \cdot \mathbf{v}_j = 0$, so every $c_j = 0$ because $\mathbf{v}_j^2 \neq 0$.

It is straightforward to extend these theorems to n or more vectors in n-dimensional Euclidean space. Thus, **the maximum number of linearly independent vectors in n-dimensional space is** n. The coordinate unit vectors are linearly independent because they span a nonzero parallelepiped in n-dimensional space and their determinant is unity.

Gram–Schmidt Procedure

In an n-dimensional vector space with an inner (or scalar) product, we can always construct an orthonormal basis of n vectors $\mathbf{w}_i$ with $\mathbf{w}_i \cdot \mathbf{w}_j = \delta_{ij}$ starting from n linearly independent vectors $\mathbf{v}_i$, $i = 0, 1, \ldots, n-1$.

We start by normalizing $\mathbf{v}_0$ to unity, defining $\mathbf{w}_0 = \dfrac{\mathbf{v}_0}{\sqrt{\mathbf{v}_0^2}}$. Then we project $\mathbf{v}_0$ from $\mathbf{v}_1$, forming $\mathbf{u}_1 = \mathbf{v}_1 + a_{10}\mathbf{w}_0$, with the admixture coefficient a_{10} chosen so that $\mathbf{v}_0 \cdot \mathbf{u}_1 = 0$. Dotting $\mathbf{v}_0$ into $\mathbf{u}_1$ yields $a_{10} = -\frac{\mathbf{v}_0 \cdot \mathbf{v}_1}{\sqrt{\mathbf{v}_0^2}} = -\mathbf{v}_1 \cdot \mathbf{w}_0$. Again, we normalize $\mathbf{u}_1$ defining $\mathbf{w}_1 = \frac{\mathbf{u}_1}{\sqrt{\mathbf{u}_1^2}}$. Here, $\mathbf{u}_1^2 \neq 0$ because $\mathbf{v}_0$, $\mathbf{v}_1$ are linearly independent. This first step generalizes to

$$\mathbf{u}_j = \mathbf{v}_j + a_{j0}\mathbf{w}_0 + a_{j1}\mathbf{w}_1 + \cdots + a_{jj-1}\mathbf{w}_{j-1},$$

with coefficients $a_{ji} = -\mathbf{v}_j \cdot \mathbf{w}_i$. Normalizing $\mathbf{w}_j = \frac{\mathbf{u}_j}{\sqrt{\mathbf{u}_j^2}}$ completes our construction.

It will be noticed that although this Gram–Schmidt procedure is one possible way of constructing an orthogonal or orthonormal set, the vectors $\mathbf{w}_i$ are not unique. There is an infinite number of possible orthonormal sets.

As an illustration of the freedom involved, consider two (nonparallel) vectors $\mathbf{A}$ and $\mathbf{B}$ in the xy-plane. We may normalize $\mathbf{A}$ to unit magnitude and then form $\mathbf{B}' = a\mathbf{A} + \mathbf{B}$ so that $\mathbf{B}'$ is perpendicular to $\mathbf{A}$. By normalizing $\mathbf{B}'$ we have completed the Gram–Schmidt orthogonalization for two vectors. But any two perpendicular unit vectors, such as $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$, could have been chosen as our orthonormal set. Again, with an infinite number of possible rotations of $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ about the z-axis, we have an infinite number of possible orthonormal sets.

Example 3.1.2 VECTORS BY GRAM–SCHMIDT ORTHOGONALIZATION

To illustrate the method, we consider two vectors

$$\mathbf{v}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{v}_1 = \begin{pmatrix} 1 \\ -2 \end{pmatrix},$$

which are neither orthogonal nor normalized. Normalizing the first vector $\mathbf{w}_0 = \mathbf{v}_0/\sqrt{2}$, we then construct $\mathbf{u}_1 = \mathbf{v}_1 + a_{10}\mathbf{w}_0$ so as to be orthogonal to $\mathbf{v}_0$. This yields

$$\mathbf{u}_1 \cdot \mathbf{v}_0 = 0 = \mathbf{v}_1 \cdot \mathbf{v}_0 + \frac{a_{10}}{\sqrt{2}}\mathbf{v}_0^2 = -1 + a_{10}\sqrt{2},$$

so the adjustable admixture coefficient $a_{10} = 1/\sqrt{2}$. As a result,

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ -2 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{3}{2}\begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

so the second orthonormal vector becomes

$$\mathbf{w}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

We check that $\mathbf{w}_0 \cdot \mathbf{w}_1 = 0$. The two vectors $\mathbf{w}_0$, $\mathbf{w}_1$ form an orthonormal set of vectors, a basis of two-dimensional Euclidean space. ■

Exercises

3.1.1 Evaluate the following determinants:

$$\text{(a)} \begin{vmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{vmatrix}, \qquad \text{(b)} \begin{vmatrix} 1 & 2 & 0 \\ 3 & 1 & 2 \\ 0 & 3 & 1 \end{vmatrix}, \qquad \text{(c)} \frac{1}{\sqrt{2}} \begin{vmatrix} 0 & \sqrt{3} & 0 & 0 \\ \sqrt{3} & 0 & 2 & 0 \\ 0 & 2 & 0 & \sqrt{3} \\ 0 & 0 & \sqrt{3} & 0 \end{vmatrix}.$$

3.1.2 Test the set of linear homogeneous equations

$$x + 3y + 3z = 0, \qquad x - y + z = 0, \qquad 2x + y + 3z = 0$$

to see if it possesses a nontrivial solution, and find one.

3.1.3 Given the pair of equations

$$x + 2y = 3, \qquad 2x + 4y = 6,$$

(a) Show that the determinant of the coefficients vanishes.
(b) Show that the numerator determinants (Eq. (3.18)) also vanish.
(c) Find at least two solutions.

3.1.4 Express the **components** of $\mathbf{A} \times \mathbf{B}$ as 2×2 determinants. Then show that the dot product $\mathbf{A} \cdot (\mathbf{A} \times \mathbf{B})$ yields a Laplacian expansion of a 3×3 determinant. Finally, note that two rows of the 3×3 determinant are identical and hence that $\mathbf{A} \cdot (\mathbf{A} \times \mathbf{B}) = 0$.

3.1.5 If C_{ij} is the cofactor of element a_{ij} (formed by striking out the ith row and jth column and including a sign $(-1)^{i+j}$), show that

(a) $\sum_i a_{ij} C_{ij} = \sum_i a_{ji} C_{ji} = |A|$, where $|A|$ is the determinant with the elements a_{ij},
(b) $\sum_i a_{ij} C_{ik} = \sum_i a_{ji} C_{ki} = 0$, $j \neq k$.

3.1.6 A determinant with all elements of order unity may be surprisingly small. The Hilbert determinant $H_{ij} = (i + j - 1)^{-1}$, $i, j = 1, 2, \ldots, n$ is notorious for its small values.

(a) Calculate the value of the Hilbert determinants of order n for $n = 1$, 2, and 3.
(b) If an appropriate subroutine is available, find the Hilbert determinants of order n for $n = 4$, 5, and 6.

ANS.

n	Det(H_n)
1	1.
2	8.33333×10^{-2}
3	4.62963×10^{-4}
4	1.65344×10^{-7}
5	3.74930×10^{-12}
6	5.36730×10^{-18}

3.1.7 Solve the following set of linear simultaneous equations. Give the results to five decimal places.

$$\begin{aligned}
1.0x_1 + 0.9x_2 + 0.8x_3 + 0.4x_4 + 0.1x_5 &= 1.0 \\
0.9x_1 + 1.0x_2 + 0.8x_3 + 0.5x_4 + 0.2x_5 + 0.1x_6 &= 0.9 \\
0.8x_1 + 0.8x_2 + 1.0x_3 + 0.7x_4 + 0.4x_5 + 0.2x_6 &= 0.8 \\
0.4x_1 + 0.5x_2 + 0.7x_3 + 1.0x_4 + 0.6x_5 + 0.3x_6 &= 0.7 \\
0.1x_1 + 0.2x_2 + 0.4x_3 + 0.6x_4 + 1.0x_5 + 0.5x_6 &= 0.6 \\
0.1x_2 + 0.2x_3 + 0.3x_4 + 0.5x_5 + 1.0x_6 &= 0.5.
\end{aligned}$$

Note. These equations may also be solved by matrix inversion, Section 3.2.

3.1.8 Solve the linear equations $\mathbf{a}\cdot\mathbf{x}=c$, $\mathbf{a}\times\mathbf{x}+\mathbf{b}=0$ for $\mathbf{x}=(x_1,x_2,x_3)$ with constant vectors $\mathbf{a}\neq 0$, $\mathbf{b}$ and constant c.

ANS. $\mathbf{x}=\frac{c}{a^2}\mathbf{a}+(\mathbf{a}\times\mathbf{b})/a^2$.

3.1.9 Solve the linear equations $\mathbf{a}\cdot\mathbf{x}=d$, $\mathbf{b}\cdot\mathbf{x}=e$, $\mathbf{c}\cdot\mathbf{x}=f$, for $\mathbf{x}=(x_1,x_2,x_3)$ with constant vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ and constants d, e, f such that $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}\neq 0$.

ANS. $[(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}]\mathbf{x}=d(\mathbf{b}\times\mathbf{c})+e(\mathbf{c}\times\mathbf{a})+f(\mathbf{a}\times\mathbf{b})$.

3.1.10 Express in vector form the solution (x_1,x_2,x_3) of $\mathbf{a}x_1+\mathbf{b}x_2+\mathbf{c}x_3+\mathbf{d}=0$ with constant vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ so that $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}\neq 0$.

3.2 MATRICES

Matrix analysis belongs to linear algebra because matrices are linear operators or maps such as rotations. Suppose, for instance, we rotate the Cartesian coordinates of a two-dimensional space, as in Section 1.2, so that, in vector notation,

$$\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} x_1\cos\varphi + x_2\sin\varphi \\ -x_2\sin\varphi + x_2\cos\varphi \end{pmatrix} = \begin{pmatrix} \sum_j a_{1j}x_j \\ \sum_j a_{2j}x_j \end{pmatrix}. \tag{3.26}$$

We label the array of elements $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ a 2×2 matrix A consisting of two rows and two columns and consider the vectors x, x' as 2×1 matrices. **We take the summation of products in Eq. (3.26) as a definition of matrix multiplication involving the scalar product of each row vector of A with the column vector** x. Thus, in matrix notation Eq. (3.26) becomes

$$x' = \mathsf{A}x. \tag{3.27}$$

To extend this definition of multiplication of a matrix times a column vector to the product of two 2×2 matrices, let the coordinate rotation be followed by a second rotation given by matrix B such that

$$x'' = \mathsf{B}x'. \tag{3.28}$$

In component form,

$$x_i'' = \sum_j b_{ij}x_j' = \sum_j b_{ij}\sum_k a_{jk}x_k = \sum_k \left(\sum_j b_{ij}a_{jk}\right)x_k. \tag{3.29}$$

The summation over j is matrix multiplication defining a matrix $\mathsf{C}=\mathsf{BA}$ such that

$$x_i'' = \sum_k c_{ik}x_k, \tag{3.30}$$

or $x''=\mathsf{C}x$ in matrix notation. Again, this definition involves the scalar products of row vectors of B with column vectors of A. This definition of matrix multiplication generalizes to $m\times n$ matrices and is found useful; indeed, **this usefulness is the justification for its existence**. The geometrical interpretation is that the matrix product of the two matrices BA is the rotation that carries the unprimed system directly into the double-primed coordinate

system. Before passing to formal definitions, the your should note that operator A is described by its effect on the coordinates or basis vectors. The matrix elements a_{ij} constitute a **representation** of the operator, a representation that depends on the choice of a basis.

The special case where a matrix has one column and n rows is called a column vector, $|x\rangle$, with components x_i, $i = 1, 2, \ldots, n$. If A is an $n \times n$ matrix, $|x\rangle$ an n-component column vector, $A|x\rangle$ is defined as in Eqs. (3.27) and (3.26). Similarly, if a matrix has one row and n columns, it is called a row vector, $\langle x|$ with components x_i, $i = 1, 2, \ldots, n$. Clearly, $\langle x|$ results from $|x\rangle$ by interchanging rows and columns, a matrix operation called **transposition**, and transposition for any matrix A, $\tilde{\mathsf{A}}$ is called[6] "A transpose" with matrix elements $(\tilde{\mathsf{A}})_{ik} = A_{ki}$. Transposing a product of matrices AB reverses the order and gives $\tilde{\mathsf{B}}\tilde{\mathsf{A}}$; similarly $\mathsf{A}|x\rangle$ transpose is $\langle x|\mathsf{A}$. The scalar product takes the form $\langle x|y\rangle = \sum_i x_i y_i$ (x_i^* in a complex vector space). This **Dirac bra-ket notation** is used in quantum mechanics extensively and in Chapter 10 and here subsequently.

More abstractly, we can define the **dual space $\tilde{V}$ of linear functionals** F on a vector space V, where each linear functional F of $\tilde{V}$ assigns a number $F(\mathbf{v})$ so that

$$F(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1F(\mathbf{v}_1) + c_2F(\mathbf{v}_2)$$

for any vectors $\mathbf{v}_1$, $\mathbf{v}_2$ from our vector space V and numbers c_1, c_2. If we define the sum of two functionals by linearity as

$$(F_1 + F_2)(\mathbf{v}) = F_1(\mathbf{v}) + F_2(\mathbf{v}),$$

then $\tilde{V}$ is a linear space by construction.

Riesz' theorem says that there is a one-to-one correspondence between linear functionals F in $\tilde{V}$ and vectors $\mathbf{f}$ in a vector space V that has an inner (or scalar) product $\langle \mathbf{f}|\mathbf{v}\rangle$ defined for any pair of vectors $\mathbf{f}$, $\mathbf{v}$.

The proof relies on the scalar product by defining a linear functional F for any vector $\mathbf{f}$ of V as $F(\mathbf{v}) = \langle \mathbf{f}|\mathbf{v}\rangle$ for any $\mathbf{v}$ of V. The linearity of the scalar product in $\mathbf{f}$ shows that these functionals form a vector space (contained in $\tilde{V}$ necessarily). Note that a linear functional is completely specified when it is defined for every vector $\mathbf{v}$ of a given vector space.

On the other hand, starting from any nontrivial linear functional F of $\tilde{V}$ we now construct a unique vector $\mathbf{f}$ of V so that $F(\mathbf{v}) = \mathbf{f} \cdot \mathbf{v}$ is given by an inner product. We start from an orthonormal basis $\mathbf{w}_i$ of vectors in V using the Gram–Schmidt procedure (see Section 3.2). Take any vector $\mathbf{v}$ from V and expand it as $\mathbf{v} = \sum_i \mathbf{w}_i \cdot \mathbf{v}\mathbf{w}_i$. Then the linear functional $F(\mathbf{v}) = \sum_i \mathbf{w}_i \cdot \mathbf{v}F(\mathbf{w}_i)$ is well defined on V. If we define the specific vector $\mathbf{f} = \sum_i F(\mathbf{w}_i)\mathbf{w}_i$, then its inner product with an arbitrary vector $\mathbf{v}$ is given by $\langle \mathbf{f}|\mathbf{v}\rangle = \mathbf{f} \cdot \mathbf{v} = \sum_i F(\mathbf{w}_i)\mathbf{w}_i \cdot \mathbf{v} = F(\mathbf{v})$, which proves Riesz' theorem.

Basic Definitions

A matrix is defined as a square or rectangular array of numbers or functions that obeys certain laws. This is a perfectly logical extension of familiar mathematical concepts. In arithmetic we deal with single numbers. In the theory of complex variables (Chapter 6) we deal with ordered pairs of numbers, $(1, 2) = 1 + 2i$, in which the ordering is important. We

[6] Some texts (including ours sometimes) denote A transpose by A^T.

now consider numbers (or functions) ordered in a square or rectangular array. For convenience in later work the numbers are distinguished by two subscripts, the first indicating the row (horizontal) and the second indicating the column (vertical) in which the number appears. For instance, a_{13} is the matrix element in the first row, third column. Hence, if A is a matrix with m rows and n columns,

$$\mathsf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}. \tag{3.31}$$

Perhaps the most important fact to note is that the elements a_{ij} are not combined with one another. A matrix is not a determinant. It is an ordered array of numbers, not a single number.

The matrix A, so far just an array of numbers, has the properties we assign to it. Literally, this means constructing a new form of mathematics. We define that matrices A, B, and C, with elements a_{ij}, b_{ij}, and c_{ij}, respectively, combine according to the following rules.

Rank

Looking back at the homogeneous linear Eqs. (3.1), we note that the matrix of coefficients, A, is made up of three row vectors that each represent one linear equation of the set. If their triple scalar product is not zero, than they span a nonzero volume and are linearly independent, and the homogeneous linear equations have only the trivial solution. In this case the matrix is said to have **rank** 3. In n dimensions the volume represented by the triple scalar product becomes the determinant, $\det(\mathsf{A})$, for a square matrix. If $\det(\mathsf{A}) \neq 0$, the $n \times n$ matrix A has **rank** n. The case of Eqs. (3.1), where the vector $\mathbf{c}$ lies in the plane spanned by $\mathbf{a}$ and $\mathbf{b}$, corresponds to rank 2 of the matrix of coefficients, because only two of its row vectors ($\mathbf{a}$, $\mathbf{b}$ corresponding to two equations) are independent. In general, the **rank r of a matrix is the maximal number of linearly independent row or column vectors** it has, with $0 \leq r \leq n$.

Equality

Matrix A = Matrix B if and only if $a_{ij} = b_{ij}$ for all values of i and j. This, of course, requires that A and B each be $m \times n$ arrays (m rows, n columns).

Addition, Subtraction

$\mathsf{A} \pm \mathsf{B} = \mathsf{C}$ if and only if $a_{ij} \pm b_{ij} = c_{ij}$ for all values of i and j, the elements combining according to the laws of ordinary algebra (or arithmetic if they are simple numbers). This means that $\mathsf{A} + \mathsf{B} = \mathsf{B} + \mathsf{A}$, commutation. Also, an associative law is satisfied $(\mathsf{A} + \mathsf{B}) + \mathsf{C} = \mathsf{A} + (\mathsf{B} + \mathsf{C})$. If all elements are zero, the matrix, called the **null matrix**, is denoted by O. For all A,

$$\mathsf{A} + \mathsf{O} = \mathsf{O} + \mathsf{A} = \mathsf{A},$$

with

$$\mathsf{O} = \begin{pmatrix} 0 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}. \tag{3.32}$$

Such $m \times n$ matrices form a linear space with respect to addition and subtraction.

Multiplication (by a Scalar)

The multiplication of matrix A by the scalar quantity α is defined as

$$\alpha\mathsf{A} = (\alpha\mathsf{A}), \tag{3.33}$$

in which the elements of $\alpha\mathsf{A}$ are αa_{ij}; that is, each element of matrix A is multiplied by the scalar factor. This is in striking contrast to the behavior of determinants in which the factor α multiplies only one column or one row and not every element of the entire determinant. A consequence of this scalar multiplication is that

$$\alpha\mathsf{A} = \mathsf{A}\alpha, \qquad \text{commutation.}$$

If A is a square matrix, then

$$\det(\alpha\mathsf{A}) = \alpha^n \det(\mathsf{A}).$$

Matrix Multiplication, Inner Product

$$\boxed{\mathsf{AB} = \mathsf{C}} \qquad \text{if and only if}^7 \qquad \boxed{c_{ij} = \sum_k a_{ik} b_{kj}.} \tag{3.34}$$

The ij element of C is formed as a scalar product of the ith row of A with the jth column of B (which demands that A have the same number of columns (n) as B has rows). The dummy index k takes on all values $1, 2, \ldots, n$ in succession; that is,

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} \tag{3.35}$$

for $n = 3$. Obviously, the dummy index k may be replaced by any other symbol that is not already in use without altering Eq. (3.34). Perhaps the situation may be clarified by stating that Eq. (3.34) defines the method of combining certain matrices. This method of combination, to give it a label, is called **matrix multiplication**. To illustrate, consider two (so-called Pauli) matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \text{and} \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.36}$$

[7] Some authors follow the summation convention here (compare Section 2.6).

The $_{11}$ element of the product, $(\sigma_1\sigma_3)_{11}$ is given by the sum of the products of elements of the first **row** of σ_1 with the corresponding elements of the first **column** of σ_3:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \to 0\cdot 1 + 1\cdot 0 = 0.$$

Continuing, we have

$$\sigma_1\sigma_3 = \begin{pmatrix} 0\cdot 1+1\cdot 0 & 0\cdot 0+1\cdot(-1) \\ 1\cdot 1+0\cdot 0 & 1\cdot 0+0\cdot(-1) \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{3.37}$$

Here

$$(\sigma_1\sigma_3)_{ij} = \sigma_{1_{i1}}\sigma_{3_{1j}} + \sigma_{1_{i2}}\sigma_{3_{2j}}.$$

Direct application of the definition of matrix multiplication shows that

$$\sigma_3\sigma_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{3.38}$$

and by Eq. (3.37)

$$\sigma_3\sigma_1 = -\sigma_1\sigma_3. \tag{3.39}$$

Except in special cases, matrix multiplication is not commutative:[8]

$$\mathsf{AB} \neq \mathsf{BA}. \tag{3.40}$$

However, from the definition of matrix multiplication we can show[9] that an associative law holds, $(\mathsf{AB})\mathsf{C} = \mathsf{A}(\mathsf{BC})$. There is also a distributive law, $\mathsf{A}(\mathsf{B}+\mathsf{C}) = \mathsf{AB} + \mathsf{AC}$.

The unit matrix 1 has elements δ_{ij}, Kronecker delta, and the property that $1\mathsf{A} = \mathsf{A}1 = \mathsf{A}$ for all A,

$$1 = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}. \tag{3.41}$$

It should be noted that it is possible for the product of two matrices to be the null matrix without either one being the null matrix. For example, if

$$\mathsf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathsf{B} = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix},$$

$\mathsf{AB} = \mathsf{O}$. This differs from the multiplication of real or complex numbers, which form a **field**, whereas the additive and multiplicative structure of matrices is called a **ring** by mathematicians. See also Exercise 3.2.6(a), from which it is evident that, if $\mathsf{AB} = 0$, at

[8] Commutation or the lack of it is conveniently described by the commutator bracket symbol, $[\mathsf{A}, \mathsf{B}] = \mathsf{AB} - \mathsf{BA}$. Equation (3.40) becomes $[\mathsf{A}, \mathsf{B}] \neq 0$.

[9] Note that the basic definitions of equality, addition, and multiplication are given in terms of the matrix elements, the a_{ij}. All our matrix operations can be carried out in terms of the matrix elements. However, we can also treat a matrix as a single algebraic operator, as in Eq. (3.40). Matrix elements and single operators each have their advantages, as will be seen in the following section. We shall use both approaches.

least one of the matrices must have a zero determinant (that is, be singular as defined after Eq. (3.50) in this section).

If A is an $n \times n$ matrix with determinant $|\mathsf{A}| \neq 0$, then it has a unique inverse A^{-1} satisfying $\mathsf{A}\mathsf{A}^{-1} = \mathsf{A}^{-1}\mathsf{A} = 1$. If B is also an $n \times n$ matrix with inverse B^{-1}, then the product AB has the inverse

$$(\mathsf{AB})^{-1} = \mathsf{B}^{-1}\mathsf{A}^{-1} \tag{3.42}$$

because $\mathsf{ABB}^{-1}\mathsf{A}^{-1} = 1 = \mathsf{B}^{-1}\mathsf{A}^{-1}\mathsf{AB}$ (see also Exercises 3.2.31 and 3.2.32).

The **product theorem**, which says that the determinant of the product, $|\mathsf{AB}|$, of two $n \times n$ matrices A and B is equal to the product of the determinants, $|\mathsf{A}||\mathsf{B}|$, links matrices with determinants. To prove this, consider the n column vectors $\mathbf{c}_k = (\sum_j a_{ij} b_{jk}, i = 1, 2, \ldots, n)$ of the product matrix $\mathsf{C} = \mathsf{AB}$ for $k = 1, 2, \ldots, n$. Each $\mathbf{c}_k = \sum_{j_k} b_{j_k k} \mathbf{a}_{j_k}$ is a sum of n column vectors $\mathbf{a}_{j_k} = (a_{i j_k}, i = 1, 2, \ldots, n)$. Note that we are now using a different product summation index j_k for each column $\mathbf{c}_k$. Since any determinant $D(b_1\mathbf{a}_1 + b_2\mathbf{a}_2) = b_1 D(\mathbf{a}_1) + b_2 D(\mathbf{a}_2)$ is linear in its column vectors, we can pull out the summation sign in front of the determinant from each column vector in C together with the common column factor $b_{j_k k}$ so that

$$|\mathsf{C}| = \sum_{j_k' s} b_{j_1 1} b_{j_2 2} \cdots b_{j_n n} \det(\mathbf{a}_{j_1} \mathbf{a}_{j_2}, \ldots, \mathbf{a}_{j_n}). \tag{3.43}$$

If we rearrange the column vectors $\mathbf{a}_{j_k}$ of the determinant factor in Eq. (3.43) in the proper order, then we can pull the common factor $\det(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) = |\mathsf{A}|$ in front of the n summation signs in Eq. (3.43). These column permutations generate just the right sign $\varepsilon_{j_1 j_2 \cdots j_n}$ to produce in Eq. (3.43) the expression in Eq. (3.8) for $|\mathsf{B}|$ so

$$|\mathsf{C}| = |\mathsf{A}| \sum_{j_k' s} \varepsilon_{j_1 j_2 \cdots j_n} b_{j_1 1} b_{j_2 2} \cdots b_{j_n n} = |\mathsf{A}||\mathsf{B}|, \tag{3.44}$$

which proves the product theorem.

Direct Product

A second procedure for multiplying matrices, known as the **direct** tensor or Kronecker **product**, follows. If A is an $m \times m$ matrix and B is an $n \times n$ matrix, then the direct product is

$$\mathsf{A} \otimes \mathsf{B} = \mathsf{C}. \tag{3.45}$$

C is an $mn \times mn$ matrix with elements

$$C_{\alpha\beta} = A_{ij} B_{kl}, \tag{3.46}$$

with

$$\alpha = m(i-1) + k, \qquad \beta = n(j-1) + l.$$

For instance, if A and B are both 2×2 matrices,

$$\mathsf{A} \otimes \mathsf{B} = \begin{pmatrix} a_{11}\mathsf{B} & a_{12}\mathsf{B} \\ a_{21}\mathsf{B} & a_{22}\mathsf{B} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{pmatrix}. \tag{3.47}$$

The direct product is associative but not commutative. As an example of the direct product, the Dirac matrices of Section 3.4 may be developed as direct products of the Pauli matrices and the unit matrix. Other examples appear in the construction of groups (see Chapter 4) and in vector or Hilbert space in quantum theory.

Example 3.2.1 DIRECT PRODUCT OF VECTORS

The direct product of two two-dimensional vectors is a four-component vector,

$$\begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \otimes \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 y_0 \\ x_0 y_1 \\ x_1 y_0 \\ x_1 y_1 \end{pmatrix};$$

while the direct product of three such vectors,

$$\begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \otimes \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} \otimes \begin{pmatrix} z_0 \\ z_1 \end{pmatrix} = \begin{pmatrix} x_0 y_0 z_0 \\ x_0 y_0 z_1 \\ x_0 y_1 z_0 \\ x_0 y_1 z_1 \\ x_1 y_0 z_0 \\ x_1 y_0 z_1 \\ x_1 y_1 z_0 \\ x_1 y_1 z_1 \end{pmatrix},$$

is a $(2^3 = 8)$-dimensional vector. ■

Diagonal Matrices

An important special type of matrix is the square matrix in which all the nondiagonal elements are zero. Specifically, if a 3×3 matrix A is diagonal, then

$$\mathsf{A} = \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}.$$

A physical interpretation of such diagonal matrices and the method of reducing matrices to this diagonal form are considered in Section 3.5. Here we simply note a significant property of diagonal matrices — multiplication of diagonal matrices is commutative,

$$\mathsf{AB} = \mathsf{BA}, \qquad \text{if A and B are each diagonal.}$$

Multiplication by a diagonal matrix $[d_1, d_2, \ldots, d_n]$ that has only nonzero elements in the diagonal is particularly simple:

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2\cdot 3 & 2\cdot 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 6 & 8 \end{pmatrix};$$

while the opposite order gives

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2\cdot 2 \\ 3 & 2\cdot 4 \end{pmatrix} = \begin{pmatrix} 1 & 4 \\ 3 & 8 \end{pmatrix}.$$

Thus, **a diagonal matrix does not commute with another matrix unless both are diagonal, or the diagonal matrix is proportional to the unit matrix.** This is borne out by the more general form

$$\begin{aligned}[d_1, d_2, \ldots, d_n]\mathsf{A} &= \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \cdots & \cdots & \cdot & \\ 0 & 0 & \cdots & d_n \end{pmatrix}\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdot & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \\ &= \begin{pmatrix} d_1a_{11} & d_1a_{12} & \cdots & d_1a_{1n} \\ d_2a_{21} & d_2a_{22} & \cdots & d_2a_{2n} \\ \cdots & \cdots & \cdot & \\ d_na_{n1} & d_na_{n2} & \cdots & d_na_{nn} \end{pmatrix},\end{aligned}$$

whereas

$$\begin{aligned}\mathsf{A}[d_1, d_2, \ldots, d_n] &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdot & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \cdots & \cdots & \cdot & \\ 0 & 0 & \cdots & d_n \end{pmatrix} \\ &= \begin{pmatrix} d_1a_{11} & d_2a_{12} & \cdots & d_na_{1n} \\ d_1a_{21} & d_2a_{22} & \cdots & d_na_{2n} \\ \cdots & \cdots & \cdot & \\ d_1a_{n1} & d_2a_{n2} & \cdots & d_na_{nn} \end{pmatrix}.\end{aligned}$$

Here we have denoted by $[d_1, \ldots, d_n]$ a diagonal matrix with diagonal elements $d_1, \ldots, d_n$. In the special case of multiplying two diagonal matrices, we simply multiply the corresponding diagonal matrix elements, which obviously is commutative.

Trace

In any square matrix the sum of the diagonal elements is called the **trace**.

Clearly the trace is a linear operation:

$$\operatorname{trace}(\mathsf{A} - \mathsf{B}) = \operatorname{trace}(\mathsf{A}) - \operatorname{trace}(\mathsf{B}).$$

One of its interesting and useful properties is that the trace of a product of two matrices A and B is independent of the order of multiplication:

$$\begin{aligned}\text{trace}(\mathsf{AB}) &= \sum_i (\mathsf{AB})_{ii} = \sum_i \sum_j a_{ij} b_{ji} \\ &= \sum_j \sum_i b_{ji} a_{ij} = \sum_j (\mathsf{BA})_{jj} \\ &= \text{trace}(\mathsf{BA}).\end{aligned} \tag{3.48}$$

This holds even though $\mathsf{AB} \neq \mathsf{BA}$. Equation (3.48) means that the trace of any commutator $[\mathsf{A}, \mathsf{B}] = \mathsf{AB} - \mathsf{BA}$ is zero. From Eq. (3.48) we obtain

$$\text{trace}(\mathsf{ABC}) = \text{trace}(\mathsf{BCA}) = \text{trace}(\mathsf{CAB}),$$

which shows that the trace is invariant under cyclic permutation of the matrices in a product.

For a real symmetric or a complex Hermitian matrix (see Section 3.4) the trace is the sum, and the determinant the product, of its eigenvalues, and both are coefficients of the characteristic polynomial. In Exercise 3.4.23 the operation of taking the trace selects one term out of a sum of 16 terms. The trace will serve a similar function relative to matrices as orthogonality serves for vectors and functions.

In terms of tensors (Section 2.7) the trace is a contraction and, like the contracted second-rank tensor, is a scalar (invariant).

Matrices are used extensively to represent the elements of groups (compare Exercise 3.2.7 and Chapter 4). The trace of the matrix representing the group element is known in group theory as the **character**. The reason for the special name and special attention is that, the trace or character remains invariant under similarity transformations (compare Exercise 3.3.9).

Matrix Inversion

At the beginning of this section matrix A is introduced as the representation of an operator that (linearly) transforms the coordinate axes. A rotation would be one example of such a linear transformation. Now we look for the inverse transformation A^{-1} that will restore the original coordinate axes. This means, as either a matrix or an operator equation,[10]

$$\mathsf{A}\mathsf{A}^{-1} = \mathsf{A}^{-1}\mathsf{A} = 1. \tag{3.49}$$

With $(\mathsf{A}^{-1})_{ij} \equiv a_{ij}^{(-1)}$,

$$a_{ij}^{(-1)} \equiv \frac{C_{ji}}{|\mathsf{A}|}, \tag{3.50}$$

[10] Here and throughout this chapter our matrices have finite rank. If A is an infinite-rank matrix ($n \times n$ with $n \to \infty$), then life is more difficult. For A^{-1} to be the inverse we must demand that both

$$\mathsf{A}\mathsf{A}^{-1} = 1 \qquad \text{and} \qquad \mathsf{A}^{-1}\mathsf{A} = 1.$$

one relation no longer implies the other.

with C_{ji} the cofactor (see discussion preceding Eq. (3.11)) of a_{ij} and the assumption that the determinant of A, $|\mathsf{A}| \neq 0$. If it is zero, we label A singular. No inverse exists.

There is a wide variety of alternative techniques. One of the best and most commonly used is the Gauss–Jordan matrix inversion technique. The theory is based on the results of Exercises 3.2.34 and 3.2.35, which show that there exist matrices M_L such that the product $\mathsf{M}_L\mathsf{A}$ will be A but with

a. one row multiplied by a constant, or

b. one row replaced by the original row minus a multiple of another row, or

c. rows interchanged.

Other matrices M_R operating on the right ($\mathsf{A}\mathsf{M}_R$) can carry out the same operations on the **columns** of A.

This means that the matrix rows and columns may be altered (by matrix multiplication) as though we were dealing with determinants, so we can apply the Gauss–Jordan elimination techniques of Section 3.1 to the matrix elements. Hence there exists a matrix M_L (or M_R) such that[11]

$$\mathsf{M}_L\mathsf{A} = 1. \tag{3.51}$$

Then $\mathsf{M}_L = \mathsf{A}^{-1}$. We determine M_L by carrying out the identical elimination operations on the unit matrix. Then

$$\mathsf{M}_L 1 = \mathsf{M}_L. \tag{3.52}$$

To clarify this, we consider a specific example.

Example 3.2.2 GAUSS–JORDAN MATRIX INVERSION

We want to invert the matrix

$$\mathsf{A} = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{pmatrix}. \tag{3.53}$$

For convenience we write A and 1 side by side and carry out the identical operations on each:

$$\begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.54}$$

To be systematic, we multiply each row to get $a_{k1} = 1$,

$$\begin{pmatrix} 1 & \frac{2}{3} & \frac{1}{3} \\ 1 & \frac{3}{2} & \frac{1}{2} \\ 1 & 1 & 4 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.55}$$

[11] Remember that $\det(\mathsf{A}) \neq 0$.

Subtracting the first row from the second and third rows, we obtain

$$\begin{pmatrix} 1 & \frac{2}{3} & \frac{1}{3} \\ 0 & \frac{5}{6} & \frac{1}{6} \\ 0 & \frac{1}{3} & \frac{11}{3} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ -\frac{1}{3} & \frac{1}{2} & 0 \\ -\frac{1}{3} & 0 & 1 \end{pmatrix}. \tag{3.56}$$

Then we divide the second row (of **both** matrices) by $\frac{5}{6}$ and subtract $\frac{2}{3}$ times it from the first row and $\frac{1}{3}$ times it from the third row. The results for both matrices are

$$\begin{pmatrix} 1 & 0 & \frac{1}{5} \\ 0 & 1 & \frac{1}{5} \\ 0 & 0 & \frac{18}{5} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \frac{3}{5} & -\frac{2}{5} & 0 \\ -\frac{2}{5} & \frac{3}{5} & 0 \\ -\frac{1}{5} & -\frac{1}{5} & 1 \end{pmatrix}. \tag{3.57}$$

We divide the third row (of **both** matrices) by $\frac{18}{5}$. Then as the last step $\frac{1}{5}$ times the third row is subtracted from each of the first two rows (of both matrices). Our final pair is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathsf{A}^{-1} = \begin{pmatrix} \frac{11}{18} & -\frac{7}{18} & -\frac{1}{18} \\ -\frac{7}{18} & \frac{11}{18} & -\frac{1}{18} \\ -\frac{1}{18} & -\frac{1}{18} & \frac{5}{18} \end{pmatrix}. \tag{3.58}$$

The check is to multiply the original A by the calculated A^{-1} to see if we really do get the unit matrix 1. ■

As with the Gauss–Jordan solution of simultaneous linear algebraic equations, this technique is well adapted to computers. Indeed, this Gauss–Jordan matrix inversion technique will probably be available in the program library as a subroutine (see Sections 2.3 and 2.4 of Press *et al.*, loc. cit.).

For **matrices of special form, the inverse matrix can be given in closed form**. For example, for

$$\mathsf{A} = \begin{pmatrix} a & b & c \\ b & d & b \\ c & b & e \end{pmatrix},$$

the inverse matrix has a similar but slightly more general form,

$$\mathsf{A}^{-1} = \begin{pmatrix} \alpha & \beta_1 & \gamma \\ \beta_1 & \delta & \beta_2 \\ \gamma & \beta_2 & \epsilon \end{pmatrix},$$

where $D = \det(\mathsf{A})$ is the determinant of the matrix A. If $e = a$ in A, then the inverse matrix A^{-1} also simplifies to

$$\beta_1 = \beta_2, \qquad \epsilon = \alpha, \qquad D = \left(a^2 - c^2\right)d + 2(c-a)b^2.$$

As a check, let us work out the 11-matrix element of the product $\mathsf{A}\mathsf{A}^{-1} = 1$. We find

$$a\alpha + b\beta_1 + c\gamma = \frac{1}{D}\big[a\big(ed - b^2\big) + b^2(c - e) - c\big(cd - b^2\big)\big]$$
$$= \frac{1}{D}\big(-ab^2 + aed + 2b^2c - b^2e - c^2d\big) = \frac{D}{D} = 1.$$

Similarly we check that the 12-matrix element vanishes,

$$a\beta_1 + b\delta + c\beta_2 = \frac{1}{D}\big[ab(c - e) + b\big(ae - c^2\big) + cb(c - a)\big] = 0,$$

and so on.

Note though that we cannot always find an inverse of A^{-1} by solving for the matrix elements $a, b, \ldots$ of A, because not every inverse matrix A^{-1} of the form in Eq. (3.60) has a corresponding A of the special form in Eq. (3.59), as Example 3.2.2 clearly shows.

Matrices are square or rectangular arrays of numbers that define linear transformations, such as rotations of a coordinate system. As such, they are linear operators. Square matrices may be inverted when their determinant is nonzero. When a matrix defines a system of linear equations, the inverse matrix solves it. Matrices with the same number of rows and columns may be added and subtracted. They form what mathematicians call a ring with a unit and a zero matrix. Matrices are also useful for representing group operations and operators in Hilbert spaces.

Exercises

3.2.1 Show that matrix multiplication is associative, $(\mathsf{A}\mathsf{B})\mathsf{C} = \mathsf{A}(\mathsf{B}\mathsf{C})$.

3.2.2 Show that

$$(\mathsf{A} + \mathsf{B})(\mathsf{A} - \mathsf{B}) = \mathsf{A}^2 - \mathsf{B}^2$$

if and only if A and B commute,

$$[\mathsf{A}, \mathsf{B}] = 0.$$

3.2.3 Show that matrix A is a **linear operator** by showing that

$$\mathsf{A}(c_1\mathbf{r}_1 + c_2\mathbf{r}_2) = c_1\mathsf{A}\mathbf{r}_1 + c_2\mathsf{A}\mathbf{r}_2.$$

It can be shown that an $n \times n$ matrix is the **most general** linear operator in an n-dimensional vector space. This means that every linear operator in this n-dimensional vector space is equivalent to a matrix.

3.2.4 (a) Complex numbers, $a + ib$, with a and b real, may be represented by (or are isomorphic with) 2×2 matrices:

$$a + ib \leftrightarrow \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Show that this matrix representation is valid for (i) addition and (ii) multiplication.

(b) Find the matrix corresponding to $(a + ib)^{-1}$.

3.2.5 If A is an $n \times n$ matrix, show that

$$\det(-\mathsf{A}) = (-1)^n \det \mathsf{A}.$$

3.2.6 (a) The matrix equation $\mathsf{A}^2 = 0$ does not imply $\mathsf{A} = 0$. Show that the most general 2×2 matrix whose square is zero may be written as

$$\begin{pmatrix} ab & b^2 \\ -a^2 & -ab \end{pmatrix},$$

where a and b are real or complex numbers.

(b) If $\mathsf{C} = \mathsf{A} + \mathsf{B}$, in general

$$\det \mathsf{C} \neq \det \mathsf{A} + \det \mathsf{B}.$$

Construct a specific numerical example to illustrate this inequality.

3.2.7 Given the three matrices

$$\mathsf{A} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \mathsf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \mathsf{C} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix},$$

find all possible products of A, B, and C, two at a time, including squares. Express your answers in terms of A, B, and C, and 1, the unit matrix. These three matrices, together with the unit matrix, form a representation of a mathematical group, the **vierergruppe** (see Chapter 4).

3.2.8 Given

$$\mathsf{K} = \begin{pmatrix} 0 & 0 & i \\ 0 & 1 & 0 \\ 0 & -1 & 0 \end{pmatrix},$$

show that

$$\mathsf{K}^n = \mathsf{K}\mathsf{K}\mathsf{K}\cdots(n \text{ factors}) = 1$$

(with the proper choice of n, $n \neq 0$).

3.2.9 Verify the **Jacobi identity**,

$$\big[\mathsf{A}, [\mathsf{B}, \mathsf{C}]\big] = \big[\mathsf{B}, [\mathsf{A}, \mathsf{C}]\big] - \big[\mathsf{C}, [\mathsf{A}, \mathsf{B}]\big].$$

This is useful in matrix descriptions of elementary particles (see Eq. (4.16)). As a mnemonic aid, the you might note that the Jacobi identity has the same form as the *BAC–CAB* rule of Section 1.5.

3.2.10 Show that the matrices

$$\mathsf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathsf{B} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathsf{C} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

satisfy the commutation relations

$$[\mathsf{A}, \mathsf{B}] = \mathsf{C}, \qquad [\mathsf{A}, \mathsf{C}] = 0, \qquad \text{and} \qquad [\mathsf{B}, \mathsf{C}] = 0.$$

3.2.11 Let

$$\mathsf{i} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \qquad \mathsf{j} = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

and

$$\mathsf{k} = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}.$$

Show that

(a) $\mathsf{i}^2 = \mathsf{j}^2 = \mathsf{k}^2 = -1$, where 1 is the unit matrix.
(b) $\mathsf{ij} = -\mathsf{ji} = \mathsf{k}$,
$\mathsf{jk} = -\mathsf{kj} = \mathsf{i}$,
$\mathsf{ki} = -\mathsf{ik} = \mathsf{j}$.

These three matrices (i, j, and k) plus the unit matrix 1 form a basis for **quaternions**. An alternate basis is provided by the four 2×2 matrices, $i\sigma_1$, $i\sigma_2$, $-i\sigma_3$, and 1, where the σ are the Pauli spin matrices of Exercise 3.2.13.

3.2.12 A matrix with elements $a_{ij} = 0$ for $j < i$ may be called upper right triangular. The elements in the lower left (below and to the left of the main diagonal) vanish. Examples are the matrices in Chapters 12 and 13, Exercise 13.1.21, relating power series and eigenfunction expansions.
Show that the product of two upper right triangular matrices is an upper right triangular matrix.

3.2.13 The three Pauli spin matrices are

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \text{and} \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Show that

(a) $(\sigma_i)^2 = 1_2$,
(b) $\sigma_j \sigma_k = i\sigma_l$, $(j, k, l) = (1, 2, 3), (2, 3, 1), (3, 1, 2)$ (cyclic permutation),
(c) $\sigma_i \sigma_j + \sigma_j \sigma_i = 2\delta_{ij} 1_2$; 1_2 is the 2×2 unit matrix.

These matrices were used by Pauli in the nonrelativistic theory of electron spin.

3.2.14 Using the Pauli σ_i of Exercise 3.2.13, show that

$$(\boldsymbol{\sigma} \cdot \mathbf{a})(\boldsymbol{\sigma} \cdot \mathbf{b}) = \mathbf{a} \cdot \mathbf{b}\, 1_2 + i\boldsymbol{\sigma} \cdot (\mathbf{a} \times \mathbf{b}).$$

Here

$$\boldsymbol{\sigma} \equiv \hat{\mathbf{x}}\sigma_1 + \hat{\mathbf{y}}\sigma_2 + \hat{\mathbf{z}}\sigma_3,$$

$\mathbf{a}$ and $\mathbf{b}$ are ordinary vectors, and 1_2 is the 2×2 unit matrix.

3.2.15 One description of spin 1 particles uses the matrices and

$$\mathsf{M}_x = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathsf{M}_y = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix},$$

$$\mathsf{M}_z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Show that

(a) $[\mathsf{M}_x, \mathsf{M}_y] = i\mathsf{M}_z$, and so on[12] (cyclic permutation of indices). Using the Levi-Civita symbol of Section 2.9, we may write

$$[\mathsf{M}_p, \mathsf{M}_q] = i\varepsilon_{pqr}\mathsf{M}_r.$$

(b) $\mathsf{M}^2 \equiv \mathsf{M}_x^2 + \mathsf{M}_y^2 + \mathsf{M}_z^2 = 2\,1_3$, where 1_3 is the 3×3 unit matrix.

(c) $[\mathsf{M}^2, \mathsf{M}_i] = 0$,

$[\mathsf{M}_z, \mathsf{L}^+] = \mathsf{L}^+$,

$[\mathsf{L}^+, \mathsf{L}^-] = 2\mathsf{M}_z$,

where

$\mathsf{L}^+ \equiv \mathsf{M}_x + i\mathsf{M}_y$,

and $\mathsf{L}^- \equiv \mathsf{M}_x - i\mathsf{M}_y$.

In Chapter 4 these matrices appear as the **generators** of the rotation group.

3.2.16 Repeat Exercise 3.2.15 using an alternate representation,

$$\mathsf{M}_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \mathsf{M}_y = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix},$$

$$\mathsf{M}_z = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Show that the matrix-vector equation

$$\left(\mathsf{M} \cdot \nabla + 1_3 \frac{\partial}{c\,\partial t}\right)\psi = 0$$

reproduces Maxwell's equations in vacuum. Here ψ is a column vector with components $\psi_j = B_j - iE_j/c$, $j = x, y, z$. M is a vector whose elements are the angular momentum **matrices** of Exercise 3.2.16. Note that $\varepsilon_0\mu_0 = 1/c^2$, 1_3 is the 3×3 unit matrix.

[12] $[\mathsf{A}, \mathsf{B}] = \mathsf{AB} - \mathsf{BA}$.

From Exercise 3.2.15(b),

$$\mathsf{M}^2\psi = 2\psi.$$

A comparison with the Dirac relativistic electron equation suggests that the "particle" of electromagnetic radiation, the photon, has zero rest mass and a spin of 1 (in units of h).

3.2.18 Repeat Exercise 3.2.15, using the matrices for a spin of $3/2$,

$$\mathsf{M}_x = \frac{1}{2}\begin{pmatrix} 0 & \sqrt{3} & 0 & 0 \\ \sqrt{3} & 0 & 2 & 0 \\ 0 & 2 & 0 & \sqrt{3} \\ 0 & 0 & \sqrt{3} & 0 \end{pmatrix}, \qquad \mathsf{M}_y = \frac{i}{2}\begin{pmatrix} 0 & -\sqrt{3} & 0 & 0 \\ \sqrt{3} & 0 & -2 & 0 \\ 0 & 2 & 0 & -\sqrt{3} \\ 0 & 0 & \sqrt{3} & 0 \end{pmatrix},$$

and

$$\mathsf{M}_z = \frac{1}{2}\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -3 \end{pmatrix}.$$

3.2.19 An operator P commutes with J_x and J_y, the x and y components of an angular momentum operator. Show that P commutes with the third component of angular momentum, that is, that

$$[\mathsf{P}, \mathsf{J}_z] = 0.$$

Hint. The angular momentum components must satisfy the commutation relation of Exercise 3.2.15(a).

3.2.20 The L^+ and L^- matrices of Exercise 3.2.15 are *ladder operators* (see Chapter 4): L^+ operating on a system of spin projection m will raise the spin projection to $m+1$ if m is below its maximum. L^+ operating on $m_{\max}$ yields zero. L^- reduces the spin projection in unit steps in a similar fashion. Dividing by $\sqrt{2}$, we have

Show that

$$\mathsf{L}^+ = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathsf{L}^- = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

$$\mathsf{L}^+|-1\rangle = |0\rangle, \mathsf{L}^-|-1\rangle = \text{null column vector},$$

$$\mathsf{L}^+|0\rangle = |1\rangle, \mathsf{L}^-|0\rangle = |-1\rangle,$$

$$\mathsf{L}^+|1\rangle = \text{null column vector}, \mathsf{L}^-|1\rangle = |0\rangle,$$

represent states of spin projection $+1$, 0, and -1, respectively.

Note. Differential operator analogs of these ladder operators appear in Exercise 12.6.7.

where

$$|-1\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \qquad |0\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad \text{and} \qquad |1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

3.2.21 Vectors **A** and **B** are related by the tensor T,

$$\mathbf{B} = \mathsf{T}\mathbf{A}.$$

Given **A** and **B**, show that there is **no unique solution** for the components of T. This is why vector division $\mathbf{B}/\mathbf{A}$ is undefined (apart from the special case of **A** and **B** parallel and T then a scalar).

3.2.22 We might ask for a vector $\mathbf{A}^{-1}$, an inverse of a given vector **A** in the sense that

$$\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{A}^{-1} \cdot \mathbf{A} = 1.$$

Show that this relation does not suffice to define $\mathbf{A}^{-1}$ uniquely; **A** would then have an infinite number of inverses.

3.2.23 If A is a diagonal matrix, with all diagonal elements different, and A and B commute, show that B is diagonal.

3.2.24 If A and B are diagonal, show that A and B commute.

3.2.25 Show that $\text{trace}(\mathsf{ABC}) = \text{trace}(\mathsf{CBA})$ if any two of the three matrices commute.

3.2.26 Angular momentum matrices satisfy a commutation relation

$$[\mathsf{M}_j, \mathsf{M}_k] = i\mathsf{M}_l, \qquad j, k, l \text{ cyclic}.$$

Show that the trace of each angular momentum matrix vanishes.

3.2.27 (a) The operator trace replaces a matrix A by its trace; that is,

$$\text{trace}(\mathsf{A}) = \sum_i a_{ii}.$$

Show that trace is a **linear** operator.

(b) The operator det replaces a matrix A by its determinant; that is,

$$\det(\mathsf{A}) = \text{determinant of } \mathsf{A}.$$

Show that det is **not** a linear operator.

3.2.28 A and B anticommute: $\mathsf{BA} = -\mathsf{AB}$. Also, $\mathsf{A}^2 = 1$, $\mathsf{B}^2 = 1$. Show that $\text{trace}(\mathsf{A}) = \text{trace}(\mathsf{B}) = 0$.
Note. The Pauli and Dirac (Section 3.4) matrices are specific examples.

3.2.29 With $|x\rangle$ an N-dimensional column vector and $\langle y|$ an N-dimensional row vector, show that

$$\text{trace}\big(|x\rangle\langle y|\big) = \langle y|x\rangle.$$

Note. $|x\rangle\langle y|$ means direct product of column vector $|x\rangle$ with row vector $\langle y|$. The result is a square $N \times N$ matrix.

3.2.30 (a) If two nonsingular matrices anticommute, show that the trace of each one is zero. (*Nonsingular* means that the determinant of the matrix nonzero.)

(b) For the conditions of part (a) to hold, A and B must be $n \times n$ matrices with n **even**. Show that if n is **odd**, a contradiction results.

3.2.31 If a matrix has an inverse, show that the inverse is unique.

3.2.32 If A^{-1} has elements

$$\left(\mathsf{A}^{-1}\right)_{ij} = a_{ij}^{(-1)} = \frac{C_{ji}}{|\mathsf{A}|},$$

where C_{ji} is the jith cofactor of $|\mathsf{A}|$, show that

$$\mathsf{A}^{-1}\mathsf{A} = 1.$$

Hence A^{-1} is the inverse of A (if $|\mathsf{A}| \neq 0$).

3.2.33 Show that $\det \mathsf{A}^{-1} = (\det \mathsf{A})^{-1}$.
Hint. Apply the product theorem of Section 3.2.
Note. If $\det \mathsf{A}$ is zero, then A has no inverse. A is singular.

3.2.34 Find the matrices M_L such that the product $\mathsf{M}_L\mathsf{A}$ will be A but with:

(a) the ith row multiplied by a constant k ($a_{ij} \to ka_{ij}$, $j = 1, 2, 3, \ldots$);
(b) the ith row replaced by the original ith row minus a multiple of the mth row ($a_{ij} \to a_{ij} - Ka_{mj}$, $i = 1, 2, 3, \ldots$);
(c) the ith and mth rows interchanged ($a_{ij} \to a_{mj}$, $a_{mj} \to a_{ij}$, $j = 1, 2, 3, \ldots$).

3.2.35 Find the matrices M_R such that the product $\mathsf{A}\mathsf{M}_R$ will be A but with:

(a) the ith column multiplied by a constant k ($a_{ji} \to ka_{ji}$, $j = 1, 2, 3, \ldots$);
(b) the ith column replaced by the original ith column minus a multiple of the mth column ($a_{ji} \to a_{ji} - ka_{jm}$, $j = 1, 2, 3, \ldots$);
(c) the ith and mth columns interchanged ($a_{ji} \to a_{jm}$, $a_{jm} \to a_{ji}$, $j = 1, 2, 3, \ldots$).

3.2.36 Find the inverse of

$$\mathsf{A} = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 4 \end{pmatrix}.$$

3.2.37 (a) Rewrite Eq. (2.4) of Chapter 2 (and the corresponding equations for dy and dz) as a single matrix equation

$$|dx_k\rangle = \mathsf{J}|dq_j\rangle.$$

J is a matrix of derivatives, the **Jacobian** matrix. Show that

$$\langle dx_k|dx_k\rangle = \langle dq_i|\mathsf{G}|dq_j\rangle,$$

with the metric (matrix) G having elements g_{ij} given by Eq. (2.6).

(b) Show that

$$\det(\mathsf{J})\, dq_1\, dq_2\, dq_3 = dx\, dy\, dz,$$

with $\det(\mathsf{J})$ the usual Jacobian.

3.2.38 Matrices are far too useful to remain the exclusive property of physicists. They may appear wherever there are linear relations. For instance, in a study of population movement the initial fraction of a fixed population in each of n areas (or industries or religions, etc.) is represented by an n-component column vector $\mathbf{P}$. The movement of people from one area to another in a given time is described by an $n \times n$ (stochastic) matrix T. Here T_{ij} is the fraction of the population in the jth area that moves to the ith area. (Those not moving are covered by $i = j$.) With $\mathbf{P}$ describing the initial population distribution, the final population distribution is given by the matrix equation $\mathsf{T}\mathbf{P} = \mathbf{Q}$.
From its definition, $\sum_{i=1}^{n} P_i = 1$.

(a) Show that conservation of people requires that

$$\sum_{i=1}^{n} T_{ij} = 1, \quad j = 1, 2, \ldots, n.$$

(b) Prove that

$$\sum_{i=1}^{n} Q_i = 1$$

continues the conservation of people.

3.2.39 Given a 6×6 matrix A with elements $a_{ij} = 0.5^{|i-j|}$, $i = 0, 1, 2, \ldots, 5$; $i = 0, 1, 2, \ldots, 5$, find A^{-1}. List its matrix elements to five decimal places.

$$\text{ANS.} \quad \mathsf{A}^{-1} = \frac{1}{3}\begin{pmatrix} 4 & -2 & 0 & 0 & 0 & 0 \\ -2 & 5 & -2 & 0 & 0 & 0 \\ 0 & -2 & 5 & -2 & 0 & 0 \\ 0 & 0 & -2 & 5 & -2 & 0 \\ 0 & 0 & 0 & -2 & 5 & -2 \\ 0 & 0 & 0 & 0 & -2 & 4 \end{pmatrix}.$$

3.2.40 Exercise 3.1.7 may be written in matrix form:

$$\mathsf{A}\mathbf{X} = \mathbf{C}.$$

Find A^{-1} and calculate $\mathbf{X}$ as $\mathsf{A}^{-1}\mathbf{C}$.

3.2.41 (a) Write a **subroutine** that will multiply **complex** matrices. Assume that the complex matrices are in a general rectangular form.
(b) Test your subroutine by multiplying pairs of the Dirac 4×4 matrices, Section 3.4.

3.2.42 (a) Write a subroutine that will call the complex matrix multiplication subroutine of Exercise 3.2.41 and will calculate the commutator bracket of two complex matrices.
(b) Test your complex commutator bracket subroutine with the matrices of Exercise 3.2.16.

3.2.43 *Interpolating polynomial* is the name given to the $(n-1)$-degree polynomial determined by (and passing through) n points, (x_i, y_i) with all the x_i distinct. This interpolating polynomial forms a basis for numerical quadratures.

(a) Show that the requirement that an $(n-1)$-degree polynomial in x pass through each of the n points (x_i, y_i) with all x_i distinct leads to n simultaneous equations of the form

$$\sum_{j=0}^{n-1} a_j x_i^j = y_i, \qquad i = 1, 2, \ldots, n.$$

(b) Write a computer program that will read in n data points and return the n coefficients a_j. Use a subroutine to solve the simultaneous equations if such a subroutine is available.

(c) Rewrite the set of simultaneous equations as a matrix equation

$$\mathsf{X}\mathbf{A} = \mathbf{Y}.$$

(d) Repeat the computer calculation of part (b), but this time solve for vector $\mathbf{A}$ by inverting matrix X (again, using a subroutine).

3.2.44 A calculation of the values of electrostatic potential inside a cylinder leads to

$$V(0.0) = 52.640 \quad V(0.6) = 25.844$$
$$V(0.2) = 48.292 \quad V(0.8) = 12.648$$
$$V(0.4) = 38.270 \quad V(1.0) = 0.0.$$

The problem is to determine the values of the argument for which $V = 10$, 20, 30, 40, and 50. Express $V(x)$ as a series $\sum_{n=0}^{5} a_{2n} x^{2n}$. (Symmetry requirements in the original problem require that $V(x)$ be an even function of x.) Determine the coefficients a_{2n}. With $V(x)$ now a known function of x, find the root of $V(x) - 10 = 0$, $0 \le x \le 1$. Repeat for $V(x) - 20$, and so on.

ANS. $a_0 = 52.640$,
$a_2 = -117.676$,
$V(0.6851) = 20$.

3.3 ORTHOGONAL MATRICES

Ordinary three-dimensional space may be described with the Cartesian coordinates (x_1, x_2, x_3). We consider a second set of Cartesian coordinates (x_1', x_2', x_3'), whose origin and handedness coincides with that of the first set but whose orientation is different (Fig. 3.1). We can say that the primed coordinate **axes** have been **rotated** relative to the initial, unprimed coordinate axes. Since this rotation is a **linear** operation, we expect a matrix equation relating the primed basis to the unprimed basis.

This section repeats portions of Chapters 1 and 2 in a slightly different context and with a different emphasis. Previously, attention was focused on the vector or tensor. In the case of the tensor, transformation properties were strongly stressed and were critical. Here emphasis is placed on the description of the coordinate rotation itself — the matrix. Transformation properties, the behavior of the matrix when the basis is changed, appear at the end of this section. Sections 3.4 and 3.5 continue with transformation properties in complex vector spaces.

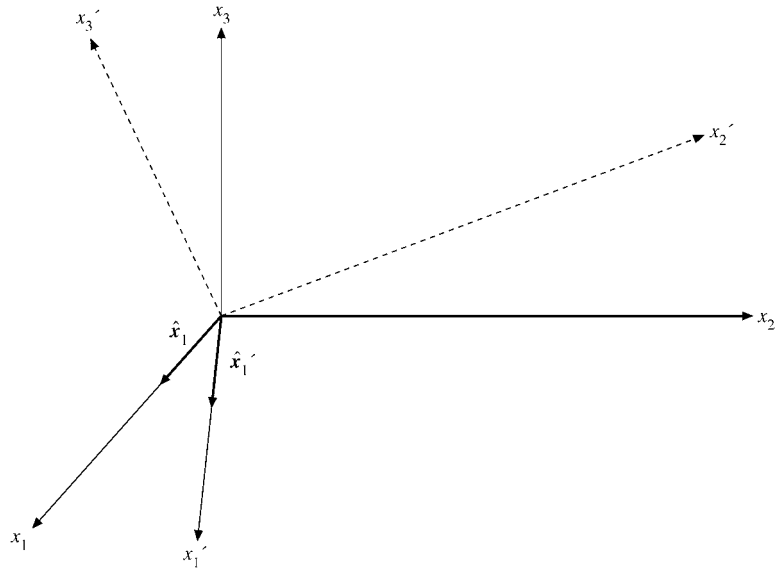

FIGURE 3.1 Cartesian coordinate systems.

Direction Cosines

A unit vector along the x_1'-axis ($\hat{\mathbf{x}}_1'$) may be resolved into components along the x_1-, x_2-, and x_3-axes by the usual projection technique:

$$\hat{\mathbf{x}}_1' = \hat{\mathbf{x}}_1 \cos(x_1', x_1) + \hat{\mathbf{x}}_2 \cos(x_1', x_2) + \hat{\mathbf{x}}_3 \cos(x_1', x_3). \tag{3.61}$$

Equation (3.61) is a specific example of the linear relations discussed at the beginning of Section 3.2.

For convenience these cosines, which are the direction cosines, are labeled

$$\begin{aligned} \cos(x_1', x_1) &= \hat{\mathbf{x}}_1' \cdot \hat{\mathbf{x}}_1 = a_{11}, \\ \cos(x_1', x_2) &= \hat{\mathbf{x}}_1' \cdot \hat{\mathbf{x}}_2 = a_{12}, \\ \cos(x_1', x_3) &= \hat{\mathbf{x}}_1' \cdot \hat{\mathbf{x}}_3 = a_{13}. \end{aligned} \tag{3.62a}$$

Continuing, we have

$$\begin{aligned} \cos(x_2', x_1) &= \hat{\mathbf{x}}_2' \cdot \hat{\mathbf{x}}_1 = a_{21}, \\ \cos(x_2', x_2) &= \hat{\mathbf{x}}_2' \cdot \hat{\mathbf{x}}_2 = a_{22}, \end{aligned} \tag{3.62b}$$

and so on, where $a_{21} \neq a_{12}$ in general. Now, Eq. (3.62) may be rewritten

$$\hat{\mathbf{x}}_1' = \hat{\mathbf{x}}_1 a_{11} + \hat{\mathbf{x}}_2 a_{12} + \hat{\mathbf{x}}_3 a_{13}, \tag{3.62c}$$

and also

$$\begin{aligned} \hat{\mathbf{x}}_2' &= \hat{\mathbf{x}}_1 a_{21} + \hat{\mathbf{x}}_2 a_{22} + \hat{\mathbf{x}}_3 a_{23}, \\ \hat{\mathbf{x}}_3' &= \hat{\mathbf{x}}_1 a_{31} + \hat{\mathbf{x}}_2 a_{32} + \hat{\mathbf{x}}_3 a_{33}. \end{aligned} \tag{3.62d}$$

Applications to Vectors

Associating $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ with the subscript 1, $\hat{\mathbf{x}}_2$ and $\hat{\mathbf{x}}_3$ with the subscript 2, $\hat{\mathbf{x}}_3$ and $\hat{\mathbf{x}}_1$ with the subscript 3, we see that in each case the first subscript is the primed unit vector ($\hat{\mathbf{x}}'_1, \hat{\mathbf{x}}'_2, \hat{\mathbf{x}}'_3$), whereas the second subscript refers to the unprimed unit vector ($\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \hat{\mathbf{x}}_3$).

We may also go the other way by resolving $\hat{\mathbf{x}}_1$, $\hat{\mathbf{x}}_2$, and $\hat{\mathbf{x}}_3$ into components in the primed system. Then

$$\begin{aligned}
\hat{\mathbf{x}}_1 &= \hat{\mathbf{x}}'_1 a_{11} + \hat{\mathbf{x}}'_2 a_{21} + \hat{\mathbf{x}}'_3 a_{31}, \\
\hat{\mathbf{x}}_2 &= \hat{\mathbf{x}}'_1 a_{12} + \hat{\mathbf{x}}'_2 a_{22} + \hat{\mathbf{x}}'_3 a_{32}, \\
\hat{\mathbf{x}}_3 &= \hat{\mathbf{x}}'_1 a_{13} + \hat{\mathbf{x}}'_2 a_{23} + \hat{\mathbf{x}}'_3 a_{33}.
\end{aligned} \tag{3.63}$$

If we consider a vector whose components are functions of the position in space, then

$$\begin{aligned}
\mathbf{V}(x_1, x_2, x_3) &= \hat{\mathbf{x}}_1 V_1 + \hat{\mathbf{x}}_2 V_2 + \hat{\mathbf{x}}_3 V_3; \\
\mathbf{V}'(x_1, x_2, x_3) &= \hat{\mathbf{x}}'_1 V'_1 + \hat{\mathbf{x}}'_2 V'_2 + \hat{\mathbf{x}}'_3 V'_3;
\end{aligned} \tag{3.64}$$

since the point may be given both by the coordinates (x_1, x_2, x_3) and by the coordinates (x'_1, x'_2, x'_3). Note that $\mathbf{V}$ and $\mathbf{V}'$ are geometrically the same vector (but with different components). The coordinate axes are being rotated; the vector stays fixed. Using Eqs. (3.62) to eliminate $\hat{\mathbf{x}}_1$, $\hat{\mathbf{x}}_2$, and $\hat{\mathbf{x}}_3$, we may separate Eq. (3.64) into three scalar equations,

$$\begin{aligned}
V'_1 &= a_{11} V_1 + a_{12} V_2 + a_{13} V_3, \\
V'_2 &= a_{21} V_1 + a_{22} V_2 + a_{23} V_3, \\
V'_3 &= a_{31} V_1 + a_{32} V_2 + a_{33} V_3.
\end{aligned} \tag{3.65}$$

In particular, these relations will hold for the coordinates of a point (x_1, x_2, x_3) and (x'_1, x'_2, x'_3), giving

$$\begin{aligned}
x'_1 &= a_{11} x_1 + a_{12} x_2 + a_{13} x_3, \\
x'_2 &= a_{21} x_1 + a_{22} x_2 + a_{23} x_3, \\
x'_3 &= a_{31} x_1 + a_{32} x_2 + a_{33} x_3,
\end{aligned} \tag{3.66}$$

and similarly for the primed coordinates. In this notation the set of three equations (3.66) may be written as

$$x'_i = \sum_{j=1}^{3} a_{ij} x_j, \tag{3.67}$$

where i takes on the values 1, 2, and 3 and the result is three **separate** equations.

Now let us set aside these results and try a different approach to the same problem. We consider two coordinate systems (x_1, x_2, x_3) and (x'_1, x'_2, x'_3) with a common origin and one point (x_1, x_2, x_3) in the unprimed system, (x'_1, x'_2, x'_3) in the primed system. Note the usual ambiguity. The same symbol x denotes both the coordinate axis and a particular

distance along that axis. Since our system is linear, x_i' must be a linear combination of the x_i. Let

$$x_i' = \sum_{j=1}^{3} a_{ij} x_j. \tag{3.68}$$

The a_{ij} may be identified as the direction cosines. This identification is carried out for the two-dimensional case later.

If we have two sets of quantities (V_1, V_2, V_3) in the unprimed system and (V_1', V_2', V_3') in the primed system, related in the same way as the coordinates of a point in the two different systems (Eq. (3.68)),

$$V_i' = \sum_{j=1}^{3} a_{ij} V_j, \tag{3.69}$$

then, as in Section 1.2, the quantities (V_1, V_2, V_3) are defined as the components of a vector that stays fixed while the coordinates rotate; that is, a vector is defined in terms of transformation properties of its components under a rotation of the coordinate axes. In a sense the coordinates of a point have been taken as a prototype vector. The power and usefulness of this definition became apparent in Chapter 2, in which it was extended to define pseudovectors and tensors.

From Eq. (3.67) we can derive interesting information about the a_{ij} that describe the orientation of coordinate system (x_1', x_2', x_3') relative to the system (x_1, x_2, x_3). The length from the origin to the point is the same in both systems. Squaring, for convenience,[13]

$$\begin{aligned} \sum_i x_i^2 = \sum_i x_i'^2 &= \sum_i \left(\sum_j a_{ij} x_j \right) \left(\sum_k a_{ik} x_k \right) \\ &= \sum_{j,k} x_j x_k \sum_i a_{ij} a_{ik}. \end{aligned} \tag{3.70}$$

This can be true for all points if and only if

$$\sum_i a_{ij} a_{ik} = \delta_{jk}, \qquad j, k = 1, 2, 3. \tag{3.71}$$

Note that Eq. (3.71) is equivalent to the matrix equation (3.83); see also Eqs. (3.87a) to (3.87d).

Verification of Eq. (3.71), if needed, may be obtained by returning to Eq. (3.70) and setting $\mathbf{r} = (x_1, x_2, x_3) = (1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 0)$, and so on to evaluate the nine relations given by Eq. (3.71). This process is valid, since Eq. (3.70) must hold for all $\mathbf{r}$ for a given set of a_{ij}. Equation (3.71), a consequence of requiring that the length remain constant (invariant) under rotation of the coordinate system, is called the **orthogonality condition**. The a_{ij}, written as a matrix A subject to Eq. (3.71), form an orthogonal matrix, a first definition of an orthogonal matrix. Note that Eq. (3.71) is **not** matrix multiplication. Rather, it is interpreted later as a scalar product of two columns of A.

[13] Note that **two** independent indices j and k are used.

In matrix notation Eq. (3.67) becomes

$$|x'\rangle = \mathsf{A}|x\rangle. \tag{3.72}$$

Orthogonality Conditions — Two-Dimensional Case

A better understanding of the a_{ij} and the orthogonality condition may be gained by considering rotation in two dimensions in detail. (This can be thought of as a three-dimensional system with the x_1-, x_2-axes rotated about x_3.) From Fig. 3.2,

$$\begin{aligned} x_1' &= x_1 \cos\varphi + x_2 \sin\varphi, \\ x_2' &= -x_1 \sin\varphi + x_2 \cos\varphi. \end{aligned} \tag{3.73}$$

Therefore by Eq. (3.72)

$$\mathsf{A} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}. \tag{3.74}$$

Notice that A reduces to the unit matrix for $\varphi = 0$. Zero angle rotation means nothing has changed. It is clear from Fig. 3.2 that

$$\begin{aligned} a_{11} &= \cos\varphi = \cos(x_1', x_1), \\ a_{12} &= \sin\varphi = \cos\left(\tfrac{\pi}{2} - \varphi\right) = \cos(x_1', x_2), \end{aligned} \tag{3.75}$$

and so on, thus identifying the matrix elements a_{ij} as the direction cosines. Equation (3.71), the orthogonality condition, becomes

$$\begin{aligned} \sin^2\varphi + \cos^2\varphi &= 1, \\ \sin\varphi\cos\varphi - \sin\varphi\cos\varphi &= 0. \end{aligned} \tag{3.76}$$

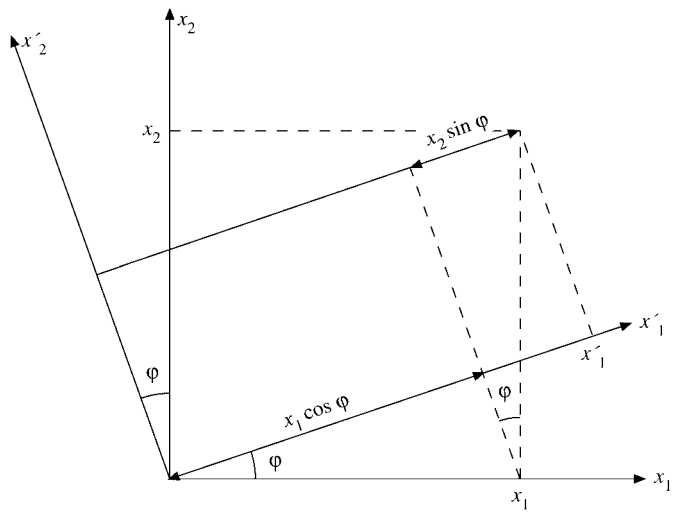

FIGURE 3.2 Rotation of coordinates.

$$\mathsf{A} = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.77}$$

The extension to three dimensions (rotation of the coordinates through an angle φ counterclockwise about x_3) is simply

The $a_{33} = 1$ expresses the fact that $x_3' = x_3$, since the rotation has been about the x_3-axis. The zeros guarantee that x_1' and x_2' do not depend on x_3 and that x_3' does not depend on x_1 and x_2.

Inverse Matrix, A^{-1}

Returning to the general transformation matrix A, the inverse matrix A^{-1} is defined such that

$$|x\rangle = \mathsf{A}^{-1}|x'\rangle. \tag{3.78}$$

That is, A^{-1} describes the reverse of the rotation given by A and returns the coordinate system to its original position. Symbolically, Eqs. (3.72) and (3.78) combine to give

$$|x\rangle = \mathsf{A}^{-1}\mathsf{A}|x\rangle, \tag{3.79}$$

and since $|x\rangle$ is arbitrary,

$$\mathsf{A}^{-1}\mathsf{A} = 1, \tag{3.80}$$

the unit matrix. Similarly,

$$\mathsf{A}\mathsf{A}^{-1} = 1, \tag{3.81}$$

using Eqs. (3.72) and (3.78) and eliminating $|x\rangle$ instead of $|x'\rangle$.

Transpose Matrix, $\tilde{\mathsf{A}}$

We can determine the elements of our postulated inverse matrix A^{-1} by employing the orthogonality condition. Equation (3.71), the orthogonality condition, does not conform to our definition of matrix multiplication, but it can be put in the required form by **defining** a new matrix $\tilde{\mathsf{A}}$ such that

$$\boxed{\tilde{a}_{ji} = a_{ij}.} \tag{3.82}$$

Equation (3.71) becomes

$$\boxed{\tilde{\mathsf{A}}\mathsf{A} = 1.} \tag{3.83}$$

This is a restatement of the orthogonality condition and may be taken as the constraint defining an orthogonal matrix, a second definition of an orthogonal matrix. Multiplying Eq. (3.83) by A^{-1} from the right and using Eq. (3.81), we have

$$\boxed{\tilde{\mathsf{A}} = \mathsf{A}^{-1},} \tag{3.84}$$

a third definition of an orthogonal matrix. This important result, that the inverse equals the transpose, holds only for orthogonal matrices and indeed may be taken as a further restatement of the orthogonality condition.

Multiplying Eq. (3.84) by A from the left, we obtain

$$\mathsf{A}\tilde{\mathsf{A}} = 1 \tag{3.85}$$

or

$$\boxed{\sum_i a_{ji}a_{ki} = \delta_{jk},} \tag{3.86}$$

which is still another form of the orthogonality condition.

Summarizing, the orthogonality condition may be stated in several equivalent ways:

$$\sum_i a_{ij}a_{ik} = \delta_{jk}, \tag{3.87a}$$

$$\sum_i a_{ji}a_{ki} = \delta_{jk}, \tag{3.87b}$$

$$\tilde{\mathsf{A}}\mathsf{A} = \mathsf{A}\tilde{\mathsf{A}} = 1, \tag{3.87c}$$

$$\tilde{\mathsf{A}} = \mathsf{A}^{-1}. \tag{3.87d}$$

Any one of these relations is a necessary and a sufficient condition for A to be orthogonal.

It is now possible to see and understand why the term **orthogonal** is appropriate for these matrices. We have the general form

$$\mathsf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \tag{3.88}$$

a matrix of direction cosines in which a_{ij} is the cosine of the angle between x_i' and x_j. Therefore a_{11}, a_{12}, a_{13} are the direction cosines of x_1' relative to x_1, x_2, x_3. These three elements of A **define** a unit length along x_1', that is, a unit vector $\hat{\mathbf{x}}_1'$,

$$\hat{\mathbf{x}}_1' = \hat{\mathbf{x}}_1 a_{11} + \hat{\mathbf{x}}_2 a_{12} + \hat{\mathbf{x}}_3 a_{13}.$$

The orthogonality relation (Eq. (3.86)) is simply a statement that the unit vectors $\hat{\mathbf{x}}_1'$, $\hat{\mathbf{x}}_2'$, and $\hat{\mathbf{x}}_3'$ are mutually perpendicular, or orthogonal. Our orthogonal transformation matrix A transforms one orthogonal coordinate system into a second orthogonal coordinate system by rotation and/or reflection.

As an example of the use of matrices, the unit vectors in spherical polar coordinates may be written as

$$\begin{pmatrix} \hat{\mathbf{r}} \\ \hat{\boldsymbol{\theta}} \\ \hat{\boldsymbol{\varphi}} \end{pmatrix} = \mathsf{C} \begin{pmatrix} \hat{\mathbf{x}} \\ \hat{\mathbf{y}} \\ \hat{\mathbf{z}} \end{pmatrix},$$

Our transformation matrix A contains nine direction cosines. Clearly, only three of these are independent, Eq. (3.71) providing six constraints. Equivalently, we may say that two parameters (θ and φ in spherical polar coordinates) are required to fix the axis of rotation. Then one additional parameter describes the amount of rotation about the specified axis, replaced by $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\varphi}}$. From the preceding analysis C is orthogonal. Therefore the inverse relation becomes

$$\begin{pmatrix} \hat{\mathbf{x}} \\ \hat{\mathbf{y}} \\ \hat{\mathbf{z}} \end{pmatrix} = \mathsf{C}^{-1} \begin{pmatrix} \hat{\mathbf{r}} \\ \hat{\boldsymbol{\theta}} \\ \hat{\boldsymbol{\varphi}} \end{pmatrix}$$

where C is given in Exercise 2.5.1. This is equivalent to Eqs. (3.62) with $\mathbf{x}'$, $\mathbf{x}'$ and $\mathbf{x}'$ relay in becomes

(In the Lagrangian formulation of mechanics (Section 17.3) it is necessary to describe the A by C and C^{-1}.)

The goal is to describe the orientation of a final rotated system (x_1''', x_2''', x_3''') relative to some initial coordinate system (x_1, x_2, x_3). The final system is developed in three steps, with each step involving one rotation described by one Euler angle (Fig. 3.3):

1. The coordinates are rotated about the x_3-axis through an angle α counterclockwise into new axes denoted by x_1'-, x_2'-, x_3'. (The x_3- and x_3'-axes coincide.)

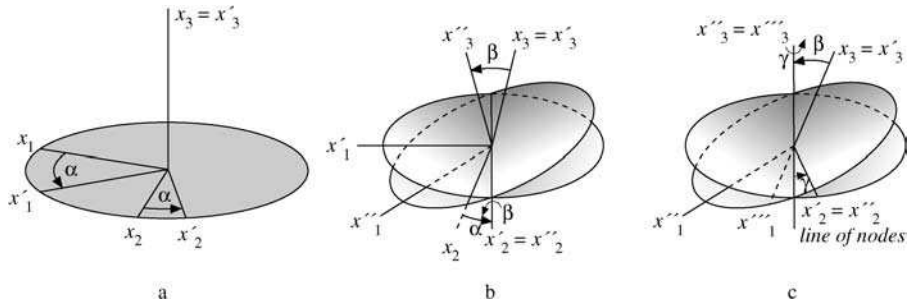

FIGURE 3.3 (a) Rotation about x_3 through angle α; (b) rotation about x_2' through angle β; (c) rotation about x_3'' through angle γ.

[14]There are almost as many definitions of the Euler angles as there are authors. Here we follow the choice generally made by workers in the area of group theory and the quantum theory of angular momentum (compare Sections 4.3, 4.4).

2. The coordinates are rotated about the x_2'-axis[15] through an angle β counterclockwise into new axes denoted by x_1''-, x_2''-, x_3''. (The x_2'- and x_2''-axes coincide.)
3. The third and final rotation is through an angle γ counterclockwise about the x_3''-axis, yielding the x_1''', x_2''', x_3''' system. (The x_3''- and x_3'''-axes coincide.)

The three matrices describing these rotations are

$$\mathsf{R}_z(\alpha) = \begin{pmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{3.90}$$

exactly like Eq. (3.77),

$$\mathsf{R}_y(\beta) = \begin{pmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{pmatrix}, \tag{3.91}$$

and

$$\mathsf{R}_z(\gamma) = \begin{pmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.92}$$

The total rotation is described by the triple matrix product,

$$\mathsf{R}(\alpha,\beta,\gamma) = \mathsf{R}_z(\gamma)\mathsf{R}_y(\beta)\mathsf{R}_z(\alpha). \tag{3.93}$$

Note the order: $\mathsf{R}_z(\alpha)$ operates first, then $\mathsf{R}_y(\beta)$, and finally $\mathsf{R}_z(\gamma)$. Direct multiplication gives

$$\mathsf{A}(\alpha,\beta,\gamma) = \begin{pmatrix} \cos\gamma\cos\beta\cos\alpha - \sin\gamma\sin\alpha & \cos\gamma\cos\beta\sin\alpha + \sin\gamma\cos\alpha & -\cos\gamma\sin\beta \\ -\sin\gamma\cos\beta\cos\alpha - \cos\gamma\sin\alpha & -\sin\gamma\cos\beta\sin\alpha + \cos\gamma\cos\alpha & \sin\gamma\sin\beta \\ \sin\beta\cos\alpha & \sin\beta\sin\alpha & \cos\beta \end{pmatrix}. \tag{3.94}$$

Equating $\mathsf{A}(a_{ij})$ with $\mathsf{A}(\alpha,\beta,\gamma)$, element by element, yields the direction cosines in terms of the three Euler angles. We could use this Euler angle identification to verify the direction cosine identities, Eq. (1.46) of Section 1.4, but the approach of Exercise 3.3.3 is much more elegant.

Symmetry Properties

Our matrix description leads to the rotation group $\mathsf{SO}(3)$ in three-dimensional space $\mathbb{R}^3$, and the Euler angle description of rotations forms a basis for developing the rotation group in Chapter 4. Rotations may also be described by the unitary group $\mathsf{SU}(2)$ in two-dimensional space $\mathbb{C}^2$ over the complex numbers. The concept of groups such as $\mathsf{SU}(2)$ and its generalizations and group theoretical techniques are often encountered in modern

[15] Some authors choose this second rotation to be about the x_1'-axis.

particle physics, where symmetry properties play an important role. The SU(2) group is also considered in Chapter 4. The power and flexibility of matrices pushed quaternions into obscurity early in the 20th century.[16]

It will be noted that matrices have been handled in two ways in the foregoing discussion: by their components and as single entities. Each technique has its own advantages and both are useful.

The transpose matrix is useful in a discussion of symmetry properties. If

$$\mathsf{A} = \tilde{\mathsf{A}}, \qquad a_{ij} = a_{ji}, \tag{3.95}$$

the matrix is called **symmetric**, whereas if

$$\mathsf{A} = -\tilde{\mathsf{A}}, \qquad a_{ij} = -a_{ji}, \tag{3.96}$$

it is called **antisymmetric** or **skewsymmetric**. The diagonal elements vanish. It is easy to show that any (square) matrix may be written as the sum of a symmetric matrix and an antisymmetric matrix. Consider the identity

$$\mathsf{A} = \tfrac{1}{2}[\mathsf{A} + \tilde{\mathsf{A}}] + \tfrac{1}{2}[\mathsf{A} - \tilde{\mathsf{A}}]. \tag{3.97}$$

$[\mathsf{A} + \tilde{\mathsf{A}}]$ is **clearly symmetric**, whereas $[\mathsf{A} - \tilde{\mathsf{A}}]$ is **clearly antisymmetric**. This is the matrix analog of Eq. (2.75), Chapter 2, for tensors. Similarly, a function may be broken up into its even and odd parts.

So far we have interpreted the orthogonal matrix as rotating the coordinate system. This changes the components of a fixed vector (not rotating with the coordinates) (Fig. 1.6, Chapter 1). However, an orthogonal matrix A may be interpreted equally well as a rotation of the **vector** in the **opposite** direction (Fig. 3.4).

These two possibilities, (1) rotating the vector keeping the coordinates fixed and (2) rotating the coordinates (in the opposite sense) keeping the vector fixed, have a direct analogy in quantum theory. Rotation (a time transformation) of the state vector gives the Schrödinger picture. Rotation of the basis keeping the state vector fixed yields the Heisenberg picture.

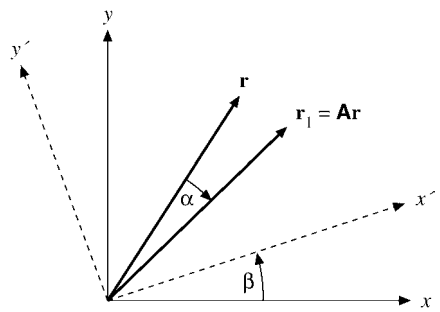

FIGURE 3.4 Fixed coordinates—rotated vector.

[16] R. J. Stephenson, Development of vector analysis from quaternions. *Am. J. Phys.* **34**: 194 (1966).

Suppose we interpret matrix A as rotating a **vector r** into the position shown by $\mathbf{r}_1$; that is, in a particular coordinate system we have a relation

$$\mathbf{r}_1 = \mathsf{A}\mathbf{r}. \tag{3.98}$$

Now let us rotate the **coordinates** by applying matrix B, which rotates (x, y, z) into (x', y', z'),

$$\begin{aligned}\mathbf{r}_1' &= \mathsf{B}\mathbf{r}_1 = \mathsf{BA}\mathbf{r} = (\mathsf{A}\mathbf{r})' = \mathsf{BA}\big(\mathsf{B}^{-1}\mathsf{B}\big)\mathbf{r} \\ &= \big(\mathsf{BAB}^{-1}\big)\mathsf{B}\mathbf{r} = \big(\mathsf{BAB}^{-1}\big)\mathbf{r}'.\end{aligned} \tag{3.99}$$

$\mathsf{B}\mathbf{r}_1$ is just $\mathbf{r}_1$ in the new coordinate system, with a similar interpretation holding for $\mathsf{B}\mathbf{r}$. Hence **in this new system** $(\mathsf{B}\mathbf{r})$ is rotated into position $(\mathsf{B}\mathbf{r}_1)$ by the matrix BAB^{-1}:

$$\begin{array}{ccccc} \mathsf{B}\mathbf{r}_1 & = & (\mathsf{BAB}^{-1}) & & \mathsf{B}\mathbf{r} \\ \downarrow & & \downarrow & & \downarrow \\ \mathbf{r}_1' & = & \mathsf{A}' & & \mathbf{r}' \end{array}$$

In the new system the coordinates have been rotated by matrix B; A has the form A', in which

$$\mathsf{A}' = \mathsf{BAB}^{-1}. \tag{3.100}$$

A' operates in the x', y', z' space as A operates in the x, y, z space.

The transformation defined by Eq. (3.100) with B any matrix, not necessarily orthogonal, is known as a **similarity transformation**. In component form Eq. (3.100) becomes

$$a_{ij}' = \sum_{k,l} b_{ik} a_{kl} \big(\mathsf{B}^{-1}\big)_{lj}. \tag{3.101}$$

Now, if B is orthogonal,

$$\big(\mathsf{B}^{-1}\big)_{lj} = (\tilde{\mathsf{B}})_{lj} = b_{jl}, \tag{3.102}$$

and we have

$$a_{ij}' = \sum_{k,l} b_{ik} b_{jl} a_{kl}. \tag{3.103}$$

It may be helpful to think of A again as an operator, possibly as rotating coordinate axes, relating angular momentum and angular velocity of a rotating solid (Section 3.5). Matrix A is the representation in a given coordinate system — or basis. But there are directions associated with A — crystal axes, symmetry axes in the rotating solid, and so on — so that the representation A depends on the basis. The similarity transformation shows just how the representation changes with a change of basis.

Relation to Tensors

Comparing Eq. (3.103) with the equations of Section 2.6, we see that it is the definition of a tensor of second rank. Hence a matrix that transforms by an **orthogonal** similarity transformation is, by definition, a tensor. Clearly, then, any **orthogonal** matrix A, interpreted as rotating a vector (Eq. (3.98)), may be called a tensor. If, however, we consider the orthogonal matrix as a collection of fixed direction cosines, giving the new orientation of a coordinate system, there is no tensor property involved.

The symmetry and antisymmetry properties defined earlier are preserved under **orthogonal** similarity transformations. Let A be a symmetric matrix, $\mathsf{A} = \tilde{\mathsf{A}}$, and

$$\mathsf{A}' = \mathsf{B}\mathsf{A}\mathsf{B}^{-1}. \tag{3.104}$$

Now,

$$\tilde{\mathsf{A}}' = \tilde{\mathsf{B}}^{-1}\tilde{\mathsf{A}}\tilde{\mathsf{B}} = \mathsf{B}\tilde{\mathsf{A}}\mathsf{B}^{-1}, \tag{3.105}$$

since B is orthogonal. But $\mathsf{A} = \tilde{\mathsf{A}}$. Therefore

$$\tilde{\mathsf{A}}' = \mathsf{B}\mathsf{A}\mathsf{B}^{-1} = \mathsf{A}', \tag{3.106}$$

showing that the property of symmetry is invariant under an orthogonal similarity transformation. In general, symmetry is **not** preserved under a nonorthogonal similarity transformation.

Exercises

Note. Assume all matrix elements are real.

3.3.1 Show that the product of two orthogonal matrices is orthogonal.
Note. This is a key step in showing that all $n \times n$ orthogonal matrices form a group (Section 4.1).

3.3.2 If A is orthogonal, show that its determinant $= \pm 1$.

3.3.3 If A is orthogonal and $\det \mathsf{A} = +1$, show that $(\det \mathsf{A})a_{ij} = C_{ij}$, where C_{ij} is the **cofactor** of a_{ij}. This yields the identities of Eq. (1.46), used in Section 1.4 to show that a cross product of vectors (in three-space) is itself a vector.
Hint. Note Exercise 3.2.32.

3.3.4 Another set of Euler rotations in common use is

(1) a rotation about the x_3-axis through an angle φ, counterclockwise,
(2) a rotation about the x_1'-axis through an angle θ, counterclockwise,
(3) a rotation about the x_3''-axis through an angle ψ, counterclockwise.

If

$$\begin{aligned} \alpha &= \varphi - \pi/2 \qquad & \varphi &= \alpha + \pi/2 \\ \beta &= \theta & \theta &= \beta \\ \gamma &= \psi + \pi/2 & \psi &= \gamma - \pi/2, \end{aligned}$$

show that the final systems are identical.

3.3.5 Suppose the Earth is moved (rotated) so that the north pole goes to 30° north, 20° west (original latitude and longitude system) and the 10° west meridian points due south.

(a) What are the Euler angles describing this rotation?
(b) Find the corresponding direction cosines.

ANS. (b) $\mathsf{A} = \begin{pmatrix} 0.9551 & -0.2552 & -0.1504 \\ 0.0052 & 0.5221 & -0.8529 \\ 0.2962 & 0.8138 & 0.5000 \end{pmatrix}$

3.3.6 Verify that the Euler angle rotation matrix, Eq. (3.94), is invariant under the transformation

$$\alpha \to \alpha + \pi, \qquad \beta \to -\beta, \qquad \gamma \to \gamma - \pi.$$

3.3.7 Show that the Euler angle rotation matrix $\mathsf{A}(\alpha, \beta, \gamma)$ satisfies the following relations:

(a) $\mathsf{A}^{-1}(\alpha, \beta, \gamma) = \tilde{\mathsf{A}}(\alpha, \beta, \gamma)$,
(b) $\mathsf{A}^{-1}(\alpha, \beta, \gamma) = \mathsf{A}(-\gamma, -\beta, -\alpha)$.

3.3.8 Show that the trace of the product of a symmetric and an antisymmetric matrix is zero.

3.3.9 Show that the trace of a matrix remains invariant under similarity transformations.

3.3.10 Show that the determinant of a matrix remains invariant under similarity transformations.
Note. Exercises (3.3.9) and (3.3.10) show that the trace and the determinant are independent of the Cartesian coordinates. They are characteristics of the matrix (operator) itself.

3.3.11 Show that the property of antisymmetry is invariant under orthogonal similarity transformations.

3.3.12 A is 2×2 and orthogonal. Find the most general form of

$$\mathsf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Compare with two-dimensional rotation.

3.3.13 $|\mathbf{x}\rangle$ and $|\mathbf{y}\rangle$ are column vectors. Under an orthogonal transformation S, $|\mathbf{x}'\rangle = \mathsf{S}|\mathbf{x}\rangle$, $|\mathbf{y}'\rangle = \mathsf{S}|\mathbf{y}\rangle$. Show that the scalar product $\langle \mathbf{x} \mid \mathbf{y} \rangle$ is invariant under this orthogonal transformation.
Note. This is equivalent to the invariance of the dot product of two vectors, Section 1.3.

3.3.14 Show that the sum of the squares of the elements of a matrix remains invariant under orthogonal similarity transformations.

3.3.15 As a generalization of Exercise 3.3.14, show that

$$\sum_{jk} S_{jk} T_{jk} = \sum_{l,m} S'_{lm} T'_{lm},$$

where the primed and unprimed elements are related by an orthogonal similarity transformation. This result is useful in deriving invariants in electromagnetic theory (compare Section 4.6).
Note. This product $M_{jk} = \sum S_{jk} T_{jk}$ is sometimes called a *Hadamard product*. In the framework of tensor analysis, Chapter 2, this exercise becomes a double contraction of two second-rank tensors and therefore is clearly a scalar (invariant).

3.3.16 A rotation $\varphi_1 + \varphi_2$ about the z-axis is carried out as two successive rotations φ_1 and φ_2, each about the z-axis. Use the matrix representation of the rotations to derive the trigonometric identities

$$\cos(\varphi_1 + \varphi_2) = \cos\varphi_1 \cos\varphi_2 - \sin\varphi_1 \sin\varphi_2,$$
$$\sin(\varphi_1 + \varphi_2) = \sin\varphi_1 \cos\varphi_2 + \cos\varphi_1 \sin\varphi_2.$$

3.3.17 A column vector $\mathbf{V}$ has components V_1 and V_2 in an initial (unprimed) system. Calculate V_1' and V_2' for a

(a) rotation of the coordinates through an angle of θ **counterclockwise**,
(b) rotation of the vector through an angle of θ **clockwise**.

The results for parts (a) and (b) should be identical.

3.3.18 Write a subroutine that will test whether a real $N \times N$ matrix is symmetric. Symmetry may be defined as

$$0 \le |a_{ij} - a_{ji}| \le \varepsilon,$$

where ε is some small tolerance (which allows for truncation error, and so on in the computer).

3.4 HERMITIAN MATRICES, UNITARY MATRICES

Definitions

Thus far it has generally been assumed that our linear vector space is a real space and that the matrix elements (the representations of the linear operators) are real. For many calculations in classical physics, real matrix elements will suffice. However, in quantum mechanics complex variables are unavoidable because of the form of the basic commutation relations (or the form of the time-dependent Schrödinger equation). With this in mind, we generalize to the case of complex matrix elements. To handle these elements, let us define, or label, some new properties.

1. Complex conjugate, A^*, formed by taking the complex conjugate ($i \to -i$) of each element, where $i = \sqrt{-1}$.
2. Adjoint, $\mathsf{A}^\dagger$, formed by transposing A^*,

$$\boxed{\mathsf{A}^\dagger = \widetilde{\mathsf{A}^*} = \tilde{\mathsf{A}}^*.} \tag{3.107}$$

3. Hermitian matrix: The matrix A is labeled **Hermitian** (or **self-adjoint**) if

$$\mathsf{A} = \mathsf{A}^\dagger. \tag{3.108}$$

If A is real, then $\mathsf{A}^\dagger = \tilde{\mathsf{A}}$ and real Hermitian matrices are real symmetric matrices. In quantum mechanics (or matrix mechanics) matrices are usually constructed to be Hermitian, or unitary.

4. Unitary matrix: Matrix U is labeled **unitary** if

$$\boxed{\mathsf{U}^\dagger = \mathsf{U}^{-1}.} \tag{3.109}$$

If U is real, then $\mathsf{U}^{-1} = \tilde{\mathsf{U}}$, so real unitary matrices are orthogonal matrices. This represents a generalization of the concept of orthogonal matrix (compare Eq. (3.84)).

5. $(\mathsf{AB})^* = \mathsf{A}^*\mathsf{B}^*$, $(\mathsf{AB})^\dagger = \mathsf{B}^\dagger\mathsf{A}^\dagger$.

If the matrix elements are complex, the physicist is almost always concerned with Hermitian and unitary matrices. Unitary matrices are especially important in quantum mechanics because they leave the length of a (complex) vector unchanged — analogous to the operation of an orthogonal matrix on a real vector. It is for this reason that the S matrix of scattering theory is a unitary matrix. One important exception to this interest in unitary matrices is the group of Lorentz matrices, Chapter 4. Using Minkowski space, we see that these matrices are not unitary.

In a complex n-dimensional linear space the square of the length of a point $\tilde{x} = x^T(x_1, x_2, \ldots, x_n)$, or the square of its distance from the origin 0, is defined as $x^\dagger x = \sum x_i^* x_i = \sum |x_i|^2$. If a coordinate transformation $y = \mathsf{U}x$ leaves the distance unchanged, then $x^\dagger x = y^\dagger y = (\mathsf{U}x)^\dagger \mathsf{U}x = x^\dagger \mathsf{U}^\dagger \mathsf{U}x$. Since x is arbitrary it follows that $\mathsf{U}^\dagger\mathsf{U} = 1_n$; that is, U is a unitary $n \times n$ matrix. If $x' = \mathsf{A}x$ is a linear map, then its matrix in the new coordinates becomes the unitary (analog of a similarity) transformation

$$\mathsf{A}' = \mathsf{U}\mathsf{A}\mathsf{U}^\dagger, \tag{3.110}$$

because $\mathsf{U}x' = y' = \mathsf{U}\mathsf{A}x = \mathsf{U}\mathsf{A}\mathsf{U}^{-1}y = \mathsf{U}\mathsf{A}\mathsf{U}^\dagger y$.

Pauli and Dirac Matrices

The set of three 2×2 Pauli matrices $\boldsymbol{\sigma}$,

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{3.111}$$

were introduced by W. Pauli to describe a particle of spin $1/2$ in nonrelativistic quantum mechanics. It can readily be shown that (compare Exercises 3.2.13 and 3.2.14) the Pauli $\boldsymbol{\sigma}$ satisfy

$$\sigma_i\sigma_j + \sigma_j\sigma_i = 2\delta_{ij}1_2, \qquad \text{anticommutation} \tag{3.112}$$

$$\sigma_i\sigma_j = i\sigma_k, \qquad i, j, k \text{ a cyclic permutation of } 1, 2, 3 \tag{3.113}$$

$$(\sigma_i)^2 = 1_2, \tag{3.114}$$

where 1_2 is the 2×2 unit matrix. Thus, the vector $\boldsymbol{\sigma}/2$ satisfies the same commutation relations,

$$[\sigma_i, \sigma_j] \equiv \sigma_i \sigma_j - \sigma_j \sigma_i = 2i\varepsilon_{ijk}\sigma_k, \tag{3.115}$$

as the orbital angular momentum $\mathbf{L}$ ($\mathbf{L} \times \mathbf{L} = i\mathbf{L}$, see Exercise 2.5.15 and the $\mathsf{SO}(3)$ and $\mathsf{SU}(2)$ groups in Chapter 4).

The three Pauli matrices $\boldsymbol{\sigma}$ and the unit matrix form a complete set, so any Hermitian 2×2 matrix M may be expanded as

$$\mathsf{M} = m_0 1_2 + m_1\sigma_1 + m_2\sigma_2 + m_3\sigma_3 = m_0 + \mathbf{m} \cdot \boldsymbol{\sigma}, \tag{3.116}$$

where the m_i form a constant vector $\mathbf{m}$. Using $(\sigma_i)^2 = 1_2$ and $\text{trace}(\sigma_i) = 0$ we obtain from Eq. (3.116) the expansion coefficients m_i by forming traces,

$$2m_0 = \text{trace}(\mathsf{M}), \qquad 2m_i = \text{trace}(\mathsf{M}\sigma_i), \qquad i = 1, 2, 3. \tag{3.117}$$

Adding and multiplying such 2×2 matrices we generate the Pauli algebra.[17] Note that $\text{trace}(\sigma_i) = 0$ for $i = 1, 2, 3$.

In 1927 P. A. M. Dirac extended this formalism to fast-moving particles of spin $\frac{1}{2}$, such as electrons (and neutrinos). To include special relativity he started from Einstein's energy, $E^2 = \mathbf{p}^2c^2 + m^2c^4$, instead of the nonrelativistic kinetic and potential energy, $E = \mathbf{p}^2/2m + V$. The key to the Dirac equation is to factorize

$$E^2 - \mathbf{p}^2c^2 = E^2 - (c\boldsymbol{\sigma} \cdot \mathbf{p})^2 = (E - c\boldsymbol{\sigma} \cdot \mathbf{p})(E + c\boldsymbol{\sigma} \cdot \mathbf{p}) = m^2c^4 \tag{3.118}$$

using the 2×2 matrix identity

$$(\boldsymbol{\sigma} \cdot \mathbf{p})^2 = \mathbf{p}^2 1_2. \tag{3.119}$$

The 2×2 unit matrix 1_2 is not written explicitly in Eq. (3.118), and Eq. (3.119) follows from Exercise 3.2.14 for $\mathbf{a} = \mathbf{b} = \mathbf{p}$. Equivalently, we can introduce two matrices γ' and γ to factorize $E^2 - \mathbf{p}^2c^2$ directly:

$$\begin{aligned} &\left[E\gamma' \otimes 1_2 - c(\gamma \otimes \boldsymbol{\sigma}) \cdot \mathbf{p}\right]^2 \\ &\quad = E^2\gamma'^2 \otimes 1_2 + c^2\gamma^2 \otimes (\boldsymbol{\sigma} \cdot \mathbf{p})^2 - Ec(\gamma'\gamma + \gamma\gamma') \otimes \boldsymbol{\sigma} \cdot \mathbf{p} \\ &\quad = E^2 - \mathbf{p}^2c^2 = m^2c^4. \end{aligned} \tag{3.119$'$}$$

For Eq. (3.119$'$) to hold, the conditions

$$\gamma'^2 = 1 = -\gamma^2, \qquad \gamma'\gamma + \gamma\gamma' = 0 \tag{3.120}$$

must be satisfied. Thus, the matrices γ' and γ anticommute, just like the three Pauli matrices; therefore they cannot be real or complex numbers. Because the conditions (3.120) can be met by 2×2 matrices, we have written direct product signs (see Example 3.2.1) in Eq. (3.119$'$) because γ', γ are multiplied by 1_2, $\boldsymbol{\sigma}$ matrices, respectively, with

$$\gamma' = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \gamma = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{3.121}$$

[17] For its geometrical significance, see W. E. Baylis, J. Huschilt, and Jiansu Wei, *Am. J. Phys.* **60**: 788 (1992).

The direct-product 4×4 matrices in Eq. (3.119′) are the four conventional Dirac γ-matrices,

$$\gamma^0 = \gamma' \otimes 1_2 = \begin{pmatrix} 1_2 & 0 \\ 0 & -1_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix},$$

$$\gamma^1 = \gamma \otimes \sigma_1 = \begin{pmatrix} 0 & \sigma_1 \\ -\sigma_1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix},$$

$$\gamma^3 = \gamma \otimes \sigma_3 = \begin{pmatrix} 0 & \sigma_3 \\ -\sigma_3 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \tag{3.122}$$

and similarly for $\gamma^2 = \gamma \otimes \sigma_2$. In vector notation $\boldsymbol{\gamma} = \gamma \otimes \boldsymbol{\sigma}$ is a vector with three components, each a 4×4 matrix, a generalization of the vector of Pauli matrices to a vector of 4×4 matrices. The four matrices γ^i are the components of the four-vector $\gamma^\mu = (\gamma^0, \gamma^1, \gamma^2, \gamma^3)$. If we recognize in Eq. (1.119′)

$$E\gamma' \otimes 1_2 - c(\gamma \otimes \boldsymbol{\sigma}) \cdot \mathbf{p} = \gamma^\mu p_\mu = \gamma \cdot p = (\gamma_0, \boldsymbol{\gamma}) \cdot (E, c\mathbf{p}) \tag{3.123}$$

as a scalar product of two four-vectors γ^μ and p^μ (see Lorentz group in Chapter 4), then Eq. (3.119′) with $p^2 = p \cdot p = E^2 - \mathbf{p}^2c^2$ may be regarded as a four-vector generalization of Eq. (3.119).

Summarizing the relativistic treatment of a spin $1/2$ *particle, it leads to* 4×4 *matrices, while the spin* $1/2$ *of a nonrelativistic particle is described by the* 2×2 *Pauli matrices* $\boldsymbol{\sigma}$.

By analogy with the Pauli algebra, we can form products of the basic γ^μ matrices and linear combinations of them and the unit matrix $1 = 1_4$, thereby generating a 16-dimensional (so-called **Clifford**[18]) algebra. A basis (with convenient Lorentz transformation properties, see Chapter 4) is given (in 2×2 matrix notation of Eq. (3.122)) by

$$1_4, \gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} 0 & 1_2 \\ 1_2 & 0 \end{pmatrix}, \qquad \gamma^\mu, \gamma^5\gamma^\mu, \sigma^{\mu\nu} = i\left(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu\right)/2. \tag{3.124}$$

The γ-matrices anticommute; that is, their symmetric combinations

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}1_4, \tag{3.125}$$

where $g^{00} = 1 = -g^{11} = -g^{22} = -g^{33}$, and $g^{\mu\nu} = 0$ for $\mu \neq \nu$, are zero or proportional to the 4×4 unit matrix 1_4, while the six antisymmetric combinations in Eq. (3.124) give new basis elements that transform like a tensor under Lorentz transformations (see Chapter 4). Any 4×4 matrix can be expanded in terms of these 16 elements, and the expansion coefficients are given by forming traces similar to the 2×2 case in Eq. (3.117) us-

[18] D. Hestenes and G. Sobczyk, *loc.cit.*; D. Hestenes, *Am. J. Phys.* **39**: 1013 (1971); and *J. Math. Phys.* **16**: 556 (1975).

ing $\text{trace}(1_4) = 4$, $\text{trace}(\gamma_5) = 0$, $\text{trace}(\gamma^\mu) = 0 = \text{trace}(\gamma_5\gamma^\mu)$, $\text{trace}(\sigma^{\mu\nu}) = 0$ for $\mu, \nu = 0, 1, 2, 3$ (see Exercise 3.4.23). In Chapter 4 we show that γ_5 is odd under parity, so $\gamma_5\gamma^\mu$ transform like an axial vector that has even parity.

The spin algebra generated by the Pauli matrices is just a matrix representation of the four-dimensional Clifford algebra, while Hestenes and coworkers (loc. cit.) have developed in their **geometric calculus** a representation-free (that is, "coordinate-free") algebra that contains complex numbers, vectors, the quaternion subalgebra, and generalized cross products as directed areas (called *bivectors*). This algebraic-geometric framework is tailored to nonrelativistic quantum mechanics, where spinors acquire geometric aspects and the Gauss and Stokes theorems appear as components of a unified theorem. Their geometric algebra corresponding to the 16-dimensional Clifford algebra of Dirac γ-matrices is the appropriate coordinate-free framework for relativistic quantum mechanics and electrodynamics.

The discussion of orthogonal matrices in Section 3.3 and unitary matrices in this section is only a beginning. Further extensions are of vital concern in "elementary" particle physics. With the Pauli and Dirac matrices, we can develop **spinor** wave functions for electrons, protons, and other (relativistic) spin $\frac{1}{2}$ particles. The coordinate system rotations lead to $\mathbf{D}^j(\alpha, \beta, \gamma)$, the rotation group usually represented by matrices in which the elements are functions of the Euler angles describing the rotation. The special unitary group $\mathsf{SU}(3)$ (composed of 3×3 unitary matrices with determinant $+1$) has been used with considerable success to describe mesons and baryons involved in the strong interactions, a gauge theory that is now called *quantum chromodynamics*. These extensions are considered further in Chapter 4.

Exercises

3.4.1 Show that

$$\det(\mathsf{A}^*) = (\det \mathsf{A})^* = \det\left(\mathsf{A}^\dagger\right).$$

3.4.2 Three angular momentum matrices satisfy the basic commutation relation

$$[\mathsf{J}_x, \mathsf{J}_y] = i\mathsf{J}_z$$

(and cyclic permutation of indices). If two of the matrices have real elements, show that the elements of the third must be pure imaginary.

3.4.3 Show that $(\mathsf{AB})^\dagger = \mathsf{B}^\dagger\mathsf{A}^\dagger$.

3.4.4 A matrix $\mathsf{C} = \mathsf{S}^\dagger\mathsf{S}$. Show that the trace is positive definite unless S is the null matrix, in which case trace $(\mathsf{C}) = 0$.

3.4.5 If A and B are Hermitian matrices, show that $(\mathsf{AB} + \mathsf{BA})$ and $i(\mathsf{AB} - \mathsf{BA})$ are also Hermitian.

3.4.6 The matrix C is **not** Hermitian. Show that then $\mathsf{C} + \mathsf{C}^\dagger$ and $i(\mathsf{C} - \mathsf{C}^\dagger)$ are Hermitian. This means that a non-Hermitian matrix may be resolved into two Hermitian parts,

$$\mathsf{C} = \frac{1}{2}\left(\mathsf{C} + \mathsf{C}^\dagger\right) + \frac{1}{2i}i\left(\mathsf{C} - \mathsf{C}^\dagger\right).$$

This decomposition of a matrix into two Hermitian matrix parts parallels the decomposition of a complex number z into $x + iy$, where $x = (z + z^*)/2$ and $y = (z - z^*)/2i$.

3.4.7 A and B are two noncommuting Hermitian matrices:

$$\mathsf{AB} - \mathsf{BA} = i\mathsf{C}.$$

Prove that C is Hermitian.

3.4.8 Show that a Hermitian matrix remains Hermitian under unitary similarity transformations.

3.4.9 Two matrices A and B are each Hermitian. Find a necessary and sufficient condition for their product AB to be Hermitian.

ANS. $[\mathsf{A}, \mathsf{B}] = 0$.

3.4.10 Show that the reciprocal (that is, inverse) of a unitary matrix is unitary.

3.4.11 A particular similarity transformation yields

$$\mathsf{A}' = \mathsf{UAU}^{-1},$$
$$\mathsf{A}'^{\dagger} = \mathsf{UA}^{\dagger}\mathsf{U}^{-1}.$$

If the adjoint relationship is preserved ($\mathsf{A}^{\dagger'} = \mathsf{A}'^{\dagger}$) and $\det \mathsf{U} = 1$, show that U must be unitary.

3.4.12 Two matrices U and H are related by

$$\mathsf{U} = e^{ia\mathsf{H}},$$

with a real. (The exponential function is defined by a Maclaurin expansion. This will be done in Section 5.6.)

(a) If H is Hermitian, show that U is unitary.
(b) If U is unitary, show that H is Hermitian. (H is independent of a.)

Note. With H the Hamiltonian,

$$\psi(x, t) = \mathsf{U}(x, t)\psi(x, 0) = \exp(-it\mathsf{H}/\hbar)\psi(x, 0)$$

is a solution of the time-dependent Schrödinger equation. $\mathsf{U}(x, t) = \exp(-it\mathsf{H}/\hbar)$ is the "evolution operator."

3.4.13 An operator $T(t + \varepsilon, t)$ describes the change in the wave function from t to $t + \varepsilon$. For ε real and small enough so that ε^2 may be neglected,

$$T(t + \varepsilon, t) = 1 - \frac{i}{\hbar}\varepsilon\mathsf{H}(t).$$

(a) If T is unitary, show that H is Hermitian.
(b) If H is Hermitian, show that T is unitary.

Note. When $\mathsf{H}(t)$ is independent of time, this relation may be put in exponential form — Exercise 3.4.12.

3.4.14 Show that an alternate form,

$$T(t+\varepsilon,t)=\frac{1-i\varepsilon \mathsf{H}(t)/2\hbar}{1+i\varepsilon \mathsf{H}(t)/2\hbar},$$

agrees with the T of part (a) of Exercise 3.4.13, neglecting ε^2, and is exactly unitary (for H Hermitian).

3.4.15 Prove that the direct product of two unitary matrices is unitary.

3.4.16 Show that γ_5 anticommutes with all four γ^μ.

3.4.17 Use the four-dimensional Levi-Civita symbol $\varepsilon_{\lambda\mu\nu\rho}$ with $\varepsilon_{0123}=-1$ (generalizing Eqs. (2.93) in Section 2.9 to four dimensions) and show that (i) $2\gamma_5\sigma_{\mu\nu}=-i\varepsilon_{\mu\nu\alpha\beta}\sigma^{\alpha\beta}$ using the summation convention of Section 2.6 and (ii) $\gamma_\lambda\gamma_\mu\gamma_\nu=g_{\lambda\mu}\gamma_\nu-g_{\lambda\nu}\gamma_\mu+g_{\mu\nu}\gamma_\lambda+i\varepsilon_{\lambda\mu\nu\rho}\gamma^\rho\gamma_5$. Define $\gamma_\mu=g_{\mu\nu}\gamma^\nu$ using $g^{\mu\nu}=g_{\mu\nu}$ to raise and lower indices.

3.4.18 Evaluate the following traces: (see Eq. (3.123) for the notation)

(i) $\text{trace}(\gamma\cdot a\gamma\cdot b)=4a\cdot b$,
(ii) $\text{trace}(\gamma\cdot a\gamma\cdot b\gamma\cdot c)=0$,
(iii) $\text{trace}(\gamma\cdot a\gamma\cdot b\gamma\cdot c\gamma\cdot d)=4(a\cdot bc\cdot d-a\cdot cb\cdot d+a\cdot db\cdot c)$,
(iv) $\text{trace}(\gamma_5\gamma\cdot a\gamma\cdot b\gamma\cdot c\gamma\cdot d)=4i\varepsilon_{\alpha\beta\mu\nu}a^\alpha b^\beta c^\mu d^\nu$.

3.4.19 Show that (i) $\gamma_\mu\gamma^\alpha\gamma^\mu=-2\gamma^\alpha$, (ii) $\gamma_\mu\gamma^\alpha\gamma^\beta\gamma^\mu=4g^{\alpha\beta}$, and (iii) $\gamma_\mu\gamma^\alpha\gamma^\beta\gamma^\nu\gamma^\mu=-2\gamma^\nu\gamma^\beta\gamma^\alpha$.

3.4.20 If $\mathsf{M}=\frac{1}{2}(1+\gamma_5)$, show that

$$\mathsf{M}^2=\mathsf{M}.$$

Note that γ_5 may be replaced by any other Dirac matrix (any Γ_i of Eq. (3.124)). If M is Hermitian, then this result, $\mathsf{M}^2=\mathsf{M}$, is the defining equation for a quantum mechanical projection operator.

3.4.21 Show that

$$\boldsymbol{\alpha}\times\boldsymbol{\alpha}=2i\boldsymbol{\sigma}\otimes 1_2,$$

where $\boldsymbol{\alpha}=\gamma_0\boldsymbol{\gamma}$ is a vector

$$\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\alpha_3).$$

Note that if $\boldsymbol{\alpha}$ is a polar vector (Section 2.4), then $\boldsymbol{\sigma}$ is an axial vector.

3.4.22 Prove that the 16 Dirac matrices form a linearly independent set.

3.4.23 If we assume that a given 4×4 matrix A (with constant elements) can be written as a linear combination of the 16 Dirac matrices

$$\mathsf{A}=\sum_{i=1}^{16}c_i\Gamma_i,$$

show that

$$c_i\sim\text{trace}(\mathsf{A}\Gamma_i).$$

3.4.24 If $C = i\gamma^2\gamma^0$ is the charge conjugation matrix, show that $C\gamma^\mu C^{-1} = -\tilde{\gamma}^\mu$, where $\tilde{}$ indicates transposition.

3.4.25 Let $x'_\mu = \Lambda^\nu_\mu x_\nu$ be a rotation by an angle θ about the 3-axis,

$$x'_0 = x_0, \qquad x'_1 = x_1\cos\theta + x_2\sin\theta,$$
$$x'_2 = -x_1\sin\theta + x_2\cos\theta, \qquad x'_3 = x_3.$$

Use $R = \exp(i\theta\sigma^{12}/2) = \cos\theta/2 + i\sigma^{12}\sin\theta/2$ (see Eq. (3.170b)) and show that the γ's transform just like the coordinates x^μ, that is, $\Lambda^\nu_\mu\gamma_\nu = R^{-1}\gamma_\mu R$. (Note that $\gamma_\mu \neq g_{\mu\nu}\gamma^\nu$ and that the γ^μ are well defined only up to a similarity transformation.) Similarly, if $x' = \Lambda x$ is a boost (pure Lorentz transformation) along the 1-axis, that is,

$$x'_0 = x_0\cosh\zeta - x_1\sinh\zeta, \qquad x'_1 = -x_0\sinh\zeta + x_1\cosh\zeta,$$
$$x'_2 = x_2, \qquad x'_3 = x_3,$$

with $\tanh\zeta = v/c$ and $B = \exp(-i\zeta\sigma^{01}/2) = \cosh\zeta/2 - i\sigma^{01}\sinh\zeta/2$ (see Eq. (3.170b)), show that $\Lambda^\nu_\mu\gamma_\nu = B\gamma_\mu B^{-1}$.

3.4.26 (a) Given $\mathbf{r}' = U\mathbf{r}$, with U a unitary matrix and $\mathbf{r}$ a (column) vector with complex elements, show that the norm (magnitude) of $\mathbf{r}$ is invariant under this operation.
(b) The matrix U transforms any column vector $\mathbf{r}$ with complex elements into $\mathbf{r}'$, leaving the magnitude invariant: $\mathbf{r}^\dagger\mathbf{r} = \mathbf{r}'^\dagger\mathbf{r}'$. Show that U is unitary.

3.4.27 Write a subroutine that will test whether a complex $n \times n$ matrix is self-adjoint. In demanding equality of matrix elements $a_{ij} = a^\dagger_{ij}$, allow some small tolerance ε to compensate for truncation error of the computer.

3.4.28 Write a subroutine that will form the adjoint of a complex $M \times N$ matrix.

3.4.29 (a) Write a subroutine that will take a complex $M \times N$ matrix A and yield the product $A^\dagger A$.
Hint. This subroutine can call the subroutines of Exercises 3.2.41 and 3.4.28.
(b) Test your subroutine by taking A to be one or more of the Dirac matrices, Eq. (3.124).

3.5 DIAGONALIZATION OF MATRICES

Moment of Inertia Matrix

In many physical problems involving real symmetric or complex Hermitian matrices it is desirable to carry out a (real) orthogonal similarity transformation or a unitary transformation (corresponding to a rotation of the coordinate system) to reduce the matrix to a diagonal form, nondiagonal elements all equal to zero. One particularly direct example of this is the moment of inertia matrix I of a rigid body. From the definition of angular momentum $\mathbf{L}$ we have

$$\mathbf{L} = I\boldsymbol{\omega}, \tag{3.126}$$

$\boldsymbol{\omega}$ being the angular velocity.[19] The inertia matrix I is found to have diagonal components

$$I_{xx} = \sum_i m_i\left(r_i^2 - x_i^2\right), \qquad \text{and so on,} \tag{3.127}$$

the subscript i referring to mass m_i located at $\mathbf{r}_i = (x_i, y_i, z_i)$. For the nondiagonal components we have

$$I_{xy} = -\sum_i m_i x_i y_i = I_{yx}. \tag{3.128}$$

By inspection, matrix I is symmetric. Also, since I appears in a physical equation of the form (3.126), which holds for all orientations of the coordinate system, it may be considered to be a tensor (quotient rule, Section 2.3).

The key now is to orient the coordinate axes (along a body-fixed frame) so that the I_{xy} and the other nondiagonal elements will vanish. As a consequence of this orientation and an indication of it, if the angular velocity is along one such realigned **principal axis**, the angular velocity and the angular momentum will be parallel. As an illustration, the stability of rotation is used by football players when they throw the ball spinning about its long principal axis.

Eigenvectors, Eigenvalues

It is instructive to consider a geometrical picture of this problem. If the inertia matrix I is multiplied from each side by a unit vector of variable direction, $\hat{\mathbf{n}} = (\alpha, \beta, \gamma)$, then in the Dirac bracket notation of Section 3.2,

$$\langle\hat{\mathbf{n}}|\mathsf{I}|\hat{\mathbf{n}}\rangle = I, \tag{3.129}$$

where I is the moment of inertia about the direction $\hat{\mathbf{n}}$ and a positive number (scalar). Carrying out the multiplication, we obtain

$$I = I_{xx}\alpha^2 + I_{yy}\beta^2 + I_{zz}\gamma^2 + 2I_{xy}\alpha\beta + 2I_{xz}\alpha\gamma + 2I_{yz}\beta\gamma, \tag{3.130}$$

a positive definite quadratic form that must be an ellipsoid (see Fig. 3.5). From analytic geometry it is known that the coordinate axes can always be rotated to coincide with the axes of our ellipsoid. In many elementary cases, especially when symmetry is present, these new axes, called the *principal axes*, can be found by inspection. We can find the axes by locating the local extrema of the ellipsoid in terms of the variable components of $\mathbf{n}$, subject to the constraint $\hat{\mathbf{n}}^2 = 1$. To deal with the constraint, we introduce a Lagrange multiplier λ (Section 17.6). Differentiating $\langle\hat{\mathbf{n}}|\mathsf{I}|\hat{\mathbf{n}}\rangle - \lambda\langle\hat{\mathbf{n}}|\hat{\mathbf{n}}\rangle$,

$$\frac{\partial}{\partial n_j}\left(\langle\hat{\mathbf{n}}|\mathsf{I}|\hat{\mathbf{n}}\rangle - \lambda\langle\hat{\mathbf{n}}|\hat{\mathbf{n}}\rangle\right) = 2\sum_k I_{jk}n_k - 2\lambda n_j = 0, \qquad j = 1, 2, 3 \tag{3.131}$$

yields the eigenvalue equations

$$\mathsf{I}|\hat{\mathbf{n}}\rangle = \lambda|\hat{\mathbf{n}}\rangle. \tag{3.132}$$

[19] The moment of inertia matrix may also be developed from the kinetic energy of a rotating body, $T = 1/2\langle\boldsymbol{\omega}|\mathsf{I}|\boldsymbol{\omega}\rangle$.

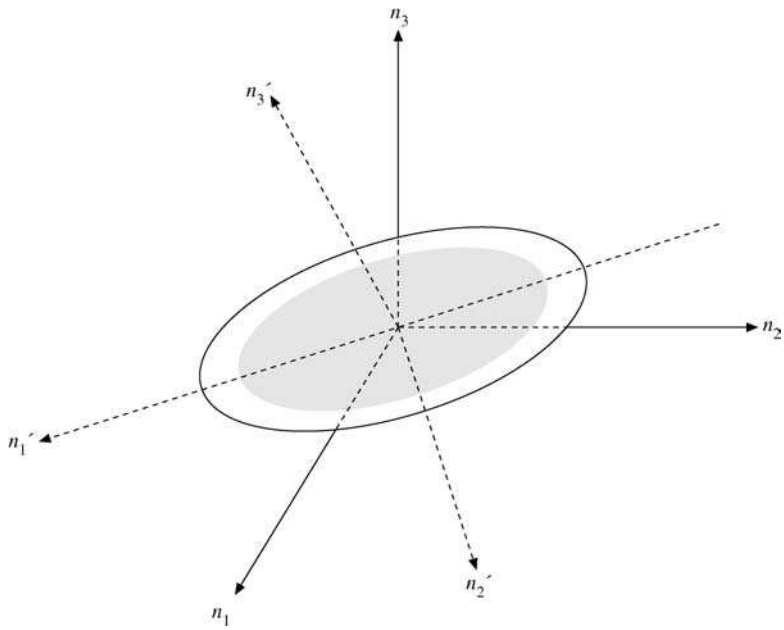

FIGURE 3.5 Moment of inertia ellipsoid.

The same result can be found by purely geometric methods. We now proceed to develop a general method of finding the diagonal elements and the principal axes.

If $\mathsf{R}^{-1} = \tilde{\mathsf{R}}$ is the real orthogonal matrix such that $\mathbf{n}' = \mathsf{R}\mathbf{n}$, or $|\mathbf{n}'\rangle = \mathsf{R}|\mathbf{n}\rangle$ in Dirac notation, are the new coordinates, then we obtain, using $\langle\mathbf{n}'|\mathsf{R} = \langle\mathbf{n}|$ in Eq. (3.132),

$$\langle\mathbf{n}|\mathsf{I}|\mathbf{n}\rangle = \langle\mathbf{n}'|\mathsf{R}\mathsf{I}\tilde{\mathsf{R}}|\mathbf{n}'\rangle = I_1'n_1'^2 + I_2'n_2'^2 + I_3'n_3'^2, \tag{3.133}$$

where the $I_i' > 0$ are the principal moments of inertia. The inertia matrix I' in Eq. (3.133) is diagonal in the new coordinates,

$$\mathsf{I}' = \mathsf{R}\mathsf{I}\tilde{\mathsf{R}} = \begin{pmatrix} I_1' & 0 & 0 \\ 0 & I_2' & 0 \\ 0 & 0 & I_3' \end{pmatrix}. \tag{3.134}$$

If we rewrite Eq. (3.134) using $\mathsf{R}^{-1} = \tilde{\mathsf{R}}$ in the form

$$\mathsf{I}\tilde{\mathsf{R}} = \tilde{\mathsf{R}}\mathsf{I}', \tag{3.135}$$

and take $\tilde{\mathsf{R}} = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ to consist of three column vectors, then Eq. (3.135) splits up into three eigenvalue equations,

$$\mathsf{I}\mathbf{v}_i = I_i'\mathbf{v}_i, \qquad i = 1, 2, 3 \tag{3.136}$$

with **eigenvalues** I_i' and **eigenvectors** $\mathbf{v}_i$. The names were introduced from the German literature on quantum mechanics. Because these equations are linear and homogeneous

$$\left| \quad \right| = 0. \tag{3.137}$$

Replacing the eigenvalue I_i' by a variable λ times the unit matrix $\mathsf{1}$, we may rewrite Eq. (3.136) as

(for fixed i), by Section 3.1 their determinants have to vanish:

$$\begin{vmatrix} I_{11} - I_i' & I_{12} & I_{13} \\ I_{12} & I_{22} - I_i' & I_{23} \\ I_{13} & I_{23} & I_{33} - I_i' \end{vmatrix}$$

$$\boxed{(\mathsf{I} - \lambda \mathsf{1})|\mathbf{v}\rangle = 0.} \tag{3.136$'$}$$

The determinant set to zero,

$$\boxed{|\mathsf{I} - \lambda \mathsf{1}| = 0,} \tag{3.137$'$}$$

is a cubic polynomial in λ; its three roots, of course, are the I_i'. Substituting one root at a time back into Eq. (3.136) (or (3.136$'$)), we can find the corresponding eigenvectors. Because of its applications in astronomical theories, Eq. (3.137) (or (3.137$'$)) is known as the **secular equation**.[20] The same treatment applies to any real symmetric matrix I, except that its eigenvalues need not all be positive. Also, the orthogonality condition in Eq. (3.87) for R say that, in geometric terms, the eigenvectors v_i are mutually orthogonal unit vectors. Indeed they form the new coordinate system. The fact that any two eigenvectors $\mathbf{v}_i, \mathbf{v}_j$ are orthogonal if $I_i' \neq I_j'$ follows from Eq. (3.136) in conjunction with the symmetry of I by multiplying with $\mathbf{v}_i$ and $\mathbf{v}_j$, respectively,

$$\langle \mathbf{v}_j | \mathsf{I} | \mathbf{v}_i \rangle = I_i' \mathbf{v}_j \cdot \mathbf{v}_i = \langle \mathbf{v}_i | \mathsf{I} | \mathbf{v}_j \rangle = I_j' \mathbf{v}_i \cdot \mathbf{v}_j. \tag{3.138a}$$

Since $I_i' \neq I_j'$ and Eq. (3.138a) implies that $(I_j' - I_i')\mathbf{v}_i \cdot \mathbf{v}_j = 0$, so $\mathbf{v}_i \cdot \mathbf{v}_j = 0$.

We can write the quadratic forms in Eq. (3.133) as a sum of squares in the original coordinates $|\mathbf{n}\rangle$,

$$\boxed{\langle \mathbf{n} | \mathsf{I} | \mathbf{n} \rangle = \langle \mathbf{n}' | \mathsf{R} \mathsf{I} \tilde{\mathsf{R}} | \mathbf{n}' \rangle = \sum_i I_i' (\mathbf{n} \cdot \mathbf{v}_i)^2,} \tag{3.138b}$$

because the rows of the rotation matrix in $\mathbf{n}' = \mathsf{R}\mathbf{n}$, i.e.,

$$\begin{pmatrix} n_1' \\ n_2' \\ n_3' \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1 \cdot \mathbf{n} \\ \mathbf{v}_2 \cdot \mathbf{n} \\ \mathbf{v}_3 \cdot \mathbf{n} \end{pmatrix}$$

componentwise, are made up of the eigenvectors $\mathbf{v}_i$. The underlying matrix identity,

$$\boxed{\mathsf{I} = \sum_i I_i' |\mathbf{v}_i\rangle\langle \mathbf{v}_i|,} \tag{3.138c}$$

[20] Equation (3.126) will take on this form when ω is along one of the principal axes. Then $\mathbf{L} = \lambda\boldsymbol{\omega}$ and $\mathsf{I}\boldsymbol{\omega} = \lambda\boldsymbol{\omega}$. In the mathematics literature λ is usually called a **characteristic value**, $\boldsymbol{\omega}$ a **characteristic vector**.

may be viewed as the **spectral decomposition** of the inertia tensor (or any real symmetric matrix). Here, the word *spectral* is just another term for expansion in terms of its eigenvalues. When we multiply this eigenvalue expansion by $\langle \mathbf{n}|$ on the left and $|\mathbf{n}\rangle$ on the right we reproduce the previous relation between quadratic forms. The operator $P_i = |\mathbf{v}_i\rangle\langle\mathbf{v}_i|$ is a projection operator satisfying $P_i^2 = P_i$ that projects the ith component w_i of any vector $|\mathbf{w}\rangle = \sum_j w_j|\mathbf{v}_j\rangle$ that is expanded in terms of the eigenvector basis $|\mathbf{v}_j\rangle$. This is verified by

$$P_i|\mathbf{w}\rangle = \sum_j w_j|\mathbf{v}_i\rangle\langle\mathbf{v}_i|\mathbf{v}_j\rangle = w_i|\mathbf{v}_i\rangle = \mathbf{v}_i \cdot \mathbf{w}|\mathbf{v}_i\rangle.$$

Finally, the identity

$$\sum_i |\mathbf{v}_i\rangle\langle\mathbf{v}_i| = 1$$

expresses the completeness of the eigenvector basis according to which any vector $|\mathbf{w}\rangle = \sum_i w_i|\mathbf{v}_i\rangle$ can be expanded in terms of the eigenvectors. Multiplying the completeness relation by $|\mathbf{w}\rangle$ proves the expansion $|\mathbf{w}\rangle = \sum_i \langle\mathbf{v}_i|\mathbf{w}\rangle|\mathbf{v}_i\rangle$.

An important extension of the spectral decomposition theorem applies to commuting symmetric (or Hermitian) matrices A, B: If $[\mathsf{A}, \mathsf{B}] = 0$, then there is an orthogonal (unitary) matrix that diagonalizes both A and B; that is, both matrices have common eigenvectors if the eigenvalues are nondegenerate. The reverse of this *theorem* is also valid.

To prove this theorem we diagonalize A: $\mathsf{A}\mathbf{v}_i = a_i\mathbf{v}_i$. Multiplying each eigenvalue equation by B we obtain $\mathsf{B}\mathsf{A}\mathbf{v}_i = a_i\mathsf{B}\mathbf{v}_i = \mathsf{A}(\mathsf{B}\mathbf{v}_i)$, which says that $\mathsf{B}\mathbf{v}_i$ is an eigenvector of A with eigenvalue a_i. Hence $\mathsf{B}\mathbf{v}_i = b_i\mathbf{v}_i$ with real b_i. Conversely, if the vectors $\mathbf{v}_i$ are common eigenvectors of A and B, then $\mathsf{A}\mathsf{B}\mathbf{v}_i = \mathsf{A}b_i\mathbf{v}_i = a_ib_i\mathbf{v}_i = \mathsf{B}\mathsf{A}\mathbf{v}_i$. Since the eigenvectors $\mathbf{v}_i$ are complete, this implies $\mathsf{A}\mathsf{B} = \mathsf{B}\mathsf{A}$.

Hermitian Matrices

For complex vector spaces, Hermitian and unitary matrices play the same role as symmetric and orthogonal matrices over real vector spaces, respectively. First, let us generalize the important theorem about the diagonal elements and the principal axes for the eigenvalue equation

$$\boxed{\mathsf{A}|\mathbf{r}\rangle = \lambda|\mathbf{r}\rangle,} \tag{3.139}$$

We now show that if A is a Hermitian matrix,[21] its eigenvalues are real and its eigenvectors orthogonal.

Let λ_i and λ_j be two eigenvalues and $|\mathbf{r}_i\rangle$ and $|\mathbf{r}_j\rangle$, the corresponding eigenvectors of A, a Hermitian matrix. Then

$$\mathsf{A}|\mathbf{r}_i\rangle = \lambda_i|\mathbf{r}_i\rangle, \tag{3.140}$$

$$\mathsf{A}|\mathbf{r}_j\rangle = \lambda_j|\mathbf{r}_j\rangle. \tag{3.141}$$

[21] If A is real, the Hermitian requirement reduces to a requirement of symmetry.

Equation (3.140) is multiplied by $\langle \mathbf{r}_j |$:

$$\langle \mathbf{r}_j | \mathsf{A} | \mathbf{r}_i \rangle = \lambda_i \langle \mathbf{r}_j | \mathbf{r}_i \rangle. \tag{3.142}$$

Equation (3.141) is multiplied by $\langle \mathbf{r}_i |$ to give

$$\langle \mathbf{r}_i | \mathsf{A} | \mathbf{r}_j \rangle = \lambda_j \langle \mathbf{r}_i | \mathbf{r}_j \rangle. \tag{3.143}$$

Taking the adjoint[22] of this equation, we have

$$\langle \mathbf{r}_j | \mathsf{A}^\dagger | \mathbf{r}_i \rangle = \lambda_j^* \langle \mathbf{r}_j | \mathbf{r}_i \rangle, \tag{3.144}$$

or

$$\langle \mathbf{r}_j | \mathsf{A} | \mathbf{r}_i \rangle = \lambda_j^* \langle \mathbf{r}_j | \mathbf{r}_i \rangle \tag{3.145}$$

since A is Hermitian. Subtracting Eq. (3.145) from Eq. (3.142), we obtain

$$(\lambda_i - \lambda_j^*) \langle \mathbf{r}_j | \mathbf{r}_i \rangle = 0. \tag{3.146}$$

This is a general result for all possible combinations of i and j. First, let $j = i$. Then Eq. (3.146) becomes

$$(\lambda_i - \lambda_i^*) \langle \mathbf{r}_i | \mathbf{r}_i \rangle = 0. \tag{3.147}$$

Since $\langle \mathbf{r}_i | \mathbf{r}_i \rangle = 0$ would be a trivial solution of Eq. (3.147), we conclude that

$$\lambda_i = \lambda_i^*, \tag{3.148}$$

or λ_i is real, for all i.

Second, for $i \neq j$ and $\lambda_i \neq \lambda_j$,

$$(\lambda_i - \lambda_j) \langle \mathbf{r}_j | \mathbf{r}_i \rangle = 0, \tag{3.149}$$

or

$$\langle \mathbf{r}_j | \mathbf{r}_i \rangle = 0, \tag{3.150}$$

which means that the eigenvectors of **distinct** eigenvalues are orthogonal, Eq. (3.150) being our generalization of orthogonality in this complex space.[23]

If $\lambda_i = \lambda_j$ (degenerate case), $|\mathbf{r}_i\rangle$ is not automatically orthogonal to $|\mathbf{r}_j\rangle$, but it may be **made** orthogonal.[24] Consider the physical problem of the moment of inertia matrix again. If x_1 is an axis of rotational symmetry, then we will find that $\lambda_2 = \lambda_3$. Eigenvectors $|\mathbf{r}_2\rangle$ and $|\mathbf{r}_3\rangle$ are each perpendicular to the symmetry axis, $|\mathbf{r}_1\rangle$, but they lie anywhere in the plane perpendicular to $|\mathbf{r}_1\rangle$; that is, any linear combination of $|\mathbf{r}_2\rangle$ and $|\mathbf{r}_3\rangle$ is also an eigenvector. Consider $(a_2|\mathbf{r}_2\rangle + a_3|\mathbf{r}_3\rangle)$ with a_2 and a_3 constants. Then

$$\begin{aligned} \mathsf{A}\big(a_2|\mathbf{r}_2\rangle + a_3|\mathbf{r}_3\rangle\big) &= a_2\lambda_2|\mathbf{r}_2\rangle + a_3\lambda_3|\mathbf{r}_3\rangle \\ &= \lambda_2\big(a_2|\mathbf{r}_2\rangle + a_3|\mathbf{r}_3\rangle\big), \end{aligned} \tag{3.151}$$

[22] Note $\langle \mathbf{r}_j | = |\mathbf{r}_j\rangle^\dagger$ for complex vectors.

[23] The corresponding theory for differential operators (Sturm–Liouville theory) appears in Section 10.2. The integral equation analog (Hilbert–Schmidt theory) is given in Section 16.4.

[24] We are assuming here that the eigenvectors of the n-fold degenerate λ_i span the corresponding n-dimensional space. This may be shown by including a parameter ε in the original matrix to remove the degeneracy and then letting ε approach zero (compare Exercise 3.5.30). This is analogous to breaking a degeneracy in atomic spectroscopy by applying an external magnetic field (Zeeman effect).

as is to be expected, for x_1 is an axis of rotational symmetry. Therefore, if $|\mathbf{r}_1\rangle$ and $|\mathbf{r}_2\rangle$ are fixed, $|\mathbf{r}_3\rangle$ may simply be chosen to lie in the plane perpendicular to $|\mathbf{r}_1\rangle$ and also perpendicular to $|\mathbf{r}_2\rangle$. A general method of orthogonalizing solutions, the Gram–Schmidt process (Section 3.1), is applied to functions in Section 10.3.

The set of n orthogonal eigenvectors $|\mathbf{r}_i\rangle$ of our $n \times n$ Hermitian matrix A forms a **complete** set, spanning the n-dimensional (complex) space, $\sum_i |\mathbf{r}_i\rangle\langle\mathbf{r}_i| = 1$. This fact is useful in a variational calculation of the eigenvalues, Section 17.8.

The spectral decomposition of any Hermitian matrix A is proved by analogy with real symmetric matrices

$$\boxed{\mathsf{A} = \sum_i \lambda_i |\mathbf{r}_i\rangle\langle\mathbf{r}_i|,}$$

with real eigenvalues λ_i and orthonormal eigenvectors $|\mathbf{r}_i\rangle$.

Eigenvalues and eigenvectors are not limited to Hermitian matrices. All matrices have at least one eigenvalue and eigenvector. However, only Hermitian matrices have all eigenvectors orthogonal and all eigenvalues real.

Anti-Hermitian Matrices

Occasionally in quantum theory we encounter anti-Hermitian matrices:

$$\mathsf{A}^\dagger = -\mathsf{A}.$$

Following the analysis of the first portion of this section, we can show that

a. The eigenvalues are pure imaginary (or zero). $= 0,$ (3.153)
b. The eigenvectors corresponding to distinct eigenvalues are orthogonal.

or

The matrix R formed from the normalized eigenvectors is unitary. This anti-Hermitian property is preserved under unitary transformations. $-\lambda(\lambda^2 - 1) = 0,$ (3.154)

Example 3.5.1 EIGENVALUES AND EIGENVECTORS OF A REAL SYMMETRIC MATRIX

Let

$$\mathsf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{3.152}$$

The secular equation is

$$\begin{vmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 0 \\ 0 & 0 & -\lambda \end{vmatrix}$$

expanding by minors. The roots are $\lambda = -1, 0, 1$. To find the eigenvector corresponding to $\lambda = -1$, we substitute this value back into the eigenvalue equation, Eq. (3.139),

$$\begin{pmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 0 \\ 0 & 0 & -\lambda \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

With $\lambda = -1$, this yields

$$x + y = 0, \qquad z = 0. \tag{3.156}$$

Within an arbitrary scale factor and an arbitrary sign (or phase factor), $\langle \mathbf{r}_1| = (1, -1, 0)$. Note that (for real $|\mathbf{r}\rangle$ in ordinary space) the eigenvector singles out a line in space. The positive or negative sense is not determined. This indeterminancy could be expected if we noted that Eq. (3.139) is homogeneous in $|\mathbf{r}\rangle$. For convenience we will require that the eigenvectors be normalized to unity, $\langle \mathbf{r}_1|\mathbf{r}_1\rangle = 1$. With this condition,

$$\langle \mathbf{r}_1| = \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0 \right) \tag{3.157}$$

is fixed except for an overall sign. For $\lambda = 0$, Eq. (3.139) yields

$$y = 0, \qquad x = 0, \tag{3.158}$$

$\langle \mathbf{r}_2| = (0, 0, 1)$ is a suitable eigenvector. Finally, for $\lambda = 1$, we get

$$-x + y = 0, \qquad z = 0, \tag{3.159}$$

or

$$\langle \mathbf{r}_3| = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0 \right). \tag{3.160}$$

The orthogonality of $\mathbf{r}_1, \mathbf{r}_2$, and $\mathbf{r}_3$, corresponding to three distinct eigenvalues, may be easily verified.

The corresponding spectral decomposition gives

$$\mathsf{A} = (-1)\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0 \right) \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} + (+1)\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0 \right) \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} + 0(0, 0, 1) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$= -\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

■

Example 3.5.2 DEGENERATE EIGENVALUES

Consider

$$\mathsf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{3.161}$$

The secular equation is

$$\begin{vmatrix} 1-\lambda & 0 & 0 \\ 0 & -\lambda & 1 \\ 0 & 1 & -\lambda \end{vmatrix} = 0 \tag{3.162}$$

or

$$(1-\lambda)(\lambda^2 - 1) = 0, \qquad \lambda = -1, 1, 1, \tag{3.163}$$

a degenerate case. If $\lambda = -1$, the eigenvalue equation (3.139) yields

$$2x = 0, \qquad y + z = 0. \tag{3.164}$$

A suitable normalized eigenvector is

$$\langle \mathbf{r}_1 | = \left(0, \frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right). \tag{3.165}$$

For $\lambda = 1$, we get

$$-y + z = 0. \tag{3.166}$$

Any eigenvector satisfying Eq. (3.166) is perpendicular to $\mathbf{r}_1$. We have an infinite number of choices. Suppose, as one possible choice, $\mathbf{r}_2$ is taken as

$$\langle \mathbf{r}_2 | = \left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \tag{3.167}$$

which clearly satisfies Eq. (3.166). Then $\mathbf{r}_3$ must be perpendicular to $\mathbf{r}_1$ and may be made perpendicular to $\mathbf{r}_2$ by[25]

$$\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2 = (1, 0, 0). \tag{3.168}$$

The corresponding spectral decomposition gives

$$\mathsf{A} = -\left(0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)\begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} + \left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)\begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} + (1, 0, 0)\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$= -\begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & -\frac{1}{2} & \frac{1}{2} \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \end{pmatrix}$$ ■

[25] The use of the cross product is limited to three-dimensional space (see Section 1.4).

Functions of Matrices

Polynomials with one or more matrix arguments are well defined and occur often. Power series of a matrix may also be defined, provided the series converge (see Chapter 5) for each matrix element. For example, if A is any $n \times n$ matrix, then the power series

$$\exp(\mathsf{A}) = \sum_{j=0}^{\infty} \frac{1}{j!} \mathsf{A}^j, \tag{3.169a}$$

$$\sin(\mathsf{A}) = \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j+1)!} \mathsf{A}^{2j+1}, \tag{3.169b}$$

$$\cos(\mathsf{A}) = \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j)!} \mathsf{A}^{2j} \tag{3.169c}$$

are well defined $n \times n$ matrices. For the Pauli matrices σ_k the **Euler identity** for real θ and $k = 1$, 2, or 3

$$\boxed{\exp(i\sigma_k\theta) = \mathsf{1}_2 \cos\theta + i\sigma_k \sin\theta,} \tag{3.170a}$$

follows from collecting all even and odd powers of θ in separate series using $\sigma_k^2 = 1$. For the 4×4 Dirac matrices $\sigma^{jk} = 1$ with $(\sigma^{jk})^2 = 1$ if $j \neq k = 1, 2$ or 3 we obtain similarly (without writing the obvious unit matrix $\mathsf{1}_4$ anymore)

$$\boxed{\exp\bigl(i\sigma^{jk}\theta\bigr) = \cos\theta + i\sigma^{jk} \sin\theta,} \tag{3.170b}$$

while

$$\boxed{\exp\bigl(i\sigma^{0k}\zeta\bigr) = \cosh\zeta + i\sigma^{0k} \sinh\zeta} \tag{3.170c}$$

holds for real ζ because $(i\sigma^{0k})^2 = 1$ for $k = 1$, 2, or 3.

For a Hermitian matrix A there is a unitary matrix U that diagonalizes it; that is, $\mathsf{UAU}^\dagger = [a_1, a_2, \ldots, a_n]$. Then the **trace formula**

$$\boxed{\det\bigl(\exp(\mathsf{A})\bigr) = \exp\bigl(\mathrm{trace}(\mathsf{A})\bigr)} \tag{3.171}$$

is obtained (see Exercises 3.5.2 and 3.5.9) from

$$\begin{aligned}
\det\bigl(\exp(\mathsf{A})\bigr) &= \det\bigl(\mathsf{U}\exp(\mathsf{A})\mathsf{U}^\dagger\bigr) = \det\bigl(\exp\bigl(\mathsf{UAU}^\dagger\bigr)\bigr) \\
&= \det\exp[a_1, a_2, \ldots, a_n] = \det\bigl[e^{a_1}, e^{a_2}, \ldots, e^{a_n}\bigr] \\
&= \prod e^{a_i} = \exp\Bigl(\sum a_i\Bigr) = \exp\bigl(\mathrm{trace}(\mathsf{A})\bigr),
\end{aligned}$$

using $\mathsf{UA}^i\mathsf{U}^\dagger = (\mathsf{UAU}^\dagger)^i$ in the power series Eq. (3.169a) for $\exp(\mathsf{UAU}^\dagger)$ and the product theorem for determinants in Section 3.2.

This trace formula is a special case of the **spectral decomposition law** for any (infinitely differentiable) function $f(\mathsf{A})$ for Hermitian A:

$$\boxed{f(\mathsf{A}) = \sum_i f(\lambda_i)|\mathbf{r}_i\rangle\langle\mathbf{r}_i|,}$$

where $|\mathbf{r}_i\rangle$ are the common eigenvectors of A and A^j. This eigenvalue expansion follows from $\mathsf{A}^j|\mathbf{r}_i\rangle = \lambda_i^j|\mathbf{r}_i\rangle$, multiplied by $f^{(j)}(0)/j!$ and summed over j to form the Taylor expansion of $f(\lambda_i)$ and yield $f(\mathsf{A})|\mathbf{r}_i\rangle = f(\lambda_i)|\mathbf{r}_i\rangle$. Finally, summing over i and using completeness we obtain $f(\mathsf{A})\sum_i |\mathbf{r}_i\rangle\langle\mathbf{r}_i| = \sum_i f(\lambda_i)|\mathbf{r}_i\rangle\langle\mathbf{r}_i| = f(\mathsf{A})$, q.e.d.

Example 3.5.3 EXPONENTIAL OF A DIAGONAL MATRIX

If the matrix A is diagonal like

$$\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

then its nth power is also diagonal with its diagonal, matrix elements raised to the nth power:

$$(\sigma_3)^n = \begin{pmatrix} 1 & 0 \\ 0 & (-1)^n \end{pmatrix}.$$

Then summing the exponential series, element for element, yields

$$e^{\sigma_3} = \begin{pmatrix} \sum_{n=0}^{\infty} \frac{1}{n!} & 0 \\ 0 & \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \end{pmatrix} = \begin{pmatrix} e & 0 \\ 0 & \frac{1}{e} \end{pmatrix}.$$

If we write the general diagonal matrix as $\mathsf{A} = [a_1, a_2, \ldots, a_n]$ with diagonal elements a_j, then $\mathsf{A}^m = [a_1^m, a_2^m, \ldots, a_n^m]$, and summing the exponentials elementwise again we obtain $e^{\mathsf{A}} = [e^{a_1}, e^{a_2}, \ldots, e^{a_n}]$.

Using the spectral decomposition law we obtain directly

$$e^{\sigma_3} = e^{+1}(1,0)\begin{pmatrix} 1 \\ 0 \end{pmatrix} + e^{-1}(0,1)\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} e & 0 \\ 0 & e^{-1} \end{pmatrix}. \qquad \blacksquare$$

Another important relation is the **Baker–Hausdorff formula**,

$$\exp(i\mathsf{G})\mathsf{H}\exp(-i\mathsf{G}) = \mathsf{H} + [i\mathsf{G}, \mathsf{H}] + \frac{1}{2}\big[i\mathsf{G}, [i\mathsf{G}, \mathsf{H}]\big] + \cdots, \tag{3.172}$$

which follows from multiplying the power series for $\exp(i\mathsf{G})$ and collecting the terms with the same powers of $i\mathsf{G}$. Here we define

$$[\mathsf{G}, \mathsf{H}] = \mathsf{GH} - \mathsf{HG}$$

as the **commutator** of G and H.

The preceding analysis has the advantage of exhibiting and clarifying conceptual relationships in the diagonalization of matrices. However, for matrices larger than 3×3, or perhaps 4×4, the process rapidly becomes so cumbersome that we turn to computers and

iterative techniques.[26] One such technique is the Jacobi method for determining eigenvalues and eigenvectors of real symmetric matrices. This Jacobi technique for determining eigenvalues and eigenvectors and the Gauss–Seidel method of solving systems of simultaneous linear equations are examples of relaxation methods. They are iterative techniques in which the errors may decrease or relax as the iterations continue. Relaxation methods are used extensively for the solution of partial differential equations.

Exercises

3.5.1 (a) Starting with the orbital angular momentum of the ith element of mass,

$$\mathbf{L}_i = \mathbf{r}_i \times \mathbf{p}_i = m_i \mathbf{r}_i \times (\boldsymbol{\omega} \times \mathbf{r}_i),$$

derive the inertia matrix such that $\mathbf{L} = \mathsf{I}\boldsymbol{\omega}$, $|\mathbf{L}\rangle = \mathsf{I}|\boldsymbol{\omega}\rangle$.

(b) Repeat the derivation starting with kinetic energy

$$T_i = \frac{1}{2} m_i (\boldsymbol{\omega} \times \mathbf{r}_i)^2 \qquad \left(T = \frac{1}{2} \langle \boldsymbol{\omega} | \mathsf{I} | \boldsymbol{\omega} \rangle \right).$$

3.5.2 Show that the eigenvalues of a matrix are unaltered if the matrix is transformed by a similarity transformation.

This property is not limited to symmetric or Hermitian matrices. It holds for any matrix satisfying the eigenvalue equation, Eq. (3.139). If our matrix can be brought into diagonal form by a similarity transformation, then two immediate consequences are

1. The trace (sum of eigenvalues) is invariant under a similarity transformation.
2. The determinant (product of eigenvalues) is invariant under a similarity transformation.

Note. The invariance of the trace and determinant are often demonstrated by using the Cayley–Hamilton theorem: A matrix satisfies its own characteristic (secular) equation.

3.5.3 As a converse of the theorem that Hermitian matrices have real eigenvalues and that eigenvectors corresponding to distinct eigenvalues are orthogonal, show that if

(a) the eigenvalues of a matrix are real and
(b) the eigenvectors satisfy $\mathbf{r}_i^\dagger \mathbf{r}_j = \delta_{ij} = \langle \mathbf{r}_i | \mathbf{r}_j \rangle$,

then the matrix is Hermitian.

3.5.4 Show that a real matrix that is not symmetric cannot be diagonalized by an orthogonal similarity transformation.

Hint. Assume that the nonsymmetric real matrix can be diagonalized and develop a contradiction.

[26] In higher-dimensional systems the secular equation may be strongly ill-conditioned with respect to the determination of its roots (the eigenvalues). Direct solution by computer may be very inaccurate. Iterative techniques for diagonalizing the original matrix are usually preferred. See Sections 2.7 and 2.9 of Press *et al.*, loc. cit.

3.5.5 The matrices representing the angular momentum components J_x, J_y, and J_z are all Hermitian. Show that the eigenvalues of $\mathbf{J}^2$, where $\mathbf{J}^2 = J_x^2 + J_y^2 + J_z^2$, are real and nonnegative.

3.5.6 A has eigenvalues λ_i and corresponding eigenvectors $|\mathbf{x}_i\rangle$. Show that A^{-1} has the same eigenvectors but with eigenvalues λ_i^{-1}.

3.5.7 A square matrix with zero determinant is labeled **singular**.

(a) If A is singular, show that there is at least one nonzero column vector $\mathbf{v}$ such that

$$\mathsf{A}|\mathbf{v}\rangle = 0.$$

(b) If there is a nonzero vector $|\mathbf{v}\rangle$ such that

$$\mathsf{A}|\mathbf{v}\rangle = 0,$$

show that A is a singular matrix. This means that if a matrix (or operator) has zero as an eigenvalue, the matrix (or operator) has no inverse and its determinant is zero.

3.5.8 The same similarity transformation diagonalizes each of two matrices. Show that the original matrices must commute. (This is particularly important in the matrix (Heisenberg) formulation of quantum mechanics.)

3.5.9 Two Hermitian matrices A and B have the same eigenvalues. Show that A and B are related by a unitary similarity transformation.

3.5.10 Find the eigenvalues and an orthonormal (orthogonal and normalized) set of eigenvectors for the matrices of Exercise 3.2.15.

3.5.11 Show that the inertia matrix for a single particle of mass m at (x, y, z) has a zero determinant. Explain this result in terms of the invariance of the determinant of a matrix under similarity transformations (Exercise 3.3.10) and a possible rotation of the coordinate system.

3.5.12 A certain rigid body may be represented by three point masses: $m_1 = 1$ at $(1, 1, -2)$, $m_2 = 2$ at $(-1, -1, 0)$, and $m_3 = 1$ at $(1, 1, 2)$.

(a) Find the inertia matrix.
(b) Diagonalize the inertia matrix, obtaining the eigenvalues and the principal axes (as orthonormal eigenvectors).

3.5.13 Unit masses are placed as shown in Fig. 3.6.

(a) Find the moment of inertia matrix.
(b) Find the eigenvalues and a set of orthonormal eigenvectors.
(c) Explain the degeneracy in terms of the symmetry of the system.

ANS. $\mathsf{I} = \begin{pmatrix} 4 & -1 & -1 \\ -1 & 4 & -1 \\ -1 & -1 & 4 \end{pmatrix}$ $\lambda_1 = 2$ $\mathbf{r}_1 = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$ $\lambda_2 = \lambda_3 = 5.$

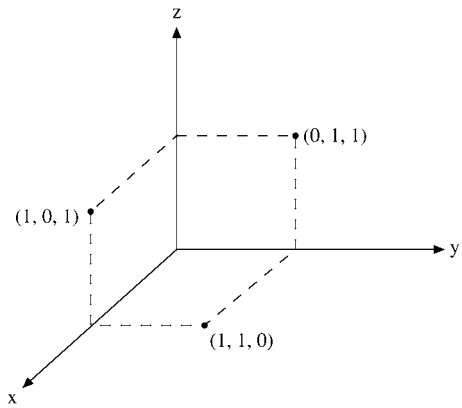

FIGURE 3.6 Mass sites for inertia tensor.

3.5.14 A mass $m_1 = 1/2$ kg is located at $(1, 1, 1)$ (meters), a mass $m_2 = 1/2$ kg is at $(-1, -1, -1)$. The two masses are held together by an ideal (weightless, rigid) rod.

(a) Find the inertia tensor of this pair of masses.
(b) Find the eigenvalues and eigenvectors of this inertia matrix.
(c) Explain the meaning, the physical significance of the $\lambda = 0$ eigenvalue. What is the significance of the corresponding eigenvector?
(d) Now that you have solved this problem by rather sophisticated matrix techniques, explain how you could obtain

(1) $\lambda = 0$ and $\lambda =?$ — by inspection (that is, using common sense).
(2) $\mathbf{r}_{\lambda=0} =?$ — by inspection (that is, using freshman physics).

3.5.15 Unit masses are at the eight corners of a cube $(\pm 1, \pm 1, \pm 1)$. Find the moment of inertia matrix and show that there is a triple degeneracy. This means that so far as moments of inertia are concerned, the cubic structure exhibits spherical symmetry.
Find the eigenvalues and corresponding orthonormal eigenvectors of the following matrices (as a numerical check, note that the sum of the eigenvalues equals the sum of the diagonal elements of the original matrix, Exercise 3.3.9). Note also the correspondence between $\det \mathsf{A} = 0$ and the existence of $\lambda = 0$, as required by Exercises 3.5.2 and 3.5.7.

3.5.16 $$\mathsf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

ANS. $\lambda = 0, 1, 2$.

3.5.17 $$\mathsf{A} = \begin{pmatrix} 0 & \sqrt{2} & 0 \\ \sqrt{2} & 0 & \sqrt{2} \\ 0 & \sqrt{2} & 0 \end{pmatrix}.$$

ANS. $\lambda = -1, 0, 2$.

3.5.18 $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$.

ANS. $\lambda = -1, 1, 2$.

3.5.19 $A = \begin{pmatrix} \sqrt{8} & 1 & \sqrt{8} \\ 0 & \sqrt{8} & 1 \end{pmatrix}$.

ANS. $\lambda = -3, 1, 5$.

3.5.20 $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$.

ANS. $\lambda = 0, 1, 2$.

3.5.21 $A = \begin{pmatrix} 0 & \sqrt{2} & 0 \end{pmatrix}$.

ANS. $\lambda = -1, 1, 2$.

3.5.22 $A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$.

ANS. $\lambda = -\sqrt{2}, 0, \sqrt{2}$.

3.5.23 $A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$.

ANS. $\lambda = 1, 2, 2$.

3.5.24 $A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$.

ANS. $\lambda = 0, 2, 2$.

3.5.25 $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$.

ANS. $\lambda = -1, -1, 2$.

3.5.26 $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$.

ANS. $\lambda = 0, 0, 3$.

3.5.27 $$A = \begin{pmatrix} 5 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 0 & 2 \end{pmatrix}.$$

ANS. $\lambda = 1, 1, 6.$

3.5.28 $$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

ANS. $\lambda = 0, 0, 2.$

3.5.29 $$A = \begin{pmatrix} 5 & 0 & \sqrt{3} \\ 0 & 3 & 0 \\ \sqrt{3} & 0 & 3 \end{pmatrix}.$$

ANS. $\lambda = 2, 3, 6.$

3.5.30 (a) Determine the eigenvalues and eigenvectors of

$$\begin{pmatrix} 1 & \varepsilon \\ \varepsilon & 1 \end{pmatrix}.$$

Note that the eigenvalues are degenerate for $\varepsilon = 0$ but that the eigenvectors are orthogonal for all $\varepsilon \neq 0$ and $\varepsilon \to 0$.

(b) Determine the eigenvalues and eigenvectors of

$$\begin{pmatrix} 1 & 1 \\ \varepsilon^2 & 1 \end{pmatrix}.$$

Note that the eigenvalues are degenerate for $\varepsilon = 0$ and that for this (nonsymmetric) matrix the eigenvectors ($\varepsilon = 0$) do **not** span the space.

(c) Find the cosine of the angle between the two eigenvectors as a function of ε for $0 \leq \varepsilon \leq 1$.

3.5.31 (a) Take the coefficients of the simultaneous linear equations of Exercise 3.1.7 to be the matrix elements a_{ij} of matrix A (symmetric). Calculate the eigenvalues and eigenvectors.

(b) Form a matrix R whose columns are the eigenvectors of A, and calculate the triple matrix product $\tilde{\mathsf{R}}\mathsf{A}\mathsf{R}$.

ANS. $\lambda = 3.33163.$

3.5.32 Repeat Exercise 3.5.31 by using the matrix of Exercise 3.2.39.

3.5.33 Describe the geometric properties of the surface

$$x^2 + 2xy + 2y^2 + 2yz + z^2 = 1.$$

How is it oriented in three-dimensional space? Is it a conic section? If so, which kind?

Table 3.1

Matrix	Eigenvalues	Eigenvectors (for different eigenvalues)
Hermitian	Real	Orthogonal
Anti-Hermitian	Pure imaginary (or zero)	Orthogonal
Unitary	Unit magnitude	Orthogonal
Normal	If A has eigenvalue λ, $\mathsf{A}^\dagger$ has eigenvalue λ^*	Orthogonal A and $\mathsf{A}^\dagger$ have the same eigenvectors

3.5.34 For a Hermitian $n \times n$ matrix A with distinct eigenvalues λ_j and a function f, show that the spectral decomposition law may be expressed as

$$f(\mathsf{A}) = \sum_{j=1}^{n} f(\lambda_j) \frac{\prod_{i \neq j} (\mathsf{A} - \lambda_i)}{\prod_{i \neq j} (\lambda_j - \lambda_i)}.$$

This formula is due to Sylvester.

3.6 NORMAL MATRICES

In Section 3.5 we concentrated primarily on Hermitian or real symmetric matrices and on the actual process of finding the eigenvalues and eigenvectors. In this section[27] we generalize to normal matrices, with Hermitian and unitary matrices as special cases. The physically important problem of normal modes of vibration and the numerically important problem of ill-conditioned matrices are also considered.

A normal matrix is a matrix that commutes with its adjoint,

$$\left[\mathsf{A}, \mathsf{A}^\dagger\right] = 0.$$

Obvious and important examples are Hermitian and unitary matrices. We will show that normal matrices have orthogonal eigenvectors (see Table 3.1). We proceed in two steps.

I. Let A have an eigenvector $|\mathbf{x}\rangle$ and corresponding eigenvalue λ. Then

$$\mathsf{A}|\mathbf{x}\rangle = \lambda|\mathbf{x}\rangle \tag{3.173}$$

or

$$(\mathsf{A} - \lambda 1)|\mathbf{x}\rangle = 0. \tag{3.174}$$

For convenience the combination $\mathsf{A} - \lambda 1$ will be labeled B. Taking the adjoint of Eq. (3.174), we obtain

$$\langle \mathbf{x}|(\mathsf{A} - \lambda 1)^\dagger = 0 = \langle \mathbf{x}|\mathsf{B}^\dagger. \tag{3.175}$$

Because

$$\left[(\mathsf{A} - \lambda 1)^\dagger, (\mathsf{A} - \lambda 1)\right] = \left[\mathsf{A}, \mathsf{A}^\dagger\right] = 0,$$

[27] Normal matrices are the largest class of matrices that can be diagonalized by unitary transformations. For an extensive discussion of normal matrices, see P. A. Macklin, Normal matrices for physicists. *Am. J. Phys.* **52**: 513 (1984).

we have

$$\left[\mathsf{B}, \mathsf{B}^\dagger\right] = 0. \tag{3.176}$$

The matrix B is also normal.

From Eqs. (3.174) and (3.175) we form

$$\langle \mathbf{x}|\mathsf{B}^\dagger \mathsf{B}|\mathbf{x}\rangle = 0. \tag{3.177}$$

This equals

$$\langle \mathbf{x}|\mathsf{B}\mathsf{B}^\dagger|\mathbf{x}\rangle = 0 \tag{3.178}$$

by Eq. (3.176). Now Eq. (3.178) may be rewritten as

$$\left(\mathsf{B}^\dagger|\mathbf{x}\rangle\right)^\dagger \left(\mathsf{B}^\dagger|\mathbf{x}\rangle\right) = 0. \tag{3.179}$$

Thus

$$\mathsf{B}^\dagger|\mathbf{x}\rangle = \left(\mathsf{A}^\dagger - \lambda^* 1\right)|\mathbf{x}\rangle = 0. \tag{3.180}$$

We see that for normal matrices, $\mathsf{A}^\dagger$ has the same eigenvectors as A but the complex conjugate eigenvalues.

II. Now, considering more than one eigenvector–eigenvalue, we have

$$\mathsf{A}|\mathbf{x}_i\rangle = \lambda_i|\mathbf{x}_i\rangle, \tag{3.181}$$

$$\mathsf{A}|\mathbf{x}_j\rangle = \lambda_j|\mathbf{x}_j\rangle. \tag{3.182}$$

Multiplying Eq. (3.182) from the left by $\langle \mathbf{x}_i|$ yields

$$\langle \mathbf{x}_i|\mathsf{A}|\mathbf{x}_j\rangle = \lambda_j \langle \mathbf{x}_i|\mathbf{x}_j\rangle. \tag{3.183}$$

Taking the transpose of Eq. (3.181), we obtain

$$\langle \mathbf{x}_i|\mathsf{A} = \left(\mathsf{A}^\dagger|\mathbf{x}_i\rangle\right)^\dagger. \tag{3.184}$$

From Eq. (3.180), with $\mathsf{A}^\dagger$ having the same eigenvectors as A but the complex conjugate eigenvalues,

$$\left(\mathsf{A}^\dagger|\mathbf{x}_i\rangle\right)^\dagger = \left(\lambda_i^*|\mathbf{x}_i\rangle\right)^\dagger = \lambda_i \langle \mathbf{x}_i|. \tag{3.185}$$

Substituting into Eq. (3.183) we have

$$\lambda_i \langle \mathbf{x}_i|\mathbf{x}_j\rangle = \lambda_j \langle \mathbf{x}_i|\mathbf{x}_j\rangle$$

or

$$(\lambda_i - \lambda_j)\langle \mathbf{x}_i|\mathbf{x}_j\rangle = 0. \tag{3.186}$$

This is the same as Eq. (3.149).

For $\lambda_i \neq \lambda_j$,

$$\langle \mathbf{x}_j|\mathbf{x}_i\rangle = 0.$$

The eigenvectors corresponding to different eigenvalues of a normal matrix are **orthogonal**. This means that a normal matrix may be diagonalized by a unitary transformation. The required unitary matrix may be constructed from the orthonormal eigenvectors as shown earlier, in Section 3.5.

The converse of this result is also true. If A can be diagonalized by a unitary transformation, then A is normal.

Normal Modes of Vibration

We consider the vibrations of a classical model of the CO_2 molecule. It is an illustration of the application of matrix techniques to a problem that does not start as a matrix problem. It also provides an example of the eigenvalues and eigenvectors of an asymmetric real matrix.

Example 3.6.1 NORMAL MODES

Consider three masses on the x-axis joined by springs as shown in Fig. 3.7. The spring forces are assumed to be linear (small displacements, Hooke's law), and the mass is constrained to stay on the x-axis.

Using a different coordinate for each mass, Newton's second law yields the set of equations

$$\left.\begin{aligned}\ddot{x}_1 &= -\frac{k}{M}(x_1 - x_2)\\ \ddot{x}_2 &= -\frac{k}{m}(x_2 - x_1) - \frac{k}{m}(x_2 - x_3)\\ \ddot{x}_3 &= -\frac{k}{M}(x_3 - x_2).\end{aligned}\right\} \tag{3.188}$$

The system of masses is vibrating. We seek the common frequencies, ω, such that all masses vibrate at this same frequency. These are the **normal** modes. Let

$$x_i = x_{i0}e^{i\omega t}, \qquad i = 1, 2, 3. \tag{3.187}$$

with the common factor $e^{i\omega t}$ divided out. We have a matrix–eigenvalue equation with the matrix asymmetric. The secular equation is

Substituting this set into Eq. (3.187), we may rewrite it as

$$\begin{pmatrix} \frac{k}{M} & -\frac{k}{M} & 0 \\ -\frac{k}{m} & \frac{2k}{m} & -\frac{k}{m} \\ 0 & -\frac{k}{M} & \frac{k}{M} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$\begin{vmatrix} \frac{k}{M} - \omega^2 & -\frac{k}{M} & 0 \\ -\frac{k}{m} & \frac{2k}{m} - \omega^2 & -\frac{k}{m} \\ 0 & -\frac{k}{M} & \frac{k}{M} - \omega^2 \end{vmatrix} = 0. \tag{3.189}$$

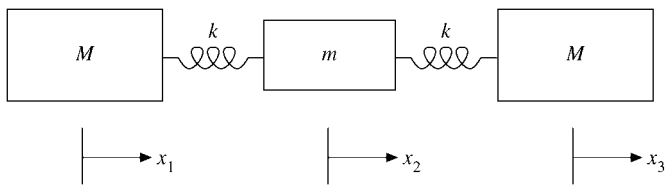

FIGURE 3.7 Double oscillator.

This leads to

$$\omega^2\left(\frac{k}{M}-\omega^2\right)\left(\omega^2-\frac{2k}{m}-\frac{k}{M}\right)=0.$$

The eigenvalues are

$$\omega^2=0,\qquad \frac{k}{M},\qquad \frac{k}{M}+\frac{2k}{m},$$

all real.

The corresponding eigenvectors are determined by substituting the eigenvalues back into Eq. (3.188) one eigenvalue at a time. For $\omega^2=0$, Eq. (3.188), yields

$$x_1-x_2=0,\qquad -x_1+2x_2-x_3=0,\qquad -x_2+x_3=0.$$

Then we get

$$x_1=x_2=x_3.$$

This describes pure translation with no relative motion of the masses and no vibration.

For $\omega^2=k/M$, Eq. (3.188) yields

$$x_1=-x_3,\qquad x_2=0.$$

The two outer masses are moving in opposite direction. The central mass is stationary.

For $\omega^2=k/M+2k/m$, the eigenvector components are

$$x_1=x_3,\qquad x_2=-\frac{2M}{m}x_1.$$

The two outer masses are moving together. The central mass is moving opposite to the two outer ones. The net momentum is zero.

Any displacement of the three masses along the x-axis can be described as a linear combination of these three types of motion: translation plus two forms of vibration. ■

Ill-Conditioned Systems

A system of simultaneous linear equations may be written as

$$\mathsf{A}|\mathbf{x}\rangle=|\mathbf{y}\rangle\qquad\text{or}\qquad \mathsf{A}^{-1}|\mathbf{y}\rangle=|\mathbf{x}\rangle, \tag{3.190}$$

with A and $|\mathbf{y}\rangle$ known and $|\mathbf{x}\rangle$ unknown. When a small error in $|\mathbf{y}\rangle$ results in a larger error in $|\mathbf{x}\rangle$, then the matrix A is called *ill-conditioned*. With $|\delta\mathbf{x}\rangle$ an error in $|\mathbf{x}\rangle$ and $|\delta\mathbf{x}\rangle$ an error in $|\mathbf{y}\rangle$, the relative errors may be written as

$$\left[\frac{\langle\delta\mathbf{x}|\delta\mathbf{x}\rangle}{\langle\mathbf{x}|\mathbf{x}\rangle}\right]^{1/2}\le K(\mathsf{A})\left[\frac{\langle\delta\mathbf{y}|\delta\mathbf{y}\rangle}{\langle\mathbf{y}|\mathbf{y}\rangle}\right]^{1/2}. \tag{3.191}$$

Here $K(\mathsf{A})$, a property of matrix A, is labeled the **condition number**. For A Hermitian one form of the condition number is given by[28]

$$K(\mathsf{A})=\frac{|\lambda|_{\max}}{|\lambda|_{\min}}. \tag{3.192}$$

[28]G. E. Forsythe, and C. B. Moler, *Computer Solution of Linear Algebraic Systems*. Englewood Cliffs, NJ, Prentice Hall (1967).

An approximate form due to Turing[29] is

$$K(\mathsf{A}) = n[A_{ij}]_{\max}\left[A_{ij}^{-1}\right]_{\max}, \tag{3.193}$$

in which n is the order of the matrix and $[A_{ij}]_{\max}$ is the maximum element in A.

Example 3.6.1 AN ILL-CONDITIONED MATRIX

A common example of an ill-conditioned matrix is the Hilbert matrix, $H_{ij} = (i + j - 1)^{-1}$. The Hilbert matrix of order 4, H_4, is encountered in a least-squares fit of data to a third-degree polynomial. We have

$$\mathsf{H}_4 = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{pmatrix}. \tag{3.194}$$

The elements of the inverse matrix (order n) are given by

$$\left(\mathsf{H}_n^{-1}\right)_{ij} = \frac{(-1)^{i+j}}{i+j-1} \cdot \frac{(n+i-1)!(n+j-1)!}{[(i-1)!(j-1)!]^2(n-i)!(n-j)!}. \tag{3.195}$$

For $n = 4$,

$$\mathsf{H}_4^{-1} = \begin{pmatrix} 16 & -120 & 240 & -140 \\ -120 & 1200 & -2700 & 1680 \\ 240 & -2700 & 6480 & -4200 \\ -140 & 1680 & -4200 & 2800 \end{pmatrix}. \tag{3.196}$$

From Eq. (3.193) the Turing estimate of the condition number for H_4 becomes

$$\begin{aligned} K_{\text{Turing}} &= 4 \times 1 \times 6480 \\ &= 2.59 \times 10^4. \end{aligned}$$

This is a warning that an input error may be multiplied by 26,000 in the calculation of the output result. It is a statement that H_4 is ill-conditioned. If you encounter a highly ill-conditioned system, you have two alternatives (besides abandoning the problem).

(a) Try a different mathematical attack.
(b) Arrange to carry more significant figures and push through by brute force.

As previously seen, matrix eigenvector–eigenvalue techniques are not limited to the solution of strictly matrix problems. A further example of the transfer of techniques from one area to another is seen in the application of matrix techniques to the solution of Fredholm eigenvalue integral equations, Section 16.3. In turn, these matrix techniques are strengthened by a variational calculation of Section 17.8. ■

[29] Compare J. Todd, *The Condition of the Finite Segments of the Hilbert Matrix*, Applied Mathematics Series No. 313. Washington, DC: National Bureau of Standards.

Exercises

3.6.1 Show that every 2×2 matrix has two eigenvectors and corresponding eigenvalues. The eigenvectors are not necessarily orthogonal and may be degenerate. The eigenvalues are not necessarily real.

3.6.2 As an illustration of Exercise 3.6.1, find the eigenvalues and corresponding eigenvectorsfor

$$\begin{pmatrix} 2 & 4 \\ 1 & 2 \end{pmatrix}.$$

Note that the eigenvectors are **not** orthogonal.

ANS. $\lambda_1 = 0$, $\mathbf{r}_1 = (2, -1)$;
$\lambda_2 = 4$, $\mathbf{r}_2 = (2, 1)$.

3.6.3 If A is a 2×2 matrix, show that its eigenvalues λ satisfy the secular equation

$$\lambda^2 - \lambda \,\mathrm{trace}(\mathsf{A}) + \det \mathsf{A} = 0.$$

3.6.4 Assuming a unitary matrix U to satisfy an eigenvalue equation $\mathsf{U}\mathbf{r} = \lambda\mathbf{r}$, show that the eigenvalues of the unitary matrix have unit magnitude. This same result holds for real orthogonal matrices.

3.6.5 Since an orthogonal matrix describing a rotation in real three-dimensional space is a special case of a unitary matrix, such an orthogonal matrix can be diagonalized by a unitary transformation.

(a) Show that the sum of the three eigenvalues is $1 + 2\cos\varphi$, where φ is the net angle of rotation about a single fixed axis.
(b) Given that one eigenvalue is 1, show that the other two eigenvalues must be $e^{i\varphi}$ and $e^{-i\varphi}$.

Our orthogonal rotation matrix (real elements) has complex eigenvalues.

3.6.6 A is an nth-order Hermitian matrix with orthonormal eigenvectors $|\mathbf{x}_i\rangle$ and real eigenvalues $\lambda_1 \le \lambda_2 \le \lambda_3 \le \cdots \le \lambda_n$. Show that for a unit magnitude vector $|\mathbf{y}\rangle$,

$$\lambda_1 \le \langle \mathbf{y}|\mathsf{A}|\mathbf{y}\rangle \le \lambda_n.$$

3.6.7 A particular matrix is both Hermitian and unitary. Show that its eigenvalues are all ± 1. *Note.* The Pauli and Dirac matrices are specific examples.

3.6.8 For his relativistic electron theory Dirac required a set of **four** anticommuting matrices. Assume that these matrices are to be Hermitian and unitary. If these are $n \times n$ matrices, show that n must be even. With 2×2 matrices inadequate (why?), this demonstrates that the smallest possible matrices forming a set of four anticommuting, Hermitian, unitary matrices are 4×4.

3.6.9 A is a normal matrix with eigenvalues λ_n and orthonormal eigenvectors $|\mathbf{x}_n\rangle$. Show that A may be written as

$$\mathsf{A} = \sum_n \lambda_n |\mathbf{x}_n\rangle\langle\mathbf{x}_n|.$$

Hint. Show that both this eigenvector form of A and the original A give the same result acting on an arbitrary vector $|\mathbf{y}\rangle$.

3.6.10 A has eigenvalues 1 and -1 and corresponding eigenvectors $\binom{1}{0}$ and $\binom{0}{1}$. Construct A.

ANS. $\mathsf{A} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

3.6.11 A non-Hermitian matrix A has eigenvalues λ_i and corresponding eigenvectors $|\mathbf{u}_i\rangle$. The adjoint matrix $\mathsf{A}^\dagger$ has the same set of eigenvalues but **different** corresponding eigenvectors, $|\mathbf{v}_i\rangle$. Show that the eigenvectors form a **biorthogonal** set, in the sense that

$$\langle\mathbf{v}_i|\mathbf{u}_j\rangle = 0 \qquad \text{for} \quad \lambda_i^* \neq \lambda_j.$$

3.6.12 You are given a pair of equations:

$$\mathsf{A}|\mathbf{f}_n\rangle = \lambda_n|\mathbf{g}_n\rangle$$
$$\tilde{\mathsf{A}}|\mathbf{g}_n\rangle = \lambda_n|\mathbf{f}_n\rangle \qquad \text{with } \mathsf{A} \text{ real.}$$

(a) Prove that $|\mathbf{f}_n\rangle$ is an eigenvector of $(\tilde{\mathsf{A}}\mathsf{A})$ with eigenvalue λ_n^2.
(b) Prove that $|\mathbf{g}_n\rangle$ is an eigenvector of $(\mathsf{A}\tilde{\mathsf{A}})$ with eigenvalue λ_n^2.
(c) State how you know that

(1) The $|\mathbf{f}_n\rangle$ form an orthogonal set.
(2) The $|\mathbf{g}_n\rangle$ form an orthogonal set.
(3) λ_n^2 is real.

3.6.13 Prove that A of the preceding exercise may be written as

$$\mathsf{A} = \sum_n \lambda_n |\mathbf{g}_n\rangle\langle\mathbf{f}_n|,$$

with the $|\mathbf{g}_n\rangle$ and $\langle\mathbf{f}_n|$ normalized to unity.
Hint. Expand your arbitrary vector as a linear combination of $|\mathbf{f}_n\rangle$.

3.6.14 Given

$$\mathsf{A} = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 2 \\ 1 & -4 \end{pmatrix},$$

(a) Construct the transpose $\tilde{\mathsf{A}}$ and the symmetric forms $\tilde{\mathsf{A}}\mathsf{A}$ and $\mathsf{A}\tilde{\mathsf{A}}$.
(b) From $\mathsf{A}\tilde{\mathsf{A}}|\mathbf{g}_n\rangle = \lambda_n^2|\mathbf{g}_n\rangle$ find λ_n and $|\mathbf{g}_n\rangle$. Normalize the $|\mathbf{g}_n\rangle$.
(c) From $\tilde{\mathsf{A}}\mathsf{A}|\mathbf{f}_n\rangle = \lambda_n^2|\mathbf{f}_n\rangle$ find λ_n [same as (b)] and $|\mathbf{f}_n\rangle$. Normalize the $|\mathbf{f}_n\rangle$.
(d) Verify that $\mathsf{A}|\mathbf{f}_n\rangle = \lambda_n|\mathbf{g}_n\rangle$ and $\tilde{\mathsf{A}}|\mathbf{g}_n\rangle = \lambda_n|\mathbf{f}_n\rangle$.
(e) Verify that $\mathsf{A} = \sum_n \lambda_n |\mathbf{g}_n\rangle\langle\mathbf{f}_n|$.

3.6.15 Given the eigenvalues $\lambda_1 = 1$, $\lambda_2 = -1$ and the corresponding eigenvectors

$$|\mathbf{f}_1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |\mathbf{g}_1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad |\mathbf{f}_2\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad \text{and} \qquad |\mathbf{g}_2\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

(a) construct A;
(b) verify that $\mathsf{A}|\mathbf{f}_n\rangle = \lambda_n|\mathbf{g}_n\rangle$;
(c) verify that $\tilde{\mathsf{A}}|\mathbf{g}_n\rangle = \lambda_n|\mathbf{f}_n\rangle$.

$$\text{ANS. } \mathsf{A} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

3.6.16 This is a continuation of Exercise 3.4.12, where the unitary matrix U and the Hermitian matrix H are related by

$$\mathsf{U} = e^{ia\mathsf{H}}.$$

(a) If trace $\mathsf{H} = 0$, show that $\det \mathsf{U} = +1$.
(b) If $\det \mathsf{U} = +1$, show that trace $\mathsf{H} = 0$.

Hint. H may be diagonalized by a similarity transformation. Then interpreting the exponential by a Maclaurin expansion, U is also diagonal. The corresponding eigenvalues are given by $u_j = \exp(iah_j)$.
Note. These properties, and those of Exercise 3.4.12, are vital in the development of the concept of generators in group theory — Section 4.2.

3.6.17 An $n \times n$ matrix A has n eigenvalues A_i. If $\mathsf{B} = e^{\mathsf{A}}$, show that B has the same eigenvectors as A, with the corresponding eigenvalues B_i given by $B_i = \exp(A_i)$.
Note. e^{A} is defined by the Maclaurin expansion of the exponential:

$$e^{\mathsf{A}} = 1 + \mathsf{A} + \frac{\mathsf{A}^2}{2!} + \frac{\mathsf{A}^3}{3!} + \cdots.$$

3.6.18 A matrix P is a projection operator (see the discussion following Eq. (3.138c)) satisfying the condition

$$\mathsf{P}^2 = \mathsf{P}.$$

Show that the corresponding eigenvalues $(\rho^2)_\lambda$ and ρ_λ satisfy the relation

$$\left(\rho^2\right)_\lambda = (\rho_\lambda)^2 = \rho_\lambda.$$

This means that the eigenvalues of P are 0 and 1.

3.6.19 In the **matrix** eigenvector–eigenvalue equation

$$\mathsf{A}|\mathbf{r}_i\rangle = \lambda_i|\mathbf{r}_i\rangle,$$

A is an $n \times n$ Hermitian matrix. For simplicity assume that its n real eigenvalues are distinct, λ_1 being the largest. If $|\mathbf{r}\rangle$ is an approximation to $|\mathbf{r}_1\rangle$,

$$|\mathbf{r}\rangle = |\mathbf{r}_1\rangle + \sum_{i=2}^{n} \delta_i|\mathbf{r}_i\rangle,$$

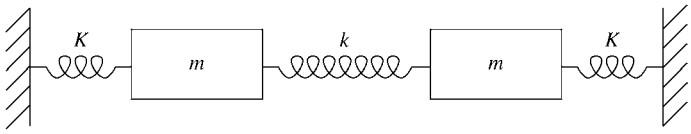

FIGURE 3.8 Triple oscillator.

show that

$$\frac{\langle \mathbf{r}|\mathsf{A}|\mathbf{r}\rangle}{\langle \mathbf{r}|\mathbf{r}\rangle} \leq \lambda_1$$

and that the error in λ_1 is of the order $|\delta_i|^2$. Take $|\delta_i| \ll 1$.
Hint. The n $|\mathbf{r}_i\rangle$ form a **complete** orthogonal set spanning the n-dimensional (complex) space.

3.6.20 Two equal masses are connected to each other and to walls by springs as shown in Fig. 3.8. The masses are constrained to stay on a horizontal line.

(a) Set up the Newtonian acceleration equation for each mass.
(b) Solve the secular equation for the eigenvectors.
(c) Determine the eigenvectors and thus the normal modes of motion.

3.6.21 Given a normal matrix A with eigenvalues λ_j, show that $\mathsf{A}^\dagger$ has eigenvalues λ_j^*, its real part $(\mathsf{A}+\mathsf{A}^\dagger)/2$ has eigenvalues $\Re(\lambda_j)$, and its imaginary part $(\mathsf{A}-\mathsf{A}^\dagger)/2i$ has eigenvalues $\Im(\lambda_j)$.

Additional Readings

Aitken, A. C., *Determinants and Matrices*. New York: Interscience (1956). Reprinted, Greenwood (1983). A readable introduction to determinants and matrices.

Barnett, S., *Matrices: Methods and Applications*. Oxford: Clarendon Press (1990).

Bickley, W. G., and R. S. H. G. Thompson, *Matrices—Their Meaning and Manipulation*. Princeton, NJ: Van Nostrand (1964). A comprehensive account of matrices in physical problems, their analytic properties, and numerical techniques.

Brown, W. C., *Matrices and Vector Spaces*. New York: Dekker (1991).

Gilbert, J. and L., *Linear Algebra and Matrix Theory*. San Diego: Academic Press (1995).

Heading, J., *Matrix Theory for Physicists*. London: Longmans, Green and Co. (1958). A readable introduction to determinants and matrices, with applications to mechanics, electromagnetism, special relativity, and quantum mechanics.

Vein, R., and P. Dale, *Determinants and Their Applications in Mathematical Physics*. Berlin: Springer (1998).

Watkins, D. S., *Fundamentals of Matrix Computations*. New York: Wiley (1991).

This page intentionally left blank

CHAPTER 4

GROUP THEORY

Disciplined judgment, about what is neat and symmetrical and elegant has time and time again proved an excellent guide to how nature works

MURRAY GELL-MANN

4.1 INTRODUCTION TO GROUP THEORY

In classical mechanics the **symmetry** of a physical system leads to **conservation laws**. Conservation of angular momentum is a direct consequence of rotational symmetry, which means **invariance** under spatial rotations. In the first third of the 20th century, Wigner and others realized that invariance was a key concept in understanding the new quantum phenomena and in developing appropriate theories. Thus, in quantum mechanics the concept of angular momentum and spin has become even more central. Its generalizations, **isospin** in nuclear physics and the **flavor symmetry** in particle physics, are indispensable tools in building and solving theories. Generalizations of the concept of **gauge invariance** of classical electrodynamics to the isospin symmetry lead to the electroweak gauge theory.

In each case the set of these symmetry operations forms a group. Group theory is the mathematical tool to treat invariants and symmetries. It brings unification and formalization of principles, such as spatial reflections, or parity, angular momentum, and geometry, that are widely used by physicists.

In geometry the fundamental role of group theory was recognized more than a century ago by mathematicians (e.g., Felix Klein's Erlanger Program). In Euclidean geometry the distance between two points, the scalar product of two vectors or metric, does not change under rotations or translations. These symmetries are characteristic of this geometry. In special relativity the metric, or scalar product of four-vectors, differs from that of

Euclidean geometry in that it is no longer positive definite and is invariant under Lorentz transformations.

For a crystal the symmetry group contains only a finite number of rotations at discrete values of angles or reflections. The theory of such **discrete** or **finite** groups, developed originally as a branch of pure mathematics, now is a useful tool for the development of crystallography and condensed matter physics. A brief introduction to this area appears in Section 4.7. When the rotations depend on continuously varying angles (the Euler angles of Section 3.3) the rotation groups have an infinite number of elements. Such continuous (or **Lie**[1]) **groups** are the topic of Sections 4.2–4.6. In Section 4.8 we give an introduction to differential forms, with applications to Maxwell's equations and topics of Chapters 1 and 2, which allows seeing these topics from a different perspective.

Definition of a Group

A group G may be defined as a set of objects or operations, rotations, transformations, called the elements of G, that may be combined, or "multiplied," to form a well-defined product in G, denoted by a *, that satisfies the following four conditions.

1. If a and b are any two elements of G, then the product $a * b$ is also an element of G, where b acts before a; or $(a, b) \to a * b$ associates (or maps) an element $a * b$ of G with the pair (a, b) of elements of G. This property is known as "G is closed under multiplication of its own elements."
2. This multiplication is associative: $(a * b) * c = a * (b * c)$.
3. There is a unit element[2] 1 in G such that $1 * a = a * 1 = a$ for every element a in G. The unit is unique: $1 = 1' * 1 = 1'$.
4. There is an inverse, or reciprocal, of each element a of G, labeled a^{-1}, such that $a * a^{-1} = a^{-1} * a = 1$. The inverse is unique: If a^{-1} and a'^{-1} are both inverses of a, then $a'^{-1} = a'^{-1} * (a * a'^{-1}) = (a'^{-1} * a) * a^{-1} = a^{-1}$.

Since the * for multiplication is tedious to write, it is customary to drop it and simply let it be understood. From now on, we write ab instead of $a * b$.

- If a subset G' of G is closed under multiplication, it is a group and called a **subgroup** of G; that is, G' is closed under the multiplication of G. The unit of G always forms a subgroup of G.
- If $gg'g^{-1}$ is an element of G' for any g of G and g' of G', then G' is called an **invariant subgroup** of G. The subgroup consisting of the unit is invariant. If the group elements are square matrices, then $gg'g^{-1}$ corresponds to a similarity transformation (see Eq. (3.100)).
- If $ab = ba$ for all a, b of G, the group is called **abelian**, that is, the order in products does not matter; commutative multiplication is often denoted by a $+$ sign. Examples are vector spaces whose unit is the zero vector and $-a$ is the inverse of a for all elements a in G.

[1] After the Norwegian mathematician Sophus Lie.

[2] Following E. Wigner, the unit element of a group is often labeled E, from the German **Einheit**, that is, unit, or just 1, or I for identity.

Example 4.1.1 ORTHOGONAL AND UNITARY GROUPS

Orthogonal $n \times n$ matrices form the group $\mathsf{O}(n)$, and $\mathsf{SO}(n)$ if their determinants are $+1$ (S stands for "special"). If $\tilde{\mathsf{O}}_i = \mathsf{O}_i^{-1}$ for $i = 1$ and 2 (see Section 3.3 for orthogonal matrices) are elements of $\mathsf{O}(n)$, then the product

$$\widetilde{\mathsf{O}_1\mathsf{O}_2} = \tilde{\mathsf{O}}_2\tilde{\mathsf{O}}_1 = \mathsf{O}_2^{-1}\mathsf{O}_1^{-1} = (\mathsf{O}_1\mathsf{O}_2)^{-1}$$

is also an orthogonal matrix in $\mathsf{O}(n)$, thus proving closure under (matrix) multiplication. The inverse is the transpose (orthogonal) matrix. The unit of the group is the n-dimensional unit matrix 1_n. A real orthogonal $n \times n$ matrix has $n(n-1)/2$ independent parameters. For $n = 2$, there is only one parameter: one angle. For $n = 3$, there are three independent parameters: the three Euler angles of Section 3.3.

If $\tilde{\mathsf{O}}_i = \mathsf{O}_i^{-1}$ (for $i = 1$ and 2) are elements of $\mathsf{SO}(n)$, then closure requires proving in addition that their product has determinant $+1$, which follows from the product theorem in Chapter 3.

Likewise, unitary $n \times n$ matrices form the group $\mathsf{U}(n)$, and $\mathsf{SU}(n)$ if their determinants are $+1$. If $\mathsf{U}_i^\dagger = \mathsf{U}_i^{-1}$ (see Section 3.4 for unitary matrices) are elements of $\mathsf{U}(n)$, then

$$(\mathsf{U}_1\mathsf{U}_2)^\dagger = \mathsf{U}_2^\dagger\mathsf{U}_1^\dagger = \mathsf{U}_2^{-1}\mathsf{U}_1^{-1} = (\mathsf{U}_1\mathsf{U}_2)^{-1},$$

so the product is unitary and an element of $\mathsf{U}(n)$, thus proving closure under multiplication. Each unitary matrix has an inverse (its Hermitian adjoint), which again is unitary.

If $\mathsf{U}_i^\dagger = \mathsf{U}_i^{-1}$ are elements of $\mathsf{SU}(n)$, then closure requires us to prove that their product also has determinant $+1$, which follows from the product theorem in Chapter 3. ■

- Orthogonal groups are called **Lie groups**; that is, they depend on continuously varying parameters (the Euler angles and their generalization for higher dimensions); they are **compact** because the angles vary over closed, finite intervals (containing the limit of any converging sequence of angles). Unitary groups are also compact. Translations form a noncompact group because the limit of translations with distance $d \to \infty$ is not part of the group. The Lorentz group is not compact either.

Homomorphism, Isomorphism

There may be a correspondence between the elements of two groups: one-to-one, two-to-one, or many-to-one. If this correspondence preserves the group multiplication, we say that the two groups are **homomorphic**. A most important homomorphic correspondence between the rotation group $\mathsf{SO}(3)$ and the unitary group $\mathsf{SU}(2)$ is developed in Section 4.2. If the correspondence is one-to-one, still preserving the group multiplication,[3] then the groups are **isomorphic**.

- If a group G is homomorphic to a group of matrices G', then G' is called a **representation** of G. If G and G' are isomorphic, the representation is called **faithful**. There are many representations of groups; they are not unique.

[3] Suppose the elements of one group are labeled g_i, the elements of a second group h_i. Then $g_i \leftrightarrow h_i$ is a one-to-one correspondence for all values of i. If $g_i g_j = g_k$ and $h_i h_j = h_k$, then g_k and h_k must be the corresponding group elements.

involves only one angle of the rotation about the z-axis. As shown in Chapter 3, the linear transformation of two successive rotations involves the product of the matrices corresponding to the sum of the angles. The product corresponds to two rotations, $\mathsf{R}_z(\varphi_1)\mathsf{R}_z(\varphi_2)$, and is defined by rotating first by the angle φ_2 and then by φ_1. According to Eq. (3.29), this corresponds to the product of the orthogonal 2×2 submatrices,

$$\begin{pmatrix} \cos\varphi_1 & \sin\varphi_1 \\ -\sin\varphi_1 & \cos\varphi_1 \end{pmatrix} \begin{pmatrix} \cos\varphi_2 & \sin\varphi_2 \\ -\sin\varphi_2 & \cos\varphi_2 \end{pmatrix} \equiv \begin{pmatrix} \cos(\varphi_1+\varphi_2) & \sin(\varphi_1+\varphi_2) \\ -\sin(\varphi_1+\varphi_2) & \cos(\varphi_1+\varphi_2) \end{pmatrix} \tag{4.2}$$

$$= \mathsf{R}_z(\varphi_1+\varphi_2).$$

Eq. (1.8), Fig. 1.6, and Section 3.3.

Example 4.1.2 ROTATIONS

Another instructive example for a group is the set of counterclockwise coordinate rotations of three-dimensional Euclidean space about its z-axis. From Chapter 3 we know that such a rotation is described by a linear transformation of the coordinates involving a 3×3 matrix made up of three rotations depending on the Euler angles. If the z-axis is fixed, the linear transformation is through an angle φ of the xy-coordinate system to a new orientation in using the addition formulas for the trigonometric functions. The unity in the lower right-hand corner of the matrix in Eq. (4.1) is also reproduced upon multiplication. The product is clearly a rotation, represented by the orthogonal matrix with angle $\varphi_1+\varphi_2$. The associative group multiplication corresponds to the associative matrix multiplication. It is **commutative**, or **abelian**, because the order in which these rotations are performed does not matter. The inverse of the rotation with angle φ is that with angle $-\varphi$. The unit corresponds to the angle $\varphi = 0$. Striking off the coordinate vectors in Eq. (4.1), we can associate the matrix of the linear transformation with each rotation, which is a group multiplication preserving one-to-one mapping, an isomorphism: The matrices form a faithful representation of the rotation group. The unity in the right-hand corner is superfluous as well, like the coordinate vectors, and may be deleted. This defines another isomorphism and representation by the 2×2 submatrices:

$$\mathsf{R}_z(\varphi) = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \to \mathsf{R}(\varphi) = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}. \tag{4.3}$$

The group's name is $\mathsf{SO}(2)$, if the angle φ varies continuously from 0 to 2π; $\mathsf{SO}(2)$ has infinitely many elements and is compact.

The group of rotations R_z is obviously isomorphic to the group of rotations in Eq. (4.3). The unity with angle $\varphi = 0$ and the rotation with $\varphi = \pi$ form a finite subgroup. The finite subgroups with angles $2\pi m/n$, n an integer and $m = 0, 1, \ldots, n-1$ are **cyclic**; that is, the rotations $\mathsf{R}(2\pi m/n) = \mathsf{R}(2\pi/n)^m$. ■

In the following we shall discuss only the rotation groups $\mathsf{SO}(n)$ and unitary groups $\mathsf{SU}(n)$ among the classical Lie groups. (More examples of finite groups will be given in Section 4.7.)

Representations — Reducible and Irreducible

The representation of group elements by matrices is a very powerful technique and has been almost universally adopted by physicists. The use of matrices imposes no significant restriction. It can be shown that the elements of any finite group and of the continuous groups of Sections 4.2–4.4 may be represented by matrices. Examples are the rotations described in Eq. (4.3).

To illustrate how matrix representations arise from a symmetry, consider the stationary Schrödinger equation (or some other eigenvalue equation, such as $\mathsf{I}v_i = I_i v_i$ for the principal moments of inertia of a rigid body in classical mechanics, say),

$$H\psi = E\psi. \tag{4.4}$$

Let us assume that the Hamiltonian H stays invariant under a group G of transformations R in G (coordinate rotations, for example, for a central potential $V(r)$ in the Hamiltonian H); that is,

$$H_R = \mathsf{R}H\mathsf{R}^{-1} = H, \qquad \mathsf{R}H = H\mathsf{R}. \tag{4.5}$$

Now take a solution ψ of Eq. (4.4) and "rotate" it: $\psi \to \mathsf{R}\psi$. Then $\mathsf{R}\psi$ has the **same energy E** because multiplying Eq. (4.4) by R and using Eq. (4.5) yields

$$\boxed{\mathsf{R}H\psi = E(\mathsf{R}\psi) = \left(\mathsf{R}H\mathsf{R}^{-1}\right)R\psi = H(\mathsf{R}\psi).} \tag{4.6}$$

In other words, all rotated solutions $\mathsf{R}\psi$ are **degenerate** in energy or form what physicists call a **multiplet**. For example, the spin-up and -down states of a bound electron in the ground state of hydrogen form a doublet, and the states with projection quantum numbers $m = -l, -l+1, \ldots, l$ of orbital angular momentum l form a multiplet with $2l+1$ basis states.

Let us assume that this vector space V_ψ of transformed solutions has a finite dimension n. Let $\psi_1, \psi_2, \ldots, \psi_n$ be a basis. Since $\mathsf{R}\psi_j$ is a member of the multiplet, we can expand it in terms of its basis,

$$\mathsf{R}\psi_j = \sum_k r_{jk}\psi_k. \tag{4.7}$$

Thus, with each R in G we can associate a matrix (r_{jk}). Just as in Example 4.1.2, two successive rotations correspond to the product of their matrices, so this map $\mathsf{R} \to (r_{jk})$ is a representation of G. It is necessary for a representation to be **irreducible** that we can take any element of V_ψ and, by rotating with **all** elements R of G, transform it into **all** other elements of V_ψ. If not all elements of V_ψ are reached, then V_ψ splits into a direct sum of two or more vector subspaces, $V_\psi = V_1 \oplus V_2 \oplus \cdots$, which are mapped into themselves by rotating their elements. For example, the $2s$ state and $2p$ states of principal quantum number $n = 2$ of the hydrogen atom have the same energy (that is, are degenerate) and form

a reducible representation, because the $2s$ state cannot be rotated into the $2p$ states, and vice versa (angular momentum is conserved under rotations). In this case the representation is called **reducible**. Then we can find a basis in V_ψ (that is, there is a unitary matrix U) so that

$$\mathsf{U}(r_{jk})\mathsf{U}^\dagger = \begin{pmatrix} \mathbf{r}_1 & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{r}_2 & \cdots \\ \vdots & \vdots & \end{pmatrix} \tag{4.8}$$

for **all** R of G, and **all** matrices (r_{jk}) have **similar** block-diagonal shape. Here $\mathbf{r}_1, \mathbf{r}_2, \ldots$ are matrices of lower dimension than (r_{jk}) that are lined up along the diagonal and the $\mathbf{0}$'s are matrices made up of zeros. We may say that the representation has been decomposed into $\mathbf{r}_1 + \mathbf{r}_2 + \cdots$ along with $V_\psi = V_1 \oplus V_2 \oplus \cdots$.

The irreducible representations play a role in group theory that is roughly analogous to the unit vectors of vector analysis. They are the simplest representations; all others can be built from them. (See Section 4.4 on Clebsch–Gordan coefficients and Young tableaux.)

Exercises

4.1.1 Show that an $n \times n$ orthogonal matrix has $n(n-1)/2$ independent parameters.
Hint. The orthogonality condition, Eq. (3.71), provides constraints.

4.1.2 Show that an $n \times n$ unitary matrix has $n^2 - 1$ independent parameters.
Hint. Each element may be complex, doubling the number of possible parameters. Some of the constraint equations are likewise complex and count as two constraints.

4.1.3 The special linear group $\mathsf{SL}(2)$ consists of all 2×2 matrices (with complex elements) having a determinant of $+1$. Show that such matrices form a group.
Note. The $\mathsf{SL}(2)$ group can be related to the full Lorentz group in Section 4.4, much as the $\mathsf{SU}(2)$ group is related to $\mathsf{SO}(3)$.

4.1.4 Show that the rotations about the z-axis form a subgroup of $\mathsf{SO}(3)$. Is it an invariant subgroup?

4.1.5 Show that if R, S, T are elements of a group G so that $RS = T$ and $R \to (r_{ik})$, $S \to (s_{ik})$ is a representation according to Eq. (4.7), then

$$(r_{ik})(s_{ik}) = \left(t_{ik} = \sum_n r_{in} s_{nk}\right),$$

that is, group multiplication translates into matrix multiplication for any group representation.

4.2 Generators of Continuous Groups

A characteristic property of continuous groups known as Lie groups is that the parameters of a product element are analytic functions[4] of the parameters of the factors. The analytic

[4] *Analytic* here means having derivatives of all orders.

nature of the functions (differentiability) allows us to develop the concept of generator and to reduce the study of the whole group to a study of the group elements in the neighborhood of the identity element.

Lie's essential idea was to study elements R in a group G that are infinitesimally close to the unity of G. Let us consider the $\mathsf{SO}(2)$ group as a simple example. The 2×2 rotation matrices in Eq. (4.2) can be written in exponential form using the Euler identity, Eq. (3.170a), as

$$\mathsf{R}(\varphi) = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} = 1_2 \cos\varphi + i\sigma_2 \sin\varphi = \exp(i\sigma_2\varphi). \tag{4.9}$$

From the exponential form it is obvious that multiplication of these matrices is equivalent to addition of the arguments

$$\mathsf{R}(\varphi_2)\mathsf{R}(\varphi_1) = \exp(i\sigma_2\varphi_2)\exp(i\sigma_2\varphi_1) = \exp\big(i\sigma_2(\varphi_1+\varphi_2)\big) = \mathsf{R}(\varphi_1+\varphi_2).$$

Rotations close to 1 have small angle $\varphi \approx 0$.

This suggests that we look for an exponential representation

$$\mathsf{R} = \exp(i\varepsilon\mathsf{S}) = 1 + i\varepsilon\mathsf{S} + \mathsf{O}\big(\varepsilon^2\big), \qquad \varepsilon \to 0, \tag{4.10}$$

for group elements R in G close to the unity 1. The infinitesimal transformations are $\varepsilon\mathsf{S}$, and the S are called generators of G. They form a linear space because multiplication of the group elements R translates into addition of generators S. The dimension of this vector space (over the complex numbers) is the **order** of G, that is, the number of linearly independent generators of the group.

If R is a rotation, it does not change the volume element of the coordinate space that it rotates, that is, $\det(\mathsf{R}) = 1$, and we may use Eq. (3.171) to see that

$$\det(\mathsf{R}) = \exp\big(\mathrm{trace}(\ln\mathsf{R})\big) = \exp\big(i\varepsilon\,\mathrm{trace}(\mathsf{S})\big) = 1$$

implies $\varepsilon\,\mathrm{trace}(\mathsf{S}) = 0$ and, upon dividing by the small but nonzero parameter ε, that **generators are traceless**,

$$\mathrm{trace}(\mathsf{S}) = 0. \tag{4.11}$$

This is the case not only for the rotation groups $\mathsf{SO}(n)$ but also for unitary groups $\mathsf{SU}(n)$.

If R of G in Eq. (4.10) is unitary, then $S^\dagger = S$ is Hermitian, which is also the case for $\mathsf{SO}(n)$ and $\mathsf{SU}(n)$. This explains why the extra i has been inserted in Eq. (4.10).

Next we go around the unity in four steps, similar to parallel transport in differential geometry. We expand the group elements

$$\begin{aligned} \mathsf{R}_i &= \exp(i\varepsilon_i\mathsf{S}_i) = 1 + i\varepsilon_i\mathsf{S}_i - \tfrac{1}{2}\varepsilon_i^2\mathsf{S}_i^2 + \cdots, \\ \mathsf{R}_i^{-1} &= \exp(-i\varepsilon_i\mathsf{S}_i) = 1 - i\varepsilon_i\mathsf{S}_i - \tfrac{1}{2}\varepsilon_i^2\mathsf{S}_i^2 + \cdots, \end{aligned} \tag{4.12}$$

to second order in the small group parameter ε_i because the linear terms and several quadratic terms all cancel in the product (Fig. 4.1)

$$\begin{aligned} \mathsf{R}_i^{-1}\mathsf{R}_j^{-1}\mathsf{R}_i\mathsf{R}_j &= 1 + \varepsilon_i\varepsilon_j[\mathsf{S}_j, \mathsf{S}_i] + \cdots, \\ &= 1 + \varepsilon_i\varepsilon_j \sum_k c_{ji}^k \mathsf{S}_k + \cdots, \end{aligned} \tag{4.13}$$

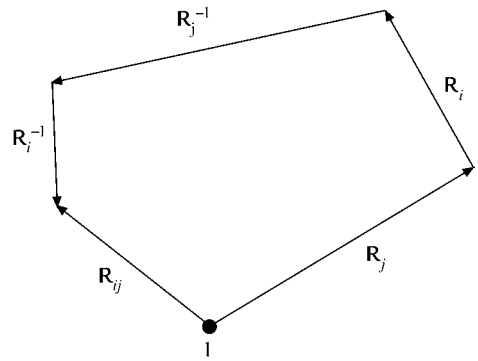

FIGURE 4.1 Illustration of Eq. (4.13).

when Eq. (4.12) is substituted into Eq. (4.13). The last line holds because the product in Eq. (4.13) is again a group element, R_{ij}, close to the unity in the group G. Hence its exponent must be a linear combination of the generators S_k, and its infinitesimal group parameter has to be proportional to the product $\varepsilon_i \varepsilon_j$. Comparing both lines in Eq. (4.13) we find the **closure** relation of the generators of the Lie group G,

$$[\mathsf{S}_i, \mathsf{S}_j] = \sum_k c_{ij}^k \mathsf{S}_k. \tag{4.14}$$

The coefficients c_{ij}^k are the structure constants of the group G. Since the commutator in Eq. (4.14) is antisymmetric in i and j, so are the structure constants in the lower indices,

$$c_{ij}^k = -c_{ji}^k. \tag{4.15}$$

If the commutator in Eq. (4.14) is taken as a multiplication law of generators, we see that the vector space of generators becomes an algebra, the **Lie algebra** $\mathcal{G}$ of the group G. An **algebra** has two group structures, a commutative product denoted by a $+$ symbol (this is the addition of infinitesimal generators of a Lie group) and a multiplication (the commutator of generators). Often an algebra is a vector space with a multiplication, such as a ring of square matrices. For $\mathsf{SU}(l+1)$ the Lie algebra is called $\mathcal{A}_l$, for $\mathsf{SO}(2l+1)$ it is $\mathcal{B}_l$, and for $\mathsf{SO}(2l)$ it is $\mathcal{D}_l$, where $l = 1, 2, \ldots$ is a positive integer, later called the **rank** of the Lie group G or of its algebra $\mathcal{G}$.

Finally, the **Jacobi identity** holds for all double commutators

$$\big[[\mathsf{S}_i, \mathsf{S}_j], \mathsf{S}_k\big] + \big[[\mathsf{S}_j, \mathsf{S}_k], \mathsf{S}_i\big] + \big[[\mathsf{S}_k, \mathsf{S}_i], \mathsf{S}_j\big] = 0, \tag{4.16}$$

which is easily verified using the definition of any commutator $[A, B] \equiv AB - BA$. When Eq. (4.14) is substituted into Eq. (4.16) we find another constraint on structure constants,

$$\sum_m \big\{c_{ij}^m [\mathsf{S}_m, \mathsf{S}_k] + c_{jk}^m [\mathsf{S}_m, \mathsf{S}_i] + c_{ki}^m [\mathsf{S}_m, \mathsf{S}_j]\big\} = 0. \tag{4.17}$$

Upon inserting Eq. (4.14) again, Eq. (4.17) implies that

$$\sum_{mn} \big\{c_{ij}^m c_{mk}^n \mathsf{S}_n + c_{jk}^m c_{mi}^n \mathsf{S}_n + c_{ki}^m c_{mj}^n \mathsf{S}_n\big\} = 0, \tag{4.18}$$

where the common factor S_n (and the sum over n) may be dropped because the generators are linearly independent. Hence

$$\sum_m \left\{ c_{ij}^m c_{mk}^n + c_{jk}^m c_{mi}^n + c_{ki}^m c_{mj}^n \right\} = 0. \tag{4.19}$$

The relations (4.14), (4.15), and (4.19) form the basis of Lie algebras from which finite elements of the Lie group near its unity can be reconstructed.

Returning to Eq. (4.5), the inverse of R is $\mathsf{R}^{-1} = \exp(-i\varepsilon \mathsf{S})$. We expand H_R according to the Baker–Hausdorff formula, Eq. (3.172),

$$H = H_R = \exp(i\varepsilon \mathsf{S}) H \exp(-i\varepsilon \mathsf{S}) = H + i\varepsilon[\mathsf{S}, H] - \tfrac{1}{2}\varepsilon^2 \big[\mathsf{S}[\mathsf{S}, H]\big] + \cdots \tag{4.20}$$

We drop H from Eq. (4.20), divide by the small (but nonzero), ε, and let $\varepsilon \to 0$. Then Eq. (4.20) implies that the commutator

$$\boxed{[\mathsf{S}, H] = 0.} \tag{4.21}$$

If S and H are Hermitian matrices, Eq. (4.21) implies that S and H can be simultaneously diagonalized and have common eigenvectors (for matrices, see Section 3.5; for operators, see Schur's lemma in Section 4.3). If S and H are differential operators like the Hamiltonian and orbital angular momentum in quantum mechanics, then Eq. (4.21) implies that S and H have common eigenfunctions and that the degenerate eigenvalues of H can be distinguished by the eigenvalues of the generators S. These eigenfunctions and eigenvalues, s, are solutions of separate differential equations, $\mathsf{S}\psi_s = s\psi_s$, so group theory (that is, symmetries) leads to a separation of variables for a partial differential equation that is invariant under the transformations of the group.

For example, let us take the single-particle Hamiltonian

$$H = -\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} + \frac{\hbar^2}{2mr^2}\mathbf{L}^2 + V(r)$$

that is invariant under $\mathsf{SO}(3)$ and, therefore, a function of the radial distance r, the radial gradient, and the rotationally invariant operator $\mathbf{L}^2$ of $\mathsf{SO}(3)$. Upon replacing the orbital angular momentum operator $\mathbf{L}^2$ by its eigenvalue $l(l+1)$ we obtain the radial Schrödinger equation (ODE),

$$H R_l(r) = \left[-\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{d}{dr} r^2 \frac{d}{dr} + \frac{\hbar^2 l(l+1)}{2mr^2} + V(r) \right] R_l(r) = E_l R_l(r),$$

where $R_l(r)$ is the radial wave function.

For cylindrical symmetry, the invariance of H under rotations about the z-axis would require H to be independent of the rotation angle φ, leading to the ODE

$$H R_m(z, \rho) = E_m R_m(z, \rho),$$

with m the eigenvalue of $L_z = -i\,\partial/\partial\varphi$, the z-component of the orbital angular momentum operator. For more examples, see the separation of variables method for partial differential equations in Section 9.3 and special functions in Chapter 12. This is by far the most important application of group theory in quantum mechanics.

In the next subsections we shall study orthogonal and unitary groups as examples to understand better the general concepts of this section.

Rotation Groups SO(2) and SO(3)

For SO(2) as defined by Eq. (4.3) there is only one linearly independent generator, σ_2, and the order of SO(2) is 1. We get σ_2 from Eq. (4.9) by differentiation at the unity of SO(2), that is, $\varphi = 0$,

$$-i\,d\mathsf{R}(\varphi)/d\varphi|_{\varphi=0} = -i\begin{pmatrix} -\sin\varphi & \cos\varphi \\ -\cos\varphi & -\sin\varphi \end{pmatrix}\Bigg|_{\varphi=0} = -i\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \sigma_2. \tag{4.22}$$

For the rotations $\mathsf{R}_z(\varphi)$ about the z-axis described by Eq. (4.1), the generator is given by

$$-i\,d\mathsf{R}_z(\varphi)/d\varphi|_{\varphi=0} = \mathsf{S}_z = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{4.23}$$

where the factor i is inserted to make S_z Hermitian. The rotation $\mathsf{R}_z(\delta\varphi)$ through an infinitesimal angle $\delta\varphi$ may then be expanded to first order in the small $\delta\varphi$ as

$$\mathsf{R}_z(\delta\varphi) = 1_3 + i\delta\varphi\mathsf{S}_z. \tag{4.24}$$

A finite rotation $\mathsf{R}(\varphi)$ may be compounded of successive infinitesimal rotations

$$\mathsf{R}_z(\delta\varphi_1 + \delta\varphi_2) = (1 + i\delta\varphi_1\mathsf{S}_z)(1 + i\delta\varphi_2\mathsf{S}_z). \tag{4.25}$$

Let $\delta\varphi = \varphi/N$ for N rotations, with $N \to \infty$. Then

$$\mathsf{R}_z(\varphi) = \lim_{N\to\infty}\left[1 + (i\varphi/N)\mathsf{S}_z\right]^N = \exp(i\varphi\mathsf{S}_z). \tag{4.26}$$

This form identifies S_z as the generator of the group R_z, an abelian subgroup of SO(3), the group of rotations in three dimensions with determinant $+1$. Each 3×3 matrix $\mathsf{R}_z(\varphi)$ is orthogonal, hence unitary, and trace$(\mathsf{S}_z) = 0$, in accord with Eq. (4.11).

By differentiation of the coordinate rotations

$$\mathsf{R}_x(\psi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & \sin\psi \\ 0 & -\sin\psi & \cos\psi \end{pmatrix}, \qquad \mathsf{R}_y(\theta) = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}, \tag{4.27}$$

we get the generators

$$\mathsf{S}_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \qquad \mathsf{S}_y = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} \tag{4.28}$$

of $\mathsf{R}_x(\mathsf{R}_y)$, the subgroup of rotations about the x- (y-)axis.

Rotation of Functions and Orbital Angular Momentum

In the foregoing discussion the group elements are matrices that rotate the coordinates. Any physical system being described is held fixed. Now let us hold the coordinates fixed and rotate a function $\psi(x, y, z)$ relative to our fixed coordinates. With R to rotate the coordinates,

$$\mathbf{x}' = \mathsf{R}\mathbf{x}, \tag{4.29}$$

we define R on ψ by

$$\mathsf{R}\psi(x, y, z) = \psi'(x, y, z) \equiv \psi(\mathbf{x}'). \tag{4.30}$$

In words, R operates on the function ψ, creating a **new function** ψ' that is numerically equal to $\psi(\mathbf{x}')$, where $\mathbf{x}'$ are the coordinates rotated by R. If R rotates the coordinates counterclockwise, the effect of R is to rotate the pattern of the function ψ clockwise.

Returning to Eqs. (4.30) and (4.1), consider an infinitesimal rotation again, $\varphi \to \delta\varphi$. Then, using R_z Eq. (4.1), we obtain

$$\mathsf{R}_z(\delta\varphi)\psi(x, y, z) = \psi(x + y\delta\varphi, y - x\delta\varphi, z). \tag{4.31}$$

The right side may be expanded to first order in the small $\delta\varphi$ to give

$$\begin{aligned}\mathsf{R}_z(\delta\varphi)\psi(x, y, z) &= \psi(x, y, z) - \delta\varphi\{x\partial\psi/\partial y - y\partial\psi/\partial x\} + O(\delta\varphi)^2 \\ &= (1 - i\delta\varphi L_z)\psi(x, y, z),\end{aligned} \tag{4.32}$$

the differential expression in curly brackets being the orbital angular momentum iL_z (Exercise 1.8.7). Since a rotation of first φ and then $\delta\varphi$ about the z-axis is given by

$$\mathsf{R}_z(\varphi + \delta\varphi)\psi = \mathsf{R}_z(\delta\varphi)\mathsf{R}_z(\varphi)\psi = (1 - i\delta\varphi L_z)\mathsf{R}_z(\varphi)\psi, \tag{4.33}$$

we have (as an operator equation)

$$\frac{d\mathsf{R}_z}{d\varphi} = \lim_{\delta\varphi\to 0} \frac{\mathsf{R}_z(\varphi + \delta\varphi) - \mathsf{R}_z(\varphi)}{\delta\varphi} = -iL_z\mathsf{R}_z(\varphi). \tag{4.34}$$

In this form Eq. (4.34) integrates immediately to

$$\mathsf{R}_z(\varphi) = \exp(-i\varphi L_z). \tag{4.35}$$

Note that $\mathsf{R}_z(\varphi)$ rotates functions (clockwise) relative to fixed coordinates and that L_z is the z component of the orbital angular momentum $\mathbf{L}$. The constant of integration is fixed by the boundary condition $\mathsf{R}_z(0) = 1$.

As suggested by Eq. (4.32), L_z is connected to S_z by

$$L_z = (x, y, z)\mathsf{S}_z \begin{pmatrix} \partial/\partial x \\ \partial/\partial y \\ \partial/\partial z \end{pmatrix} = -i\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right), \tag{4.36}$$

so L_x, L_y, and L_z satisfy the same commutation relations,

$$\boxed{[L_i, L_j] = i\varepsilon_{ijk}L_k,} \tag{4.37}$$

as S_x, S_y, and S_z and yield the same structure constants $i\varepsilon_{ijk}$ of $\mathsf{SO}(3)$.

SU(2) — SO(3) Homomorphism

Since unitary 2×2 matrices transform complex two-dimensional vectors preserving their norm, they represent the most general transformations of (a basis in the Hilbert space of) spin $\frac{1}{2}$ wave functions in nonrelativistic quantum mechanics. The basis states of this system are conventionally chosen to be

$$|\uparrow\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |\downarrow\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

corresponding to spin $\frac{1}{2}$ up and down states, respectively. We can show that the **special unitary** group SU(2) of unitary 2×2 matrices with determinant $+1$ has all three Pauli matrices σ_i as generators (while the rotations of Eq. (4.3) form a one-dimensional abelian subgroup). So SU(2) is of order 3 and depends on three real continuous parameters ξ, η, ζ, which are often called the **Cayley–Klein** parameters. To construct its general element, we start with the observation that orthogonal 2×2 matrices are real unitary matrices, so they form a subgroup of SU(2). We also see that

$$\begin{pmatrix} e^{i\alpha} & 0 \\ 0 & e^{-i\alpha} \end{pmatrix}$$

is unitary for real angle α with determinant $+1$. So these simple and manifestly unitary matrices form another subgroup of SU(2) from which we can obtain all elements of SU(2), that is, the general 2×2 unitary matrix of determinant $+1$. For a two-component spin $\frac{1}{2}$ wave function of quantum mechanics this diagonal unitary matrix corresponds to multiplication of the spin-up wave function with a phase factor $e^{i\alpha}$ and the spin-down component with the inverse phase factor. Using the real angle η instead of φ for the rotation matrix and then multiplying by the diagonal unitary matrices, we construct a 2×2 unitary matrix that depends on three parameters and clearly is a more general element of SU(2):

$$\begin{aligned}
&\begin{pmatrix} e^{i\alpha} & 0 \\ 0 & e^{-i\alpha} \end{pmatrix} \begin{pmatrix} \cos\eta & \sin\eta \\ -\sin\eta & \cos\eta \end{pmatrix} \begin{pmatrix} e^{i\beta} & 0 \\ 0 & e^{-i\beta} \end{pmatrix} \\
&\quad = \begin{pmatrix} e^{i\alpha}\cos\eta & e^{i\alpha}\sin\eta \\ -e^{-i\alpha}\sin\eta & e^{-i\alpha}\cos\eta \end{pmatrix} \begin{pmatrix} e^{i\beta} & 0 \\ 0 & e^{-i\beta} \end{pmatrix} \\
&\quad = \begin{pmatrix} e^{i(\alpha+\beta)}\cos\eta & e^{i(\alpha-\beta)}\sin\eta \\ -e^{-i(\alpha-\beta)}\sin\eta & e^{-i(\alpha+\beta)}\cos\eta \end{pmatrix}.
\end{aligned}$$

Defining $\alpha + \beta \equiv \xi, \alpha - \beta \equiv \zeta$, we have in fact constructed the general element of SU(2):

$$\mathsf{U}(\xi, \eta, \zeta) = \begin{pmatrix} e^{i\xi}\cos\eta & e^{i\zeta}\sin\eta \\ -e^{-i\zeta}\sin\eta & e^{-i\xi}\cos\eta \end{pmatrix} = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}. \tag{4.38}$$

To see this, we write the general SU(2) element as $\mathsf{U} = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with complex numbers a, b, c, d so that $\det(\mathsf{U}) = 1$. Writing unitarity, $\mathsf{U}^\dagger = \mathsf{U}^{-1}$, and using Eq. (3.50) for the

inverse we obtain

$$\begin{pmatrix} a^* & c^* \\ b^* & d^* \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

implying $c = -b^*, d = a^*$, as shown in Eq. (4.38). It is easy to check that the determinant $\det(\mathsf{U}) = 1$ and that $\mathsf{U}^\dagger\mathsf{U} = 1 = \mathsf{U}\mathsf{U}^\dagger$ hold.

To get the generators, we differentiate (and drop irrelevant overall factors):

$$-i\,\partial\mathsf{U}/\partial\xi_{|\xi=0,\eta=0} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \sigma_3, \tag{4.39a}$$

$$-i\,\partial\mathsf{U}/\partial\eta_{|\eta=0,\zeta=0} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \sigma_2. \tag{4.39b}$$

To avoid a factor $1/\sin\eta$ for $\eta \to 0$ upon differentiating with respect to ζ, we use instead the right-hand side of Eq. (4.38) for U for pure imaginary $b = i\beta$ with $\beta \to 0$, so $a = \sqrt{1-\beta^2}$ from $|a|^2 + |b|^2 = a^2 + \beta^2 = 1$. Differentiating such a U, we get the third generator,

$$-i\,\frac{\partial}{\partial\beta}\begin{pmatrix} \sqrt{1-\beta^2} & i\beta \\ i\beta & \sqrt{1-\beta^2} \end{pmatrix}\Bigg|_{\beta=0} = -i\begin{pmatrix} -\frac{\beta}{\sqrt{1-\beta^2}} & i \\ -i & \frac{\beta}{\sqrt{1-\beta^2}} \end{pmatrix}\Bigg|_{\beta=0} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \sigma_1. \tag{4.39c}$$

The Pauli matrices are all traceless and Hermitian.

With the Pauli matrices as generators, the elements $\mathsf{U}_1, \mathsf{U}_2, \mathsf{U}_3$ of $\mathsf{SU}(2)$ may be generated by

$$\mathsf{U}_1 = \exp(ia_1\sigma_1/2), \qquad \mathsf{U}_2 = \exp(ia_2\sigma_2/2), \qquad \mathsf{U}_3 = \exp(ia_3\sigma_3/2). \tag{4.40}$$

The three parameters a_i are real. The extra factor $1/2$ is present in the exponents to make $S_i = \sigma_i/2$ satisfy the same commutation relations,

$$\boxed{[\mathsf{S}_i, \mathsf{S}_j] = i\varepsilon_{ijk}\mathsf{S}_k,} \tag{4.41}$$

as the angular momentum in Eq. (4.37).

To connect and compare our results, Eq. (4.3) gives a rotation operator for rotating the Cartesian coordinates in the three-space $\mathbb{R}^3$. Using the angular momentum matrix S_3, we have as the corresponding rotation operator in two-dimensional (complex) space $\mathsf{R}_z(\varphi) = \exp(i\varphi\sigma_3/2)$. For rotating the two-component vector wave function (spinor) or a spin $1/2$ particle relative to fixed coordinates, the corresponding rotation operator is $\mathsf{R}_z(\varphi) = \exp(-i\varphi\sigma_3/2)$ according to Eq. (4.35).

More generally, using in Eq. (4.40) the Euler identity, Eq. (3.170a), we obtain

$$\mathsf{U}_j = \cos\left(\frac{a_j}{2}\right) + i\sigma_j \sin\left(\frac{a_j}{2}\right). \tag{4.42}$$

Here the parameter a_j appears as an angle, the coefficient of an angular momentum matrix-like φ in Eq. (4.26). The selection of Pauli matrices corresponds to the Euler angle rotations described in Section 3.3.

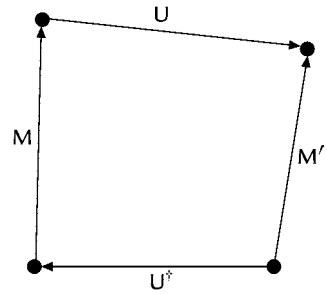

FIGURE 4.2 Illustration of $\mathsf{M}' = \mathsf{UMU}^\dagger$ in Eq. (4.43).

As just seen, the elements of SU(2) describe rotations in a two-dimensional complex space that leave $|z_1|^2 + |z_2|^2$ invariant. The determinant is $+1$. There are three independent real parameters. Our real orthogonal group SO(3) clearly describes rotations in ordinary three-dimensional space with the important characteristic of leaving $x^2 + y^2 + z^2$ **invariant**. Also, there are three independent real parameters. The rotation interpretations and the equality of numbers of parameters suggest the existence of some correspondence between the groups SU(2) and SO(3). Here we develop this correspondence.

The operation of SU(2) on a matrix is given by a unitary transformation, Eq. (4.5), with $\mathsf{R} = \mathsf{U}$ and Fig. 4.2:

$$\mathsf{M}' = \mathsf{UMU}^\dagger. \tag{4.43}$$

Taking M to be a 2×2 matrix, we note that any 2×2 matrix may be written as a linear combination of the unit matrix and the three Pauli matrices of Section 3.4. Let M be the zero-trace matrix,

$$\mathsf{M} = x\sigma_1 + y\sigma_2 + z\sigma_3 = \begin{pmatrix} z & x - iy \\ x + iy & -z \end{pmatrix}, \tag{4.44}$$

the unit matrix not entering. Since the trace is invariant under a unitary similarity transformation (Exercise 3.3.9), M' must have the same form,

$$\mathsf{M}' = x'\sigma_1 + y'\sigma_2 + z'\sigma_3 = \begin{pmatrix} z' & x' - iy' \\ x' + iy' & -z' \end{pmatrix}. \tag{4.45}$$

The determinant is also invariant under a unitary transformation (Exercise 3.3.10). Therefore

$$-\left(x^2 + y^2 + z^2\right) = -\left(x'^2 + y'^2 + z'^2\right), \tag{4.46}$$

or $x^2 + y^2 + z^2$ is invariant under this operation of SU(2), just as with SO(3). Operations of SU(2) on M must produce rotations of the coordinates x, y, z appearing therein. This suggests that SU(2) and SO(3) may be isomorphic or at least homomorphic.

We approach the problem of what this operation of SU(2) corresponds to by considering special cases. Returning to Eq. (4.38), let $a = e^{i\xi}$ and $b = 0$, or

$$\mathsf{U}_3 = \begin{pmatrix} e^{i\xi} & 0 \\ 0 & e^{-i\xi} \end{pmatrix}. \tag{4.47}$$

In anticipation of Eq. (4.51), this U is given a subscript 3.

Carrying out a unitary similarity transformation, Eq. (4.43), on each of the three Pauli σ's of SU(2), we have

$$\begin{aligned} \mathsf{U}_3\sigma_1\mathsf{U}_3^\dagger &= \begin{pmatrix} e^{i\xi} & 0 \\ 0 & e^{-i\xi} \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} e^{-i\xi} & 0 \\ 0 & e^{i\xi} \end{pmatrix} \\ &= \begin{pmatrix} 0 & e^{2i\xi} \\ e^{-2i\xi} & 0 \end{pmatrix}. \end{aligned} \tag{4.48}$$

We reexpress this result in terms of the Pauli σ_i, as in Eq. (4.44), to obtain

$$\mathsf{U}_3 x\sigma_1\mathsf{U}_3^\dagger = x\sigma_1\cos 2\xi - x\sigma_2\sin 2\xi. \tag{4.49}$$

Similarly,

$$\begin{aligned} \mathsf{U}_3 y\sigma_2\mathsf{U}_3^\dagger &= y\sigma_1\sin 2\xi + y\sigma_2\cos 2\xi, \\ \mathsf{U}_3 z\sigma_3\mathsf{U}_3^\dagger &= z\sigma_3. \end{aligned} \tag{4.50}$$

From these double angle expressions we see that we should start with a **halfangle**: $\xi = \alpha/2$. Then, adding Eqs. (4.49) and (4.50) and comparing with Eqs. (4.44) and (4.45), we obtain

$$\begin{aligned} x' &= x\cos\alpha + y\sin\alpha \\ y' &= -x\sin\alpha + y\cos\alpha \\ z' &= z. \end{aligned} \tag{4.51}$$

The 2×2 unitary transformation using $\mathsf{U}_3(\alpha)$ is equivalent to the rotation operator $\mathsf{R}(\alpha)$ of Eq. (4.3).

The correspondence of

$$\mathsf{U}_2(\beta) = \begin{pmatrix} \cos\beta/2 & \sin\beta/2 \\ -\sin\beta/2 & \cos\beta/2 \end{pmatrix} \tag{4.52}$$

and $\mathsf{R}_y(\beta)$ and of

$$\mathsf{U}_1(\varphi) = \begin{pmatrix} \cos\varphi/2 & i\sin\varphi/2 \\ i\sin\varphi/2 & \cos\varphi/2 \end{pmatrix} \tag{4.53}$$

and $\mathsf{R}_1(\varphi)$ follow similarly. Note that $\mathsf{U}_k(\psi)$ has the general form

$$\mathsf{U}_k(\psi) = 1_2\cos\psi/2 + i\sigma_k\sin\psi/2, \tag{4.54}$$

where $k = 1, 2, 3$.

The correspondence

$$\mathsf{U}_3(\alpha)=\begin{pmatrix} e^{i\alpha/2} & 0 \\ 0 & e^{-i\alpha/2}\end{pmatrix}\leftrightarrow\begin{pmatrix}\cos\alpha & \sin\alpha & 0\\ -\sin\alpha & \cos\alpha & 0\\ 0 & 0 & 1\end{pmatrix}=\mathsf{R}_z(\alpha) \tag{4.55}$$

is not a simple one-to-one correspondence. Specifically, as α in R_z ranges from 0 to 2π, the parameter in U_3, $\alpha/2$, goes from 0 to π. We find

$$\begin{aligned}\mathsf{R}_z(\alpha+2\pi)&=\mathsf{R}_z(\alpha)\\ \mathsf{U}_3(\alpha+2\pi)&=\begin{pmatrix}-e^{i\alpha/2} & 0\\ 0 & -e^{-i\alpha/2}\end{pmatrix}=-\mathsf{U}_3(\alpha).\end{aligned} \tag{4.56}$$

Therefore **both** $\mathsf{U}_3(\alpha)$ and $\mathsf{U}_3(\alpha+2\pi)=-\mathsf{U}_3(\alpha)$ correspond to $\mathsf{R}_z(\alpha)$. The correspondence is 2 to 1, or SU(2) and SO(3) are **homomorphic**. This establishment of the correspondence between the representations of SU(2) and those of SO(3) means that the known representations of SU(2) automatically provide us with the representations of SO(3).

Combining the various rotations, we find that a unitary transformation using

$$\mathsf{U}(\alpha,\beta,\gamma)=\mathsf{U}_3(\gamma)\mathsf{U}_2(\beta)\mathsf{U}_3(\alpha) \tag{4.57}$$

corresponds to the general Euler rotation $\mathsf{R}_z(\gamma)\mathsf{R}_y(\beta)\mathsf{R}_z(\alpha)$. By direct multiplication,

$$\begin{aligned}\mathsf{U}(\alpha,\beta,\gamma)&=\begin{pmatrix}e^{i\gamma/2} & 0\\ 0 & e^{-i\gamma/2}\end{pmatrix}\begin{pmatrix}\cos\beta/2 & \sin\beta/2\\ -\sin\beta/2 & \cos\beta/2\end{pmatrix}\begin{pmatrix}e^{i\alpha/2} & 0\\ 0 & e^{-i\alpha/2}\end{pmatrix}\\ &=\begin{pmatrix}e^{i(\gamma+\alpha)/2}\cos\beta/2 & e^{i(\gamma-\alpha)/2}\sin\beta/2\\ -e^{-i(\gamma-\alpha)/2}\sin\beta/2 & e^{-i(\gamma+\alpha)/2}\cos\beta/2\end{pmatrix}.\end{aligned} \tag{4.58}$$

This is our alternate general form, Eq. (4.38), with

$$\xi=(\gamma+\alpha)/2,\qquad \eta=\beta/2,\qquad \zeta=(\gamma-\alpha)/2. \tag{4.59}$$

Thus, from Eq. (4.58) we may identify the parameters of Eq. (4.38) as

$$\begin{aligned}a&=e^{i(\gamma+\alpha)/2}\cos\beta/2\\ b&=e^{i(\gamma-\alpha)/2}\sin\beta/2.\end{aligned} \tag{4.60}$$

SU(2)-Isospin and SU(3)-Flavor Symmetry

The application of group theory to "elementary" particles has been labeled by Wigner the third stage of group theory and physics. The first stage was the search for the 32 crystallographic point groups and the 230 space groups giving crystal symmetries — Section 4.7. The second stage was a search for representations such as of SO(3) and SU(2) — Section 4.2. Now in this stage, physicists are back to a search for groups.

In the 1930s to 1960s the study of strongly interacting particles of nuclear and high-energy physics led to the SU(2) isospin group and the SU(3) flavor symmetry. In the 1930s, after the neutron was discovered, Heisenberg proposed that the nuclear forces were charge

Table 4.1 Baryons with Spin $\frac{1}{2}$ Even Parity

		Mass (MeV)	Y	I	I_3
	Ξ^-	1321.32			$-\frac{1}{2}$
Ξ			-1	$\frac{1}{2}$	
	Ξ^0	1314.9			$+\frac{1}{2}$
	Σ^-	1197.43			-1
Σ	Σ^0	1192.55	0	1	0
	Σ^+	1189.37			$+1$
Λ	Λ	1115.63	0	0	0
	n	939.566			$-\frac{1}{2}$
N			1	$\frac{1}{2}$	
	p	938.272			$+\frac{1}{2}$

independent. The neutron mass differs from that of the proton by only 1.6%. If this tiny mass difference is ignored, the neutron and proton may be considered as two charge (or isospin) states of a doublet, called the **nucleon**. The isospin $\mathbf{I}$ has z-projection $I_3 = 1/2$ for the proton and $I_3 = -1/2$ for the neutron. Isospin has nothing to do with spin (the particle's intrinsic angular momentum), but the two-component isospin state obeys the same mathematical relations as the spin $1/2$ state. For the nucleon, $\mathbf{I} = \boldsymbol{\tau}/2$ are the usual Pauli matrices and the $\pm 1/2$ isospin states are eigenvectors of the Pauli matrix $\tau_3 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. Similarly, the three charge states of the pion (π^+, π^0, π^-) form a triplet. The pion is the lightest of all strongly interacting particles and is the carrier of the nuclear force at long distances, much like the photon is that of the electromagnetic force. The strong interaction treats alike members of these particle families, or multiplets, and conserves isospin. The symmetry is the SU(2) isospin group.

By the 1960s particles produced as resonances by accelerators had proliferated. The eight shown in Table 4.1 attracted particular attention.[5] The relevant conserved quantum numbers that are analogs and generalizations of L_z and $\mathbf{L}^2$ from SO(3) are I_3 and I^2 for isospin and Y for **hypercharge**. Particles may be grouped into charge or isospin multiplets. Then the hypercharge may be taken as twice the average charge of the multiplet. For the nucleon, that is, the neutron–proton doublet, $Y = 2 \cdot \frac{1}{2}(0 + 1) = 1$. The hypercharge and isospin values are listed in Table 4.1 for baryons like the nucleon and its (approximately degenerate) partners. They form an octet, as shown in Fig. 4.3, after which the corresponding symmetry is called the **eightfold way**. In 1961 Gell-Mann, and independently Ne'eman, suggested that the strong interaction should be (approximately) invariant under a three-dimensional special unitary group, SU(3), that is, has SU(3) **flavor symmetry**.

The choice of SU(3) was based first on the two conserved and independent quantum numbers, $H_1 = I_3$ and $H_2 = Y$ (that is, generators with $[I_3, Y] = 0$, not Casimir invariants; see the summary in Section 4.3) that call for a group of rank 2. Second, the group had to have an eight-dimensional representation to account for the nearly degenerate baryons and four similar octets for the mesons. In a sense, SU(3) is the simplest generalization of SU(2) isospin. Three of its generators are zero-trace Hermitian 3×3 matrices that contain

[5] All masses are given in energy units, 1 MeV $= 10^6$ eV.

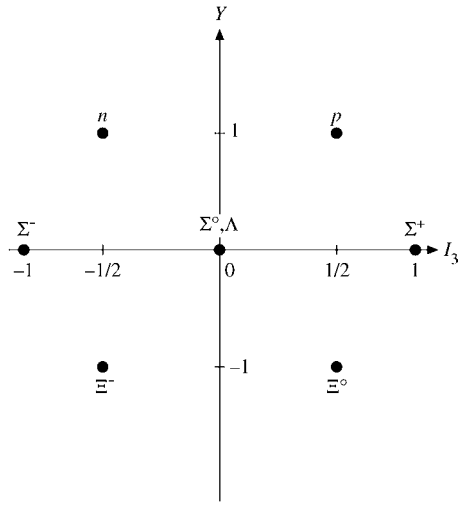

FIGURE 4.3 Baryon octet weight diagram for SU(3).

the 2×2 isospin Pauli matrices τ_i in the upper left corner,

$$\lambda_i = \begin{pmatrix} \tau_i & & 0 \\ & & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad i = 1, 2, 3. \tag{4.61a}$$

Thus, the SU(2)-isospin group is a subgroup of SU(3)-flavor with $I_3 = \lambda_3/2$. Four other generators have the off-diagonal 1's of τ_1, and $-i, i$ of τ_2 in all other possible locations to form zero-trace Hermitian 3×3 matrices,

$$\begin{aligned} \lambda_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \\ \lambda_6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}. \end{aligned} \tag{4.61b}$$

The second diagonal generator has the two-dimensional unit matrix 1_2 in the upper left corner, which makes it clearly independent of the SU(2)-isospin subgroup because of its nonzero trace in that subspace, and -2 in the third diagonal place to make it traceless,

$$\lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \tag{4.61c}$$

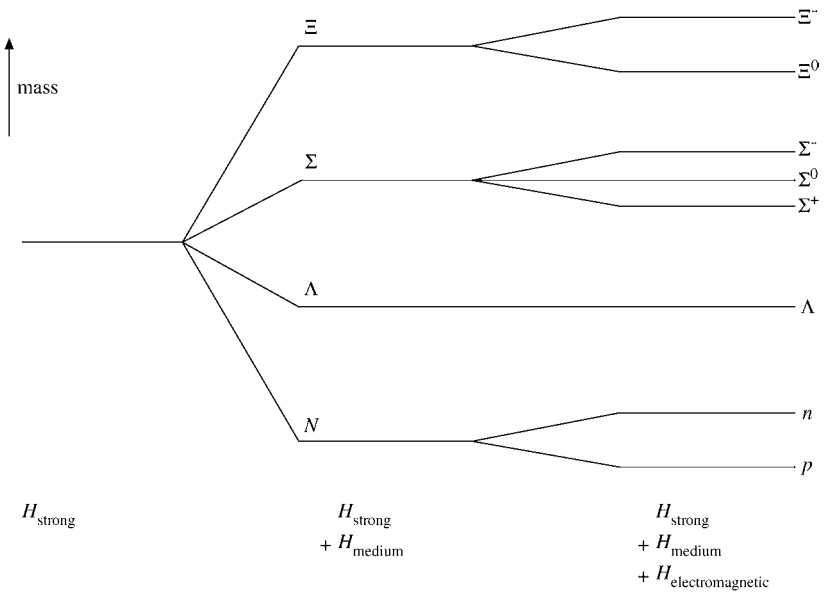

FIGURE 4.4 Baryon mass splitting.

Altogether there are $3^2 - 1 = 8$ generators for SU(3), which has order 8. From the commutators of these generators the structure constants of SU(3) can easily be obtained.

Returning to the SU(3) flavor symmetry, we imagine the Hamiltonian for our eight baryons to be composed of three parts:

$$H = H_{\text{strong}} + H_{\text{medium}} + H_{\text{electromagnetic}}. \tag{4.62}$$

The first part, H_{strong}, has the SU(3) symmetry and leads to the eightfold degeneracy. Introduction of the symmetry-breaking term, H_{medium}, removes part of the degeneracy, giving the four isospin multiplets (Ξ^-, Ξ^0), $(\Sigma^-, \Sigma^0, \Sigma^+)$, Λ, and $N = (p, n)$ different masses. These are still multiplets because H_{medium} has SU(2)-isospin symmetry. Finally, the presence of charge-dependent forces splits the isospin multiplets and removes the last degeneracy. This imagined sequence is shown in Fig. 4.4.

The octet representation is not the simplest SU(3) representation. The simplest representations are the triangular ones shown in Fig. 4.5, from which all others can be generated by generalized angular momentum coupling (see Section 4.4 on Young tableaux). The **fundamental representation** in Fig. 4.5a contains the u (up), d (down), and s (strange) quarks, and Fig. 4.5b contains the corresponding antiquarks. Since the meson octets can be obtained from the quark representations as $q\bar{q}$, with $3^2 = 8 + 1$ states, this suggests that mesons contain quarks (and antiquarks) as their constituents (see Exercise 4.4.3). The resulting quark model gives a successful description of hadronic spectroscopy. The resolution of its problem with the Pauli exclusion principle eventually led to the SU(3)-color gauge theory of the **strong interaction** called **quantum chromodynamics** (QCD).

To keep group theory and its very real accomplishment in proper perspective, we should emphasize that group theory identifies and formalizes symmetries. It classifies (and sometimes predicts) particles. But aside from saying that one part of the Hamiltonian has SU(2)

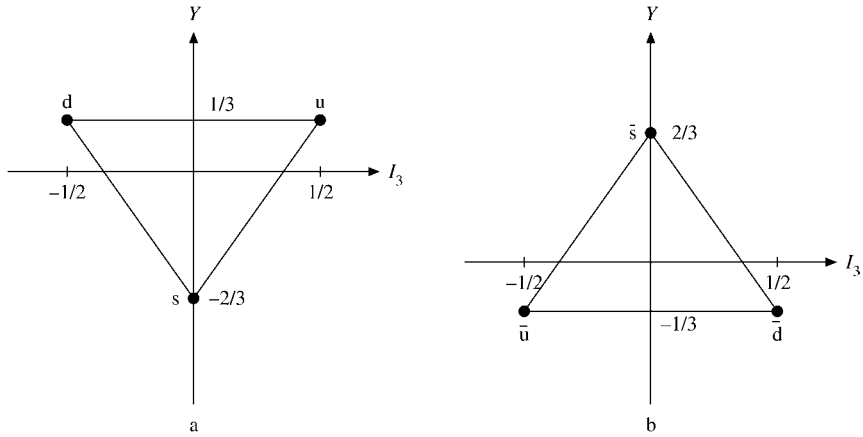

FIGURE 4.5 (a) Fundamental representation of SU(3), the weight diagram for the u, d, s quarks; (b) weight diagram for the antiquarks $\bar{u}, \bar{d}, \bar{s}$.

symmetry and another part has SU(3) symmetry, group theory says nothing about the particle interaction. Remember that the statement that the atomic potential is spherically symmetric tells us nothing about the radial dependence of the potential or of the wave function. In contrast, in a gauge theory the interaction is mediated by vector bosons (like the photon in quantum electrodynamics) and uniquely determined by the gauge covariant derivative (see Section 1.13).

Exercises

4.2.1 (i) Show that the Pauli matrices are the generators of SU(2) without using the parameterization of the general unitary 2×2 matrix in Eq. (4.38). (ii) Derive the eight independent generators λ_i of SU(3) similarly. Normalize them so that $\text{tr}(\lambda_i \lambda_j) = 2\delta_{ij}$. Then determine the structure constants of SU(3).
Hint. The λ_i are traceless and Hermitian 3×3 matrices.
(iii) Construct the quadratic Casimir invariant of SU(3).
Hint. Work by analogy with $\sigma_1^2 + \sigma_2^2 + \sigma_3^2$ of SU(2) or $\mathbf{L}^2$ of SO(3).

4.2.2 Prove that the general form of a 2×2 unitary, unimodular matrix is

$$\mathsf{U} = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}$$

with $a^*a + b^*b = 1$.

4.2.3 Determine three SU(2) subgroups of SU(3).

4.2.4 A **translation** operator $T(a)$ converts $\psi(x)$ to $\psi(x + a)$,

$$T(a)\psi(x) = \psi(x + a).$$

In terms of the (quantum mechanical) linear momentum operator $p_x = -id/dx$, show that $T(a) = \exp(iap_x)$, that is, p_x is the generator of translations.
Hint. Expand $\psi(x + a)$ as a Taylor series.

4.2.5 Consider the general SU(2) element Eq. (4.38) to be built up of three Euler rotations: (i) a rotation of $a/2$ about the z-axis, (ii) a rotation of $b/2$ about the new x-axis, and (iii) a rotation of $c/2$ about the new z-axis. (All rotations are counterclockwise.) Using the Pauli $\boldsymbol{\sigma}$ generators, show that these rotation angles are determined by

$$\begin{aligned} a &= \xi - \zeta + \tfrac{\pi}{2} = \alpha + \tfrac{\pi}{2} \\ b &= 2\eta \qquad\quad = \beta \\ c &= \xi + \zeta - \tfrac{\pi}{2} = \gamma - \tfrac{\pi}{2}. \end{aligned}$$

Note. The angles a and b here are not the a and b of Eq. (4.38).

4.2.6 Rotate a nonrelativistic wave function $\tilde{\psi} = (\psi_\uparrow, \psi_\downarrow)$ of spin $1/2$ about the z-axis by a small angle $d\theta$. Find the corresponding generator.

4.3 Orbital Angular Momentum

The classical concept of angular momentum, $\mathbf{L}_{\text{class}} = \mathbf{r} \times \mathbf{p}$, is presented in Section 1.4 to introduce the cross product. Following the usual Schrödinger representation of quantum mechanics, the classical linear momentum $\mathbf{p}$ is replaced by the operator $-i\boldsymbol{\nabla}$. The quantum mechanical orbital angular momentum **operator** becomes[6]

$$\mathbf{L}_{QM} = -i\mathbf{r} \times \boldsymbol{\nabla}. \tag{4.63}$$

This is used repeatedly in Sections 1.8, 1.9, and 2.4 to illustrate vector differential operators. From Exercise 1.8.8 the angular momentum components satisfy the commutation relations

$$[L_i, L_j] = i\varepsilon_{ijk} L_k. \tag{4.64}$$

The ε_{ijk} is the Levi-Civita symbol of Section 2.9. A summation over the index k is understood.

The differential operator corresponding to the square of the angular momentum may be determined from

$$\mathbf{L}^2 = \mathbf{L} \cdot \mathbf{L} = L_x^2 + L_y^2 + L_z^2 \tag{4.65}$$

$$\mathbf{L} \cdot \mathbf{L} = (\mathbf{r} \times \mathbf{p}) \cdot (\mathbf{r} \times \mathbf{p}), \tag{4.66}$$

which is the subject of Exercises 1.9.9 and 2.5.17(b). Since $\mathbf{L}^2$ as a scalar product is invariant under rotations, that is, a rotational scalar, we expect $[\mathbf{L}^2, L_i] = 0$, which can also be verified directly.

Equation (4.64) presents the basic commutation relations of the components of the quantum mechanical angular momentum. Indeed, within the framework of quantum mechanics and group theory, these commutation relations define an angular momentum operator. We shall use them now to construct the angular momentum eigenstates and find the eigenvalues. For the orbital angular momentum these are the spherical harmonics of Section 12.6.

[6] For simplicity, $\hbar$ is set equal to 1. This means that the angular momentum is measured in units of $\hbar$.

Ladder Operator Approach

Let us start with a general approach, where the angular momentum $\mathbf{J}$ we consider may represent an orbital angular momentum $\mathbf{L}$, a spin $\boldsymbol{\sigma}/2$, or a total angular momentum $\mathbf{L}+\boldsymbol{\sigma}/2$, etc. We assume that

1. $\mathbf{J}$ is an Hermitian operator whose components satisfy the commutation relations
$$[J_i, J_j] = i\varepsilon_{ijk}J_k, \qquad \left[\mathbf{J}^2, J_i\right] = 0. \tag{4.67}$$
Otherwise $\mathbf{J}$ is arbitrary. (See Exercise 4.3.l.)
2. $|\lambda M\rangle$ is simultaneously a normalized eigenfunction (or eigenvector) of J_z with eigenvalue M and an eigenfunction[7] of $\mathbf{J}^2$,
$$J_z|\lambda M\rangle = M|\lambda M\rangle, \qquad \mathbf{J}^2|\lambda M\rangle = \lambda|\lambda M\rangle, \qquad \langle\lambda M|\lambda M\rangle = 1. \tag{4.68}$$

We shall show that $\lambda = J(J+1)$ and then find other properties of the $|\lambda M\rangle$. The treatment will illustrate the generality and power of operator techniques, particularly the use of ladder operators.[8]

The **ladder operators** are defined as
$$J_+ = J_x + iJ_y, \qquad J_- = J_x - iJ_y. \tag{4.69}$$
In terms of these operators $\mathbf{J}^2$ may be rewritten as
$$\mathbf{J}^2 = \tfrac{1}{2}(J_+J_- + J_-J_+) + J_z^2. \tag{4.70}$$
From the commutation relations, Eq. (4.67), we find
$$\boxed{[J_z, J_+] = +J_+, \qquad [J_z, J_-] = -J_-, \qquad [J_+, J_-] = 2J_z.} \tag{4.71}$$
Since J_+ commutes with $\mathbf{J}^2$ (Exercise 4.3.1),
$$\mathbf{J}^2\big(J_+|\lambda M\rangle\big) = J_+\big(\mathbf{J}^2|\lambda M\rangle\big) = \lambda\big(J_+|\lambda M\rangle\big). \tag{4.72}$$
Therefore, $J_+|\lambda M\rangle$ is still an eigenfunction of $\mathbf{J}^2$ with eigenvalue λ, and similarly for $J_-|\lambda M\rangle$. But from Eq. (4.71),
$$J_zJ_+ = J_+(J_z+1), \tag{4.73}$$
or
$$J_z\big(J_+|\lambda M\rangle\big) = J_+(J_z+1)|\lambda M\rangle = (M+1)J_+|\lambda M\rangle. \tag{4.74}$$

[7] That $|\lambda M\rangle$ can be an eigenfunction of **both** J_z and $\mathbf{J}^2$ follows from $[J_z, \mathbf{J}^2] = 0$ in Eq. (4.67). For SU(2), $\langle\lambda M|\lambda M\rangle$ is the scalar product (of the bra and ket vector or spinors) in the bra-ket notation introduced in Section 3.1. For SO(3), $|\lambda M\rangle$ is a function $Y(\theta,\varphi)$ and $|\lambda M'\rangle$ is a function $Y'(\theta,\varphi)$ and the matrix element $\langle\lambda M|\lambda M'\rangle \equiv \int_{\varphi=0}^{2\pi}\int_{\theta=0}^{\pi} Y^*(\theta,\varphi)Y'(\theta,\varphi)\sin\theta\,d\theta\,d\varphi$ is their overlap. However, in our algebraic approach only the norm in Eq. (4.68) is used and matrix elements of the angular momentum operators are reduced to the norm by means of the eigenvalue equation for J_z, Eq. (4.68), and Eqs. (4.83) and (4.84).

[8] Ladder operators can be developed for other mathematical functions. Compare the next subsection, on other Lie groups, and Section 13.1, for Hermite polynomials.

Therefore, $J_+|\lambda M\rangle$ is still an eigenfunction of J_z but with eigenvalue $M+1$. J_+ has raised the eigenvalue by 1 and so is called a **raising operator**. Similarly, J_- lowers the eigenvalue by 1 and is called a **lowering operator**.

Taking expectation values and using $J_x^\dagger = J_x$, $J_y^\dagger = J_y$, we get

$$\langle \lambda M|\mathbf{J}^2 - J_z^2|\lambda M\rangle = \langle \lambda M|J_x^2 + J_y^2|\lambda M\rangle = |J_x|\lambda M\rangle|^2 + |J_y|\lambda M\rangle|^2$$

and see that $\lambda - M^2 \geq 0$, so M is bounded. Let J be the **largest** M. Then $J_+|\lambda J\rangle = 0$, which implies $J_-J_+|\lambda J\rangle = 0$. Hence, combining Eqs. (4.70) and (4.71) to get

$$\mathbf{J}^2 = J_-J_+ + J_z(J_z+1), \tag{4.75}$$

we find from Eq. (4.75) that

$$0 = J_-J_+|\lambda J\rangle = \big(\mathbf{J}^2 - J_z^2 - J_z\big)|\lambda J\rangle = \big(\lambda - J^2 - J\big)|\lambda J\rangle.$$

Therefore

$$\lambda = J(J+1) \geq 0, \tag{4.76}$$

with nonnegative J. We now relabel the states $|\lambda M\rangle \equiv |JM\rangle$. Similarly, let J' be the **smallest** M. Then $J_-|JJ'\rangle = 0$. From

$$\mathbf{J}^2 = J_+J_- + J_z(J_z-1), \tag{4.77}$$

we see that

$$0 = J_+J_-|JJ'\rangle = \big(\mathbf{J}^2 + J_z - J_z^2\big)|JJ'\rangle = \big(\lambda + J' - J'^2\big)|JJ'\rangle. \tag{4.78}$$

Hence

$$\lambda = J(J+1) = J'(J'-1) = (-J)(-J-1).$$

So $J' = -J$, and M runs in **integer steps** from $-J$ to $+J$,

$$-J \leq M \leq J. \tag{4.79}$$

Starting from $|JJ\rangle$ and applying J_- repeatedly, we reach all other states $|JM\rangle$. Hence the $|JM\rangle$ form an irreducible representation of SO(3) or SU(2); M varies and J is fixed.

Then using Eqs. (4.67), (4.75), and (4.77) we obtain

$$\begin{aligned} J_-J_+|JM\rangle &= \big[J(J+1) - M(M+1)\big]|JM\rangle = (J-M)(J+M+1)|JM\rangle, \\ J_+J_-|JM\rangle &= \big[J(J+1) - M(M-1)\big]|JM\rangle = (J+M)(J-M+1)|JM\rangle. \end{aligned} \tag{4.80}$$

Because J_+ and J_- are Hermitian conjugates,[9]

$$J_+^\dagger = J_-, \qquad J_-^\dagger = J_+, \tag{4.81}$$

the eigenvalues in Eq. (4.80) must be positive or zero.[10] Examples of Eq. (4.81) are provided by the matrices of Exercise 3.2.13 (spin 1/2), 3.2.15 (spin 1), and 3.2.18 (spin 3/2).

[9] The Hermitian conjugation or adjoint operation is defined for matrices in Section 3.5, and for operators in general in Section 10.1.

[10] For an excellent discussion of adjoint operators and Hilbert space see A. Messiah, *Quantum Mechanics*. New York: Wiley 1961, Chapter 7.

For the orbital angular momentum ladder operators, L_+, and L_-, explicit forms are given in Exercises 2.5.14 and 12.6.7. You can now show (see also Exercise 12.7.2) that

$$\langle JM|J_-\big(J_+|JM\rangle\big) = \big(J_+|JM\rangle\big)^\dagger J_+|JM\rangle. \tag{4.82}$$

Since J_+ raises the eigenvalue M to $M+1$, we relabel the resultant eigenfunction $|JM+1\rangle$. The normalization is given by Eq. (4.80) as

$$J_+|JM\rangle = \sqrt{(J-M)(J+M+1)}|JM+1\rangle = \sqrt{J(J+1)-M(M+1)}|JM+1\rangle, \tag{4.83}$$

taking the positive square root and not introducing any phase factor. By the same arguments,

$$J_-|JM\rangle = \sqrt{(J+M)(J-M+1)}|JM-1\rangle = \sqrt{(J(J+1)-M(M-1)}|JM-1\rangle. \tag{4.84}$$

Applying J_+ to Eq. (4.84), we obtain the second line of Eq. (4.80) and verify that Eq. (4.84) is consistent with Eq. (4.83).

Finally, since M ranges from $-J$ to $+J$ in unit steps, $2J$ must be an integer; J is either an integer or half of an odd integer. As seen later, if **J** is an orbital angular momentum **L**, the set $|LM\rangle$ for all M is a basis defining a representation of SO(3) and L will then be integral. In spherical polar coordinates θ, φ, the functions $|LM\rangle$ become the spherical harmonics $Y_L^M(\theta,\varphi)$ of Section 12.6. The sets of $|JM\rangle$ states with half-integral J define representations of SU(2) that are not representations of SO(3); we get $J = 1/2, 3/2, 5/2, \ldots$. Our angular momentum is quantized, essentially as a result of the commutation relations. All these representations are irreducible, as an application of the raising and lowering operators suggests.

Summary of Lie Groups and Lie Algebras

The general commutation relations, Eq. (4.14) in Section 4.2, for a classical Lie group [SO(n) and SU(n) in particular] can be simplified to look more like Eq. (4.71) for SO(3) and SU(2) in this section. Here we merely review and, as a rule, do not provide proofs for various theorems that we explain.

First we choose linearly independent and mutually commuting generators H_i which are generalizations of J_z for SO(3) and SU(2). Let l be the maximum number of such H_i with

$$[H_i, H_k] = 0. \tag{4.85}$$

Then l is called the **rank** of the Lie group G or its Lie algebra $\mathcal{G}$. The rank and dimension, or order, of some Lie groups are given in Table 4.2. All other generators E_α can be shown to be raising and lowering operators with respect to all the H_i, so

$$[H_i, E_\alpha] = \alpha_i E_\alpha, \qquad i = 1, 2, \ldots, l. \tag{4.86}$$

The set of so-called **root vectors** $(\alpha_1, \alpha_2, \ldots, \alpha_l)$ form the **root diagram** of $\mathcal{G}$.

When the H_i commute, they can be simultaneously diagonalized (for symmetric (or Hermitian) matrices see Chapter 3; for operators see Chapter 10). The H_i provide us with a set of eigenvalues $m_1, m_2, \ldots, m_l$ [projection or additive quantum numbers generalizing

Table 4.2 Rank and Order of Unitary and Rotational Groups

Lie algebra	$\mathcal{A}_l$	$\mathcal{B}_l$	$\mathcal{D}_l$
Lie group	SU$(l+1)$	SO$(2l+1)$	SO$(2l)$
Rank	l	l	l
Order	$l(l+2)$	$l(2l+1)$	$l(2l-1)$

M of J_z in SO(3) and SU(2)]. The set of so-called **weight vectors** $(m_1, m_2, \ldots, m_l)$ for an irreducible representation (multiplet) form a **weight diagram**.

There are l invariant operators C_i, called **Casimir** operators, that commute with all generators and are generalizations of $\mathbf{J}^2$,

$$[C_i, H_j] = 0, \qquad [C_i, E_\alpha] = 0, \qquad i = 1, 2, \ldots, l. \tag{4.87}$$

The first one, C_1, is a quadratic function of the generators; the others are more complicated. Since the C_j commute with all H_j, they can be simultaneously diagonalized with the H_j. Their eigenvalues $c_1, c_2, \ldots, c_l$ characterize irreducible representations and stay constant while the weight vector varies over any particular irreducible representation. Thus the general eigenfunction may be written as

$$|(c_1, c_2, \ldots, c_l) m_1, m_2, \ldots, m_l\rangle. \tag{4.89b}$$

We can now show that $E_\alpha|(c_1, c_2, \ldots, c_l) m_1, m_2, \ldots, m_l\rangle$ has the weight vector $(m_1 + \alpha_1, \ldots, m_l + \alpha_l)$, using the commutation relations, Eq. (4.86), in conjunction with Eqs. (4.89a) and (4.89b):

$$H_i E_\alpha |(c_1, c_2, \ldots, c_l) m_1, m_2, \ldots, m_l\rangle = \big(E_\alpha H_i + [H_i, E_\alpha]\big) |(c_1, c_2, \ldots, c_l) m_1, m_2, \ldots, m_l\rangle$$
$$= (m_i + \alpha_i) E_\alpha |(c_1, c_2, \ldots, c_l) m_1, m_2, \ldots, m_l\rangle$$
$$= (c_1, \ldots, c_l) m_1 + \alpha_1, \ldots, m_l + \alpha_l\rangle, \tag{4.90}$$

Therefore E_α generalizing the multiplet $|JM\rangle$ of SO(3) and SU(2). Their eigenvalue equations are the generalization of Eqs. (4.83) and (4.84) from SO(3). These changes of eigenvalues by the operator E_α are called its **selection rules** in quantum mechanics. They are displayed in the root diagram of a Lie algebra.

Examples of root diagrams are given in Fig. 4.6 for SU(2) and SU(3). If we attach the roots denoted by arrows in Fig. 4.6b to a weight in Figs. 4.3 or 4.5a, b, we can reach any other state (represented by a dot in the weight diagram).

Here **Schur's lemma** applies: An operator H that commutes with all group operators, and therefore with all generators H_i of a (classical) Lie group G in particular, has as eigenvectors all states of a multiplet and is degenerate with the multiplet. As a consequence, such an operator commutes with all Casimir invariants, $[H, C_i] = 0$.

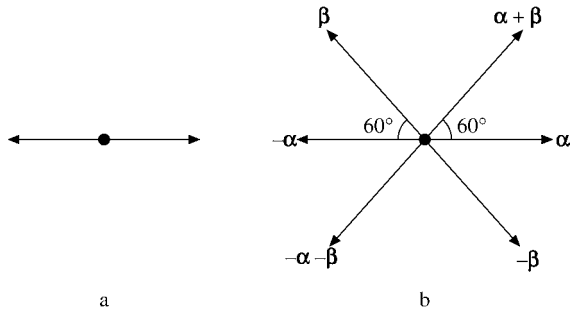

FIGURE 4.6 Root diagram for (a) SU(2) and (b) SU(3).

The last result is clear because the Casimir invariants are constructed from the generators and raising and lowering operators of the group. To prove the rest, let ψ be an eigenvector, $H\psi = E\psi$. Then, for any rotation R of G, we have $HR\psi = ER\psi$, which says that $R\psi$ is an eigenstate with the same eigenvalue E along with ψ. Since $[H, C_i] = 0$, all Casimir invariants can be diagonalized simultaneously with H and an eigenstate of H is an eigenstate of all the C_i. Since $[H_i, C_i] = 0$, the rotated eigenstates $R\psi$ are eigenstates of C_i, along with ψ belonging to the same multiplet characterized by the eigenvalues c_i of C_i.

Finally, such an operator H cannot induce transitions between different multiplets of the group because

$$\langle (c_1, c_2, \ldots, c_l)m_1, m_2, \ldots, m_l \rangle = 0.$$

Using $[H, C_i] = 0$ (for any i) we have

$$\langle (c_1, c_2, \ldots, c_l)m_1, m_2, \ldots, m_l \rangle$$

If $c_j \neq c_j'$ for some j, then the previous equation follows.

$$\langle (c_1, c_2, \ldots, c_l)m_1, m_2, \ldots, m_l \rangle$$

$$= \langle (c_1, c_2, \ldots, c_l)m_1, m_2, \ldots, m_l \rangle$$

$$0 = \langle (c_1, c_2, \ldots, c_l)m_1, m_2, \ldots, m_l \rangle$$

Exercises

4.3.1 Show that (a) $[J_+, \mathbf{J}^2] = 0$, (b) $[J_-, \mathbf{J}^2] = 0$.

4.3.2 Derive the root diagram of SU(3) in Fig. 4.6b from the generators λ_i in Eq. (4.61). *Hint*. Work out first the SU(2) case in Fig. 4.6a from the Pauli matrices.

4.4 ANGULAR MOMENTUM COUPLING

In many-body systems of classical mechanics, the total angular momentum is the sum $\mathbf{L} = \sum_i \mathbf{L}_i$ of the individual orbital angular momenta. Any isolated particle has conserved angular momentum. In quantum mechanics, conserved angular momentum arises when particles move in a central potential, such as the Coulomb potential in atomic physics, a shell model potential in nuclear physics, or a confinement potential of a quark model in

particle physics. In the relativistic Dirac equation, orbital angular momentum is no longer conserved, but $\mathbf{J} = \mathbf{L} + \mathbf{S}$ is conserved, the total angular momentum of a particle consisting of its orbital and intrinsic angular momentum, called spin $\mathbf{S} = \boldsymbol{\sigma}/2$, in units of $\hbar$.

It is readily shown that the sum of angular momentum operators obeys the same commutation relations in Eq. (4.37) or (4.41) as the individual angular momentum operators, provided those from different particles commute.

Clebsch–Gordan Coefficients: SU(2)–SO(3)

Clearly, combining two commuting angular momenta $\mathbf{J}_i$ to form their sum

$$\mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2, \qquad [J_{1i}, J_{2i}] = 0, \tag{4.91}$$

occurs often in applications, and $\mathbf{J}$ satisfies the angular momentum commutation relations

$$[J_j, J_k] = [J_{1j} + J_{2j}, J_{1k} + J_{2k}] = [J_{1j}, J_{1k}] + [J_{2j}, J_{2k}] = i\varepsilon_{jkl}(J_{1l} + J_{2l}) = i\varepsilon_{jkl} J_l.$$

For a single particle with spin $1/2$, for example, an electron or a quark, the total angular momentum is a sum of orbital angular momentum and spin. For two spinless particles their total orbital angular momentum $\mathbf{L} = \mathbf{L}_1 + \mathbf{L}_2$. For $\mathbf{J}^2$ and J_z of Eq. (4.91) to be both diagonal, $[\mathbf{J}^2, J_z] = 0$ has to hold. To show this we use the obvious commutation relations $[J_{iz}, \mathbf{J}_i^2] = 0$, and

$$\mathbf{J}^2 = \mathbf{J}_1^2 + \mathbf{J}_2^2 + 2\mathbf{J}_1 \cdot \mathbf{J}_2 = \mathbf{J}_1^2 + \mathbf{J}_2^2 + 2J_{1z}J_{2z} + J_{1+}J_{2-} + J_{1-}J_{2+}, \tag{4.91′}$$

in conjunction with Eq. (4.71), for both $\mathbf{J}_i$, to obtain

$$\begin{aligned}[\mathbf{J}^2, J_z] &= [\mathbf{J}_1^2 + \mathbf{J}_2^2, J_{1z} + J_{2z}] + [J_{1-}J_{2+} + J_{1+}J_{2-}, J_{1z} + J_{2z}] \\ &= [J_{1-}, J_{1z}]J_{2+} + J_{1-}[J_{2+}, J_{2z}] + [J_{1+}, J_{1z}]J_{2-} + J_{1+}[J_{2-}, J_{2z}] \\ &= J_{1-}J_{2+} - J_{1-}J_{2+} - J_{1+}J_{2-} + J_{1+}J_{2-} = 0.\end{aligned}$$

Similarly $[\mathbf{J}^2, \mathbf{J}_i^2] = 0$ is proved. Hence, the eigenvalues of $\mathbf{J}^2$, J_z can be used to label the total angular momentum states $|J_1 J_2 J M\rangle$.

The product states $|J_1 m_1\rangle|J_2 m_2\rangle$ obviously satisfy the eigenvalue equations

$$\begin{aligned}J_z|J_1 m_1\rangle|J_2 m_2\rangle &= (J_{1z} + J_{2z})|J_1 m_1\rangle|J_2 m_2\rangle = (m_1 + m_2)|J_1 m_1\rangle|J_2 m_2\rangle \\ &= M|J_1 m_1\rangle|J_2 m_2\rangle, \\ \mathbf{J}_i^2|J_1 m_1\rangle|J_2 m_2\rangle &= J_i(J_i + 1)|J_1 m_1\rangle|J_2 m_2\rangle,\end{aligned} \tag{4.92}$$

but will not have diagonal $\mathbf{J}^2$ except for the maximally stretched states with $M = \pm(J_1 + J_2)$ and $J = J_1 + J_2$ (see Fig. 4.7a). To see this we use Eq. (4.91′) again in conjunction with Eqs. (4.83) and (4.84) in

$$\begin{aligned}\mathbf{J}^2|J_1 m_1\rangle J_2 m_2\rangle &= \big\{J_1(J_1+1) + J_2(J_2+1) + 2m_1 m_2\big\}|J_1 m_1\rangle|J_2 m_2\rangle \\ &+ \big\{J_1(J_1+1) - m_1(m_1+1)\big\}^{1/2}\big\{J_2(J_2+1) - m_2(m_2-1)\big\}^{1/2} \\ &\times |J_1 m_1 + 1\rangle|J_2 m_2 - 1\rangle + \big\{J_1(J_1+1) - m_1(m_1-1)\big\}^{1/2} \\ &\times \big\{J_2(J_2+1) - m_2(m_2+1)\big\}^{1/2}|J_1 m_1 - 1\rangle|J_2 m_2 + 1\rangle.\end{aligned} \tag{4.93}$$

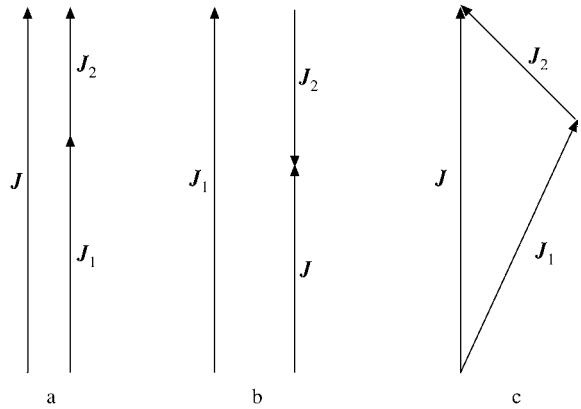

FIGURE 4.7 Coupling of two angular momenta: (a) parallel stretched, (b) antiparallel, (c) general case.

The last two terms in Eq. (4.93) vanish only when $m_1 = J_1$ and $m_2 = J_2$ or $m_1 = -J_1$ and $m_2 = -J_2$. In both cases $J = J_1 + J_2$ follows from the first line of Eq. (4.93). In general, therefore, we have to form appropriate linear combinations of product states

$$|J_1 J_2 J M\rangle = \sum_{m_1, m_2} C\big(J_1 J_2 J | m_1 m_2 M\big) |J_1 m_1\rangle |J_2 m_2\rangle, \tag{4.94}$$

so that $\mathbf{J}^2$ has eigenvalue $J(J+1)$. The quantities $C(J_1 J_2 J | m_1 m_2 M)$ in Eq. (4.94) are called **Clebsch–Gordan coefficients**. From Eq. (4.92) we see that they vanish unless $M = m_1 + m_2$, reducing the double sum to a single sum. Applying $J_\pm$ to $|JM\rangle$ shows that the eigenvalues M of J_z satisfy the usual inequalities $-J \le M \le J$.

Clearly, the maximal $J_{\max} = J_1 + J_2$ (see Fig. 4.7a). In this case Eq. (4.93) reduces to a pure product state

$$|J_1 J_2 J = J_1 + J_2 M = J_1 + J_2\rangle = |J_1 J_1\rangle |J_2 J_2\rangle, \tag{4.95a}$$

so the Clebsch–Gordan coefficient

$$C(J_1 J_2 J = J_1 + J_2 | J_1 J_2 J_1 + J_2) = 1. \tag{4.95b}$$

The minimal $J = J_1 - J_2$ (if $J_1 > J_2$, see Fig. 4.7b) and $J = J_2 - J_1$ for $J_2 > J_1$ follow if we keep in mind that there are just as many product states as $|JM\rangle$ states; that is,

$$\begin{aligned} \sum_{J=J_{\min}}^{J_{\max}} (2J+1) &= (J_{\max} - J_{\min} + 1)(J_{\max} + J_{\min} + 1) \\ &= (2J_1 + 1)(2J_2 + 1). \end{aligned} \tag{4.96}$$

This condition holds because the $|J_1 J_2 J M\rangle$ states merely rearrange all product states into irreducible representations of total angular momentum. It is equivalent to the **triangle rule**:

$$\begin{aligned} &\Delta(J_1 J_2 J) = 1, \qquad \text{if } |J_1 - J_2| \le J \le J_1 + J_2; \\ &\Delta(J_1 J_2 J) = 0, \qquad \text{else.} \end{aligned} \tag{4.97}$$

This indicates that one complete multiplet of each J value from $J_{\min}$ to $J_{\max}$ accounts for all the states and that all the $|JM\rangle$ states are necessarily orthogonal. In other words, Eq. (4.94) defines a unitary transformation from the orthogonal basis set of products of single-particle states $|J_1 m_1; J_2 m_2\rangle = |J_1 m_1\rangle|J_2 m_2\rangle$ to the two-particle states $|J_1 J_2 JM\rangle$. The Clebsch–Gordan coefficients are just the overlap matrix elements

$$C(J_1 J_2 J|m_1 m_2 M) \equiv \langle J_1 J_2 JM|J_1 m_1; J_2 m_2\rangle. \tag{4.98}$$

The explicit construction in what follows shows that they are all real. The states in Eq. (4.94) are orthonormalized, provided that the constraints

$$\sum_{m_1,m_2,\ m_1+m_2=M} C(J_1 J_2 J|m_1 m_2 M) C(J_1 J_2 J'|m_1 m_2 M')\rangle \tag{4.99a}$$

$$= \langle J_1 J_2 JM|J_1 J_2 J'M'\rangle = \delta_{JJ'}\delta_{MM'} \tag{4.99b}$$

hold.

$$\sum C(J_1 J_2 J|m_1 m_2 M) C(J_1 J_2 J|m_1' m_2' M)\rangle = \delta_{m_2 m_2'}\delta_{m_1 m_1'}$$

$$= \langle J_1 m_1|J_1 m_1'\rangle\langle J_2 m_2|J_2 m_2'\rangle$$

Now we are ready to construct more directly the total angular momentum states starting from $|J_{\max} = J_1 + J_2\ M = J_1 + J_2\rangle$ in Eq. (4.95a) and using the lowering operator $J_- = J_{1-} + J_{2-}$ repeatedly. In the first step we use Eq. (4.84) for

$$J_{i-}|J_i J_i\rangle = \big\{J_i(J_i+1) - J_i(J_i-1)\big\}^{1/2}|J_i J_i - 1\rangle = (2J_i)^{1/2}|J_i J_i - 1\rangle,$$

which we substitute into $(J_{1-} + J_{2-})|J_1 J_1\rangle|J_2 J_2\rangle$. Normalizing the resulting state with $M = J_1 + J_2 - 1$ properly to 1, we obtain

$$\begin{aligned}|J_1 J_2 J_1 + J_2 J_1 + J_2 - 1\rangle = {} & \big\{J_1/(J_1+J_2)\big\}^{1/2}|J_1 J_1 - 1\rangle|J_2 J_2\rangle \\ & + \big\{J_2/(J_1+J_2)\big\}^{1/2}|J_1 J_1\rangle|J_2 J_2 - 1\rangle.\end{aligned} \tag{4.100}$$

Equation (4.100) yields the Clebsch–Gordan coefficients

$$\begin{aligned}C(J_1 J_2 J_1 + J_2|J_1 - 1\ J_2\ J_1 + J_2 - 1) &= \big\{J_1/(J_1+J_2)\big\}^{1/2}, \\ C(J_1 J_2 J_1 + J_2|J_1\ J_2 - 1\ J_1 + J_2 - 1) &= \big\{J_2/(J_1+J_2)\big\}^{1/2}.\end{aligned} \tag{4.101}$$

Then we apply J_- again and normalize the states obtained until we reach $|J_1 J_2\ J_1 + J_2 M\rangle$ with $M = -(J_1 + J_2)$. The Clebsch–Gordan coefficients $C(J_1 J_2 J_1 + J_2|m_1 m_2 M)$ may thus be calculated step by step, and they are all real.

The next step is to realize that the only other state with $M = J_1 + J_2 - 1$ is the top of the next lower tower of $|J_1 + J_2 - 1 M\rangle$ states. Since $|J_1 + J_2 - 1\ J_1 + J_2 - 1\rangle$ is orthogonal to $|J_1 + J_2 J_1 + J_2 - 1\rangle$ in Eq. (4.100), it must be the other linear combination with a relative minus sign,

$$\begin{aligned}|J_1 + J_2 - 1\ J_1 + J_2 - 1\rangle = {} & -\big\{J_2/(J_1+J_2)\big\}^{1/2}|J_1 J_1 - 1\rangle|J_2 J_2\rangle \\ & + \big\{J_1/(J_1+J_2)\big\}^{1/2}|J_1 J_1\rangle|J_2 J_2 - 1\rangle,\end{aligned} \tag{4.102}$$

up to an overall sign.

Hence we have determined the Clebsch–Gordan coefficients (for $J_2 \geq J_1$)

$$\begin{aligned} C(J_1 J_2\ J_1+J_2-1|J_1-1\ J_2\ J_1+J_2-1) &= -\left\{J_2/(J_1+J_2)\right\}^{1/2}, \\ C(J_1 J_2\ J_1+J_2-1|J_1\ J_2-1\ J_1+J_2-1) &= \left\{J_1/(J_1+J_2)\right\}^{1/2}. \end{aligned} \tag{4.103}$$

Again we continue using J_- until we reach $M=-(J_1+J_2-1)$, and we keep normalizing the resulting states $|J_1+J_2-1M\rangle$ of the $J=J_1+J_2-1$ tower.

In order to get to the top of the next tower, $|J_1+J_2-2M\rangle$ with $M=J_1+J_2-2$, we remember that we have already constructed two states with that M. Both $|J_1+J_2 J_1+J_2-2\rangle$ and $|J_1+J_2-1\ J_1+J_2-2\rangle$ are known linear combinations of the three product states $|J_1 J_1\rangle|J_2 J_2-2\rangle$, $|J_1 J_1-1\rangle\times|J_2 J_2-1\rangle$, and $|J_1 J_1-2\rangle|J_2 J_2\rangle$. The third linear combination is easy to find from orthogonality to these two states, up to an overall phase, which is chosen by the **Condon–Shortley phase conventions**[11] so that the coefficient $C(J_1 J_2\ J_1+J_2-2|J_1\ J_2-2\ J_1+J_2-2)$ of the last product state is positive for $|J_1 J_2\ J_1+J_2-2\ J_1+J_2-2\rangle$. It is straightforward, though a bit tedious, to determine the rest of the Clebsch–Gordan coefficients.

Numerous recursion relations can be derived from matrix elements of various angular momentum operators, for which we refer to the literature.[12]

The symmetry properties of Clebsch–Gordan coefficients are best displayed in the more symmetric Wigner's $3j$-symbols, which are tabulated:[12]

$$\begin{pmatrix} J_1 J_2 J_3 \\ m_1 m_2 m_3 \end{pmatrix} = \frac{(-1)^{J_1-J_2-m_3}}{(2J_3+1)^{1/2}} C(J_1 J_2 J_3|m_1 m_2,-m_3), \tag{4.104a}$$

obeying the symmetry relations

$$\begin{pmatrix} J_1 J_2 J_3 \\ m_1 m_2 m_3 \end{pmatrix} = (-1)^{J_1+J_2+J_3} \begin{pmatrix} J_k J_l J_n \\ m_k m_l m_n \end{pmatrix} \tag{4.104b}$$

for (k,l,n) an odd permutation of $(1,2,3)$. One of the most important places where Clebsch–Gordan coefficients occur is in matrix elements of tensor operators, which are governed by the Wigner–Eckart theorem discussed in the next section, on spherical tensors. Another is coupling of operators or state vectors to total angular momentum, such as spin-orbit coupling. Recoupling of operators and states in matrix elements leads to $6j$- and $9j$-symbols.[12] Clebsch–Gordan coefficients can and have been calculated for other Lie groups, such as SU(3).

[11] E. U. Condon and G. H. Shortley, *Theory of Atomic Spectra*. Cambridge, UK: Cambridge University Press (1935).

[12] There is a rich literature on this subject, e.g., A. R. Edmonds, *Angular Momentum in Quantum Mechanics*. Princeton, NJ: Princeton University Press (1957); M. E. Rose, *Elementary Theory of Angular Momentum*. New York: Wiley (1957); A. de-Shalit and I. Talmi, *Nuclear Shell Model*. New York: Academic Press (1963); Dover (2005). Clebsch–Gordan coefficients are tabulated in M. Rotenberg, R. Bivins, N. Metropolis, and J. K. Wooten, Jr., *The 3j- and 6j-Symbols*. Cambridge, MA: Massachusetts Institute of Technology Press (1959).

Spherical Tensors

In Chapter 2 the properties of Cartesian tensors are defined using the group of nonsingular general linear transformations, which contains the three-dimensional rotations as a subgroup. A tensor of a given rank that is irreducible with respect to the full group may well become reducible for the rotation group SO(3). To explain this point, consider the second-rank tensor with components $T_{jk} = x_j y_k$ for $j, k = 1, 2, 3$. It contains the symmetric tensor $S_{jk} = (x_j y_k + x_k y_j)/2$ and the antisymmetric tensor $A_{jk} = (x_j y_k - x_k y_j)/2$, so $T_{jk} = S_{jk} + A_{jk}$. This reduces T_{jk} in SO(3). However, under rotations the scalar product $\mathbf{x} \cdot \mathbf{y}$ is invariant and is therefore irreducible in SO(3). Thus, S_{jk} can be reduced by subtraction of the multiple of $\mathbf{x} \cdot \mathbf{y}$ that makes it traceless. This leads to the SO(3)-irreducible tensor

$$S'_{jk} = \tfrac{1}{2}(x_j y_k + x_k y_j) - \tfrac{1}{3}\mathbf{x} \cdot \mathbf{y}\delta_{jk}.$$

Tensors of higher rank may be treated similarly. When we form tensors from products of the components of the coordinate vector $\mathbf{r}$ then, in polar coordinates that are tailored to SO(3) symmetry, we end up with the spherical harmonics of Chapter 12.

The form of the ladder operators for SO(3) in Section 4.3 leads us to introduce the **spherical components** (note the different normalization and signs, though, prescribed by the Y_{lm}) of a vector $\mathbf{A}$:

$$A_{+1} = -\tfrac{1}{\sqrt{2}}(A_x + iA_y), \qquad A_{-1} = \tfrac{1}{\sqrt{2}}(A_x - iA_y), \qquad A_0 = A_z. \tag{4.105}$$

Then we have for the coordinate vector $\mathbf{r}$ in polar coordinates,

$$r_{+1} = -\tfrac{1}{\sqrt{2}} r \sin\theta e^{i\varphi} = r\sqrt{\tfrac{4\pi}{3}} Y_{11}, \qquad r_{-1} = \tfrac{1}{\sqrt{2}} r \sin\theta e^{-i\varphi} = r\sqrt{\tfrac{4\pi}{3}} Y_{1,-1},$$
$$r_0 = r\sqrt{\tfrac{4\pi}{3}} Y_{10}, \tag{4.106}$$

where $Y_{lm}(\theta, \varphi)$ are the spherical harmonics of Chapter 12. Again, the spherical jm components of tensors T_{jm} of higher rank j may be introduced similarly.

An irreducible **spherical tensor operator** T_{jm} of rank j has $2j + 1$ components, just as for spherical harmonics, and m runs from $-j$ to $+j$. Under a rotation $\mathsf{R}(\alpha)$, where α stands for the Euler angles, the Y_{lm} transform as

$$Y_{lm}(\hat{\mathbf{r}}') = \sum_{m'} Y_{lm'}(\hat{\mathbf{r}}) D^l_{m'm}(\mathsf{R}), \tag{4.107a}$$

where $\hat{\mathbf{r}}' = (\theta', \varphi')$ are obtained from $\hat{\mathbf{r}} = (\theta, \varphi)$ by the rotation R and are the angles of the same point in the rotated frame, and

$$\mathsf{D}^J_{m'm}(\alpha, \beta, \gamma) = \langle Jm | \exp(i\alpha J_z) \exp(i\beta J_y) \exp(i\gamma J_z) | Jm' \rangle$$

are the rotation matrices. So, for the operator T_{jm}, we define

$$\mathsf{R} T_{jm} \mathsf{R}^{-1} = \sum_{m'} T_{jm'} D^j_{m'm}(\alpha). \tag{4.107b}$$

For an infinitesimal rotation (see Eq. (4.20) in Section 4.2 on generators) the left side of Eq. (4.107b) simplifies to a commutator and the right side to the matrix elements of **J**, the infinitesimal generator of the rotation R:

$$[J_n, T_{jm}] = \sum_{m'} T_{jm'} \langle jm'|J_n|jm\rangle. \tag{4.108}$$

If we substitute Eqs. (4.83) and (4.84) for the matrix elements of J_m we obtain the alternative transformation laws of a tensor operator,

$$[J_0, T_{jm}] = mT_{jm}, \qquad [J_\pm, T_{jm}] = T_{jm\pm1}\big\{(j-m)(j\pm m+1)\big\}^{1/2}. \tag{4.109}$$

We can use the Clebsch–Gordan coefficients of the previous subsection to couple two tensors of given rank to another rank. An example is the cross or vector product of two vectors **a** and **b** from Chapter 1. Let us write both vectors in spherical components, a_m and b_m. Then we verify that the tensor C_m of rank 1 defined as

$$C_m \equiv \sum_{m_1 m_2} C(111|m_1 m_2 m) a_{m_1} b_{m_2} = \frac{i}{\sqrt{2}}(\mathbf{a}\times\mathbf{b})_m. \tag{4.110}$$

Since C_m is a spherical tensor of rank 1 that is linear in the components of **a** and **b**, it must be proportional to the cross product, $C_m = N(\mathbf{a}\times\mathbf{b})_m$. The constant N can be determined from a special case, $\mathbf{a} = \hat{\mathbf{x}}, \mathbf{b} = \hat{\mathbf{y}}$, essentially writing $\hat{\mathbf{x}}\times\hat{\mathbf{y}} = \hat{\mathbf{z}}$ in spherical components as follows. Using

$$(\hat{\mathbf{z}})_0 = 1; \qquad (\hat{\mathbf{x}})_1 = -1/\sqrt{2}, \qquad (\hat{\mathbf{x}})_{-1} = 1/\sqrt{2};$$
$$(\hat{\mathbf{y}})_1 = -i/\sqrt{2}, \qquad (\hat{\mathbf{y}})_{-1} = -i/\sqrt{2},$$

Eq. (4.110) for $m = 0$ becomes

$$C(111|1,-1,0)\big[(\hat{\mathbf{x}})_1(\hat{\mathbf{y}})_{-1} - (\hat{\mathbf{x}})_{-1}(\hat{\mathbf{y}})_1\big] = N\big((\hat{\mathbf{z}})_0\big) = N$$
$$= \frac{1}{\sqrt{2}}\left[-\frac{1}{\sqrt{2}}\left(-\frac{i}{\sqrt{2}}\right) - \frac{1}{\sqrt{2}}\left(-\frac{i}{\sqrt{2}}\right)\right] = \frac{i}{\sqrt{2}},$$

where we have used $C(111|101) = \frac{1}{\sqrt{2}}$ from Eq. (4.103) for $J_1 = 1 = J_2$, which implies $C(111|1,-1,0) = \frac{1}{\sqrt{2}}$ using Eqs. (4.104a,b):

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \end{pmatrix} = -\frac{1}{\sqrt{3}} C(111|101) = -\frac{1}{6} = -\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix} = -\frac{1}{\sqrt{3}} C(111|1,-1,0).$$

A bit simpler is the usual scalar product of two vectors in Chapter 1, in which **a** and **b** are coupled to zero angular momentum:

$$\mathbf{a}\cdot\mathbf{b} \equiv -(\mathbf{ab})_0\sqrt{3} \equiv -\sqrt{3}\sum_m C(110|m,-m,0) a_m b_{-m}. \tag{4.111}$$

Again, the rank zero of our tensor product implies $\mathbf{a}\cdot\mathbf{b} = n(\mathbf{ab})_0$. The constant n can be determined from a special case, essentially writing $\hat{\mathbf{z}}^2 = 1$ in spherical components: $\hat{\mathbf{z}}^2 = 1 = nC(110|000) = -\frac{n}{\sqrt{3}}$.

Another often-used application of tensors is the **recoupling** that involves **6j-symbols** for three operators and $9j$ for four operators.[12] An example is the following scalar product, for which it can be shown[12] that

$$\boldsymbol{\sigma}_1 \cdot \mathbf{r}\boldsymbol{\sigma}_2 \cdot \mathbf{r} = \frac{1}{3}\mathbf{r}^2 \boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_2 + (\boldsymbol{\sigma}_1\boldsymbol{\sigma}_2)_2 \cdot (\mathbf{rr})_2, \tag{4.112}$$

but which can also be rearranged by elementary means. Here the tensor operators are defined as

$$(\boldsymbol{\sigma}_1\boldsymbol{\sigma}_2)_{2m} = \sum_{m_1 m_2} C(112|m_1 m_2 m)\sigma_{1m_1}\sigma_{2m_2}, \tag{4.113}$$

$$(\mathbf{rr})_{2m} = \sum_m C(112|m_1 m_2 m) r_{m_1} r_{m_2} = \sqrt{\frac{8\pi}{15}}\mathbf{r}^2 Y_{2m}(\hat{\mathbf{r}}), \tag{4.114}$$

and the scalar product of tensors of rank 2 as

$$(\boldsymbol{\sigma}_1\boldsymbol{\sigma}_2)_2 \cdot (\mathbf{rr})_2 = \sum_m (-1)^m (\boldsymbol{\sigma}_1\boldsymbol{\sigma}_2)_{2m}(\mathbf{rr})_{2,-m} = \sqrt{5}\big((\boldsymbol{\sigma}_1\boldsymbol{\sigma}_2)_2(\mathbf{rr})_2\big)_0. \tag{4.115}$$

One of the most important applications of spherical tensor operators is the **Wigner–Eckart theorem**. It says that a matrix element of a spherical tensor operator T_{km} of rank k between states of angular momentum j and j' factorizes into a Clebsch–Gordan coefficient and a so-called **reduced matrix element**, denoted by double bars, that no longer has any dependence on the projection quantum numbers m, m', n:

$$\langle j'm'|T_{kn}|jm\rangle = C(kjj'|nmm')(-1)^{k-j+j'}\langle j'\|T_k\|j\rangle/\sqrt{(2j'+1)}. \tag{4.116}$$

In other words, such a matrix element factors into a dynamic part, the reduced matrix element, and a geometric part, the Clebsch–Gordan coefficient that contains the rotational properties (expressed by the projection quantum numbers) from the SO(3) invariance. To see this we couple T_{kn} with the initial state to total angular momentum j':

$$|j'm'\rangle_0 \equiv \sum_{nm} C(kjj'|nmm')T_{kn}|jm\rangle. \tag{4.117}$$

Under rotations the state $|j'm'\rangle_0$ transforms just like $|j'm'\rangle$. Thus, the overlap matrix element $\langle j'm'|j'm'\rangle_0$ is a rotational scalar that has no m' dependence, so we can average over the projections,

$$\langle JM|j'm'\rangle_0 = \frac{\delta_{Jj'}\delta_{Mm'}}{2j'+1}\sum_\mu \langle j'\mu|j'\mu\rangle_0. \tag{4.118}$$

Next we substitute our definition, Eq. (4.117), into Eq. (4.118) and invert the relation Eq. (4.117) using orthogonality, Eq. (4.99b), to find that

$$\langle JM|T_{kn}|jm\rangle = \sum_{j'm'} C(kjj'|nmm')\frac{\delta_{Jj'}\delta_{Mm'}}{2J+1}\sum_\mu \langle J\mu|J\mu\rangle_0, \tag{4.119}$$

which proves the Wigner–Eckart theorem, Eq. (4.116).[13]

[13]The extra factor $(-1)^{k-j+j'}/\sqrt{(2j'+1)}$ in Eq. (4.116) is just a convention that varies in the literature.

As an application, we can write the Pauli matrix elements in terms of Clebsch–Gordan coefficients. We apply the Wigner–Eckart theorem to

$$\left\langle \tfrac{1}{2}\gamma \left| \sigma_\alpha \right| \tfrac{1}{2}\beta \right\rangle = (\sigma_\alpha)_{\gamma\beta} = -\tfrac{1}{\sqrt{2}} C\left(1\tfrac{1}{2}\tfrac{1}{2}|\alpha\beta\gamma\right) \left\langle \tfrac{1}{2} \left\| \sigma \right\| \tfrac{1}{2} \right\rangle. \tag{4.120}$$

Since $\langle \frac{1}{2}\frac{1}{2}|\sigma_0|\frac{1}{2}\frac{1}{2}\rangle = 1$ with $\sigma_0 = \sigma_3$ and $C(1\frac{1}{2}\frac{1}{2} \mid 0\frac{1}{2}\frac{1}{2}) = -1/\sqrt{3}$, we find

$$\left\langle \tfrac{1}{2} \left\| \sigma \right\| \tfrac{1}{2} \right\rangle = \sqrt{6}, \tag{4.121}$$

which, substituted into Eq. (4.120), yields

$$(\sigma_\alpha)_{\gamma\beta} = -\sqrt{3} C\left(1\tfrac{1}{2}\tfrac{1}{2}|\alpha\beta\gamma\right). \tag{4.122}$$

Note that the $\alpha = \pm 1, 0$ denote the spherical components of the Pauli matrices.

Young Tableaux for SU(n)

Young tableaux (YT) provide a powerful and elegant method for decomposing products of SU(n) group representations into sums of irreducible representations. The YT provide the dimensions and symmetry types of the irreducible representations in this so-called **Clebsch–Gordan series**, though not the Clebsch–Gordan coefficients by which the product states are coupled to the quantum numbers of each irreducible representation of the series (see Eq. (4.94)).

Products of representations correspond to multiparticle states. In this context, permutations of particles are important when we deal with several identical particles. Permutations of n identical objects form the **symmetric group** S_n. A close connection between irreducible representations of S_n, which are the YT, and those of SU(n) is provided by this **theorem**: Every N-particle state of S_n that is made up of single-particle states of the fundamental n-dimensional SU(n) multiplet belongs to an irreducible SU(n) representation. A proof is in Chapter 22 of Wybourne.[14]

For SU(2) the fundamental representation is a box that stands for the spin $+\frac{1}{2}$ (up) and $-\frac{1}{2}$ (down) states and has dimension 2. For SU(3) the box comprises the three quark states in the triangle of Fig. 4.5a; it has dimension 3.

An array of boxes shown in Fig. 4.8 with λ_1 boxes in the first row, λ_2 boxes in the second row, ..., and λ_{n-1} boxes in the last row is called a Young tableau (YT), denoted by $[\lambda_1, \ldots, \lambda_{n-1}]$, and represents an irreducible representation of SU(n) if and only if

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n-1}. \tag{4.123}$$

Boxes in the same row are symmetric representations; those in the same column are antisymmetric. A YT consisting of one row is totally symmetric. A YT consisting of a single column is totally antisymmetric.

There are at most $n-1$ rows for SU(n) YT because a column of n boxes is the totally antisymmetric (**Slater determinant** of single-particle states) singlet representation that may be struck from the YT.

An array of N boxes is an N-particle state whose boxes may be labeled by positive integers so that the (particle labels or) numbers in one row of the YT do not decrease from

[14] B. G. Wybourne, *Classical Groups for Physicists*. New York: Wiley (1974).

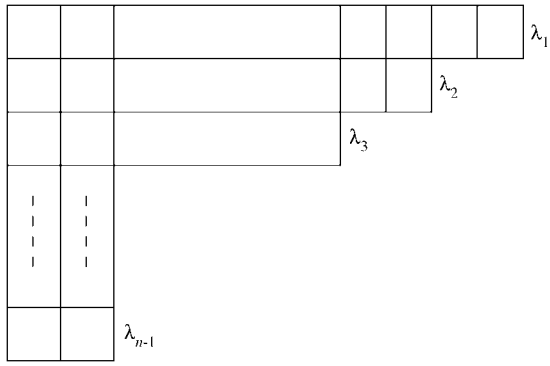

FIGURE 4.8 Young tableau (YT) for SU(n).

left to right and those in any one column increase from top to bottom. In contrast to the possible repetitions of row numbers, the numbers in any column must be different because of the antisymmetry of these states.

The product of a YT with a single box, [1], is the sum of YT formed when the box is put at the end of each row of the YT, provided the resulting YT is legitimate, that is, obeys Eq. (4.123). For SU(2) the product of two boxes, spin $1/2$ representations of dimension 2, generates

$$[1] \otimes [1] = [2] \oplus [1, 1], \tag{4.124}$$

the symmetric spin 1 representation of dimension 3 and the antisymmetric singlet of dimension 1 mentioned earlier.

The column of $n-1$ boxes is the conjugate representation of the fundamental representation; its product with a single box contains the column of n boxes, which is the singlet. For SU(3) the conjugate representation of the single box, [1] or fundamental quark representation, is the inverted triangle in Fig. 4.5b, [1, 1], which represents the three antiquarks $\bar{u}, \bar{d}, \bar{s}$, obviously of dimension 3 as well.

The dimension of a YT is given by the ratio

$$\dim \mathrm{YT} = \frac{N}{D}. \tag{4.125}$$

The numerator N is obtained by writing an n in all boxes of the YT along the diagonal, $(n+1)$ in all boxes immediately above the diagonal, $(n-1)$ immediately below the diagonal, etc. N is the product of all the numbers in the YT. An example is shown in Fig. 4.9a for the octet representation of SU(3), where $N = 2 \cdot 3 \cdot 4 = 24$. There is a closed formula that is equivalent to Eq. (4.125).[15] The denominator D is the product of all **hooks**.[16] A hook is drawn through each box of the YT by starting a horizontal line from the right to the box in question and then continuing it vertically out of the YT. The number of boxes encountered by the hook-line is the hook-number of the box. D is the product of all hook-numbers of

[15] See, for example, M. Hamermesh, *Group Theory and Its Application to Physical Problems*. Reading, MA: Addison-Wesley (1962).

[16] F. Close, *Introduction to Quarks and Partons*. New York: Academic Press (1979).

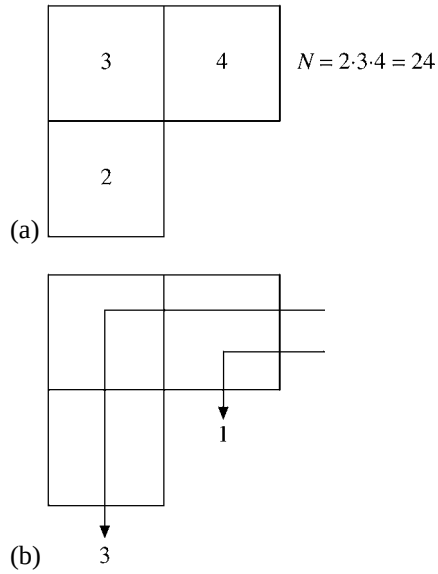

FIGURE 4.9 Illustration of (a) N and (b) D in Eq. (4.125) for the octet Young tableau of SU(3).

the YT. An example is shown in Fig. 4.9b for the octet of SU(3), whose hook-number is $D = 1 \cdot 3 \cdot 1 = 3$. Hence the dimension of the SU(3) octet is $24/3 = 8$, whence its name.

Now we can calculate the dimensions of the YT in Eq. (4.124). For SU(2) they are $2 \times 2 = 3 + 1 = 4$. For SU(3) they are $3 \cdot 3 = 3 \cdot 4/(1 \cdot 2) + 3 \cdot 2/(2 \cdot 1) = 6 + 3 = 9$. For the product of the quark times antiquark YT of SU(3) we get

$$[1, 1] \otimes [1] = [2, 1] \oplus [1, 1, 1], \tag{4.126}$$

that is, octet and singlet, which are precisely the meson multiplets considered in the subsection on the eightfold way, the SU(3) flavor symmetry, which suggest mesons are bound states of a quark and an antiquark, $q\bar{q}$ configurations. For the product of three quarks we get

$$\big([1] \otimes [1]\big) \otimes [1] = \big([2] \oplus [1, 1]\big) \otimes [1] = [3] \oplus 2[2, 1] \oplus [1, 1, 1], \tag{4.127}$$

that is, decuplet, octet, and singlet, which are the observed multiplets for the baryons, which suggests they are bound states of three quarks, q^3 configurations.

As we have seen, YT describe the decomposition of a product of SU(n) *irreducible representations into irreducible representations of* SU(n)*, which is called the Clebsch–Gordan series, while the Clebsch–Gordan coefficients considered earlier allow construction of the individual states in this series.*

Exercises

4.4.1 Derive recursion relations for Clebsch–Gordan coefficients. Use them to calculate $C(11J|m_1m_2M)$ for $J = 0, 1, 2$.
Hint. Use the known matrix elements of $J_+ = J_{1+} + J_{2+}$, J_{i+}, and $\mathbf{J}^2 = (\mathbf{J}_1 + \mathbf{J}_2)^2$, etc.

4.4.2 Show that $(Y_l\chi)^J_M = \sum C(l\tfrac{1}{2}J|m_l m_s M)Y_{lm_l}\chi_{m_s}$, where $\chi_{\pm 1/2}$ are the spin up and down eigenfunctions of $\sigma_3 = \sigma_z$, transforms like a spherical tensor of rank J.

4.4.3 When the spin of quarks is taken into account, the SU(3) flavor symmetry is replaced by the SU(6) symmetry. Why? Obtain the Young tableau for the antiquark configuration $\bar{q}$. Then decompose the product $q\bar{q}$. Which SU(3) representations are contained in the nontrivial SU(6) representation for mesons?
Hint. Determine the dimensions of all YT.

4.4.4 For $l = 1$, Eq. (4.107a) becomes

$$Y_1^m(\theta', \varphi') = \sum_{m'=-1}^{1} D^1_{m'm}(\alpha, \beta, \gamma)Y_1^{m'}(\theta, \varphi).$$

Rewrite these spherical harmonics in Cartesian form. Show that the resulting Cartesian coordinate equations are equivalent to the Euler rotation matrix $\mathsf{A}(\alpha, \beta, \gamma)$, Eq. (3.94), rotating the coordinates.

4.4.5 Assuming that $D^j(\alpha, \beta, \gamma)$ is unitary, show that

$$\sum_{m=-l}^{l} Y_l^{m*}(\theta_1, \varphi_1)Y_l^m(\theta_2, \varphi_2)$$

is a scalar quantity (invariant under rotations). This is a spherical tensor analog of a scalar product of vectors.

4.4.6 (a) Show that the α and γ dependence of $\mathsf{D}^j(\alpha, \beta, \gamma)$ may be factored out such that

$$\mathsf{D}^j(\alpha, \beta, \gamma) = \mathsf{A}^j(\alpha)\mathsf{d}^j(\beta)\mathsf{C}^j(\gamma).$$

(b) Show that $\mathsf{A}^j(\alpha)$ and $\mathsf{C}^j(\gamma)$ are diagonal. Find the explicit forms.
(c) Show that $\mathsf{d}^j(\beta) = \mathsf{D}^j(0, \beta, 0)$.

4.4.7 The angular momentum–exponential form of the Euler angle rotation operators is

$$\begin{aligned}\mathsf{R} &= \mathsf{R}_{z''}(\gamma)\mathsf{R}_{y'}(\beta)\mathsf{R}_z(\alpha)\\ &= \exp(-i\gamma J_{z''})\exp(-i\beta J_{y'})\exp(-i\alpha J_z).\end{aligned}$$

Show that in terms of the original axes

$$\mathsf{R} = \exp(i\alpha J_z)\exp(-i\beta J_y)\exp(-i\gamma J_z).$$

Hint. The R operators transform as matrices. The rotation about the y'-axis (second Euler rotation) may be referred to the original y-axis by

$$\exp(-i\beta J_{y'}) = \exp(-i\alpha J_z)\exp(-i\beta J_y)\exp(i\alpha J_z).$$

4.4.8 Using the Wigner–Eckart theorem, prove the decomposition theorem for a spherical vector operator $\langle j'm'|T_{1m}|jm\rangle = \frac{\langle jm'|\mathbf{J}\cdot\mathbf{T}_1|jm\rangle}{j(j+1)}\delta_{jj'}$.

4.4.9 Using the Wigner–Eckart theorem, prove the factorization $\langle j'm'|J_M\mathbf{J}\cdot\mathbf{T_1}|jm\rangle = \langle jm'|J_M|jm\rangle\delta_{j'j}\langle jm|\mathbf{J}\cdot\mathbf{T_1}|jm\rangle$.

4.5 HOMOGENEOUS LORENTZ GROUP

Generalizing the approach to vectors of Section 1.2, in special relativity we demand that our physical laws be covariant[17] under

a. space and time translations,
b. rotations in real, three-dimensional space, and
c. Lorentz transformations.

The demand for covariance under translations is based on the homogeneity of space and time. Covariance under rotations is an assertion of the isotropy of space. The requirement of Lorentz covariance follows from special relativity. All three of these transformations together form the inhomogeneous Lorentz group or the Poincaré group. When we exclude translations, the space rotations and the Lorentz transformations together form a group — the homogeneous Lorentz group.

We first generate a subgroup, the Lorentz transformations in which the relative velocity $\mathbf{v}$ is along the $x = x^1$-axis. The generator may be determined by considering space–time reference frames moving with a relative velocity δv, an infinitesimal.[18] The relations are similar to those for rotations in real space, Sections 1.2, 2.6, and 3.3, except that here the angle of rotation is pure imaginary (compare Section 4.6).

Lorentz transformations are linear not only in the space coordinates x_i but in the time t as well. They originate from Maxwell's equations of electrodynamics, which are invariant under Lorentz transformations, as we shall see later. Lorentz transformations leave the quadratic form $c^2t^2 - x_1^2 - x_2^2 - x_3^2 = x_0^2 - x_1^2 - x_2^2 - x_3^2$ invariant, where $x_0 = ct$. We see this if we switch on a light source at the origin of the coordinate system. At time t light has traveled the distance $ct = \sqrt{\sum x_i^2}$, so $c^2t^2 - x_1^2 - x_2^2 - x_3^2 = 0$. Special relativity requires that in all (inertial) frames that move with velocity $v \le c$ in any direction relative to the x_i-system and have the same origin at time $t = 0$, $c^2t'^2 - x_1'^2 - x_2'^2 - x_3'^2 = 0$ holds also. Four-dimensional space–time with the metric $x \cdot x = x^2 = x_0^2 - x_1^2 - x_2^2 - x_3^2$ is called Minkowski space, with the scalar product of two four-vectors defined as $a \cdot b = a_0b_0 - \mathbf{a}\cdot\mathbf{b}$. Using the metric tensor

$$(g_{\mu\nu}) = \left(g^{\mu\nu}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{4.128}$$

[17] To be covariant means to have the same form in different coordinate systems so that there is no preferred reference system (compare Sections 1.2 and 2.6).

[18] This derivation, with a slightly different metric, appears in an article by J. L. Strecker, *Am. J. Phys.* **35**: 12 (1967).

we can raise and lower the indices of a four-vector, such as the coordinates $x^\mu = (x_0, \mathbf{x})$, so that $x_\mu = g_{\mu\nu}x^\nu = (x_0, -\mathbf{x})$ and $x^\mu g_{\mu\nu}x^\nu = x_0^2 - \mathbf{x}^2$, Einstein's summation convention being understood. For the gradient, $\partial^\mu = (\partial/\partial x_0, -\nabla) = \partial/\partial x_\mu$ and $\partial_\mu = (\partial/\partial x_0, \nabla)$, so $\partial^2 = \partial^\mu\partial_\mu = (\partial/\partial x_0)^2 - \nabla^2$ is a Lorentz scalar, just like the metric $x^2 = x_0^2 - \mathbf{x}^2$.

For $v \ll c$, in the nonrelativistic limit, a Lorentz transformation must be Galilean. Hence, to derive the form of a Lorentz transformation along the x_1-axis, we start with a Galilean transformation for infinitesimal relative velocity δv:

$$x'^1 = x^1 - \delta v t = x^1 - x^0\delta\beta. \tag{4.129}$$

Here, $\beta = v/c$. By symmetry we also write

$$x'^0 = x^0 + a\delta\beta x^1, \tag{4.129$'$}$$

with the parameter a chosen so that $x_0^2 - x_1^2$ is invariant,

$$x_0'^2 - x_1'^2 = x_0^2 - x_1^2. \tag{4.130}$$

Remember, $x^\mu = (x^0, \mathbf{x})$ is the prototype four-dimensional vector in Minkowski space. Thus Eq. (4.130) is simply a statement of the invariance of the square of the magnitude of the "distance" vector under Lorentz transformation in Minkowski space. Here is where the special relativity is brought into our transformation. Squaring and subtracting Eqs. (4.129) and (4.129$'$) and discarding terms of order $(\delta\beta)^2$, we find $a = -1$. Equations (4.129) and (4.129$'$) may be combined as a matrix equation,

$$\begin{pmatrix} x'^0 \\ x'^1 \end{pmatrix} = (1_2 - \delta\beta\sigma_1)\begin{pmatrix} x^0 \\ x^1 \end{pmatrix}; \tag{4.131}$$

σ_1 happens to be the Pauli matrix, σ_1, and the parameter $\delta\beta$ represents an infinitesimal change. Using the same techniques as in Section 4.2, we repeat the transformation N times to develop a finite transformation with the velocity parameter $\rho = N\delta\beta$. Then

$$\begin{pmatrix} x'^0 \\ x'^1 \end{pmatrix} = \left(1_2 - \frac{\rho\sigma_1}{N}\right)^N \begin{pmatrix} x^0 \\ x^1 \end{pmatrix}. \tag{4.132}$$

In the limit as $N \to \infty$,

$$\lim_{N\to\infty}\left(1_2 - \frac{\rho\sigma_1}{N}\right)^N = \exp(-\rho\sigma_1). \tag{4.133}$$

As in Section 4.2, the exponential is interpreted by a Maclaurin expansion,

$$\exp(-\rho\sigma_1) = 1_2 - \rho\sigma_1 + \frac{1}{2!}(\rho\sigma_1)^2 - \frac{1}{3!}(\rho\sigma_1)^3 + \cdots. \tag{4.134}$$

Noting that $(\sigma_1)^2 = 1_2$,

$$\exp(-\rho\sigma_1) = 1_2\cosh\rho - \sigma_1\sinh\rho. \tag{4.135}$$

Hence our finite Lorentz transformation is

$$\begin{pmatrix} x'^0 \\ x'^1 \end{pmatrix} = \begin{pmatrix} \cosh\rho & -\sinh\rho \\ -\sinh\rho & \cosh\rho \end{pmatrix}\begin{pmatrix} x^0 \\ x^1 \end{pmatrix}. \tag{4.136}$$

σ_1 has generated the representations of this pure Lorentz transformation. The quantities $\cosh\rho$ and $\sinh\rho$ may be identified by considering the origin of the primed coordinate system, $x'^1 = 0$, or $x^1 = vt$. Substituting into Eq. (4.136), we have

$$0 = x^1 \cosh\rho - x^0 \sinh\rho. \tag{4.137}$$

With $x^1 = vt$ and $x^0 = ct$,

$$\tanh\rho = \beta = \frac{v}{c}.$$

Note that the **rapidity** $\rho \neq v/c$, except in the limit as $v \to 0$. The rapidity is the additive parameter for pure Lorentz transformations ("boosts") along the same axis that corresponds to angles for rotations about the same axis. Using $1 - \tanh^2\rho = (\cosh^2\rho)^{-1}$,

$$\cosh\rho = \left(1-\beta^2\right)^{-1/2} \equiv \gamma, \qquad \sinh\rho = \beta\gamma. \tag{4.138}$$

The group of Lorentz transformations is not compact, because the limit of a sequence of rapidities going to infinity is no longer an element of the group.

The preceding special case of the velocity parallel to one space axis is easy, but it illustrates the infinitesimal velocity-exponentiation-generator technique. Now, this exact technique may be applied to derive the Lorentz transformation for the relative velocity **v not** parallel to any space axis. The matrices given by Eq. (4.136) for the case of $\mathbf{v} = \hat{\mathbf{x}}v_x$ form a subgroup. The matrices in the general case do not. The product of two Lorentz transformation matrices $\mathsf{L}(\mathbf{v}_1)$ and $\mathsf{L}(\mathbf{v}_2)$ yields a third Lorentz matrix, $\mathsf{L}(\mathbf{v}_3)$, if the two velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ are parallel. The resultant velocity, $\mathbf{v}_3$, is related to $\mathbf{v}_1$ and $\mathbf{v}_2$ by the Einstein velocity addition law, Exercise 4.5.3. If $\mathbf{v}_1$ and $\mathbf{v}_2$ are not parallel, no such simple relation exists. Specifically, consider three reference frames S, S', and S'', with S and S' related by $\mathsf{L}(\mathbf{v}_1)$ and S' and S'' related by $\mathsf{L}(\mathbf{v}_2)$. If the velocity of S'' relative to the original system S is $\mathbf{v}_3$, S'' is not obtained from S by $\mathsf{L}(\mathbf{v}_3) = \mathsf{L}(\mathbf{v}_2)\mathsf{L}(\mathbf{v}_1)$. Rather, we find that

$$\mathsf{L}(\mathbf{v}_3) = \mathsf{R}\mathsf{L}(\mathbf{v}_2)\mathsf{L}(\mathbf{v}_1), \tag{4.139}$$

where R is a 3×3 space rotation matrix embedded in our four-dimensional space–time. With $\mathbf{v}_1$ and $\mathbf{v}_2$ not parallel, the final system, S'', is **rotated** relative to S. This rotation is the origin of the Thomas precession involved in spin-orbit coupling terms in atomic and nuclear physics. Because of its presence, the pure Lorentz transformations $\mathsf{L}(\mathbf{v})$ by themselves do not form a group.

Kinematics and Dynamics in Minkowski Space–Time

We have seen that the propagation of light determines the metric

$$\mathbf{r}^2 - c^2t^2 = 0 = \mathbf{r}'^2 - c^2t'^2,$$

where $x^\mu = (ct, \mathbf{r})$ is the coordinate four-vector. For a particle moving with velocity $\mathbf{v}$, the Lorentz invariant infinitesimal version

$$c\,d\tau \equiv \sqrt{dx^\mu\,dx_\mu} = \sqrt{c^2\,dt^2 - d\mathbf{r}^2} = dt\sqrt{c^2 - \mathbf{v}^2}$$

defines the invariant proper time τ on its track. Because of time dilation in moving frames, a proper-time clock rides with the particle (in its rest frame) and runs at the slowest possible

rate compared to any other inertial frame (of an observer, for example). The four-velocity of the particle can now be defined properly as

$$\frac{dx^\mu}{d\tau} = u^\mu = \left(\frac{c}{\sqrt{c^2 - \mathbf{v}^2}}, \frac{\mathbf{v}}{\sqrt{c^2 - \mathbf{v}^2}}\right),$$

so $u^2 = 1$, and the four-momentum $p^\mu = cmu^\mu = (\frac{E}{c}, \mathbf{p})$ yields Einstein's famous energy relation

$$E = \frac{mc^2}{\sqrt{1 - \mathbf{v}^2/c^2}} = mc^2 + \frac{m}{2}\mathbf{v}^2 \pm \cdots .$$

A consequence of $u^2 = 1$ and its physical significance is that the particle is on its mass shell $p^2 = m^2c^2$.

Now we formulate Newton's equation for a **single particle** of mass m in special relativity as $\frac{dp^\mu}{d\tau} = K^\mu$, with K^μ denoting the force four-vector, so its vector part of the equation coincides with the usual form. For $\mu = 1, 2, 3$ we use $d\tau = dt\sqrt{1 - \mathbf{v}^2/c^2}$ and find

$$\frac{1}{\sqrt{1 - \mathbf{v}^2/c^2}}\frac{d\mathbf{p}}{dt} = \frac{\mathbf{F}}{\sqrt{1 - \mathbf{v}^2/c^2}} = \mathbf{K},$$

determining $\mathbf{K}$ in terms of the usual force $\mathbf{F}$. We need to find K^0. We proceed by analogy with the derivation of energy conservation, multiplying the force equation into the four-velocity

$$mu_\nu \frac{du^\nu}{d\tau} = \frac{m}{2}\frac{du^2}{d\tau} = 0,$$

because $u^2 = 1 = \text{const}$. The other side of Newton's equation yields

$$0 = \frac{1}{c}u \cdot K = \frac{K^0}{\sqrt{1 - \mathbf{v}^2/c^2}} - \frac{\mathbf{F} \cdot \mathbf{v}/c}{\sqrt{1 - \mathbf{v}^2/c^{2^2}}},$$

so $K^0 = \frac{\mathbf{F}\cdot\mathbf{v}/c}{\sqrt{1-\mathbf{v}^2/c^2}}$ is related to the rate of work done by the force on the particle.

Now we turn to two-body collisions, in which energy–momentum conservation takes the form $p_1 + p_2 = p_3 + p_4$, where p_i^μ are the particle four-momenta. Because the scalar product of any four-vector with itself is an invariant under Lorentz transformations, it is convenient to define the Lorentz invariant energy squared $s = (p_1 + p_2)^2 = P^2$, where P^μ is the total four-momentum, and to use units where the velocity of light $c = 1$. The laboratory system (lab) is defined as the rest frame of the particle with four-momentum $p_2^\mu = (m_2, \mathbf{0})$ and the center of momentum frame (cms) by the total four-momentum $P^\mu = (E_1 + E_2, \mathbf{0})$. When the incident lab energy E_1^L is given, then

$$s = p_1^2 + p_2^2 + 2p_1 \cdot p_2 = m_1^2 + m^{2^2} + 2m_2E_1^L$$

is determined. Now, the cms energies of the four particles are obtained from scalar products

$$p_1 \cdot P = E_1(E_1 + E_2) = E_1\sqrt{s},$$

so

$$E_1 = \frac{p_1 \cdot (p_1 + p_2)}{\sqrt{s}} = \frac{m_1^2 + p_1 \cdot p_2}{\sqrt{s}} = \frac{m_1^2 - m_2^2 + s}{2\sqrt{s}},$$

$$E_2 = \frac{p_2 \cdot (p_1 + p_2)}{\sqrt{s}} = \frac{m_2^2 + p_1 \cdot p_2}{\sqrt{s}} = \frac{m_2^2 - m_1^2 + s}{2\sqrt{s}},$$

$$E_3 = \frac{p_3 \cdot (p_3 + p_4)}{\sqrt{s}} = \frac{m_3^2 + p_3 \cdot p_4}{\sqrt{s}} = \frac{m_3^2 - m_4^2 + s}{2\sqrt{s}},$$

$$E_4 = \frac{p_4 \cdot (p_3 + p_4)}{\sqrt{s}} = \frac{m_4^2 + p_3 \cdot p_4}{\sqrt{s}} = \frac{m_4^2 - m_3^2 + s}{2\sqrt{s}},$$

by substituting $2p_1 \cdot p_2 = s - m_1^2 - m_2^2$, $2p_3 \cdot p_4 = s - m_3^2 - m_4^2$, $2p_1 \cdot p_3 = m_1^2 + m_3^2 - t = (p_1 - p_3)^2 = m_1^2 + m_3^2 - 2p_1 \cdot p_3$

Thus, all cms energies E_i depend only on the incident energy but not on the scattering angle. For elastic scattering, $m_3 = m_1$, $m_4 = m_2$, so $E_3 = E_1$, $E_4 = E_2$. The Lorentz invariant momentum transfer squared

depends linearly on the cosine of the scattering angle.

Example 4.5.1 KAON DECAY AND PION PHOTOPRODUCTION THRESHOLD

Find the kinetic energies of the muon of mass 106 MeV and massless neutrino into which a K meson of mass 494 MeV decays in its rest frame.

Conservation of energy and momentum gives $m_K = E_\mu + E_\nu = \sqrt{\overline{s}}$. Applying the relativistic kinematics described previously yields

$$E_\mu = \frac{p_\mu \cdot (p_\mu + p_\nu)}{m_K} = \frac{m_\mu^2 + p_\mu \cdot p_\nu}{m_K},$$

$$E_\nu = \frac{p_\nu \cdot (p_\mu + p_\nu)}{m_K} = \frac{p_\mu \cdot p_\nu}{m_K}.$$

Combining both results we obtain $m_K^2 = m_\mu^2 + 2p_\mu \cdot p_\nu$, so

$$E_\mu = T_\mu + m_\mu = \frac{m_K^2 + m_\mu^2}{2m_K} = 258.4 \text{ MeV},$$

$$E_\nu = T_\nu = \frac{m_K^2 - m_\mu^2}{2m_K} = 235.6 \text{ MeV}.$$

As another example, in the production of a neutral pion by an incident photon according to $\gamma + p \to \pi^0 + p'$ at threshold, the neutral pion and proton are created at rest in the cms. Therefore,

$$s = (p_\gamma + p)^2 = m_p^2 + 2m_p E_\gamma^L = (p_\pi + p')^2 = (m_\pi + m_p)^2,$$

so $E_\gamma^L = m_\pi + \frac{m_\pi^2}{2m_p} = 144.7$ MeV. ■

Exercises

4.5.1 Two Lorentz transformations are carried out in succession: v_1 along the x-axis, then v_2 along the y-axis. Show that the resultant transformation (given by the product of these two successive transformations) **cannot** be put in the form of a single Lorentz transformation.
Note. The discrepancy corresponds to a rotation.

4.5.2 Rederive the Lorentz transformation, working entirely in the real space (x^0, x^1, x^2, x^3) with $x^0 = x_0 = ct$. Show that the Lorentz transformation may be written $\mathsf{L}(\mathbf{v}) = \exp(\rho\sigma)$, with

$$\sigma = \begin{pmatrix} 0 & -\lambda & -\mu & -\nu \\ -\lambda & 0 & 0 & 0 \\ -\mu & 0 & 0 & 0 \\ -\nu & 0 & 0 & 0 \end{pmatrix}$$

and λ, μ, ν the direction cosines of the velocity $\mathbf{v}$.

4.5.3 Using the matrix relation, Eq. (4.136), let the rapidity ρ_1 relate the Lorentz reference frames (x'^0, x'^1) and (x^0, x^1). Let ρ_2 relate (x''^0, x''^1) and (x'^0, x'^1). Finally, let ρ relate (x''^0, x''^1) and (x^0, x^1). From $\rho = \rho_1 + \rho_2$ derive the Einstein velocity addition law

$$v = \frac{v_1 + v_2}{1 + v_1 v_2/c^2}.$$

4.6 Lorentz Covariance of Maxwell's Equations

If a physical law is to hold for all orientations of our (real) coordinates (that is, to be invariant under rotations), the terms of the equation must be covariant under rotations (Sections 1.2 and 2.6). This means that we write the physical laws in the mathematical form scalar = scalar, vector = vector, second-rank tensor = second-rank tensor, and so on. Similarly, if a physical law is to hold for all inertial systems, the terms of the equation must be covariant under Lorentz transformations.

Using Minkowski space ($ct = x^0$; $x = x^1$, $y = x^2$, $z = x^3$), we have a four-dimensional space with the metric $g_{\mu\nu}$ (Eq. (4.128), Section 4.5). The Lorentz transformations are linear in space and time in this four-dimensional real space.[19]

[19] A group theoretic derivation of the Lorentz transformation in Minkowski space appears in Section 4.5. See also H. Goldstein, *Classical Mechanics*. Cambridge, MA: Addison-Wesley (1951), Chapter 6. The metric equation $x_0^2 - \mathbf{x}^2 = 0$, independent of reference frame, leads to the Lorentz transformations.

Here we consider Maxwell's equations,

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{4.140a}$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \rho \mathbf{v}, \tag{4.140b}$$

$$\nabla \cdot \mathbf{D} = \rho, \tag{4.140c}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{4.140d}$$

and the relations

$$\mathbf{D} = \varepsilon_0 \mathbf{E}, \qquad \mathbf{B} = \mu_0 \mathbf{H}. \tag{4.141}$$

The symbols have their usual meanings as given in Section 1.9. For simplicity we assume vacuum ($\varepsilon = \varepsilon_0$, $\mu = \mu_0$).

We assume that Maxwell's equations hold in all inertial systems; that is, Maxwell's equations are consistent with special relativity. (The covariance of Maxwell's equations under Lorentz transformations was actually shown by Lorentz and Poincaré before Einstein proposed his theory of special relativity.) Our immediate goal is to rewrite Maxwell's equations as tensor equations in Minkowski space. This will make the Lorentz covariance explicit, or manifest.

In terms of scalar, φ, and magnetic vector potentials, $\mathbf{A}$, we may solve[20] Eq. (4.140d) and then (4.140a) by

$$\begin{aligned} \mathbf{B} &= \nabla \times \mathbf{A} \\ \mathbf{E} &= -\frac{\partial \mathbf{A}}{\partial t} - \nabla \varphi. \end{aligned} \tag{4.142}$$

Equation (4.142) specifies the curl of $\mathbf{A}$; the divergence of $\mathbf{A}$ is still undefined (compare Section 1.16). We may, and for future convenience we do, impose a further gauge restriction on the vector potential $\mathbf{A}$:

$$\nabla \cdot \mathbf{A} + \varepsilon_0 \mu_0 \frac{\partial \varphi}{\partial t} = 0. \tag{4.143}$$

This is the Lorentz gauge relation. It will serve the purpose of uncoupling the differential equations for $\mathbf{A}$ and φ that follow. The potentials $\mathbf{A}$ and φ are not yet completely fixed. The freedom remaining is the topic of Exercise 4.6.4.

Now we rewrite the Maxwell equations in terms of the potentials $\mathbf{A}$ and φ. From Eqs. (4.140c) for $\nabla \cdot \mathbf{D}$, (4.141) and (4.142),

$$\nabla^2 \varphi + \nabla \cdot \frac{\partial \mathbf{A}}{\partial t} = -\frac{\rho}{\varepsilon_0}, \tag{4.144}$$

whereas Eqs. (4.140b) for $\nabla \times \mathbf{H}$ and (4.142) and Eq. (1.86c) of Chapter 1 yield

$$\frac{\partial^2 \mathbf{A}}{\partial t^2} + \nabla \frac{\partial \varphi}{\partial t} + \frac{1}{\varepsilon_0 \mu_0} \left\{ \nabla \nabla \cdot \mathbf{A} - \nabla^2 \mathbf{A} \right\} = \frac{\rho \mathrm{v}}{\varepsilon_0}. \tag{4.145}$$

[20] Compare Section 1.13, especially Exercise 1.13.10.

Using the Lorentz relation, Eq. (4.143), and the relation $\varepsilon_0\mu_0 = 1/c^2$, we obtain

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]\mathbf{A} = -\mu_0\rho\mathbf{v},$$
$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]\varphi = -\frac{\rho}{\varepsilon_0}. \tag{4.146}$$

Now, the differential operator (see also Exercise 2.7.3)

$$\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} \equiv -\partial^2 \equiv -\partial^\mu\partial_\mu$$

is a four-dimensional Laplacian, usually called the d'Alembertian and also sometimes denoted by $\Box$. It is a scalar by construction (see Exercise 2.7.3).

For convenience we define

$$\begin{aligned} A^1 &\equiv \frac{A_x}{\mu_0 c} = c\varepsilon_0 A_x, \qquad & A^3 &\equiv \frac{A_z}{\mu_0 c} = c\varepsilon_0 A_z, \\ A^2 &\equiv \frac{A_y}{\mu_0 c} = c\varepsilon_0 A_y, \qquad & A_0 &\equiv \varepsilon_0\varphi = A^0. \end{aligned} \tag{4.147}$$

If we further define a four-vector current density

$$\frac{\rho v_x}{c} \equiv j^1, \qquad \frac{\rho v_y}{c} \equiv j^2, \qquad \frac{\rho v_z}{c} \equiv j^3, \qquad \rho \equiv j_0 = j^0, \tag{4.148}$$

then Eq. (4.146) may be written in the form

$$\partial^2 A^\mu = j^\mu. \tag{4.149}$$

The wave equation (4.149) looks like a four-vector equation, but looks do not constitute proof. To prove that it is a four-vector equation, we start by investigating the transformation properties of the generalized current j^μ.

Since an electric charge element de is an invariant quantity, we have

$$de = \rho dx^1\,dx^2\,dx^3, \qquad \text{invariant.} \tag{4.150}$$

We saw in Section 2.9 that the four-dimensional volume element $dx^0\,dx^1\,dx^2\,dx^3$ was also invariant, a pseudoscalar. Comparing this result, Eq. (2.106), with Eq. (4.150), we see that the charge density ρ must transform the same way as dx^0, the zeroth component of a four-dimensional vector dx^λ. We put $\rho = j^0$, with j^0 now established as the zeroth component of a four-vector. The other parts of Eq. (4.148) may be expanded as

$$j^1 = \frac{\rho v_x}{c} = \frac{\rho}{c}\frac{dx^1}{dt} = j^0\frac{dx^1}{dx^0}. \tag{4.151}$$

Since we have just shown that j^0 transforms as dx^0, this means that j^1 transforms as dx^1. With similar results for j^2 and j^3, We have j^λ transforming as dx^λ, proving that j^λ is a four-vector in Minkowski space.

Equation (4.149), which follows directly from Maxwell's equations, Eqs. (4.140), is assumed to hold in all Cartesian systems (all Lorentz frames). Then, by the quotient rule, Section 2.8, A^μ is also a vector and Eq. (4.149) is a legitimate tensor equation.

Now, working backward, Eq. (4.142) may be written

$$\begin{aligned} \varepsilon_0 E_j &= -\frac{\partial A^j}{\partial x^0} - \frac{\partial A^0}{\partial x^j}, \qquad j = 1, 2, 3, \\ \frac{1}{\mu_0 c} B_i &= \frac{\partial A^k}{\partial x^j} - \frac{\partial A^j}{\partial x^k}, \qquad (i, j, k) = \text{cyclic } (1, 2, 3). \end{aligned} \tag{4.152}$$

We define a new tensor,

$$\partial^\mu A^\lambda - \partial^\lambda A^\mu = \frac{\partial A^\lambda}{\partial x_\mu} - \frac{\partial A^\mu}{\partial x_\lambda} \equiv F^{\mu\lambda} = -F^{\lambda\mu} \qquad (\mu, \lambda = 0, 1, 2, 3),$$

an antisymmetric second-rank tensor, since A^λ is a vector. Written out explicitly,

$$\frac{F_{\mu\lambda}}{\varepsilon_0} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -cB_z & cB_y \\ -E_y & cB_z & 0 & -cB_x \\ -E_z & -cB_y & cB_x & 0 \end{pmatrix}, \qquad \frac{F^{\mu\lambda}}{\varepsilon_0} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -cB_z & cB_y \\ E_y & cB_z & 0 & -cB_x \\ E_z & -cB_y & cB_x & 0 \end{pmatrix}. \tag{4.153}$$

Notice that in our four-dimensional Minkowski space **E** and **B** are no longer vectors but together form a second-rank tensor. With this tensor we may write the two nonhomogeneous Maxwell equations ((4.140b) and (4.140c)) combined as a tensor equation,

$$\frac{\partial F_{\lambda\mu}}{\partial x_\mu} = j_\lambda. \tag{4.154}$$

The left-hand side of Eq. (4.154) is a four-dimensional divergence of a tensor and therefore a vector. This, of course, is equivalent to contracting a third-rank tensor $\partial F^{\lambda\mu}/\partial x_\nu$ (compare Exercises 2.7.1 and 2.7.2). The two homogeneous Maxwell equations — (4.140a) for $\nabla \times \mathbf{E}$ and (4.140d) for $\nabla \cdot \mathbf{B}$ — may be expressed in the tensor form

$$\frac{\partial F_{23}}{\partial x_1} + \frac{\partial F_{31}}{\partial x_2} + \frac{\partial F_{12}}{\partial x_3} = 0 \tag{4.155}$$

for Eq. (4.140d) and three equations of the form

$$-\frac{\partial F_{30}}{\partial x_2} - \frac{\partial F_{02}}{\partial x_3} + \frac{\partial F_{23}}{\partial x_0} = 0 \tag{4.156}$$

for Eq. (4.140a). (A second equation permutes 120, a third permutes 130.) Since

$$\partial^\lambda F^{\mu\nu} = \frac{\partial F^{\mu\nu}}{\partial x_\lambda} \equiv t^{\lambda\mu\nu}$$

is a tensor (of third rank), Eqs. (4.140a) and (4.140d) are given by the tensor equation

$$t^{\lambda\mu\nu} + t^{\nu\lambda\mu} + t^{\mu\nu\lambda} = 0. \tag{4.157}$$

From Eqs. (4.155) and (4.156) you will understand that the indices λ, μ, and ν are supposed to be different. Actually Eq. (4.157) automatically reduces to $0 = 0$ if any two indices coincide. An alternate form of Eq. (4.157) appears in Exercise 4.6.14.

Lorentz Transformation of E and B

The construction of the tensor equations ((4.154) and (4.157)) completes our initial goal of rewriting Maxwell's equations in tensor form.[21] Now we exploit the tensor properties of our four vectors and the tensor $F_{\mu\nu}$.

For the Lorentz transformation corresponding to motion along the $z(x_3)$-axis with velocity v, the "direction cosines" are given by[22]

$$\begin{aligned} x'^0 &= \gamma\left(x^0 - \beta x^3\right) \\ x'^3 &= \gamma\left(x^3 - \beta x^0\right), \end{aligned} \tag{4.158}$$

where

$$\beta = \frac{v}{c}$$

and

$$\gamma = \left(1 - \beta^2\right)^{-1/2}. \tag{4.159}$$

Using the tensor transformation properties, we may calculate the electric and magnetic fields in the moving system in terms of the values in the original reference frame. From Eqs. (2.66), (4.153), and (4.158) we obtain

$$\begin{aligned} E'_x &= \frac{1}{\sqrt{1-\beta^2}}\left(E_x - \frac{v}{c^2}B_y\right), \\ E'_y &= \frac{1}{\sqrt{1-\beta^2}}\left(E_y + \frac{v}{c^2}B_x\right), \\ E'_z &= E_z \end{aligned} \tag{4.160}$$

and

$$\begin{aligned} B'_x &= \frac{1}{\sqrt{1-\beta^2}}\left(B_x + \frac{v}{c^2}E_y\right), \\ B'_y &= \frac{1}{\sqrt{1-\beta^2}}\left(B_y - \frac{v}{c^2}E_x\right), \\ B'_z &= B_z. \end{aligned} \tag{4.161}$$

This coupling of **E** and **B** is to be expected. Consider, for instance, the case of zero electric field in the unprimed system

$$E_x = E_y = E_z = 0.$$

[21] Modern theories of quantum electrodynamics and elementary particles are often written in this "manifestly covariant" form to guarantee consistency with special relativity. Conversely, the insistence on such tensor form has been a useful guide in the construction of these theories.

[22] A group theoretic derivation of the Lorentz transformation appears in Section 4.5. See also Goldstein, *loc. cit.*, Chapter 6.

Clearly, there will be no force on a stationary charged particle. When the particle is in motion with a small velocity v along the z-axis,[23] an observer on the particle sees fields (exerting a force on his charged particle) given by

$$E'_x = -vB_y,$$
$$E'_y = vB_x,$$

where **B** is a magnetic induction field in the unprimed system. These equations may be put in vector form,

$$\mathbf{E}' = \mathbf{v} \times \mathbf{B}$$

or (4.162)

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B},$$

which is usually taken as the operational definition of the magnetic induction **B**.

Electromagnetic Invariants

Finally, the tensor (or vector) properties allow us to construct a multitude of invariant quantities. A more important one is the scalar product of the two four-dimensional vectors or four-vectors A_λ and j_λ. We have

$$\begin{aligned} A^\lambda j_\lambda &= -c\varepsilon_0 A_x \frac{\rho v_x}{c} - c\varepsilon_0 A_y \frac{\rho v_y}{c} - c\varepsilon_0 A_z \frac{\rho v_z}{c} + \varepsilon_0 \varphi \rho \\ &= \varepsilon_0(\rho\varphi - \mathbf{A}\cdot\mathbf{J}), \qquad \text{invariant,} \end{aligned} \tag{4.163}$$

with **A** the usual magnetic vector potential and **J** the ordinary current density. The first term, $\rho\varphi$, is the ordinary static electric coupling, with dimensions of energy per unit volume. Hence our newly constructed scalar invariant is an energy density. The dynamic interaction of field and current is given by the product $\mathbf{A}\cdot\mathbf{J}$. This invariant $A^\lambda j_\lambda$ appears in the electromagnetic Lagrangians of Exercises 17.3.6 and 17.5.1.

Other possible electromagnetic invariants appear in Exercises 4.6.9 and 4.6.11.

The Lorentz group is the symmetry group of electrodynamics, of the electroweak gauge theory, and of the strong interactions described by quantum chromodynamics: It governs special relativity. The metric of Minkowski space–time is Lorentz invariant and expresses the propagation of light; that is, the velocity of light is the same in all inertial frames. Newton's equations of motion are straightforward to extend to special relativity. The kinematics of two-body collisions are important applications of vector algebra in Minkowski space–time.

[23] If the velocity is not small, a relativistic transformation of force is needed.

Exercises

4.6.1 (a) Show that every four-vector in Minkowski space may be decomposed into an ordinary three-space vector and a three-space scalar. Examples: $(ct, \mathbf{r})$, $(\rho, \rho\mathbf{v}/c)$, $(\varepsilon_0\varphi, c\varepsilon_0\mathbf{A})$, $(E/c, \mathbf{p})$, $(\omega/c, \mathbf{k})$.
Hint. Consider a rotation of the three-space coordinates with time fixed.

(b) Show that the converse of (a) is **not** true — every three-vector plus scalar does **not** form a Minkowski four-vector.

4.6.2 (a) Show that

$$\partial^\mu j_\mu = \partial \cdot j = \frac{\partial j_\mu}{\partial x_\mu} = 0.$$

(b) Show how the previous tensor equation may be interpreted as a statement of continuity of charge and current in ordinary three-dimensional space and time.

(c) If this equation is known to hold in all Lorentz reference frames, why can we not conclude that j_μ is a vector?

4.6.3 Write the Lorentz gauge condition (Eq. (4.143)) as a tensor equation in Minkowski space.

4.6.4 A gauge transformation consists of varying the scalar potential φ_1 and the vector potential $\mathbf{A}_1$ according to the relation

$$\varphi_2 = \varphi_1 + \frac{\partial \chi}{\partial t},$$
$$\mathbf{A}_2 = \mathbf{A}_1 - \boldsymbol{\nabla}\chi.$$

The new function χ is required to satisfy the homogeneous wave equation

$$\boldsymbol{\nabla}^2\chi - \frac{1}{c^2}\frac{\partial^2\chi}{\partial t^2} = 0.$$

Show the following:

(a) The Lorentz gauge relation is unchanged.

(b) The new potentials satisfy the same inhomogeneous wave equations as did the original potentials.

(c) The fields **E** and **B** are unaltered.

The invariance of our electromagnetic theory under this transformation is called **gauge invariance**.

4.6.5 A charged particle, charge q, mass m, obeys the Lorentz covariant equation

$$\frac{dp^\mu}{d\tau} = \frac{q}{\varepsilon_0 mc} F^{\mu\nu} p_\nu,$$

where p^ν is the four-momentum vector $(E/c; p^1, p^2, p^3)$, τ is the proper time, $d\tau = dt\sqrt{1 - v^2/c^2}$, a Lorentz scalar. Show that the explicit space–time forms are

$$\frac{dE}{dt} = q\mathbf{v}\cdot\mathbf{E}; \qquad \frac{d\mathbf{p}}{dt} = q(\mathbf{E} + \mathbf{v}\times\mathbf{B}).$$

4.6.6 From the Lorentz transformation matrix elements (Eq. (4.158)) derive the Einstein velocity addition law

$$u' = \frac{u - v}{1 - (uv/c^2)} \qquad \text{or} \qquad u = \frac{u' + v}{1 + (u'v/c^2)},$$

where $u = c\,dx^3/dx^0$ and $u' = c\,dx'^3/dx'^0$.
Hint. If $L_{12}(v)$ is the matrix transforming system 1 into system 2, $L_{23}(u')$ the matrix transforming system 2 into system 3, $L_{13}(u)$ the matrix transforming system 1 directly into system 3, then $L_{13}(u) = L_{23}(u')L_{12}(v)$. From this matrix relation extract the Einstein velocity addition law.

4.6.7 The dual of a four-dimensional second-rank tensor $\mathbf{B}$ may be defined by $\tilde{\mathbf{B}}$, where the elements of the dual tensor are given by

$$\tilde{B}^{ij} = \frac{1}{2!}\varepsilon^{ijkl}B_{kl}.$$

Show that $\tilde{\mathbf{B}}$ transforms as

(a) a second-rank tensor under rotations,
(b) a pseudotensor under inversions.

Note. The tilde here does **not** mean transpose.

4.6.8 Construct $\tilde{\mathbf{F}}$, the dual of $\mathbf{F}$, where $\mathbf{F}$ is the electromagnetic tensor given by Eq. (4.153).

$$\text{ANS. } \tilde{\mathbf{F}}^{\mu\nu} = \varepsilon_0 \begin{pmatrix} 0 & -cB_x & -cB_y & -cB_z \\ cB_x & 0 & E_z & -E_y \\ cB_y & -E_z & 0 & E_x \\ cB_z & E_y & -E_x & 0 \end{pmatrix}.$$

This corresponds to

$$c\mathbf{B} \to -\mathbf{E},$$

$$\mathbf{E} \to c\mathbf{B}.$$

This transformation, sometimes called a **dual transformation**, leaves Maxwell's equations in vacuum ($\rho = 0$) invariant.

4.6.9 Because the quadruple contraction of a fourth-rank pseudotensor and two second-rank tensors $\varepsilon_{\mu\lambda\nu\sigma}F^{\mu\lambda}F^{\nu\sigma}$ is clearly a pseudoscalar, evaluate it.

$$\text{ANS. } -8\varepsilon_0^2 c\mathbf{B}\cdot\mathbf{E}.$$

4.6.10 (a) If an electromagnetic field is purely electric (or purely magnetic) in one particular Lorentz frame, show that $\mathbf{E}$ and $\mathbf{B}$ will be orthogonal in other Lorentz reference systems.
(b) Conversely, if $\mathbf{E}$ and $\mathbf{B}$ are orthogonal in one particular Lorentz frame, there exists a Lorentz reference system in which $\mathbf{E}$ (or $\mathbf{B}$) vanishes. Find that reference system.

4.6.11 Show that $c^2\mathbf{B}^2 - \mathbf{E}^2$ is a Lorentz scalar.

4.6.12 Since (dx^0, dx^1, dx^2, dx^3) is a four-vector, $dx_\mu\, dx^\mu$ is a scalar. Evaluate this scalar for a moving particle in two different coordinate systems: (a) a coordinate system fixed relative to you (lab system), and (b) a coordinate system moving with a moving particle (velocity v relative to you). With the time increment labeled $d\tau$ in the particle system and dt in the lab system, show that

$$d\tau = dt\sqrt{1 - v^2/c^2}.$$

τ is the proper time of the particle, a Lorentz invariant quantity.

4.6.13 Expand the scalar expression

$$-\frac{1}{4\varepsilon_0} F_{\mu\nu} F^{\mu\nu} + \frac{1}{\varepsilon_0} j_\mu A^\mu$$

in terms of the fields and potentials. The resulting expression is the Lagrangian density used in Exercise 17.5.1.

4.6.14 Show that Eq. (4.157) may be written

$$\varepsilon_{\alpha\beta\gamma\delta} \frac{\partial F^{\alpha\beta}}{\partial x_\gamma} = 0.$$

4.7 DISCRETE GROUPS

Here we consider groups with a finite number of elements. In physics, groups usually appear as a set of operations that leave a system unchanged, invariant. This is an expression of symmetry. Indeed, a symmetry may be defined as the invariance of the Hamiltonian of a system under a group of transformations. Symmetry in this sense is important in classical mechanics, but it becomes even more important and more profound in quantum mechanics. In this section we investigate the symmetry properties of sets of objects (atoms in a molecule or crystal). This provides additional illustrations of the group concepts of Section 4.1 and leads directly to dihedral groups. The dihedral groups in turn open up the study of the 32 crystallographic point groups and 230 space groups that are of such importance in crystallography and solid-state physics. It might be noted that it was through the study of crystal symmetries that the concepts of symmetry and group theory entered physics. In physics, the abstract group conditions often take on direct physical meaning in terms of transformations of vectors, spinors, and tensors.

As a simple, but not trivial, example of a finite group, consider the set $1, a, b, c$ that combine according to the group multiplication table[24] (see Fig. 4.10). Clearly, the four conditions of the definition of "group" are satisfied. The elements a, b, c, and 1 are abstract mathematical entities, completely unrestricted except for the multiplication table of Fig. 4.10.

Now, for a specific representation of these group elements, let

$$1 \to 1, \qquad a \to i, \qquad b \to -1, \qquad c \to -i, \tag{4.164}$$

[24]The order of the factors is row–column: $ab = c$ in the indicated previous example.

	1	a	b	c
1	1	a	b	c
a	a	b	c	1
b	b	c	1	a
c	c	1	a	b

FIGURE 4.10 Group multiplication table.

combining by ordinary multiplication. Again, the four group conditions are satisfied, and these four elements form a group. We label this group C_4. Since the multiplication of the group elements is commutative, the group is labeled **commutative**, or **abelian**. Our group is also a **cyclic group**, in that the elements may be written as successive powers of one element, in this case $i^n, n = 0, 1, 2, 3$. Note that in writing out Eq. (4.164) we have selected a specific faithful representation for this group of four objects, C_4.

We recognize that the group elements $1, i, -1, -i$ may be interpreted as successive 90° rotations in the complex plane. Then, from Eq. (3.74), we create the set of four 2×2 matrices (replacing φ by $-\varphi$ in Eq. (3.74) to rotate a vector rather than rotate the coordinates):

$$\mathsf{R}(\varphi) = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix},$$

and for $\varphi = 0, \pi/2, \pi$, and $3\pi/2$ we have

$$\mathsf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \mathsf{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$
$$\mathsf{B} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \mathsf{C} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{4.165}$$

This set of four matrices forms a group, with the law of combination being matrix multiplication. Here is a second faithful representation. By matrix multiplication one verifies that this representation is also abelian and cyclic. Clearly, there is a one-to-one correspondence of the two representations

$$1 \leftrightarrow 1 \leftrightarrow \mathsf{1} \qquad a \leftrightarrow i \leftrightarrow \mathsf{A} \qquad b \leftrightarrow -1 \leftrightarrow \mathsf{B} \qquad c \leftrightarrow -i \leftrightarrow \mathsf{C}. \tag{4.166}$$

In the group C_4 the two representations $(1, i, -1, -i)$ and $(\mathsf{1}, \mathsf{A}, \mathsf{B}, \mathsf{C})$ are isomorphic.

In contrast to this, there is no such correspondence between either of these representations of group C_4 and another group of four objects, the vierergruppe (Exercise 3.2.7). The

Table 4.3

	1	V_1	V_2	V_3
1	1	V_1	V_2	V_3
V_1	V_1	1	V_3	V_2
V_2	V_2	V_3	1	V_1
V_3	V_3	V_2	V_1	1

vierergruppe has the multiplication table shown in Table 4.3. Confirming the lack of correspondence between the group represented by $(1, i, -1, -i)$ or the matrices $(\mathsf{1}, \mathsf{A}, \mathsf{B}, \mathsf{C})$ of Eq. (4.165), note that although the vierergruppe is abelian, it is not cyclic. The cyclic group C_4 and the vierergruppe are not isomorphic.

Classes and Character

Consider a group element x transformed into a group element y by a similarity transform with respect to g_i, an element of the group

$$g_i x g_i^{-1} = y. \tag{4.167}$$

The group element y is **conjugate** to x. A **class** is a set of mutually conjugate group elements. In general, this set of elements forming a class does not satisfy the group postulates and is not a group. Indeed, the unit element 1, which is always in a class by itself, is the only class that is also a subgroup. All members of a given class are equivalent, in the sense that any one element is a similarity transform of any other element. Clearly, if a group is abelian, every element is a class by itself. We find that

1. Every element of the original group belongs to one and only one class.
2. The number of elements in a class is a factor of the order of the group.

We get a possible physical interpretation of the concept of class by noting that y is a similarity transform of x. If g_i represents a rotation of the coordinate system, then y is the same operation as x but relative to the new, related coordinates.

In Section 3.3 we saw that a real matrix transforms under rotation of the coordinates by an orthogonal similarity transformation. Depending on the choice of reference frame, essentially the same matrix may take on an infinity of different forms. Likewise, our group representations may be put in an infinity of different forms by using unitary transformations. But each such transformed representation is isomorphic with the original. From Exercise 3.3.9 the trace of each element (each matrix of our representation) is invariant under unitary transformations. Just because it is invariant, the trace (relabeled the **character**) assumes a role of some importance in group theory, particularly in applications to solid-state physics. Clearly, all members of a given class (in a given representation) have the same character. Elements of different classes may have the same character, but elements with different characters cannot be in the same class.

The concept of class is important (1) because of the trace or character and (2) because **the number of nonequivalent irreducible representations of a group is equal to the number of classes.**

Subgroups and Cosets

Frequently a subset of the group elements (including the unit element I) will by itself satisfy the four group requirements and therefore is a group. Such a subset is called a **subgroup**. Every group has two trivial subgroups: the unit element alone and the group itself. The elements 1 and b of the four-element group C_4 discussed earlier form a nontrivial

subgroup. In Section 4.1 we consider SO(3), the (continuous) group of all rotations in ordinary space. The rotations about any single axis form a subgroup of SO(3). Numerous other examples of subgroups appear in the following sections.

Consider a subgroup H with elements h_i and a group element x not in H. Then xh_i and $h_i x$ are not in subgroup H. The sets generated by

$$xh_i, \quad i = 1, 2, \ldots \qquad \text{and} \qquad h_i x, \quad i = 1, 2, \ldots$$

are called **cosets**, respectively the left and right cosets of subgroup H with respect to x. It can be shown (assume the contrary and prove a contradiction) that the coset of a subgroup has the same number of distinct elements as the subgroup. Extending this result we may express the original group G as the sum of H and cosets:

$$G = H + x_1 H + x_2 H + \cdots .$$

Then the order of any subgroup is a divisor of the order of the group. It is this result that makes the concept of coset significant. In the next section the six-element group D_3 (order 6) has subgroups of order 1, 2, and 3. D_3 cannot (and does not) have subgroups of order 4 or 5.

The similarity transform of a subgroup H by a fixed group element x **not** in H, xHx^{-1}, yields a subgroup — Exercise 4.7.8. If this new subgroup is identical with H for all x, that is,

$$xHx^{-1} = H,$$

then H is called an **invariant, normal**, or **self-conjugate subgroup**. Such subgroups are involved in the analysis of multiplets of atomic and nuclear spectra and the particles discussed in Section 4.2. All subgroups of a commutative (abelian) group are automatically invariant.

Two Objects — Twofold Symmetry Axis

Consider first the two-dimensional system of two identical atoms in the xy-plane at (1, 0) and $(-1, 0)$, Fig. 4.11. What rotations[25] can be carried out (keeping both atoms in the xy-plane) that will leave this system invariant? The first candidate is, of course, the unit operator 1. A rotation of π radians about the z-axis completes the list. So we have a rather uninteresting group of two members $(1, -1)$. The z-axis is labeled a twofold symmetry axis — corresponding to the two rotation angles, 0 and π, that leave the system invariant.

Our system becomes more interesting in three dimensions. Now imagine a molecule (or part of a crystal) with atoms of element X at $\pm a$ on the x-axis, atoms of element Y at $\pm b$ on the y-axis, and atoms of element Z at $\pm c$ on the z-axis, as show in Fig. 4.12. Clearly, each axis is now a twofold symmetry axis. Using $R_x(\pi)$ to designate a rotation of π radians about the x-axis, we may

[25] Here we deliberately exclude reflections and inversions. They must be brought in to develop the full set of 32 crystallographic point groups.

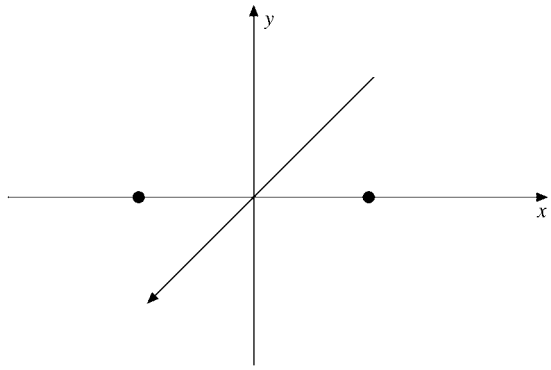

FIGURE 4.11 Diatomic molecules H_2, N_2, O_2, Cl_2.

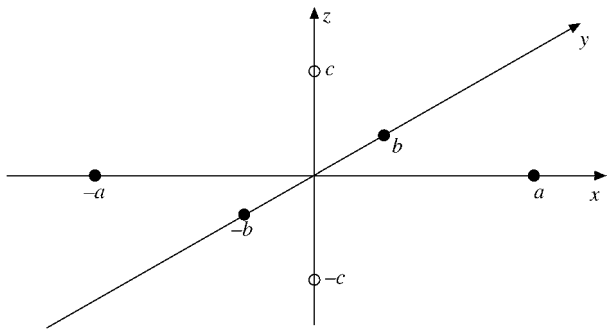

FIGURE 4.12 D_2 symmetry.

set up a matrix representation of the rotations as in Section 3.3:

$$\mathsf{R}_x(\pi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \qquad \mathsf{R}_y(\pi) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$
$$\mathsf{R}_z(\pi) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \mathsf{1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4.168}$$

These four elements $[\mathsf{1}, \mathsf{R}_x(\pi), \mathsf{R}_y(\pi), \mathsf{R}_z(\pi)]$ form an abelian group, with the group multiplication table shown in Table 4.4.

The products shown in Table 4.4 can be obtained in either of two distinct ways: (1) We may analyze the operations themselves—a rotation of π about the x-axis followed by a rotation of π about the y-axis is equivalent to a rotation of π about the z-axis: $\mathsf{R}_y(\pi)\mathsf{R}_x(\pi) = \mathsf{R}_z(\pi)$. (2) Alternatively, once a faithful representation is established, we

Table 4.4

	1	$R_x(\pi)$	$R_y(\pi)$	$R_z(\pi)$
1	1	R_x	R_y	R_x
$R_x(\pi)$	R_x	1	R_z	R_y
$R_y(\pi)$	R_y	R_z	1	R_x
$R_z(\pi)$	R_z	R_y	R_x	1

can obtain the products by matrix multiplication. This is where the power of mathematics is shown — when the system is too complex for a direct physical interpretation.

Comparison with Exercises 3.2.7, 4.7.2, and 4.7.3 shows that this group is the vierergruppe. The matrices of Eq. (4.168) are isomorphic with those of Exercise 3.2.7. Also, they are reducible, being diagonal. The subgroups are $(1, R_x)$, $(1, R_y)$, and $(1, R_z)$. They are invariant. It should be noted that a rotation of π about the y-axis and a rotation of π about the z-axis is equivalent to a rotation of π about the x-axis: $R_z(\pi)R_y(\pi) = R_x(\pi)$. In symmetry terms, if y and z are twofold symmetry axes, x is automatically a twofold symmetry axis.

This symmetry group,[26] the vierergruppe, is often labeled D_2, the D signifying a dihedral group and the subscript 2 signifying a twofold symmetry axis (and no higher symmetry axis).

Three Objects — Threefold Symmetry Axis

Consider now three identical atoms at the vertices of an equilateral triangle, Fig. 4.13. Rotations of the **triangle** of 0, $2\pi/3$, and $4\pi/3$ leave the triangle invariant. In matrix form, we have[27]

$$
\begin{aligned}
1 &= R_z(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
A &= R_z(2\pi/3) = \begin{pmatrix} \cos 2\pi/3 & -\sin 2\pi/3 \\ \sin 2\pi/3 & \cos 2\pi/3 \end{pmatrix} = \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix} \\
B &= R_z(4\pi/3) = \begin{pmatrix} -1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix}.
\end{aligned} \tag{4.169}
$$

The z-axis is a threefold symmetry axis. $(1, A, B)$ form a cyclic group, a subgroup of the complete six-element group that follows.

In the xy-plane there are three additional axes of symmetry — each atom (vertex) and the geometric center defining an axis. Each of these is a twofold symmetry axis. These rotations may most easily be described within our two-dimensional framework by introducing

[26] A **symmetry** group is a group of symmetry-preserving operations, that is, rotations, reflections, and inversions. A **symmetric** group is the group of permutations of n distinct objects — of order $n!$.

[27] Note that here we are rotating the **triangle** counterclockwise relative to fixed coordinates.

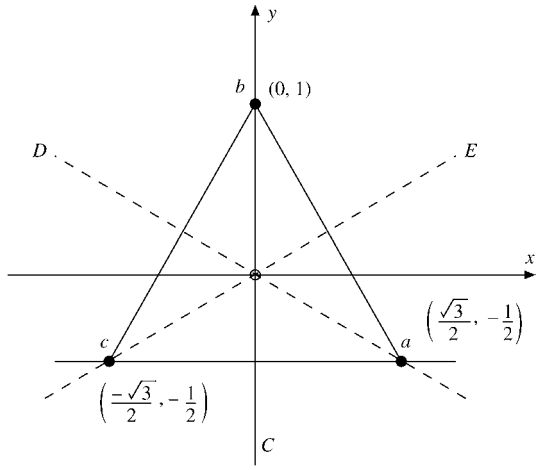

FIGURE 4.13 Symmetry operations on an equilateral triangle.

reflections. The rotation of π about the C- (or y-) axis, which means the interchanging of (structureless) atoms a and c, is just a reflection of the x-axis:

$$\mathsf{C} = \mathsf{R}_C(\pi) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.170}$$

We may replace the rotation about the D-axis by a rotation of $4\pi/3$ (about our z-axis) followed by a reflection of the x-axis ($x \to -x$) (Fig. 4.14):

$$\begin{aligned} \mathsf{D} &= \mathsf{R}_D(\pi) = \mathsf{CB} \\ &= \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix} \\ &= \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix}. \end{aligned} \tag{4.171}$$

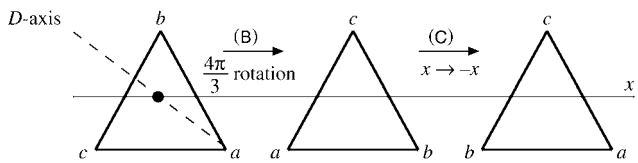

FIGURE 4.14 The triangle on the right is the triangle on the left rotated 180° about the D-axis. $\mathsf{D} = \mathsf{CB}$.

In a similar manner, the rotation of π about the E-axis, interchanging a and b, is replaced by a rotation of $2\pi/3(\mathsf{A})$ and then a reflection[28] of the x-axis:

$$\begin{aligned}\mathsf{E} &= \mathsf{R}_E(\pi) = \mathsf{CA} \\ &= \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix} \\ &= \begin{pmatrix} 1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}. \end{aligned} \tag{4.172}$$

The complete group multiplication table is

	1	A	B	C	D	E
1	1	A	B	C	D	E
A	A	B	1	D	E	C
B	B	1	A	E	C	D
C	C	E	D	1	B	A
D	D	C	E	A	1	B
E	E	D	C	B	A	1

Notice that each element of the group appears only once in each row and in each column, as required by the rearrangement theorem, Exercise 4.7.4. Also, from the multiplication table the group is not abelian. We have constructed a six-element group and a 2×2 irreducible matrix representation of it. The only other distinct six-element group is the cyclic group $[1, \mathsf{R}, \mathsf{R}^2, \mathsf{R}^3, \mathsf{R}^4, \mathsf{R}^5]$, with

$$R = e^{2\pi i/6} \qquad \text{or} \qquad \mathsf{R} = e^{-\pi i \sigma_2/3} = \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{pmatrix}. \tag{4.173}$$

Our group $[1, \mathsf{A}, \mathsf{B}, \mathsf{C}, \mathsf{D}, \mathsf{E}]$ is labeled D_3 in crystallography, the dihedral group with a threefold axis of symmetry. The three axes (C, D, and E) in the xy-plane automatically become twofold symmetry axes. As a consequence, $(1, \mathsf{C})$, $(1, \mathsf{D})$, and $(1, \mathsf{E})$ all form two-element subgroups. None of these two-element subgroups of D_3 is invariant.

A general and most important result for finite groups of h elements is that

$$\sum_i n_i^2 = h, \tag{4.174}$$

where n_i is the dimension of the matrices of the ith irreducible representation. This equality, sometimes called the **dimensionality theorem**, is very useful in establishing the irreducible representations of a group. Here for D_3 we have $1^2 + 1^2 + 2^2 = 6$ for our three representations. No other irreducible representations of this symmetry group of three objects exist. (The other representations are the identity and ± 1, depending upon whether a reflection was involved.)

[28] Note that, as a consequence of these reflections, $\det(\mathsf{C}) = \det(\mathsf{D}) = \det(\mathsf{E}) = -1$. The rotations A and B, of course, have a determinant of $+1$.

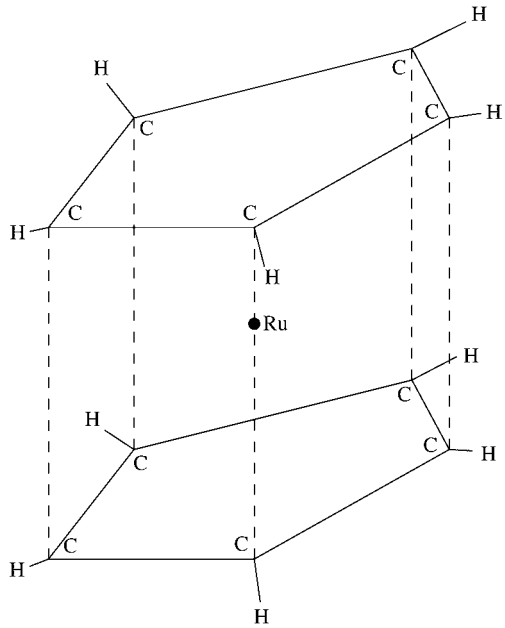

FIGURE 4.15 Ruthenocene.

Dihedral Groups, D_n

A dihedral group D_n with an n-fold symmetry axis implies n axes with angular separation of $2\pi/n$ radians, n is a positive integer, but otherwise unrestricted. If we apply the symmetry arguments to **crystal lattices**, then n is limited to 1, 2, 3, 4, and 6. The requirement of invariance of the crystal lattice under translations in the plane perpendicular to the n-fold axis excludes $n = 5, 7$, and higher values. Try to cover a plane completely with identical regular pentagons and with no overlapping.[29] For individual molecules, this constraint does not exist, although the examples with $n > 6$ are rare. $n = 5$ is a real possibility. As an example, the symmetry group for ruthenocene, $(C_5H_5)_2$Ru, illustrated in Fig. 4.15, is D_5.[30]

Crystallographic Point and Space Groups

The dihedral groups just considered are examples of the crystallographic point groups. A point group is composed of combinations of rotations and reflections (including inversions) that will leave some crystal lattice unchanged. Limiting the operations to rotations and reflections (including inversions) means that one point — the origin — remains **fixed**, hence the term **point group**. Including the cyclic groups, two cubic groups (tetrahedron and octahedron symmetries), and the improper forms (involving reflections), we come to a total of 32 crystallographic point groups.

[29] For D_6 imagine a plane covered with regular hexagons and the axis of rotation through the geometric center of one of them.
[30] Actually the full technical label is D_{5h}, with h indicating invariance under a **reflection** of the fivefold axis.

If, to the rotation and reflection operations that produced the point groups, we add the possibility of translations and still demand that some crystal lattice remain invariant, we come to the space groups. There are 230 distinct space groups, a number that is appalling except, possibly, to specialists in the field. For details (which can cover hundreds of pages) see the Additional Readings.

Exercises

4.7.1 Show that the matrices 1, A, B, and C of Eq. (4.165) are reducible. Reduce them.
Note. This means transforming A and C to diagonal form (by the same unitary transformation).
Hint. A and C are anti-Hermitian. Their eigenvectors will be orthogonal.

4.7.2 Possible operations on a crystal lattice include A_π (rotation by π), m (reflection), and i (inversion). These three operations combine as

$$A_\pi^2 = m^2 = i^2 = 1,$$
$$A_\pi \cdot m = i, \qquad m \cdot i = A_\pi, \qquad \text{and} \qquad i \cdot A_\pi = m.$$

Show that the group $(1, A_\pi, m, i)$ is isomorphic with the vierergruppe.

4.7.3 Four possible operations in the xy-plane are:

1. no change $\begin{cases} x \to x \\ y \to y \end{cases}$
2. inversion $\begin{cases} x \to -x \\ y \to -y \end{cases}$
3. reflection $\begin{cases} x \to -x \\ y \to y \end{cases}$
4. reflection $\begin{cases} x \to x \\ y \to -y. \end{cases}$

(a) Show that these four operations form a group.
(b) Show that this group is isomorphic with the vierergruppe.
(c) Set up a 2×2 matrix representation.

4.7.4 Rearrangement theorem: Given a group of n distinct elements $(I, a, b, c, \ldots, n)$, show that the set of products $(aI, a^2, ab, ac \ldots an)$ reproduces the n distinct elements in a new order.

4.7.5 Using the 2×2 matrix representation of Exercise 3.2.7 for the vierergruppe,

(a) Show that there are four classes, each with one element.

(b) Calculate the character (trace) of each class. Note that two different classes may have the same character.
(c) Show that there are three two-element subgroups. (The unit element by itself always forms a subgroup.)
(d) For any one of the two-element subgroups show that the subgroup and a single coset reproduce the original vierergruppe.

Note that subgroups, classes, and cosets are entirely different.

4.7.6 Using the 2×2 matrix representation, Eq. (4.165), of C_4,

(a) Show that there are four classes, each with one element.
(b) Calculate the character (trace) of each class.
(c) Show that there is one two-element subgroup.
(d) Show that the subgroup and a single coset reproduce the original group.

4.7.7 Prove that the number of distinct elements in a coset of a subgroup is the same as the number of elements in the subgroup.

4.7.8 A subgroup H has elements h_i. Let x be a fixed element of the original group G and **not** a member of H. The transform

$$xh_ix^{-1}, \qquad i = 1, 2, \ldots$$

generates a **conjugate subgroup** xHx^{-1}. Show that this conjugate subgroup satisfies each of the four group postulates and therefore is a group.

4.7.9 (a) A particular group is abelian. A second group is created by replacing g_i by g_i^{-1} for each element in the original group. Show that the two groups are isomorphic. *Note*. This means showing that if $a_ib_i = c_i$, then $a_i^{-1}b_i^{-1} = c_i^{-1}$.
(b) Continuing part (a), if the two groups are isomorphic, show that each must be abelian.

4.7.10 (a) Once you have a matrix representation of any group, a one-dimensional representation can be obtained by taking the determinants of the matrices. Show that the multiplicative relations are preserved in this determinant representation.
(b) Use determinants to obtain a one-dimensional representative of D_3.

4.7.11 Explain how the relation

$$\sum_i n_i^2 = h$$

applies to the vierergruppe ($h = 4$) and to the dihedral group D_3 with $h = 6$.

4.7.12 Show that the subgroup $(1, \mathsf{A}, \mathsf{B})$ of D_3 is an invariant subgroup.

4.7.13 The group D_3 may be discussed as a **permutation** group of three objects. Matrix B, for instance, rotates vertex a (originally in location 1) to the position formerly occupied by c

(location 3). Vertex b moves from location 2 to location 1, and so on. As a permutation $(abc) \to (bca)$. In three dimensions

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} b \\ c \\ a \end{pmatrix}.$$

(a) Develop analogous 3×3 representations for the other elements of D_3.
(b) Reduce your 3×3 representation to the 2×2 representation of this section.

(This 3×3 representation must be reducible or Eq. (4.174) would be violated.)
Note. The actual reduction of a reducible representation may be awkward. It is often easier to develop directly a new representation of the required dimension.

4.7.14 (a) The permutation group of four objects P_4 has $4! = 24$ elements. Treating the four elements of the cyclic group C_4 as permutations, set up a 4×4 matrix representation of C_4. C_4 that becomes a subgroup of P_4.
(b) How do you know that this 4×4 matrix representation of C_4 **must** be reducible?
Note. C_4 is abelian and every abelian group of h objects has only h one-dimensional irreducible representations.

4.7.15 (a) The objects $(abcd)$ are permuted to $(dacb)$. Write out a 4×4 matrix representation of this one permutation.
(b) Is the permutation $(abdc) \to (dacb)$ odd or even?
(c) Is this permutation a possible member of the D_4 group? Why or why not?

4.7.16 The elements of the dihedral group D_n may be written in the form

$$\mathsf{S}^{\lambda}\mathsf{R}_z^{\mu}(2\pi/n), \qquad \lambda = 0, 1$$
$$\mu = 0, 1, \ldots, n-1,$$

where $\mathsf{R}_z(2\pi/n)$ represents a rotation of $2\pi/n$ about the n-fold symmetry axis, whereas S represents a rotation of π about an axis through the center of the regular polygon and one of its vertices.
For $\mathsf{S} = \mathsf{E}$ show that this form may describe the matrices A, B, C, and D of D_3.
Note. The elements R_z and S are called the generators of this finite group. Similarly, i is the generator of the group given by Eq. (4.164).

4.7.17 Show that the cyclic group of n objects, C_n, may be represented by $r^m, m = 0, 1, 2, \ldots, n-1$. Here r is a generator given by

$$r = \exp(2\pi i s/n).$$

The parameter s takes on the values $s = 1, 2, 3, \ldots, n$, each value of s yielding a different one-dimensional (irreducible) representation of C_n.

4.7.18 Develop the irreducible 2×2 matrix representation of the group of operations (rotations and reflections) that transform a square into itself. Give the group multiplication table.
Note. This is the symmetry group of a square and also the dihedral group D_4. (See Fig. 4.16.)

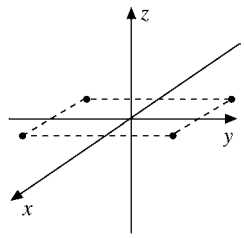

FIGURE 4.16
Square.

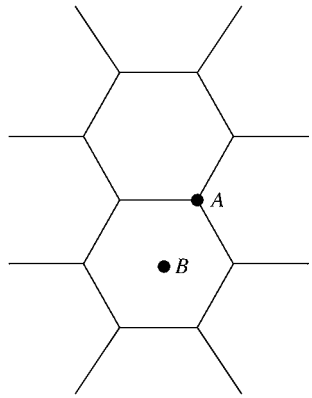

FIGURE 4.17 Hexagon.

4.7.19 The permutation group of four objects contains $4! = 24$ elements. From Exercise 4.7.18, D_4, the symmetry group for a square, has far fewer than 24 elements. Explain the relation between D_4 and the permutation group of four objects.

4.7.20 A plane is covered with regular hexagons, as shown in Fig. 4.17.

(a) Determine the dihedral symmetry of an axis perpendicular to the plane through the common vertex of three hexagons (A). That is, if the axis has n-fold symmetry, show (with careful explanation) what n is. Write out the 2×2 matrix describing the minimum (nonzero) positive rotation of the array of hexagons that is a member of your D_n group.
(b) Repeat part (a) for an axis perpendicular to the plane through the geometric center of one hexagon (B).

4.7.21 In a simple cubic crystal, we might have identical atoms at $r = (la, ma, na)$, with l, m, and n taking on all integral values.

(a) Show that each Cartesian axis is a fourfold symmetry axis.
(b) The cubic group will consist of all operations (rotations, reflections, inversion) that leave the simple cubic crystal invariant. From a consideration of the permutation

y \ x	1	A − −
1	1	1 − −
A	A	A − −

FIGURE 4.18
Multiplication table.

of the positive and negative coordinate axes, predict how many elements this cubic group will contain.

4.7.22 (a) From the D_3 multiplication table of Fig. 4.18 construct a similarity transform table showing xyx^{-1}, where x and y each range over all six elements of D_3:

(b) Divide the elements of D_3 into classes. Using the 2×2 matrix representation of Eqs. (4.169)–(4.172) note the trace (character) of each class.

4.8 DIFFERENTIAL FORMS

In Chapters 1 and 2 we adopted the view that, in n dimensions, a vector is an n-tuple of real numbers and that its components transform properly under changes of the coordinates. In this section we start from the alternative view, in which a vector is thought of as a directed line segment, an arrow. The point of the idea is this: Although the concept of a vector as a line segment does not generalize to curved space–time (manifolds of differential geometry), except by working in the flat tangent space requiring embedding in auxiliary extra dimensions, Elie Cartan's differential forms are natural in curved space–time and a very powerful tool. Calculus can be based on differential forms, as Edwards has shown by his classic textbook (see the Additional Readings). Cartan's calculus leads to a remarkable unification of concepts and theorems of vector analysis that is worth pursuing. In differential geometry and advanced analysis (on manifolds) the use of differential forms is now widespread.

Cartan's notion of vector is based on the one-to-one correspondence between the linear spaces of displacement vectors and directional differential operators (components of the gradient form a basis). A crucial advantage of the latter is that they can be generalized to curved space–time. Moreover, describing vectors in terms of directional derivatives along curves uniquely specifies the vector at a given point without the need to invoke coordinates. Ultimately, since coordinates are needed to specify points, the Cartan formalism, though an elegant mathematical tool for the efficient derivation of theorems on tensor analysis, has in principle no advantage over the component formalism.

1-Forms

We define dx, dy, dz in three-dimensional Euclidean space as functions assigning to a directed line segment PQ from the point P to the point Q the corresponding change in x, y, z. The symbol dx represents "oriented length of the projection of a curve on the

x-axis," etc. Note that dx, dy, dz can be, but need not be, infinitesimally small, and **they must not be confused with the ordinary differentials** that we associate with integrals and differential quotients. A function of the type

$$A\,dx + B\,dy + C\,dz, \qquad A, B, C \text{ real numbers} \tag{4.175}$$

is defined as a **constant 1-form**.

Example 4.8.1 CONSTANT 1-FORM

For a constant force $\mathbf{F} = (A, B, C)$, the work done along the displacement from $P = (3, 2, 1)$ to $Q = (4, 5, 6)$ is therefore given by

$$W = A(4-3) + B(5-2) + C(6-1) = A + 3B + 5C.$$

If $\mathbf{F}$ is a force field, then its rectangular components $A(x, y, z), B(x, y, z), C(x, y, z)$ will depend on the location and the (nonconstant) 1-**form** $dW = \mathbf{F} \cdot d\mathbf{r}$ corresponds to the concept of work done against the force field $\mathbf{F}(\mathbf{r})$ along $d\mathbf{r}$ on a space curve. A finite amount of work

$$W = \int_C \big[A(x, y, z)\,dx + B(x, y, z)\,dy + C(x, y, z)\,dz\big] \tag{4.176}$$

involves the familiar line integral along an oriented curve C, where the 1-form dW describes the amount of work for small displacements (segments on the path C). In this light, the integrand $f(x)\,dx$ of an integral $\int_a^b f(x)\,dx$ consisting of the function f and of the measure dx as the oriented length is here considered to be a 1-form. The value of the integral is obtained from the ordinary line integral. ■

2-Forms

Consider a unit flow of mass in the z-direction, that is, a flow in the direction of increasing z so that a unit mass crosses a unit square of the xy-plane in unit time. The orientation symbolized by the sequence of points in Fig. 4.19,

$$(0, 0, 0) \to (1, 0, 0) \to (1, 1, 0) \to (0, 1, 0) \to (0, 0, 0),$$

will be called **counterclockwise**, as usual. A unit flow in the z-direction is defined by the function $dx\,dy$[31] assigning to oriented rectangles in space the oriented area of their projections on the xy-plane. Similarly, a unit flow in the x-direction is described by $dy\,dz$ and a unit flow in the y-direction by $dz\,dx$. The reverse order, $dz\,dx$, is dictated by the orientation convention, and $dz\,dx = -dx\,dz$ by definition. This antisymmetry is consistent with the cross product of two vectors representing oriented areas in Euclidean space. This notion generalizes to polygons and curved differentiable surfaces approximated by polygons and volumes.

[31] Many authors denote this wedge product as $dx \wedge dy$ with $dy \wedge dx = -dx \wedge dy$. Note that the product $dx\,dy = dy\,dx$ for ordinary differentials.

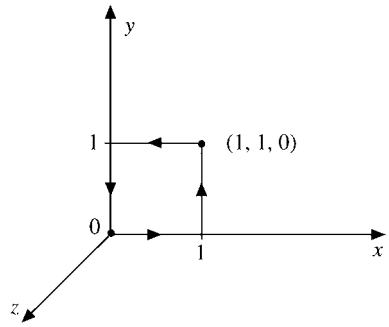

FIGURE 4.19
Counterclockwise-oriented rectangle.

Example 4.8.2 MAGNETIC FLUX ACROSS AN ORIENTED SURFACE

If $\mathbf{B} = (A, B, C)$ is a constant magnetic induction, then the constant **2-form**

$$A\,dy\,dz + B\,dz\,dx + C\,dx\,dy$$

describes the magnetic flux across an oriented rectangle. If **B** is a magnetic induction field varying across a surface S, then the flux

$$\Phi = \int_S \left[B_x(\mathbf{r})\,dy\,dz + B_y(\mathbf{r})\,dz\,dx + B_z(\mathbf{r})\,dx\,dy\right] \tag{4.177}$$

across the oriented surface S involves the familiar (Riemann) integration over approximating small oriented rectangles from which S is pieced together. ■

The definition of $\int \omega$ relies on decomposing $\omega = \sum_i \omega_i$, where the differential forms ω_i are each nonzero only in a small patch of the surface S that covers the surface. Then it can be shown that $\sum_i \int \omega_i$ converges, as the patches become smaller and more numerous, to the limit $\int \omega$, which is independent of these decompositions. For more details and proofs, we refer the reader to Edwards in the Additional Readings.

3-Forms

A 3-form $dx\,dy\,dz$ represents an oriented volume. For example, the determinant of three vectors in Euclidean space changes sign if we reverse the order of two vectors. The determinant measures the oriented volume spanned by the three vectors. In particular, $\int_V \rho(x, y, z)\,dx\,dy\,dz$ represents the total charge inside the volume V if ρ is the charge density. Higher-dimensional differential forms in higher-dimensional spaces are defined similarly and are called k-forms, with $k = 0, 1, 2, \ldots$.

If a 3-form

$$\omega = A(x_1, x_2, x_3)\,dx_1\,dx_2\,dx_3 = A'(x_1', x_2', x_3')\,dx_1'\,dx_2'\,dx_3' \tag{4.178}$$

on a 3-dimensional manifold is expressed in terms of new coordinates, then there is a one-to-one, differentiable map $x_i' = x_i'(x_1, x_2, x_3)$ between these coordinates with Jacobian

$$J = \frac{\partial(x_1', x_2', x_3')}{\partial(x_1, x_2, x_3)} = 1,$$

and $A = A'J = A'$ so that

$$\int_V \omega = \int_V A\, dx_1\, dx_2\, dx_3 = \int_{V'} A'\, dx_1'\, dx_2'\, dx_3'. \tag{4.179}$$

This statement spells out the parameter independence of integrals over differential forms, since parameterizations are essentially arbitrary. The rules governing integration of differential forms are defined on manifolds. These are continuous if we can move continuously (actually we assume them differentiable) from point to point, oriented if the orientation of curves generalizes to surfaces and volumes up to the dimension of the whole manifold. The rules on differential forms are:

- If $\omega = a\omega_1 + a'\omega_1'$, with a, a' real numbers, then $\int_S \omega = a\int_S \omega_1 + a'\int_S \omega_1'$, where S is a compact, oriented, continuous manifold with boundary.
- If the orientation is reversed, then the integral $\int_S \omega$ changes sign.

Exterior Derivative

We now introduce the **exterior derivative** d of a function f, a 0-form:

$$df \equiv \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz = \frac{\partial f}{\partial x_i}dx_i, \tag{4.180}$$

generating a 1-form $\omega_1 = df$, the differential of f (or exterior derivative), the gradient in standard vector analysis. Upon summing over the coordinates, we have used and will continue to use Einstein's summation convention. Applying the exterior derivative d to a 1-form we define

$$d(A\,dx + B\,dy + C\,dz) = dA\,dx + dB\,dy + dC\,dz \tag{4.181}$$

with functions A, B, C. This definition in conjunction with df as just given ties vectors to differential operators $\partial_i = \frac{\partial}{\partial x_i}$. Similarly, we extend d to k-forms. However, applying d twice gives zero, $ddf = 0$, because

$$\begin{aligned} d(df) &= d\frac{\partial f}{\partial x}dx + d\frac{\partial f}{\partial y}dy \\ &= \left(\frac{\partial^2 f}{\partial x^2}dx + \frac{\partial^2 f}{\partial x\,\partial y}dy\right)dx + \left(\frac{\partial^2 f}{\partial y \partial x}dx + \frac{\partial^2 f}{\partial y^2}dy\right)dy \\ &= \left(\frac{\partial^2 f}{\partial y\,\partial x} - \frac{\partial^2 f}{\partial x\,\partial y}\right)dx\,dy = 0. \end{aligned} \tag{4.182}$$

This follows from the fact that in mixed partial derivatives their order does not matter provided all functions are sufficiently differentiable. Similarly we can show $dd\omega_1 = 0$ for a 1-form ω_1, etc.

The rules governing differential forms, with ω_k denoting a k-form, that we have used so far are

- $dx\,dx = 0 = dy\,dy = dz\,dz$, $dx_i^2 = 0$;
- $dx\,dy = -dy\,dx$, $dx_i\,dx_j = -dx_j\,dx_i$, $i \neq j$,
- $dx_1\,dx_2 \cdots dx_k$ is totally antisymmetric in the dx_i, $i = 1, 2, \ldots, k$.
- $df = \frac{\partial f}{\partial x_i} dx_i$;
- $d(\omega_k + \Omega_k) = d\omega_k + d\Omega_k$, linearity;
- $dd\omega_k = 0$.

Now we apply the exterior derivative d to products of differential forms, starting with functions (0-forms). We have

$$d(fg) = \frac{\partial(fg)}{\partial x_i} dx_i = \left(f \frac{\partial g}{\partial x_i} + \frac{\partial f}{\partial x_i} g \right) dx_i = f\,dg + df g. \tag{4.183}$$

If $\omega_1 = \frac{\partial g}{\partial x_i} dx_i$ is a 1-form and f is a function, then

$$\begin{aligned} d(f\omega_1) &= d\left(f \frac{\partial g}{\partial x_i} dx_i \right) = d\left(f \frac{\partial g}{\partial x_i} \right) dx_i \\ &= \frac{\partial \left(f \frac{\partial g}{\partial x_i} \right)}{\partial x_j} dx_j\,dx_i = \left(\frac{\partial f}{\partial x_j} \frac{\partial g}{\partial x_i} + f \frac{\partial^2 g}{\partial x_i\,\partial x_j} \right) dx_j\,dx_i \\ &= df\,\omega_1 + f\,d\omega_1, \end{aligned} \tag{4.184}$$

as expected. But if $\omega_1' = \frac{\partial f}{\partial x_j} dx_j$ is another 1-form, then

$$\begin{aligned} d(\omega_1\omega_1') &= d\left(\frac{\partial g}{\partial x_i} dx_i \frac{\partial f}{\partial x_j} dx_j \right) = d\left(\frac{\partial g}{\partial x_i} \frac{\partial f}{\partial x_j} \right) dx_i\,dx_j \\ &= \frac{\partial \left(\frac{\partial g}{\partial x_i} \frac{\partial f}{\partial x_j} \right)}{\partial x_k} dx_k\,dx_i\,dx_j \\ &= \frac{\partial^2 g}{\partial x_i \partial x_k} dx_k\,dx_i \frac{\partial f}{\partial x_j} dx_j - \frac{\partial g}{\partial x_i} dx_i \frac{\partial^2 f}{\partial x_j\,\partial x_k} dx_k\,dx_j \\ &= d\omega_1\,\omega_1' - \omega_1\,d\omega_1'. \end{aligned} \tag{4.185}$$

This proof is valid for more general 1-forms $\omega = f_i\,dx_i$ with functions f_i. In general, therefore, we define for k-forms:

$$d(\omega_k\omega_k') = (d\omega_k)\omega_k' + (-1)^k \omega_k (d\omega_k'). \tag{4.186}$$

In general, the exterior derivative of a k-form is a $(k+1)$-form.

Example 4.8.3 POTENTIAL ENERGY

As an application in two dimensions (for simplicity), consider the potential $V(\mathbf{r})$, a 0-form, and dV, its exterior derivative. Integrating V along an oriented path C from $\mathbf{r}_1$ to $\mathbf{r}_2$ gives

$$V(\mathbf{r}_2) - V(\mathbf{r}_1) = \int_C dV = \int_C \left(\frac{\partial V}{\partial x} dx + \frac{\partial V}{\partial y} dy \right) = \int_C \nabla V \cdot d\mathbf{r}, \tag{4.187}$$

where the last integral is the standard formula for the potential energy difference that forms part of the energy conservation theorem. The path and parameterization independence are manifest in this special case. ■

Pullbacks

If a linear map $\mathcal{L}_2$ from the uv-plane to the xy-plane has the form

$$x = au + bv + c, \qquad y = eu + fv + g, \tag{4.188}$$

oriented polygons in the uv-plane are mapped onto similar polygons in the xy-plane, provided the determinant $af - be$ of the map $\mathcal{L}_2$ is nonzero. The 2-form

$$dx\,dy = (a\,du + b\,dv)(e\,du + f\,dv) = (af - be)du\,dv \tag{4.189}$$

can be pulled back from the xy- to the uv-plane. That is to say, an integral over a simply connected surface S becomes

$$\int_{\mathcal{L}_2(S)} dx\,dy = (af - be) \int_S du\,dv, \tag{4.190}$$

and $(af - be)\,du\,dv$ is the pullback of $dx\,dy$, opposite to the direction of the map $\mathcal{L}_2$ from the uv-plane to the xy-plane. Of course, the determinant $af - be$ of the map $\mathcal{L}_2$ is simply the Jacobian, generated without effort by the differential forms in Eq. (4.189).

Similarly, a linear map $\mathcal{L}_3$ from the $u_1u_2u_3$-space to the $x_1x_2x_3$-space

$$x_i = a_{ij}u_j + b_i, \qquad i = 1, 2, 3, \tag{4.191}$$

automatically generates its Jacobian from the 3-form

$$\begin{aligned} dx_1\,dx_2\,dx_3 &= \left(\sum_{j=1}^{3} a_{1j}\,du_j \right) \left(\sum_{j=1}^{3} a_{2j}\,du_j \right) \left(\sum_{j=1}^{3} a_{3j}\,du_j \right) \\ &= (a_{11}a_{22}a_{33} - a_{12}a_{21}a_{33} \pm \cdots) du_1\,du_2\,du_3 \\ &= \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} du_1\,du_2\,du_3. \end{aligned} \tag{4.192}$$

Thus, differential forms generate the rules governing determinants.

Given two linear maps in a row, it is straightforward to prove that the pullback under a composed map is the pullback of the pullback. This theorem is the differential-forms analog of matrix multiplication.

Let us now consider a curve C defined by a parameter t in contrast to a curve defined by an equation. For example, the circle $\{(\cos t, \sin t); 0 \le t \le 2\pi\}$ is a parameterization by t, whereas the circle $\{(x, y); x^2 + y^2 = 1\}$ is a definition by an equation. Then the line integral

$$\int_C \big[A(x, y)dx + B(x, y)\, dy\big] = \int_{t_i}^{t_f} \left[A\frac{dx}{dt} + B\frac{dy}{dt}\right] dt \tag{4.193}$$

for continuous functions $A, B, dx/dt, dy/dt$ becomes a one-dimensional integral over the oriented interval $t_i \le t \le t_f$. Clearly, the 1-form $[A\frac{dx}{dt} + B\frac{dy}{dt}]\,dt$ on the t-line is obtained from the 1-form $A\,dx + B\,dy$ on the xy-plane via the map $x = x(t), y = y(t)$ from the t-line to the curve C in the xy-plane. The 1-form $[A\frac{dx}{dt} + B\frac{dy}{dt}]\,dt$ is called the pullback of the 1-form $A\,dx + B\,dy$ under the map $x = x(t)$, $y = y(t)$. Using pullbacks we can show that integrals over 1-forms are independent of the parameterization of the path.

In this sense, the differential quotient $\frac{dy}{dx}$ can be considered as the coefficient of dx in the pullback of dy under the function $y = f(x)$, or $dy = f'(x)\,dx$. This concept of pullback readily generalizes to maps in three or more dimensions and to k-forms with $k > 1$. In particular, the chain rule can be seen to be a pullback: If

$$\begin{aligned} y_i &= f_i(x_1, x_2, \ldots, x_n), \qquad i = 1, 2, \ldots, l \quad \text{and} \\ z_j &= g_j(y_1, y_2, \ldots, y_l), \qquad j = 1, 2, \ldots, m \end{aligned} \tag{4.194}$$

are differentiable maps from $\mathbb{R}^n \to \mathbb{R}^l$ and $\mathbb{R}^l \to \mathbb{R}^m$, then the composed map $\mathbb{R}^n \to \mathbb{R}^m$ is differentiable and the pullback of any k-form under the composed map is equal to the pullback of the pullback. This theorem is useful for establishing that integrals of k-forms are parameter independent.

Similarly, we define the differential df as the pullback of the 1-form dz under the function $z = f(x, y)$:

$$dz = df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy. \tag{4.195}$$

Example 4.8.4 STOKES' THEOREM

As another application let us first sketch the standard derivation of the simplest version of Stokes' theorem for a rectangle $S = [a \le x \le b, c \le y \le d]$ oriented counterclockwise, with ∂S its boundary

$$\begin{aligned} \int_{\partial S} (A\,dx + B\,dy) &= \int_a^b A(x, c)\,dx + \int_c^d B(b, y)\,dy + \int_b^a A(x, d)\,dx + \int_d^c B(a, y)\,dy \\ &= \int_c^d \big[B(b, y) - B(a, y)\big]\,dy - \int_a^b \big[A(x, d) - A(x, c)\big]\,dx \\ &= \int_c^d \int_a^b \frac{\partial B}{\partial x}dx\,dy - \int_a^b \int_c^d \frac{\partial A}{\partial y}dy\,dx \\ &= \int_S \left(\frac{\partial B}{\partial x} - \frac{\partial A}{\partial y}\right) dx\,dy, \end{aligned} \tag{4.196}$$

which holds for any simply connected surface S that can be pieced together by rectangles.

Now we demonstrate the use of differential forms to obtain the same theorem (again in two dimensions for simplicity):

$$\begin{aligned} d(A\,dx + B\,dy) &= dA\,dx + dB\,dy \\ &= \left(\frac{\partial A}{\partial x}dx + \frac{\partial A}{\partial y}dy\right)dx + \left(\frac{\partial B}{\partial x}dx + \frac{\partial B}{\partial y}dy\right)dy = \left(\frac{\partial B}{\partial x} - \frac{\partial A}{\partial y}\right)dx\,dy, \end{aligned} \tag{4.197}$$

using the rules highlighted earlier. Integrating over a surface S and its boundary ∂S, respectively, we obtain

$$\int_{\partial S}(A\,dx + B\,dy) = \int_S d(A\,dx + B\,dy) = \int_S \left(\frac{\partial B}{\partial x} - \frac{\partial A}{\partial y}\right)dx\,dy. \tag{4.198}$$

Here contributions to the left-hand integral from inner boundaries cancel as usual because they are oriented in opposite directions on adjacent rectangles. For each oriented inner rectangle that makes up the simply connected surface S we have used,

$$\int_R dd x = \int_{\partial R} dx = 0. \tag{4.199}$$

Note that the exterior derivative automatically generates the z component of the curl.

In three dimensions, Stokes' theorem derives from the differential-form identity involving the vector potential $\mathbf{A}$ and magnetic induction $\mathbf{B} = \nabla \times \mathbf{A}$,

$$\begin{aligned} d(A_x\,dx + A_y\,dy + A_z\,dz) &= dA_x\,dx + dA_y\,dy + dA_z\,dz \\ &= \left(\frac{\partial A_x}{\partial x}dx + \frac{\partial A_x}{\partial y}dy + \frac{\partial A_x}{\partial z}dz\right)dx + \cdots \\ &= \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}\right)dy\,dz + \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}\right)dz\,dx + \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right)dx\,dy, \end{aligned} \tag{4.200}$$

generating all components of the curl in three-dimensional space. This identity is integrated over each oriented rectangle that makes up the simply connected surface S (which has no holes, that is, where every curve contracts to a point of the surface) and then is summed over all adjacent rectangles to yield the magnetic flux across S,

$$\begin{aligned} \Phi &= \int_S [B_x\,dy\,dz + B_y\,dz\,dx + B_z\,dx\,dy] \\ &= \int_{\partial S} [A_x\,dx + A_y\,dy + A_z\,dz], \end{aligned} \tag{4.201}$$

or, in the standard notation of vector analysis (Stokes' theorem, Chapter 1),

$$\int_S \mathbf{B}\cdot d\mathbf{a} = \int_S (\nabla \times \mathbf{A})\cdot d\mathbf{a} = \int_{\partial S} \mathbf{A}\cdot d\mathbf{r}. \tag{4.202}$$

■

Example 4.8.5 GAUSS' THEOREM

Consider Gauss' law, Section 1.14. We integrate the electric density $\rho = \frac{1}{\varepsilon_0}\nabla \cdot \mathbf{E}$ over the volume of a single parallelepiped $V = [a \le x \le b, c \le y \le d, e \le z \le f]$ oriented by $dx\,dy\,dz$ (right-handed), the side $x = b$ of V is oriented by $dy\,dz$ (counterclockwise, as seen from $x > b$), and so on. Using

$$E_x(b, y, z) - E_x(a, y, z) = \int_a^b \frac{\partial E_x}{\partial x} dx, \tag{4.203}$$

we have, in the notation of differential forms, summing over all adjacent parallelepipeds that make up the volume $\mathcal{V}$,

$$\int_{\partial \mathcal{V}} E_x\, dy\, dz = \int_{\mathcal{V}} \frac{\partial E_x}{\partial x} dx\, dy\, dz. \tag{4.204}$$

Integrating the electric flux (2-form) identity

$$\begin{aligned} d(E_x\, dy\, dz + E_y\, dz\, dx + E_z\, dx\, dy) &= dE_x\, dy\, dz + dE_y\, dz\, dx + dE_z\, dx\, dy \\ &= \left(\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}\right) dx\, dy\, dz \end{aligned} \tag{4.205}$$

across the simply connected surface $\partial \mathcal{V}$ we have Gauss' theorem,

$$\int_{\partial \mathcal{V}} (E_x\, dy\, dz + E_y\, dz\, dx + E_z\, dx\, dy) = \int_{\mathcal{V}} \left(\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}\right) dx\, dy\, dz, \tag{4.206}$$

or, in standard notation of vector analysis,

$$\int_{\partial \mathcal{V}} \mathbf{E} \cdot d\mathbf{a} = \int_{\mathcal{V}} \nabla \cdot \mathbf{E}\, d^3r = \frac{q}{\varepsilon_0}. \tag{4.207}$$

■

These examples are different cases of a single theorem on differential forms. To explain why, let us begin with some terminology, a preliminary **definition of a differentiable manifold M**: It is a collection of points (m-tuples of real numbers) that are smoothly (that is, differentiably) connected with each other so that the neighborhood of each point looks like a simply connected piece of an m-dimensional Cartesian space "close enough" around the point and containing it. Here, m, which stays constant from point to point, is called the dimension of the manifold. Examples are the m-dimensional Euclidean space $\mathbb{R}^m$ and the m-dimensional sphere

$$S_m = \left[\left(x^1, \ldots, x^{m+1}\right); \qquad \sum_{i=1}^{m+1} \left(x^i\right)^2 = 1 \right].$$

Any surface with sharp edges, corners, or kinks is not a manifold in our sense, that is, is not differentiable. In differential geometry, all movements, such as translation and parallel displacement, are local, that is, are defined infinitesimally. If we apply the exterior derivative d to a function $f(x^1, \ldots, x^m)$ on **M**, we generate basic 1-forms:

$$df = \frac{\partial f}{\partial x_i} dx^i, \tag{4.208}$$

where $x^i(P)$ are coordinate functions. As before we have $d(df) = 0$ because

$$d(df) = d\left(\frac{\partial f}{\partial x^i}\right) dx^i = \frac{\partial^2 f}{\partial x^j\, \partial x^i} dx^j\, dx^i$$
$$= \sum_{j<i} \left(\frac{\partial^2 f}{\partial x^j\, \partial x^i} - \frac{\partial^2 f}{\partial x^i\, \partial x^j}\right) dx^j\, dx^i = 0 \tag{4.209}$$

because the order of derivatives does not matter. Any 1-form is a linear combination $\omega = \sum_i \omega_i\, dx^i$ with functions ω_i.

Generalized Stokes' Theorem on Differential Forms

Let ω be a continuous $(k-1)$-form in $x_1x_2\cdots x_n$-space defined everywhere on a compact, oriented, differentiable k-dimensional manifold S with boundary ∂S in $x_1x_2\cdots x_n$-space. Then

$$\int_{\partial S} \omega = \int_S d\omega. \tag{4.210}$$

Here

$$d\omega = d(A\, dx_1\, dx_2 \cdots dx_{k-1} + \cdots) = dA\, dx_1\, dx_2 \cdots dx_{k-1} + \cdots. \tag{4.211}$$

The potential energy in Example 4.8.3 given this theorem for the potential $\omega = V$, a 0-form; Stokes' theorem in Example 4.8.4 is this theorem for the vector potential 1-form $\sum_i A_i\, dx_i$ (for Euclidean spaces $dx^i = dx_i$); and Gauss' theorem in Example 4.8.5 is Stokes' theorem for the electric flux 2-form in three-dimensional Euclidean space.

The method of integration by parts can be generalized to differential forms using Eq. (4.186):

$$\int_S d\omega_1\, \omega_2 = \int_{\partial S} \omega_1\omega_2 - (-1)^{k_1} \int_S \omega_1\, d\omega_2. \tag{4.212}$$

This is proved by integrating the identity

$$d(\omega_1\omega_2) = d\omega_1\, \omega_2 + (-1)^{k_1}\omega_1\, d\omega_2, \tag{4.213}$$

with the integrated term $\int_S d(\omega_1\omega_2) = \int_{\partial S} \omega_1\omega_2$.

Our next goal is to cast Sections 2.10 and 2.11 in the language of differential forms. So far we have worked in two- or three-dimensional Euclidean space.

Example 4.8.6 RIEMANN MANIFOLD

Let us look at the curved Riemann space–time of Sections 2.10–2.11 and reformulate some of this tensor analysis in curved spaces in the language of differential forms. Recall that dishinguishing between upper and lower indices is important here. The metric g_{ij} in Eq. (2.123) can be written in terms of tangent vectors, Eq. (2.114), as follows:

$$g_{ij} = \frac{\partial x^l}{\partial q^i}\frac{\partial x_l}{\partial q^j}, \tag{4.214}$$

where the sum over the index l denotes the inner product of the tangent vectors. (Here we continue to use Einstein's summation convention over repeated indices. As before, the metric tensor is used to raise and lower indices.) The key concept of connection involves the Christoffel symbols, which we address first. The exterior derivative of a tangent vector can be expanded in terms of the basis of tangent vectors (compare Eq. (2.131a)),

$$d\left(\frac{\partial x^l}{\partial q^i}\right) = \Gamma^k{}_{ij}\frac{\partial x^l}{\partial q^k}dq^j, \tag{4.215}$$

thus introducing the Christoffel symbols of the second kind. Applying d to Eq. (4.214) we obtain

$$\begin{aligned} dg_{ij} &= \frac{\partial g_{ij}}{\partial q^m}dq^m = d\left(\frac{\partial x^l}{\partial q^i}\right)\frac{\partial x_l}{\partial q^j} + \frac{\partial x^l}{\partial q^i}d\left(\frac{\partial x_l}{\partial q^j}\right) \\ &= \left(\Gamma^k{}_{im}\frac{\partial x_l}{\partial q^k}\frac{\partial x^l}{\partial q^j} + \Gamma^k{}_{jm}\frac{\partial x^l}{\partial q^i}\frac{\partial x_l}{\partial q^k}\right)dq^m = \left(\Gamma^k{}_{im}g_{kj} + \Gamma^k{}_{jm}g_{ik}\right)dq^m. \end{aligned} \tag{4.216}$$

Comparing the coefficients of dq^m yields

$$\frac{\partial g_{ij}}{\partial q^m} = \Gamma^k{}_{im}g_{kj} + \Gamma^k{}_{jm}g_{ik}. \tag{4.217}$$

Using the Christoffel symbol of the first kind,

$$[ij, m] = g_{km}\Gamma^k{}_{ij}, \tag{4.218}$$

we can rewrite Eq. (4.217) as

$$\frac{\partial g_{ij}}{\partial q^m} = [im, j] + [jm, i], \tag{4.219}$$

which corresponds to Eq. (2.136) and implies Eq. (2.137). We check that

$$[ij, m] = \frac{1}{2}\left(\frac{\partial g_{im}}{\partial q^j} + \frac{\partial g_{jm}}{\partial q^i} - \frac{\partial g_{ij}}{\partial q^m}\right) \tag{4.220}$$

is the unique solution of Eq. (4.219) and that

$$\Gamma^k{}_{ij} = g^{mk}[ij, m] = \frac{1}{2}g^{mk}\left(\frac{\partial g_{im}}{\partial q^j} + \frac{\partial g_{jm}}{\partial q^i} - \frac{\partial g_{ij}}{\partial q^m}\right) \tag{4.221}$$

follows. ■

Hodge * Operator

The differentials dx^i, $i = 1, 2, \ldots, m$, form a basis of a vector space that is (seen to be) dual to the derivatives $\partial_i = \frac{\partial}{\partial x^i}$; they are basic 1-forms. For example, the vector space $\mathcal{V} = \{(a_1, a_2, a_3)\}$ is dual to the vector space of planes (linear functions f) in three-dimensional Euclidean space $\mathcal{V}^* = \{f \equiv a_1x_1 + a_2x_2 + a_3x_3 - d = 0\}$. The gradient

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3}\right) = (a_1, a_2, a_3) \tag{4.222}$$

provides a one-to-one, differentiable map from $\mathcal{V}^*$ to $\mathcal{V}$. Such dual relationships are generalized by the Hodge * operator, based on the Levi-Civita symbol of Section 2.9.

Let the unit vectors $\hat{\mathbf{x}}_i$ be an oriented orthonormal basis of three-dimensional Euclidean space. Then the Hodge * of scalars is defined by the basis element

$$*1 \equiv \frac{1}{3!}\varepsilon^{ijk}\hat{\mathbf{x}}_i\hat{\mathbf{x}}_j\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_1\hat{\mathbf{x}}_2\hat{\mathbf{x}}_3, \tag{4.223}$$

which corresponds to $(\hat{\mathbf{x}}_1 \times \hat{\mathbf{x}}_2) \cdot \hat{\mathbf{x}}_3$ in standard vector notation. Here $\hat{\mathbf{x}}_i\hat{\mathbf{x}}_j\hat{\mathbf{x}}_k$ is the totally antisymmetric exterior product of the unit vectors that corresponds to $(\hat{\mathbf{x}}_i \times \hat{\mathbf{x}}_j) \cdot \hat{\mathbf{x}}_k$ in standard vector notation. For vectors, * is defined for the basis of unit vectors as

$$*\hat{\mathbf{x}}_i \equiv \frac{1}{2!}\varepsilon_i{}^{jk}\hat{\mathbf{x}}_j\hat{\mathbf{x}}_k. \tag{4.224}$$

In particular,

$$*\hat{\mathbf{x}}_1 = \hat{\mathbf{x}}_2\hat{\mathbf{x}}_3, \qquad *\hat{\mathbf{x}}_2 = \hat{\mathbf{x}}_3\hat{\mathbf{x}}_1, \qquad *\hat{\mathbf{x}}_3 = \hat{\mathbf{x}}_1\hat{\mathbf{x}}_2. \tag{4.225}$$

For oriented areas, * is defined on basis area elements as

$$*(\hat{\mathbf{x}}_i\hat{\mathbf{x}}_j) \equiv \varepsilon^k{}_{ij}\hat{\mathbf{x}}_k, \tag{4.226}$$

so

$$\begin{aligned} &*(\hat{\mathbf{x}}_1\hat{\mathbf{x}}_2) = \varepsilon^3{}_{12}\hat{\mathbf{x}}_3 = \hat{\mathbf{x}}_3, \qquad *(\hat{\mathbf{x}}_1\hat{\mathbf{x}}_3) = \varepsilon^2{}_{13}\hat{\mathbf{x}}_2 = -\hat{\mathbf{x}}_2, \\ &*(\hat{\mathbf{x}}_2\hat{\mathbf{x}}_3) = \varepsilon^1{}_{23}\hat{\mathbf{x}}_1 = \hat{\mathbf{x}}_1. \end{aligned} \tag{4.227}$$

For volumes, * is defined as

$$*(\hat{\mathbf{x}}_1\hat{\mathbf{x}}_2\hat{\mathbf{x}}_3) \equiv \varepsilon_{123} = 1. \tag{4.228}$$

Example 4.8.7 CROSS PRODUCT OF VECTORS

The exterior product of two vectors

$$\mathbf{a} = \sum_{i=1}^{3} a^i\hat{\mathbf{x}}_i, \qquad \mathbf{b} = \sum_{i=1}^{3} b^i\hat{\mathbf{x}}_i \tag{4.229}$$

is given by

$$\mathbf{ab} = \left(\sum_{i=1}^{3} a^i\hat{\mathbf{x}}_i\right)\left(\sum_{j=1}^{3} b^i\hat{\mathbf{x}}_j\right) = \sum_{i<j}(a^ib^j - a^jb^i)\hat{\mathbf{x}}_i\hat{\mathbf{x}}_j, \tag{4.230}$$

whereas Eq. (4.224) implies that

$$*(\mathbf{ab}) = \mathbf{a} \times \mathbf{b}. \tag{4.231}$$

■

Next, let us analyze Sections 2.1–2.2 on curvilinear coordinates in the language of differential forms.

Example 4.8.8 LAPLACIAN IN ORTHOGONAL COORDINATES

Consider orthogonal coordinates where the metric (Eq. (2.5)) leads to length elements

$$ds_i = h_i\, dq_i, \qquad \text{not summed.} \tag{4.232}$$

Here the dq_i are ordinary differentials. The 1-forms associated with the directions $\hat{\mathbf{q}}_i$ are

$$\varepsilon^i = h_i dq_i, \qquad \text{not summed.} \tag{4.233}$$

Then the gradient is defined by the 1-form

$$df = \frac{\partial f}{\partial q_i} dq_i = \left(\frac{1}{h_i}\frac{\partial f}{\partial q_i}\right)\varepsilon^i. \tag{4.234}$$

We apply the hodge star operator to df, generating the 2-form

$$\begin{aligned} *df &= \left(\frac{1}{h_i}\frac{\partial f}{\partial q_i}\right) * \varepsilon^i = \left(\frac{1}{h_1}\frac{\partial f}{\partial q_1}\right)\varepsilon^2\varepsilon^3 + \left(\frac{1}{h_2}\frac{\partial f}{\partial q_2}\right)\varepsilon^3\varepsilon^1 + \left(\frac{1}{h_3}\frac{\partial f}{\partial q_3}\right)\varepsilon^1\varepsilon^2 \\ &= \left(\frac{h_2 h_3}{h_1}\frac{\partial f}{\partial q_1}\right) dq_2\, dq_3 + \left(\frac{h_1 h_3}{h_2}\frac{\partial f}{\partial q_2}\right) dq_3\, dq_1 + \left(\frac{h_1 h_2}{h_3}\frac{\partial f}{\partial q_3}\right) dq_1\, dq_2. \end{aligned} \tag{4.235}$$

Applying another exterior derivative d, we get the Laplacian

$$\begin{aligned} d(*df) &= \frac{\partial}{\partial q_1}\left(\frac{h_2 h_3}{h_1}\frac{\partial f}{\partial q_1}\right) dq_1\, dq_2\, dq_3 + \frac{\partial}{\partial q_2}\left(\frac{h_1 h_3}{h_2}\frac{\partial f}{\partial q_2}\right) dq_2\, dq_1\, dq_2\, dq_3 \\ &\quad + \frac{\partial}{\partial q_3}\left(\frac{h_1 h_2}{h_3}\frac{\partial f}{\partial q_3}\right) dq_3\, dq_1\, dq_2\, dq_3 \\ &= \frac{1}{h_1 h_2 h_3}\left[\frac{\partial}{\partial q_1}\left(\frac{h_2 h_3}{h_1}\frac{\partial f}{\partial q_1}\right) + \frac{\partial}{\partial q_2}\left(\frac{h_1 h_3}{h_2}\frac{\partial f}{\partial q_2}\right) + \frac{\partial}{\partial q_3}\left(\frac{h_1 h_2}{h_3}\frac{\partial f}{\partial q_3}\right)\right] \\ &\quad \cdot \varepsilon^1\varepsilon^2\varepsilon^3 = \nabla^2 f\, dq_1\, dq_2\, dq_3. \end{aligned} \tag{4.236}$$

Dividing by the volume element gives Eq. (2.22). Recall that the volume elements $dx\, dy\, dz$ and $\varepsilon^1\varepsilon^2\varepsilon^3$ must be equal because ε^i and dx, dy, dz are orthonormal 1-forms and the map from the xyz to the q_i coordinates is one-to-one. ■

Example 4.8.9 MAXWELL'S EQUATIONS

We now work in four-dimensional Minkowski space, the homogeneous, flat space–time of special relativity, to discuss classical electrodynamics in terms of differential forms. We start by introducing the electromagnetic field 2-form (field tensor in standard relativistic notation):

$$\begin{aligned} F &= -E_x\, dt\, dx - E_y\, dt\, dy - E_z\, dt\, dz + B_x\, dy\, dz + B_y\, dz\, dx + B_z\, dx\, dy \\ &= \frac{1}{2} F_{\mu\nu}\, dx^\mu\, dx^\nu, \end{aligned} \tag{4.237}$$

which contains the electric 1-form $E = E_x\,dx + E_y\,dy + E_z\,dz$ and the magnetic flux 2-form. Here, terms with 1-forms in opposite order have been combined. (For Eq. (4.237) to be valid, the magnetic induction is in units of c; that is, $B_i \to cB_i$, with c the velocity of light; or we work in units where $c = 1$. Also, F is in units of $1/\varepsilon_0$, the dielectric constant of the vacuum. Moreover, the vector potential is defined as $A_0 = \varepsilon_0\phi$, with the nonstatic electric potential ϕ and $A^1 = \frac{A_x}{\mu_0 c}, \ldots$; see Section 4.6 for more details.) The field 2-form F encompasses Faraday's induction law that a moving charge is acted on by magnetic forces.

Applying the exterior derivative d generates Maxwell's homogeneous equations automatically from F:

$$\begin{aligned} dF &= -\left(\frac{\partial E_x}{\partial y}dy + \frac{\partial E_x}{\partial z}dz\right)dt\,dx - \left(\frac{\partial E_y}{\partial x}dx + \frac{\partial E_y}{\partial z}dz\right)dt\,dy \\ &\quad - \left(\frac{\partial E_z}{\partial x}dx + \frac{\partial E_z}{\partial y}dy\right)dt\,dz + \left(\frac{\partial B_x}{\partial x}dx + \frac{\partial B_x}{\partial t}dt\right)dy\,dz \\ &\quad + \left(\frac{\partial B_y}{\partial t}dt + \frac{\partial B_y}{\partial y}dy\right)dz\,dx + \left(\frac{\partial B_z}{\partial t}dt + \frac{\partial B_z}{\partial z}dz\right)dx\,dy \\ &= \left(-\frac{\partial E_x}{\partial y} + \frac{\partial E_y}{\partial x} + \frac{\partial B_z}{\partial t}\right)dt\,dx\,dy + \left(-\frac{\partial E_x}{\partial z} + \frac{\partial E_z}{\partial x} - \frac{\partial B_y}{\partial t}\right)dt\,dx\,dz \\ &\quad + \left(-\frac{\partial E_y}{\partial z} + \frac{\partial E_z}{\partial y} + \frac{\partial B_x}{\partial t}\right)dt\,dy\,dz = 0 \end{aligned} \tag{4.238}$$

which, in standard notation of vector analysis, takes the familiar vector form of Maxwell's homogeneous equations,

$$\boldsymbol{\nabla} \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0. \tag{4.239}$$

Since $dF = 0$, that is, there is no driving term so that F is closed, there must be a 1-form $\omega = A_\mu\,dx^\mu$ so that $F = d\omega$. Now,

$$d\omega = \partial_\nu A_\mu\,dx^\nu\,dx^\mu, \tag{4.240}$$

which, in standard notation, leads to the conventional relativistic form of the electromagnetic field tensor,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{4.241}$$

Maxwell's homogeneous equations, $dF = 0$, are thus equivalent to $\partial^\nu F_{\mu\nu} = 0$.

In order to derive similarly the inhomogeneous Maxwell's equations, we introduce the dual electromagnetic field tensor

$$\tilde{F}^{\mu\nu} = \frac{1}{2}\varepsilon^{\mu\nu\alpha\beta}F_{\alpha\beta}, \tag{4.242}$$

and, in terms of differential forms,

$$*F = *\left(F_{\mu\nu}\,dx^\mu\,dx^\nu\right) = F_{\mu\nu} * \left(dx^\mu\,dx^\nu\right) = \frac{1}{2}F_{\mu\nu}\varepsilon^{\mu\nu}{}_{\alpha\beta}\,dx^\alpha\,dx^\beta. \tag{4.243}$$

Applying the exterior derivative yields

$$d(*F) = \frac{1}{2}\varepsilon^{\mu\nu}{}_{\alpha\beta}(\partial_\gamma F_{\mu\nu})\,dx^\gamma\,dx^\alpha\,dx^\beta, \tag{4.244}$$

the left-hand side of Maxwell's inhomogeneous equations, a 3-form. Its driving term is the dual of the electric current density, a 3-form:

$$\begin{aligned} *J &= J_\alpha\big(*dx^\alpha\big) = J_\alpha\varepsilon^\alpha{}_{\mu\nu\lambda}\,dx^\mu\,dx^\nu\,dx^\lambda \\ &= \rho\,dx\,dy\,dz - J_x\,dt\,dy\,dz - J_y\,dt\,dz\,dx - J_z\,dt\,dx\,dy. \end{aligned} \tag{4.245}$$

Altogether Maxwell's inhomogeneous equations take the elegant form

$$d(*F) = *J. \tag{4.246}$$

■

The differential-form framework has brought considerable unification to vector algebra and to tensor analysis on manifolds more generally, such as uniting Stokes' and Gauss' theorems and providing an elegant reformulation of Maxwell's equations and an efficient derivation of the Laplacian in curved orthogonal coordinates, among others.

Exercises

4.8.1 Evaluate the 1-form $a\,dx + 2b\,dy + 4c\,dz$ on the line segment PQ, with $P = (3, 5, 7)$, $Q = (7, 5, 3)$.

4.8.2 If the force field is constant and moving a particle from the origin to $(3, 0, 0)$ requires a units of work, from $(-1, -1, 0)$ to $(-1, 1, 0)$ takes b units of work, and from $(0, 0, 4)$ to $(0, 0, 5)$ c units of work, find the 1-form of the work.

4.8.3 Evaluate the flow described by the 2-form $dx\,dy + 2\,dy\,dz + 3\,dz\,dx$ across the oriented triangle PQR with corners at

$$P = (3, 1, 4), \qquad Q = (-2, 1, 4), \qquad R = (1, 4, 1).$$

4.8.4 Are the points, in this order,

$$(0, 1, 1), \qquad (3, -1, -2), \qquad (4, 2, -2), \qquad (-1, 0, 1)$$

coplanar, or do they form an oriented volume (right-handed or left-handed)?

4.8.5 Write Oersted's law,

$$\int_{\partial S} \mathbf{H}\cdot d\mathbf{r} = \int_S \nabla\times\mathbf{H}\cdot d\mathbf{a} \sim I,$$

in differential form notation.

4.8.6 Describe the electric field by the 1-form $E_1\,dx + E_2\,dy + E_3\,dz$ and the magnetic induction by the 2-form $B_1\,dy\,dz + B_2\,dz\,dx + B_3\,dx\,dy$. Then formulate Faraday's induction law in terms of these forms.

4.8.7 Evaluate the 1-form

$$\frac{x\,dy}{x^2+y^2}-\frac{y\,dx}{x^2+y^2}$$

on the unit circle about the origin oriented counterclockwise.

4.8.8 Find the pullback of $dx\,dz$ under $x=u\cos v$, $y=u-v$, $z=u\sin v$.

4.8.9 Find the pullback of the 2-form $dy\,dz+dz\,dx+dx\,dy$ under the map $x=\sin\theta\cos\varphi$, $y=\sin\theta\sin\varphi$, $z=\cos\theta$.

4.8.10 Parameterize the surface obtained by rotating the circle $(x-2)^2+z^2=1$, $y=0$, about the z-axis in a counterclockwise orientation, as seen from outside.

4.8.11 A 1-form $A\,dx+B\,dy$ is defined as **closed** if $\frac{\partial A}{\partial y}=\frac{\partial B}{\partial x}$. It is called **exact** if there is a function f so that $\frac{\partial f}{\partial x}=A$ and $\frac{\partial f}{\partial y}=B$. Determine which of the following 1-forms are closed, or exact, and find the corresponding functions f for those that are exact:

$$y\,dx+x\,dy,\qquad \frac{y\,dx+x\,dy}{x^2+y^2},\qquad \big[\ln(xy)+1\big]dx+\frac{x}{y}dy,$$

$$-\frac{y\,dx}{x^2+y^2}+\frac{x\,dy}{x^2+y^2},\qquad f(z)\,dz \text{ with } z=x+iy.$$

4.8.12 Show that $\sum_{i=1}^{n}x_i^2=a^2$ defines a differentiable manifold of dimension $D=n-1$ if $a\neq 0$ and $D=0$ if $a=0$.

4.8.13 Show that the set of orthogonal 2×2 matrices form a differentiable manifold, and determine its dimension.

4.8.14 Determine the value of the 2-form $A\,dy\,dz+B\,dz\,dx+C\,dx\,dy$ on a parallelogram with sides $\mathbf{a}$, $\mathbf{b}$.

4.8.15 Prove Lorentz invariance of Maxwell's equations in the language of differential forms.

Additional Readings

Buerger, M. J., *Elementary Crystallography*. New York: Wiley (1956). A comprehensive discussion of crystal symmetries. Buerger develops all 32 point groups and all 230 space groups. Related books by this author include *Contemporary Crystallography*. New York: McGraw-Hill (1970); *Crystal Structure Analysis*. New York: Krieger (1979) (reprint, 1960); and *Introduction to Crystal Geometry*. New York: Krieger (1977) (reprint, 1971).

Burns, G., and A. M. Glazer, *Space Groups for Solid-State Scientists*. New York: Academic Press (1978). A well-organized, readable treatment of groups and their application to the solid state.

de-Shalit, A., and I. Talmi, *Nuclear Shell Model*. New York: Academic Press (1963). We adopt the Condon–Shortley phase conventions of this text.

Edmonds, A. R., *Angular Momentum in Quantum Mechanics*. Princeton, NJ: Princeton University Press (1957).

Edwards, H. M., *Advanced Calculus: A Differential Forms Approach*. Boston: Birkhäuser (1994).

Falicov, L. M., *Group Theory and Its Physical Applications*. Notes compiled by A. Luehrmann. Chicago: University of Chicago Press (1966). Group theory, with an emphasis on applications to crystal symmetries and solid-state physics.

Gell-Mann, M., and Y. Ne'eman, *The Eightfold Way*. New York: Benjamin (1965). A collection of reprints of significant papers on SU(3) and the particles of high-energy physics. Several introductory sections by Gell-Mann and Ne'eman are especially helpful.

Greiner, W., and B. Müller, *Quantum Mechanics Symmetries*. Berlin: Springer (1989). We refer to this textbook for more details and numerous exercises that are worked out in detail.

Hamermesh, M., *Group Theory and Its Application to Physical Problems*. Reading, MA: Addison-Wesley (1962). A detailed, rigorous account of both finite and continuous groups. The 32 point groups are developed. The continuous groups are treated, with Lie algebra included. A wealth of applications to atomic and nuclear physics.

Hassani, S., *Foundations of Mathematical Physics*. Boston: Allyn and Bacon (1991).

Heitler, W., *The Quantum Theory of Radiation*, 2nd ed. Oxford: Oxford University Press (1947). Reprinted, New York: Dover (1983).

Higman, B., *Applied Group-Theoretic and Matrix Methods*. Oxford: Clarendon Press (1955). A rather complete and unusually intelligible development of matrix analysis and group theory.

Jackson, J. D., *Classical Electrodynamics*, 3rd ed. New York: Wiley (1998).

Messiah, A., *Quantum Mechanics*, Vol. II. Amsterdam: North-Holland (1961).

Panofsky, W. K. H., and M. Phillips, *Classical Electricity and Magnetism*, 2nd ed. Reading, MA: Addison-Wesley (1962). The Lorentz covariance of Maxwell's equations is developed for both vacuum and material media. Panofsky and Phillips use contravariant and covariant tensors.

Park, D., Resource letter SP-1 on symmetry in physics. *Am. J. Phys.* **36**: 577–584 (1968). Includes a large selection of basic references on group theory and its applications to physics: atoms, molecules, nuclei, solids, and elementary particles.

Ram, B., Physics of the SU(3) symmetry model. *Am. J. Phys.* **35**: 16 (1967). An excellent discussion of the applications of SU(3) to the strongly interacting particles (baryons). For a sequel to this see R. D. Young, Physics of the quark model. *Am. J. Phys.* **41**: 472 (1973).

Rose, M. E., *Elementary Theory of Angular Momentum*. New York: Wiley (1957). Reprinted. New York: Dover (1995). As part of the development of the quantum theory of angular momentum, Rose includes a detailed and readable account of the rotation group.

Wigner, E. P., *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (translated by J. J. Griffin). New York: Academic Press (1959). This is the classic reference on group theory for the physicist. The rotation group is treated in considerable detail. There is a wealth of applications to atomic physics.

CHAPTER 5

INFINITE SERIES

5.1 FUNDAMENTAL CONCEPTS

Infinite series, literally summations of an infinite number of terms, occur frequently in both pure and applied mathematics. They may be used by the pure mathematician to define functions as a fundamental approach to the theory of functions, as well as for calculating accurate values of transcendental constants and transcendental functions. In the mathematics of science and engineering infinite series are ubiquitous, for they appear in the evaluation of integrals (Sections 5.6 and 5.7), in the solution of differential equations (Sections 9.5 and 9.6), and as Fourier series (Chapter 14) and compete with integral representations for the description of a host of special functions (Chapters 11, 12, and 13). In Section 16.3 the Neumann series solution for integral equations provides one more example of the occurrence and use of infinite series.

Right at the start we face the problem of attaching meaning to the sum of an infinite number of terms. The usual approach is by partial sums. If we have an infinite sequence of terms $u_1, u_2, u_3, u_4, u_5, \ldots$, we define the ith partial sum as

$$s_i = \sum_{n=1}^{i} u_n. \tag{5.1}$$

This is a finite summation and offers no difficulties. If the partial sums s_i converge to a (finite) limit as $i \to \infty$,

$$\lim_{i\to\infty} s_i = S, \tag{5.2}$$

the infinite series $\sum_{n=1}^{\infty} u_n$ is said to be **convergent** and to have the value S. Note that we reasonably, plausibly, but still arbitrarily **define** the infinite series as equal to S and that a necessary condition for this convergence to a limit is that $\lim_{n\to\infty} u_n = 0$. This condition, however, is not sufficient to guarantee convergence. Equation (5.2) is usually written in formal mathematical notation:

The condition for the existence of a limit S is that for each $\varepsilon > 0$, there is a fixed $N = N(\varepsilon)$ such that

$$|S - s_i| < \varepsilon, \qquad \text{for } i > N.$$

This condition is often derived from the Cauchy criterion applied to the partial sums s_i. The **Cauchy criterion** is:

A necessary and sufficient condition that a sequence (s_i) converge is that for each $\varepsilon > 0$ there is a fixed number N such that

$$|s_j - s_i| < \varepsilon, \qquad \text{for all } i, j > N.$$

This means that the individual partial sums must cluster together as we move far out in the sequence.

The Cauchy criterion may easily be extended to sequences of functions. We see it in this form in Section 5.5 in the definition of uniform convergence and in Section 10.4 in the development of Hilbert space. Our partial sums s_i may not converge to a single limit but may oscillate, as in the case

$$\sum_{n=1}^{\infty} u_n = 1 - 1 + 1 - 1 + 1 + \cdots - (-1)^n + \cdots.$$

Clearly, $s_i = 1$ for i odd but $s_i = 0$ for i even. There is no convergence to a limit, and series such as this one are labeled **oscillatory**. Whenever the sequence of partial sums diverges (approaches $\pm\infty$), the infinite series is said to **diverge**. Often the term *divergent* is extended to include oscillatory series as well. Because we evaluate the partial sums by ordinary arithmetic, the convergent series, defined in terms of a limit of the partial sums, assumes a position of supreme importance. Two examples may clarify the nature of convergence or divergence of a series and will also serve as a basis for a further detailed investigation in the next section.

Example 5.1.1 THE GEOMETRIC SERIES

The geometrical sequence, starting with a and with a ratio r $(= a_{n+1}/a_n$ independent of $n)$, is given by

$$a + ar + ar^2 + ar^3 + \cdots + ar^{n-1} + \cdots.$$

The nth partial sum is given by[1]

$$s_n = a\frac{1 - r^n}{1 - r}. \tag{5.3}$$

Taking the limit as $n \to \infty$,

$$\lim_{n\to\infty} s_n = \frac{a}{1 - r}, \quad \text{for } |r| < 1. \tag{5.4}$$

[1]Multiply and divide $s_n = \sum_{m=0}^{n-1} ar^m$ by $1 - r$.

Hence, by definition, the infinite geometric series converges for $|r| < 1$ and is given by

$$\boxed{\sum_{n=1}^{\infty} ar^{n-1} = \frac{a}{1-r}.} \tag{5.5}$$

On the other hand, if $|r| \geq 1$, the necessary condition $u_n \to 0$ is not satisfied and the infinite series diverges. ■

Example 5.1.2 THE HARMONIC SERIES

As a second and more involved example, we consider the harmonic series

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} + \cdots. \tag{5.6}$$

We have the $\lim_{n\to\infty} u_n = \lim_{n\to\infty} 1/n = 0$, but this is not sufficient to guarantee convergence. If we group the terms (no change in order) as

$$1 + \tfrac{1}{2} + \left(\tfrac{1}{3} + \tfrac{1}{4}\right) + \left(\tfrac{1}{5} + \tfrac{1}{6} + \tfrac{1}{7} + \tfrac{1}{8}\right) + \left(\tfrac{1}{9} + \cdots + \tfrac{1}{16}\right) + \cdots, \tag{5.7}$$

each pair of parentheses encloses p terms of the form

$$\frac{1}{p+1} + \frac{1}{p+2} + \cdots + \frac{1}{p+p} > \frac{p}{2p} = \frac{1}{2}. \tag{5.8}$$

Forming partial sums by adding the parenthetical groups one by one, we obtain

$$\begin{aligned} s_1 &= 1, & s_4 &> \frac{5}{2}, \\ s_2 &= \frac{3}{2}, & s_5 &> \frac{6}{2}, \cdots \\ s_3 &> \frac{4}{2}, & s_n &> \frac{n+1}{2}. \end{aligned} \tag{5.9}$$

The harmonic series considered in this way is certainly divergent.[2] An alternate and independent demonstration of its divergence appears in Section 5.2. ■

If the $u_n > 0$ are monotonically decreasing to zero, that is, $u_n > u_{n+1}$ for all n, then $\sum_n u_n$ is converging to S if, and only if, $s_n - nu_n$ converges to S. As the partial sums s_n converge to S, this theorem implies that $nu_n \to 0$, for $n \to \infty$.

To prove this **theorem**, we start by concluding from $0 < u_{n+1} < u_n$ and

$$s_{n+1} - (n+1)u_{n+1} = s_n - nu_{n+1} = s_n - nu_n + n(u_n - u_{n+1}) > s_n - nu_n$$

that $s_n - nu_n$ increases as $n \to \infty$. As a consequence of $s_n - nu_n < s_n \leq S$, $s_n - nu_n$ converges to a value $s \leq S$. Deleting the tail of positive terms $u_i - u_n$ from $i = \nu + 1$ to n,

[2] The (finite) harmonic series appears in an interesting note on the maximum stable displacement of a stack of coins. P. R. Johnson, The Leaning Tower of Lire. *Am. J. Phys.* **23**: 240 (1955).

we infer from $s_n - nu_n > u_0 + (u_1 - u_n) + \cdots + (u_\nu - u_n) = s_\nu - \nu u_n$ that $s_n - nu_n \geq s_\nu$ for $n \to \infty$. Hence also $s \geq S$, so $s = S$ and $nu_n \to 0$.

When this theorem is applied to the **harmonic series** $\sum_n \frac{1}{n}$ with $n\frac{1}{n} = 1$ it implies that it does not converge; it diverges to $+\infty$.

Addition, Subtraction of Series

If we have two convergent series $\sum_n u_n \to S$ and $\sum_n v_n \to S$, their sum and difference will also converge to $s \pm S$ because their partial sums satisfy

$$|s_j \pm S_j - (s_i \pm S_i)| \leq |s_j - s_i| + |S_j - S_i| < 2\epsilon$$

using the triangle inequality

$$|a| - |b| \leq |a + b| \leq |a| + |b|$$

for $a = s_j - s_i, b = S_j - S_i$.

A convergent series $\sum_n u_n \to S$ may be multiplied termwise by a real number a. The new series will converge to aS because

$$|as_j - as_i| = |a||s_j - s_i| < |a|\epsilon.$$

This multiplication by a constant can be generalized to a multiplication by terms c_n of a bounded sequence of numbers:

If $\sum_n u_n$ converges to S and $0 < c_n \leq M$ are bounded, then $\sum_n u_n c_n$ is convergent. If $\sum_n u_n$ is divergent and $c_n > M > 0$, then $\sum_n u_n c_n$ diverges.

To prove this **theorem** we take i, j sufficiently large so that $|s_j - s_i| < \epsilon$. Then

$$\sum_{i+1}^{j} u_n c_n \leq M \sum_{i+1}^{j} u_n = M|s_j - s_i| < M\epsilon.$$

The divergent case follows from

$$\sum_n u_n c_n > M \sum_n u_n \to \infty.$$

Using the binomial theorem[3] (Section 5.6), we may expand the function $(1 + x)^{-1}$:

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \cdots + (-x)^{n-1} + \cdots. \tag{5.10}$$

If we let $x \to 1$, this series becomes

$$1 - 1 + 1 - 1 + 1 - 1 + \cdots, \tag{5.11}$$

a series that we labeled oscillatory earlier in this section. Although it does not converge in the usual sense, meaning can be attached to this series. Euler, for example, assigned a value of $1/2$ to this oscillatory sequence on the basis of the correspondence between this series and the well-defined function $(1 + x)^{-1}$. Unfortunately, such correspondence between series and function is not unique, and this approach must be refined. Other methods

[3] Actually Eq. (5.10) may be verified by multiplying both sides by $1 + x$.

of assigning a meaning to a divergent or oscillatory series, methods of defining a sum, have been developed. See G. H. Hardy, *Divergent Series*, Chelsea Publishing Co. 2nd ed. (1992). In general, however, this aspect of infinite series is of relatively little interest to the scientist or the engineer. An exception to this statement, the very important asymptotic or semiconvergent series, is considered in Section 5.10.

Exercises

5.1.1 Show that

$$\sum_{n=1}^{\infty} \frac{1}{(2n-1)(2n+1)} = \frac{1}{2}.$$

Hint. Show (by mathematical induction) that $s_m = m/(2m+1)$.

5.1.2 Show that

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1.$$

Find the partial sum s_m and verify its correctness by mathematical induction.
Note. The method of expansion in partial fractions, Section 15.8, offers an alternative way of solving Exercises 5.1.1 and 5.1.2.

5.2 Convergence Tests

Although nonconvergent series may be useful in certain special cases (compare Section 5.10), we usually insist, as a matter of convenience if not necessity, that our series be convergent. It therefore becomes a matter of extreme importance to be able to tell whether a given series is convergent. We shall develop a number of possible tests, starting with the simple and relatively insensitive tests and working up to the more complicated but quite sensitive tests. For the present let us consider a **series of positive terms** $a_n \geq 0$, postponing negative terms until the next section.

Comparison Test

If term by term a series of terms $0 \leq u_n \leq a_n$, in which the a_n form a convergent series, the series $\sum_n u_n$ is also convergent. If $u_n \leq a_n$ for all n, then $\sum_n u_n \leq \sum_n a_n$ and $\sum_n u_n$ therefore is **convergent**. If term by term a series of terms $v_n \geq b_n$, in which the b_n, form a divergent series, the series $\sum_n v_n$ is also **divergent**. Note that comparisons of u_n with b_n or v_n with a_n yield no information. If $v_n \geq b_n$ for all n, then $\sum_n v_n \geq \sum_n b_n$ and $\sum_n v_n$ therefore is divergent.

For the convergent series a_n we already have the geometric series, whereas the harmonic series will serve as the divergent comparison series b_n. As other series are identified as either convergent or divergent, they may be used for the known series in this comparison test. All tests developed in this section are essentially comparison tests. Figure 5.1 exhibits these tests and the interrelationships.

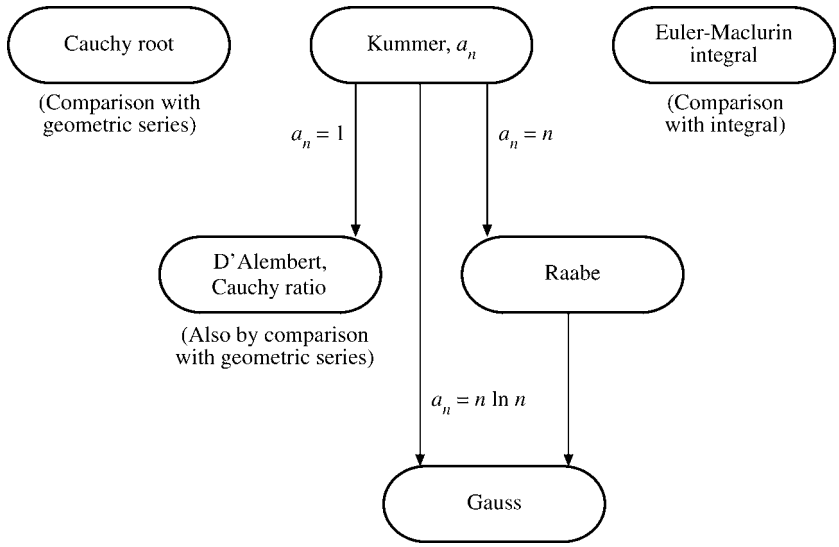

FIGURE 5.1 Comparison tests.

Example 5.2.1 A DIRICHLET SERIES

Test $\sum_{n=1}^{\infty} n^{-p}$, $p = 0.999$, for convergence. Since $n^{-0.999} > n^{-1}$ and $b_n = n^{-1}$ forms the divergent harmonic series, the comparison test shows that $\sum_n n^{-0.999}$ is divergent. Generalizing, $\sum_n n^{-p}$ is seen to be divergent for all $p \leq 1$ but convergent for $p > 1$ (see Example 5.2.3). ■

Cauchy Root Test

If $(a_n)^{1/n} \leq r < 1$ for all sufficiently large n, with r independent of n, then $\sum_n a_n$ is convergent. If $(a_n)^{1/n} \geq 1$ for all sufficiently large n, then $\sum_n a_n$ is divergent.

The first part of this test is verified easily by raising $(a_n)^{1/n} \leq r$ to the nth power. We get

$$a_n \leq r^n < 1.$$

Since r^n is just the nth term in a convergent geometric series, $\sum_n a_n$ is convergent by the comparison test. Conversely, if $(a_n)^{1/n} \geq 1$, then $a_n \geq 1$ and the series must diverge. This root test is particularly useful in establishing the properties of power series (Section 5.7).

D'Alembert (or Cauchy) Ratio Test

If $a_{n+1}/a_n \leq r < 1$ for all sufficiently large n and r is independent of n, then $\sum_n a_n$ is convergent. If $a_{n+1}/a_n \geq 1$ for all sufficiently large n, then $\sum_n a_n$ is divergent.

Convergence is proved by direct comparison with the geometric series $(1 + r + r^2 + \cdots)$. In the second part, $a_{n+1} \geq a_n$ and divergence should be reasonably obvious. Although not

quite so sensitive as the Cauchy root test, this D'Alembert ratio test is one of the easiest to apply and is widely used. An alternate statement of the ratio test is in the form of a limit: If

$$\lim_{n\to\infty}\frac{a_{n+1}}{a_n}\begin{cases}<1, & \text{convergence,}\\ >1, & \text{divergence,}\\ =1, & \text{indeterminate.}\end{cases} \tag{5.12}$$

Because of this final indeterminate possibility, the ratio test is likely to fail at crucial points, and more delicate, sensitive tests are necessary. The alert reader may wonder how this indeterminacy arose. Actually it was concealed in the first statement, $a_{n+1}/a_n \le r < 1$. We might encounter $a_{n+1}/a_n < 1$ for all **finite** n but be unable to choose an $r < 1$ **and independent of n** such that $a_{n+1}/a_n \le r$ for all sufficiently large n. An example is provided by the harmonic series

$$\frac{a_{n+1}}{a_n} = \frac{n}{n+1} < 1. \tag{5.13}$$

Since

$$\lim_{n\to\infty}\frac{a_{n+1}}{a_n} = 1, \tag{5.14}$$

no fixed ratio $r < 1$ exists and the ratio test fails.

Example 5.2.2 D'ALEMBERT RATIO TEST

Test $\sum_n n/2^n$ for convergence.

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)/2^{n+1}}{n/2^n} = \frac{1}{2}\cdot\frac{n+1}{n}. \tag{5.15}$$

Since

$$\frac{a_{n+1}}{a_n} \le \frac{3}{4} \qquad \text{for } n \ge 2, \tag{5.16}$$

we have convergence. Alternatively,

$$\lim_{n\to\infty}\frac{a_{n+1}}{a_n} = \frac{1}{2} \tag{5.17}$$

and again — convergence. ■

Cauchy (or Maclaurin) Integral Test

This is another sort of comparison test, in which we compare a series with an integral. Geometrically, we compare the area of a series of unit-width rectangles with the area under a curve.

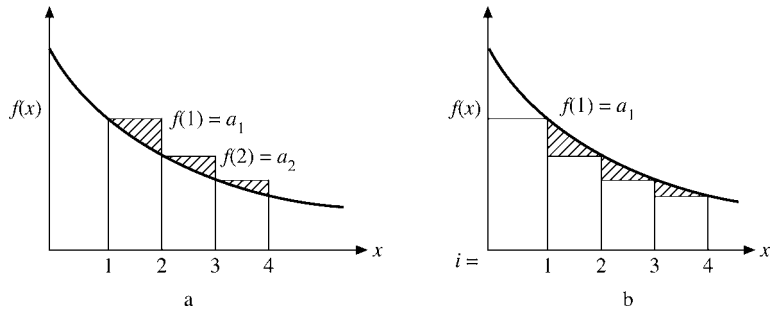

FIGURE 5.2 (a) Comparison of integral and sum-blocks leading. (b) Comparison of integral and sum-blocks lagging.

Let $f(x)$ be a continuous, **monotonic decreasing function** in which $f(n) = a_n$. Then $\sum_n a_n$ converges if $\int_1^\infty f(x)\,dx$ is finite and diverges if the integral is infinite. For the ith partial sum,

$$s_i = \sum_{n=1}^{i} a_n = \sum_{n=1}^{i} f(n). \tag{5.18}$$

But

$$s_i > \int_1^{i+1} f(x)\,dx \tag{5.19}$$

from Fig. 5.2a, $f(x)$ being monotonic decreasing. On the other hand, from Fig. 5.2b,

$$s_i - a_1 < \int_1^{i} f(x)\,dx, \tag{5.20}$$

in which the series is represented by the inscribed rectangles. Taking the limit as $i \to \infty$, we have

$$\boxed{\int_1^\infty f(x)\,dx \le \sum_{n=1}^{\infty} a_n \le \int_1^\infty f(x)\,dx + a_1.} \tag{5.21}$$

Hence the infinite series converges or diverges as the corresponding integral converges or diverges. This integral test is particularly useful in setting upper and lower bounds on the remainder of a series after some number of initial terms have been summed. That is,

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{N} a_n + \sum_{n=N+1}^{\infty} a_n,$$

where

$$\boxed{\int_{N+1}^\infty f(x)\,dx \le \sum_{n=N+1}^{\infty} a_n \le \int_{N+1}^\infty f(x)\,dx + a_{N+1}.}$$

To free the integral test from the quite restrictive requirement that the interpolating function $f(x)$ be positive and monotonic, we show for any function $f(x)$ with a continuous derivative that

$$\sum_{n=N_i+1}^{N_f} f(n) = \int_{N_i}^{N_f} f(x)\,dx + \int_{N_i}^{N_f} \big(x-[x]\big) f'(x)\,dx \tag{5.22}$$

holds. Here $[x]$ denotes the largest integer below x, so $x-[x]$ varies sawtoothlike between 0 and 1. To derive Eq. (5.22) we observe that

$$\int_{N_i}^{N_f} x f'(x)\,dx = N_f f(N_f) - N_i f(N_i) - \int_{N_i}^{N_f} f(x)\,dx, \tag{5.23}$$

using integration by parts. Next we evaluate the integral

$$\begin{aligned}\int_{N_i}^{N_f} [x] f'(x)\,dx &= \sum_{n=N_i}^{N_f-1} n \int_n^{n+1} f'(x)\,dx = \sum_{n=N_i}^{N_f-1} n\big\{f(n+1)-f(n)\big\} \\ &= -\sum_{n=N_i+1}^{N_f} f(n) - N_i f(N_i) + N_f f(N_f). \end{aligned}\tag{5.24}$$

Subtracting Eq. (5.24) from (5.23) we arrive at Eq. (5.22). Note that $f(x)$ may go up or down and even change sign, so Eq. (5.22) applies to alternating series (see Section 5.3) as well. Usually $f'(x)$ falls faster than $f(x)$ for $x\to\infty$, so the remainder term in Eq. (5.22) converges better. It is easy to improve Eq. (5.22) by replacing $x-[x]$ by $x-[x]-\frac{1}{2}$, which varies between $-\frac{1}{2}$ and $\frac{1}{2}$:

$$\begin{aligned}\sum_{N_i<n\le N_f} f(n) = \int_{N_i}^{N_f} f(x)\,dx + \int_{N_i}^{N_f} \big(x-[x]-\tfrac{1}{2}\big) f'(x)\,dx \\ + \tfrac{1}{2}\big\{f(N_f)-f(N_i)\big\}.\end{aligned}\tag{5.25}$$

Then the $f'(x)$-integral becomes even smaller, if $f'(x)$ does not change sign too often. For an application of this integral test to an alternating series see Example 5.3.1.

Example 5.2.3 RIEMANN ZETA FUNCTION

The Riemann zeta function is defined by

$$\boxed{\zeta(p) = \sum_{n=1}^{\infty} n^{-p},} \tag{5.26}$$

provided the series converges. We may take $f(x)=x^{-p}$, and then

$$\begin{aligned}\int_1^{\infty} x^{-p}\,dx &= \frac{x^{-p+1}}{-p+1}\bigg|_1^{\infty}, \qquad p\neq 1 \\ &= \ln x\,|_{x=1}^{\infty}, \qquad p=1.\end{aligned}\tag{5.27}$$

The integral and therefore the series are divergent for $p \le 1$, convergent for $p > 1$. Hence Eq. (5.26) should carry the condition $p > 1$. This, incidentally, is an independent proof that the harmonic series ($p = 1$) diverges logarithmically. The sum of the first million terms $\sum^{1,000,000} n^{-1}$ is only 14.392 726.... ■

This integral comparison may also be used to set an upper limit to the Euler–Mascheroni constant,[4] defined by

$$\boxed{\gamma = \lim_{n\to\infty}\left(\sum_{m=1}^{n} m^{-1} - \ln n\right).} \tag{5.28}$$

Returning to partial sums, Eq. (5.20) yields

$$s_n = \sum_{m=1}^{n} m^{-1} - \ln n \le \int_1^n \frac{dx}{x} - \ln n + 1. \tag{5.29}$$

Evaluating the integral on the right, $s_n < 1$ for all n and therefore $\gamma \le 1$. Exercise 5.2.12 leads to more restrictive bounds. Actually the Euler–Mascheroni constant is 0.57721566....

Kummer's Test

This is the first of three tests that are somewhat more difficult to apply than the preceding tests. Their importance lies in their power and sensitivity. Frequently, at least one of the three will work when the simpler, easier tests are indecisive. It must be remembered, however, that these tests, like those previously discussed, are ultimately based on comparisons. It can be shown that there is no most slowly converging series and no most slowly diverging series. This means that all convergence tests given here, including Kummer's, may fail sometime.

We consider a series of positive terms u_i and a sequence of finite positive constants a_i. If

$$a_n \frac{u_n}{u_{n+1}} - a_{n+1} \ge C > 0 \tag{5.30}$$

for all $n \ge N$, where N is some fixed number,[5] then $\sum_{i=1}^{\infty} u_i$ **converges**. If

$$a_n \frac{u_n}{u_{n+1}} - a_{n+1} \le 0 \tag{5.31}$$

and $\sum_{i=1}^{\infty} a_i^{-1}$ diverges, then $\sum_{i=1}^{\infty} u_i$ **diverges**.

[4]This is the notation of National Bureau of Standards, *Handbook of Mathematical Functions,* Applied Mathematics Series-55 (AMS-55). New York: Dover (1972).

[5]With u_m finite, the partial sum s_N will always be finite for N finite. The convergence or divergence of a series depends on the behavior of the last infinity of terms, not on the first N terms.

The proof of this powerful test is remarkably simple. From Eq. (5.30), with C some positive constant,

$$\begin{aligned} Cu_{N+1} &\le a_N u_N - a_{N+1}u_{N+1} \\ Cu_{N+2} &\le a_{N+1}u_{N+1} - a_{N+2}u_{N+2} \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ Cu_n &\le a_{n-1}u_{n-1} - a_n u_n. \end{aligned} \tag{5.32}$$

Adding and dividing by C (and recalling that $C \neq 0$), we obtain

$$\sum_{i=N+1}^{n} u_i \le \frac{a_N u_N}{C} - \frac{a_n u_n}{C}. \tag{5.33}$$

Hence for the partial sum s_n,

$$\begin{aligned} s_n &\le \sum_{i=1}^{N} u_i + \frac{a_N u_N}{C} - \frac{a_n u_n}{C} \\ &< \sum_{i=1}^{N} u_i + \frac{a_N u_N}{C}, \qquad \text{a constant, independent of } n. \end{aligned} \tag{5.34}$$

The partial sums therefore have an upper bound. With zero as an obvious lower bound, the series $\sum u_i$ must converge.

Divergence is shown as follows. From Eq. (5.31) for $u_{n+1} > 0$,

$$a_n u_n \ge a_{n-1}u_{n-1} \ge \cdots \ge a_N u_N, \qquad n > N. \tag{5.35}$$

Thus, for $a_n > 0$,

$$u_n \ge \frac{a_N u_N}{a_n} \tag{5.36}$$

and

$$\sum_{i=N+1}^{\infty} u_i \ge a_N u_N \sum_{i=N+1}^{\infty} a_i^{-1}. \tag{5.37}$$

If $\sum_{i=1}^{\infty} a_i^{-1}$ diverges, then by the comparison test $\sum_i u_i$ diverges. Equations (5.30) and (5.31) are often given in a limit form:

$$\lim_{n\to\infty}\left(a_n \frac{u_n}{u_{n+1}} - a_{n+1}\right) = C. \tag{5.38}$$

Thus for $C > 0$ we have convergence, whereas for $C < 0$ (and $\sum_i a_i^{-1}$ divergent) we have divergence. It is perhaps useful to show the close relation of Eq. (5.38) and Eqs. (5.30) and (5.31) and to show why indeterminacy creeps in when the limit $C = 0$. From the definition of limit,

$$\left| a_n \frac{u_n}{u_{n+1}} - a_{n+1} - C \right| < \varepsilon \tag{5.39}$$

for all $n \geq N$ and all $\varepsilon > 0$, no matter how small ε may be. When the absolute value signs are removed,

$$C - \varepsilon < a_n \frac{u_n}{u_{n+1}} - a_{n+1} < C + \varepsilon. \tag{5.40}$$

Now, if $C > 0$, Eq. (5.30) follows from ε sufficiently small. On the other hand, if $C < 0$, Eq. (5.31) follows. However, if $C = 0$, the center term, $a_n(u_n/u_{n+1}) - a_{n+1}$, may be either positive or negative and the proof fails. The primary use of Kummer's test is to prove other tests, such as Raabe's (compare also Exercise 5.2.3).

If the positive constants a_n of Kummer's test are chosen $a_n = n$, we have Raabe's test.

Raabe's Test

If $u_n > 0$ and if

$$n\left(\frac{u_n}{u_{n+1}} - 1\right) \geq P > 1 \tag{5.41}$$

for all $n \geq N$, where N is a positive integer independent of n, **then** $\sum_i u_i$ **converges**. Here, $P = C + 1$ of Kummer's test. If

$$n\left(\frac{u_n}{u_{n+1}} - 1\right) \leq 1, \tag{5.42}$$

then $\sum_i u_i$ diverges (as $\sum_n n^{-1}$ diverges). The limit form of Raabe's test is

$$\lim_{n\to\infty} n\left(\frac{u_n}{u_{n+1}} - 1\right) = P. \tag{5.43}$$

We have convergence for $P > 1$, divergence for $P < 1$, and no conclusion for $P = 1$, exactly as with the Kummer test. This indeterminacy is pointed up by Exercise 5.2.4, which presents a convergent series and a divergent series, with both series yielding $P = 1$ in Eq. (5.43).

Raabe's test is more sensitive than the d'Alembert ratio test (Exercise 5.2.3) because $\sum_{n=1}^{\infty} n^{-1}$ diverges more slowly than $\sum_{n=1}^{\infty} 1$. We obtain a more sensitive test (and one that is still fairly easy to apply) by choosing $a_n = n \ln n$. This is Gauss' test.

Gauss' Test

If $u_n > 0$ for all finite n and

$$\frac{u_n}{u_{n+1}} = 1 + \frac{h}{n} + \frac{B(n)}{n^2}, \tag{5.44}$$

in which $B(n)$ is a bounded function of n for $n \to \infty$, then $\sum_i u_i$ converges for $h > 1$ and diverges for $h \leq 1$: There is no indeterminate case here.

The Gauss test is an extremely sensitive test of series convergence. It will work for all series the physicist is likely to encounter. For $h > 1$ or $h < 1$ the proof follows directly from Raabe's test

$$\lim_{n\to\infty} n\left[1+\frac{h}{n}+\frac{B(n)}{n^2}-1\right]=\lim_{n\to\infty}\left[h+\frac{B(n)}{n}\right]=h. \tag{5.45}$$

If $h = 1$, Raabe's test fails. However, if we return to Kummer's test and use $a_n = n\ln n$, Eq. (5.38) leads to

$$\begin{aligned}\lim_{n\to\infty}&\left\{n\ln n\left[1+\frac{1}{n}+\frac{B(n)}{n^2}\right]-(n+1)\ln(n+1)\right\}\\ &=\lim_{n\to\infty}\left[n\ln n\cdot\frac{n+1}{n}-(n+1)\ln(n+1)\right]\\ &=\lim_{n\to\infty}(n+1)\left[\ln n-\ln n-\ln\left(1+\frac{1}{n}\right)\right].\end{aligned} \tag{5.46}$$

Borrowing a result from Section 5.6 (which is not dependent on Gauss' test), we have

$$\begin{aligned}\lim_{n\to\infty}-(n+1)\ln\left(1+\frac{1}{n}\right)&=\lim_{n\to\infty}-(n+1)\left(\frac{1}{n}-\frac{1}{2n^2}+\frac{1}{3n^3}\cdots\right)\\ &=-1<0.\end{aligned} \tag{5.47}$$

Hence we have divergence for $h = 1$. This is an example of a successful application of Kummer's test when Raabe's test had failed.

Example 5.2.4 LEGENDRE SERIES

The recurrence relation for the series solution of Legendre's equation (Exercise 9.5.5) may be put in the form

$$\frac{a_{2j+2}}{a_{2j}}=\frac{2j(2j+1)-l(l+1)}{(2j+1)(2j+2)}. \tag{5.48}$$

For $u_j = a_{2j}$ and $B(j) = O(1/j^2) \to 0$ (that is, $|B(j)j^2| \le C$, $C > 0$, a constant) as $j \to \infty$ in Gauss' test we apply Eq. (5.45). Then, for $j \gg l$,[6]

$$\frac{u_j}{u_{j+1}}\to\frac{(2j+1)(2j+2)}{2j(2j+1)}=\frac{2j+2}{2j}=1+\frac{1}{j}. \tag{5.49}$$

By Eq. (5.44) the series is divergent. ■

[6]The l dependence enters $B(j)$ but does not affect h in Eq. (5.45).

Improvement of Convergence

This section so far has been concerned with establishing convergence as an abstract mathematical property. In practice, the **rate** of convergence may be of considerable importance. Here we present one method of improving the rate of convergence of a convergent series. Other techniques are given in Sections 5.4 and 5.9.

The basic principle of this method, due to Kummer, is to form a linear combination of our slowly converging series and one or more series whose sum is known. For the known series the collection

$$\alpha_1 = \sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1$$

$$\alpha_2 = \sum_{n=1}^{\infty} \frac{1}{n(n+1)(n+2)} = \frac{1}{4}$$

$$\alpha_3 = \sum_{n=1}^{\infty} \frac{1}{n(n+1)(n+2)(n+3)} = \frac{1}{18}$$

$$\alpha_p = \sum_{n=1}^{\infty} \frac{1}{n(n+1)\cdots(n+p)} = \frac{1}{p \cdot p!}$$

is particularly useful.[7] The series are combined term by term and the coefficients in the linear combination chosen to cancel the most slowly converging terms.

Example 5.2.5 RIEMANN ZETA FUNCTION, $\zeta(3)$

Let the series to be summed be $\sum_{n=1}^{\infty} n^{-3}$. In Section 5.9 this is identified as the Riemann zeta function, $\zeta(3)$. We form a linear combination

$$\sum_{n=1}^{\infty} n^{-3} + a_2\alpha_2 = \sum_{n=1}^{\infty} n^{-3} + \frac{a_2}{4}.$$

α_1 is not included since it converges more slowly than $\zeta(3)$. Combining terms, we obtain on the left-hand side

$$\sum_{n=1}^{\infty} \left\{ \frac{1}{n^3} + \frac{a_2}{n(n+1)(n+2)} \right\} = \sum_{n=1}^{\infty} \frac{n^2(1+a_2) + 3n + 2}{n^3(n+1)(n+2)}.$$

If we choose $a_2 = -1$, the preceding equations yield

$$\zeta(3) = \sum_{n=1}^{\infty} n^{-3} = \frac{1}{4} + \sum_{n=1}^{\infty} \frac{3n+2}{n^3(n+1)(n+2)}. \tag{5.50}$$

[7] These series sums may be verified by expanding the forms by partial fractions, writing out the initial terms, and inspecting the pattern of cancellation of positive and negative terms.

The resulting series may not be beautiful but it does converge as n^{-4}, faster than n^{-3}. A more convenient form comes from Exercise 5.2.21. There, the symmetry leads to convergence as n^{-5}. ■

The method can be extended, including $a_3\alpha_3$ to get convergence as n^{-5}, $a_4\alpha_4$ to get convergence as n^{-6}, and so on. Eventually, you have to reach a compromise between how much algebra you do and how much arithmetic the computer does. As computers get faster, the balance is steadily shifting to less algebra for you and more arithmetic for them.

Exercises

5.2.1 (a) Prove that if

$$\lim_{n\to\infty} n^p u_n = A < \infty, \qquad p > 1,$$

the series $\sum_{n=1}^{\infty} u_n$ converges.

(b) Prove that if

$$\lim_{n\to\infty} n u_n = A > 0,$$

the series diverges. (The test fails for $A = 0$.)

These two tests, known as **limit tests**, are often convenient for establishing the convergence of a series. They may be treated as comparison tests, comparing with

$$\sum_n n^{-q}, \qquad 1 \le q < p.$$

5.2.2 If

$$\lim_{n\to\infty} \frac{b_n}{a_n} = K,$$

a constant with $0 < K < \infty$, show that $\sum_n b_n$ converges or diverges with $\sum a_n$.
Hint. If $\sum a_n$ converges, use $b'_n = \frac{1}{2K} b_n$. If $\sum_n a_n$ diverges, use $b''_n = \frac{2}{K} b_n$.

5.2.3 Show that the complete d'Alembert ratio test follows directly from Kummer's test with $a_i = 1$.

5.2.4 Show that Raabe's test is indecisive for $P = 1$ by establishing that $P = 1$ for the series

(a) $u_n = \dfrac{1}{n \ln n}$ and that this series diverges.

(b) $u_n = \dfrac{1}{n(\ln n)^2}$ and that this series converges.

Note. By direct addition $\sum_2^{100,000} [n(\ln n)^2]^{-1} = 2.02288$. The remainder of the series $n > 10^5$ yields 0.08686 by the integral comparison test. The total, then, 2 to ∞, is 2.1097.

5.2.5 Gauss' test is often given in the form of a test of the ratio

$$\frac{u_n}{u_{n+1}} = \frac{n^2 + a_1 n + a_0}{n^2 + b_1 n + b_0}.$$

For what values of the parameters a_1 and b_1 is there convergence? divergence?

ANS. Convergent for $a_1 - b_1 > 1$,
divergent for $a_1 - b_1 \leq 1$.

5.2.6 Test for convergence

(a) $\displaystyle\sum_{n=2}^{\infty} (\ln n)^{-1}$ (d) $\displaystyle\sum_{n=1}^{\infty} \big[n(n+1)\big]^{-1/2}$

(b) $\displaystyle\sum_{n=1}^{\infty} \frac{n!}{10^n}$ (e) $\displaystyle\sum_{n=0}^{\infty} \frac{1}{2n+1}$.

(c) $\displaystyle\sum_{n=1}^{\infty} \frac{1}{2n(2n+1)}$

5.2.7 Test for convergence

(a) $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$ (d) $\displaystyle\sum_{n=1}^{\infty} \ln\left(1 + \frac{1}{n}\right)$

(b) $\displaystyle\sum_{n=2}^{\infty} \frac{1}{n \ln n}$ (e) $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n \cdot n^{1/n}}$.

(c) $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n 2^n}$

5.2.8 For what values of p and q will the following series converge? $\sum_{n=2}^{\infty} \frac{1}{n^p (\ln n)^q}$.

$$\text{ANS. Convergent for } \begin{cases} p > 1, & \text{all } q, \\ p = 1, & q > 1, \end{cases} \qquad \text{divergent for } \begin{cases} p < 1, & \text{all } q, \\ p = 1, & q \leq 1. \end{cases}$$

5.2.9 Determine the range of convergence for Gauss's **hypergeometric series**

$$F(\alpha, \beta, \gamma; x) = 1 + \frac{\alpha\beta}{1!\gamma} x + \frac{\alpha(\alpha+1)\beta(\beta+1)}{2!\gamma(\gamma+1)} x^2 + \cdots.$$

Hint. Gauss developed his test for the specific purpose of establishing the convergence of this series.

ANS. Convergent for $-1 < x < 1$ and $x = \pm 1$ if $\gamma > \alpha + \beta$.

5.2.10 A pocket calculator yields

$$\sum_{n=1}^{100} n^{-3} = 1.202\ 007.$$

Show that

$$1.202\,056 \le \sum_{n=1}^{\infty} n^{-3} \le 1.202\,057.$$

Hint. Use integrals to set upper and lower bounds on $\sum_{n=101}^{\infty} n^{-3}$.
Note. A more exact value for summation of $\zeta(3) = \sum_{n=1}^{\infty} n^{-3}$ is 1.202 056 903...; $\zeta(3)$ is known to be an irrational number, but it is not linked to known constants such as $e, \pi, \gamma, \ln 2$.

5.2.11 Set upper and lower bounds on $\sum_{n=1}^{1,000,000} n^{-1}$, assuming that

(a) the Euler–Mascheroni constant is known.

$$\text{ANS. } 14.392\,726 < \sum_{n=1}^{1,000,000} n^{-1} < 14.392\,727.$$

(b) The Euler–Mascheroni constant is unknown.

5.2.12 Given $\sum_{n=1}^{1,000} n^{-1} = 7.485\,470\ldots$ set upper and lower bounds on the Euler–Mascheroni constant.

$$\text{ANS. } 0.5767 < \gamma < 0.5778.$$

5.2.13 (From **Olbers' paradox**.) Assume a static universe in which the stars are uniformly distributed. Divide all space into shells of constant thickness; the stars in any one shell by themselves subtend a solid angle of ω_0. **Allowing for the blocking out of distant stars by nearer stars**, show that the total net solid angle subtended by all stars, shells extending to infinity, is **exactly** 4π. [Therefore the night sky should be ablaze with light. For more details, see E. Harrison, *Darkness at Night: A Riddle of the Universe*. Cambridge, MA: Harvard University Press (1987).]

5.2.14 Test for convergence

$$\sum_{n=1}^{\infty} \left[\frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots (2n)} \right]^2 = \frac{1}{4} + \frac{9}{64} + \frac{25}{256} + \cdots.$$

5.2.15 The Legendre series $\sum_{j \text{ even}} u_j(x)$ satisfies the recurrence relations

$$u_{j+2}(x) = \frac{(j+1)(j+2) - l(l+1)}{(j+2)(j+3)} x^2 u_j(x),$$

in which the index j is even and l is some constant (but, in this problem, **not** a non-negative odd integer). Find the range of values of x for which this Legendre series is convergent. Test the endpoints.

$$\text{ANS. } -1 < x < 1.$$

5.2.16 A series solution (Section 9.5) of the Chebyshev equation leads to successive terms having the ratio

$$\frac{u_{j+2}(x)}{u_j(x)} = \frac{(k+j)^2 - n^2}{(k+j+1)(k+j+2)} x^2,$$

with $k = 0$ and $k = 1$. Test for convergence at $x = \pm 1$.

ANS. Convergent.

5.2.17 A series solution for the ultraspherical (Gegenbauer) function $C_n^{\alpha}(x)$ leads to the recurrence

$$a_{j+2} = a_j \frac{(k+j)(k+j+2\alpha) - n(n+2\alpha)}{(k+j+1)(k+j+2)}.$$

Investigate the convergence of each of these series at $x = \pm 1$ as a function of the parameter α.

ANS. Convergent for $\alpha < 1$,
divergent for $\alpha \geq 1$.

5.2.18 A series expansion of the incomplete beta function (Section 8.4) yields

$$B_x(p,q) = x^p \left\{ \frac{1}{p} + \frac{1-q}{p+1} x + \frac{(1-q)(2-q)}{2!(p+2)} x^2 + \cdots \right.$$
$$\left. + \frac{(1-q)(2-q)\cdots(n-q)}{n!(p+n)} x^n + \cdots \right\}.$$

Given that $0 \leq x \leq 1$, $p > 0$, and $q > 0$, test this series for convergence. What happens at $x = 1$?

5.2.19 Show that the following series is convergent.

$$\sum_{s=0}^{\infty} \frac{(2s-1)!!}{(2s)!!(2s+1)}.$$

Note. $(2s-1)!! = (2s-1)(2s-3)\cdots 3\cdot 1$ with $(-1)!! = 1$; $(2s)!! = (2s)(2s-2)\cdots 4\cdot 2$ with $0!! = 1$. The series appears as a series expansion of $\sin^{-1}(1)$ and equals $\pi/2$, and $\sin^{-1} x \equiv \arcsin x \neq (\sin x)^{-1}$.

5.2.20 Show how to combine $\zeta(2) = \sum_{n=1}^{\infty} n^{-2}$ with α_1 and α_2 to obtain a series converging as n^{-4}.
Note. $\zeta(2)$ is known: $\zeta(2) = \pi^2/6$ (see Section 5.9).

5.2.21 The convergence improvement of Example 5.2.5 may be carried out more expediently (in this special case) by putting α_2 into a more symmetric form: Replacing n by $n-1$, we have

$$\alpha_2' = \sum_{n=2}^{\infty} \frac{1}{(n-1)n(n+1)} = \frac{1}{4}.$$

(a) Combine $\zeta(3)$ and α_2' to obtain convergence as n^{-5}.
(b) Let α_4' be α_4 with $n \to n-2$. Combine $\zeta(3)$, α_2', and α_4' to obtain convergence as n^{-7}.
(c) If $\zeta(3)$ is to be calculated to six = decimal = place accuracy (error 5×10^{-7}), how many terms are required for $\zeta(3)$ alone? combined as in part (a)? combined as in part (b)?

Note. The error may be estimated using the corresponding integral.

ANS. (a) $\zeta(3) = \dfrac{5}{4} - \displaystyle\sum_{n=2}^{\infty} \frac{1}{n^3(n^2-1)}$.

5.2.22 Catalan's constant ($\beta(2)$ of M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables (AMS-55), Wash, D. C. National Bureau of Standards (1972); reprinted Dover (1974), Chapter 23) is defined by

$$\beta(2) = \sum_{k=0}^{\infty} (-1)^k (2k+1)^{-2} = \frac{1}{1^2} - \frac{1}{3^2} + \frac{1}{5^2} \cdots .$$

Calculate $\beta(2)$ to six-digit accuracy.
Hint. The rate of convergence is enhanced by pairing the terms:

$$(4k-1)^{-2} - (4k+1)^{-2} = \frac{16k}{(16k^2-1)^2}.$$

If you have carried enough digits in your series summation, $\sum_{1 \le k \le N} 16k/(16k^2-1)^2$, additional significant figures may be obtained by setting upper and lower bounds on the tail of the series, $\sum_{k=N+1}^{\infty}$. These bounds may be set by comparison with integrals, as in the Maclaurin integral test.

ANS. $\beta(2) = 0.9159\ 6559\ 4177 \ldots$.

5.3 ALTERNATING SERIES

In Section 5.2 we limited ourselves to series of positive terms. Now, in contrast, we consider infinite series in which the signs alternate. The partial cancellation due to alternating signs makes convergence more rapid and much easier to identify. We shall prove the Leibniz criterion, a general condition for the convergence of an alternating series. For series with more irregular sign changes, like Fourier series of Chapter 14 (see Example 5.3.1), the integral test of Eq. (5.25) is often helpful.

Leibniz Criterion

Consider the series $\sum_{n=1}^{\infty} (-1)^{n+1} a_n$ with $a_n > 0$. If a_n, is **monotonically decreasing** (for sufficiently large n) and $\lim_{n\to\infty} a_n = 0$, then the series converges. To prove this theorem,

we examine the even partial sums

$$\begin{aligned} s_{2n} &= a_1 - a_2 + a_3 - \cdots - a_{2n}, \\ s_{2n+2} &= s_{2n} + (a_{2n+1} - a_{2n+2}). \end{aligned} \tag{5.51}$$

Since $a_{2n+1} > a_{2n+2}$, we have

$$s_{2n+2} > s_{2n}. \tag{5.52}$$

On the other hand,

$$s_{2n+2} = a_1 - (a_2 - a_3) - (a_4 - a_5) - \cdots - a_{2n+2}. \tag{5.53}$$

Hence, with each pair of terms $a_{2p} - a_{2p+1} > 0$,

$$s_{2n+2} < a_1. \tag{5.54}$$

With the even partial sums bounded $s_{2n} < s_{2n+2} < a_1$ and the terms a_n decreasing monotonically and approaching zero, this alternating series converges.

One further important result can be extracted from the partial sums of the same alternating series. From the difference between the series limit S and the partial sum s_n,

$$\begin{aligned} S - s_n &= a_{n+1} - a_{n+2} + a_{n+3} - a_{n+4} + \cdots \\ &= a_{n+1} - (a_{n+2} - a_{n+3}) - (a_{n+4} - a_{n+5}) - \cdots, \end{aligned} \tag{5.55}$$

or

$$S - s_n < a_{n+1}. \tag{5.56}$$

Equation (5.56) says that the error in cutting off an alternating series whose terms are monotonically decreasing after n terms is less than a_{n+1}, the first term dropped. A knowledge of the error obtained this way may be of great practical importance.

Absolute Convergence

Given a series of terms u_n in which u_n may vary in sign, if $\sum |u_n|$ converges, then $\sum u_n$ is said to be absolutely convergent. If $\sum u_n$ converges but $\sum |u_n|$ diverges, the convergence is called **conditional**.

The alternating harmonic series is a simple example of this conditional convergence. We have

$$\sum_{n=1}^{\infty} (-1)^{n-1} n^{-1} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots + \frac{(-1)^{n-1}}{n} + \cdots, \tag{5.57}$$

convergent by the Leibniz criterion; but

$$\sum_{n=1}^{\infty} n^{-1} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} + \cdots$$

has been shown to be divergent in Sections 5.1 and 5.2.

Note that most tests developed in Section 5.2 assume a series of positive terms. Therefore these tests in that section guarantee absolute convergence.

Example 5.3.1 SERIES WITH IRREGULAR SIGN CHANGES

For $0 < x < 2\pi$ the Fourier series (see Chapter 14.1)

$$\sum_{n=1}^{\infty} \frac{\cos(nx)}{n} = -\ln\left(2\sin\frac{x}{2}\right) \tag{5.58}$$

converges, having coefficients that change sign often, but not so that the Leibniz convergence criterion applies easily. Let us apply the integral test of Eq. (5.22). Using integration by parts we see immediately that

$$\int_1^{\infty} \frac{\cos(nx)}{n}\,dn = \left[\frac{\sin(nx)}{nx}\right]_1^{\infty} + \frac{1}{x}\int_{n=1}^{\infty} \frac{\sin(nx)}{n^2}\,dn$$

converges, and the integral on the right-hand side even converges absolutely. The derivative term in Eq. (5.22) has the form

$$\int_1^{\infty} \big(n - [n]\big)\left\{-\frac{x}{n}\sin(nx) - \frac{\cos(nx)}{n^2}\right\} dn,$$

where the second term converges absolutely and need not be considered further. Next we observe that $g(N) = \int_1^N (n - [n])\sin(nx)\,dn$ is bounded for $N \to \infty$, just as $\int^N \sin(nx)\,dn$ is bounded because of the periodic nature of $\sin(nx)$ and its regular sign changes. Using integration by parts again,

$$\int_1^{\infty} \frac{g'(n)}{n}\,dn = \left[\frac{g(n)}{n}\right]_{n=1}^{\infty} + \int_1^{\infty} \frac{g(n)}{n^2}\,dn,$$

we see that the second term is absolutely convergent and that the first goes to zero at the upper limit. Hence the series in Eq. (5.58) converges, which is hard to see from other convergence tests.

Alternatively, we may apply the $q = 1$ case of the Euler–Maclaurin integration formula in Eq. (5.168b),

$$\sum_{\nu=1}^{n} f(\nu) = \int_1^n f(x)\,dx + \frac{1}{2}\big\{f(n) + f(1)\big\} + \frac{1}{12}\big\{f'(n) - f'(1)\big\}$$
$$- \frac{1}{2}\int_0^1 \left(x^2 - x + \frac{1}{6}\right)\sum_{\nu=1}^{n-1} f''(x+\nu)\,dx,$$

which is straightforward but more tedious because of the second derivative. ■

Exercises

5.3.1 (a) From the electrostatic two-hemisphere problem (Exercise 12.3.20) we obtain the series

$$\sum_{s=0}^{\infty}(-1)^s(4s+3)\frac{(2s-1)!!}{(2s+2)!!}.$$

Test it for convergence.

(b) The corresponding series for the surface charge density is

$$\sum_{s=0}^{\infty}(-1)^s(4s+3)\frac{(2s-1)!!}{(2s)!!}.$$

Test it for convergence.

The !! notation is explained in Section 8.1 and Exercise 5.2.19.

5.3.2 Show by direct numerical computation that the sum of the first 10 terms of

$$\lim_{x\to 1}\ln(1+x)=\ln 2=\sum_{n=1}^{\infty}(-1)^{n-1}n^{-1}$$

differs from $\ln 2$ by less than the eleventh term: $\ln 2 = 0.69314\,71806\ldots$.

5.3.3 In Exercise 5.2.9 the hypergeometric series is shown convergent for $x=\pm 1$, if $\gamma > \alpha+\beta$. Show that there is conditional convergence for $x=-1$ for γ down to $\gamma > \alpha+\beta-1$.

Hint. The asymptotic behavior of the factorial function is given by Stirling's series, Section 8.3.

5.4 Algebra of Series

The establishment of absolute convergence is important because it can be proved that absolutely convergent series may be reordered according to the familiar rules of algebra or arithmetic.

- If an infinite series is absolutely convergent, the series sum is independent of the order in which the terms are added.
- The series may be multiplied with another absolutely convergent series. The limit of the product will be the product of the individual series limits. The product series, a double series, will also converge absolutely.

No such guarantees can be given for conditionally convergent series. Again consider the alternating harmonic series. If we write

$$1-\tfrac{1}{2}+\tfrac{1}{3}-\tfrac{1}{4}+\cdots=1-\left(\tfrac{1}{2}-\tfrac{1}{3}\right)-\left(\tfrac{1}{4}-\tfrac{1}{5}\right)-\cdots, \tag{5.59}$$

it is clear that the sum

$$\sum_{n=1}^{\infty}(-1)^{n-1}n^{-1} < 1. \tag{5.60}$$

However, if we rearrange the terms slightly, we may make the alternating harmonic series converge to $\frac{3}{2}$. We regroup the terms of Eq. (5.59), taking

$$\begin{aligned}&\left(1+\tfrac{1}{3}+\tfrac{1}{5}\right)-\left(\tfrac{1}{2}\right)+\left(\tfrac{1}{7}+\tfrac{1}{9}+\tfrac{1}{11}+\tfrac{1}{13}+\tfrac{1}{15}\right)-\left(\tfrac{1}{4}\right)\\&\quad+\left(\tfrac{1}{17}+\cdots+\tfrac{1}{25}\right)-\left(\tfrac{1}{6}\right)+\left(\tfrac{1}{27}+\cdots+\tfrac{1}{35}\right)-\left(\tfrac{1}{8}\right)+\cdots.\end{aligned} \tag{5.61}$$

Treating the terms grouped in parentheses as single terms for convenience, we obtain the partial sums

$$\begin{aligned}s_1 &= 1.5333 & s_2 &= 1.0333\\ s_3 &= 1.5218 & s_4 &= 1.2718\\ s_5 &= 1.5143 & s_6 &= 1.3476\\ s_7 &= 1.5103 & s_8 &= 1.3853\\ s_9 &= 1.5078 & s_{10} &= 1.4078.\end{aligned}$$

From this tabulation of s_n and the plot of s_n versus n in Fig. 5.3, the convergence to $\frac{3}{2}$ is fairly clear. We have rearranged the terms, taking positive terms until the partial sum was equal to or greater than $\frac{3}{2}$ and then adding in negative terms until the partial sum just fell below $\frac{3}{2}$ and so on. As the series extends to infinity, all original terms will eventually appear, but the partial sums of this rearranged alternating harmonic series converge to $\frac{3}{2}$.

By a suitable rearrangement of terms, a conditionally convergent series may be made to converge to any desired value or even to diverge. This statement is sometimes given

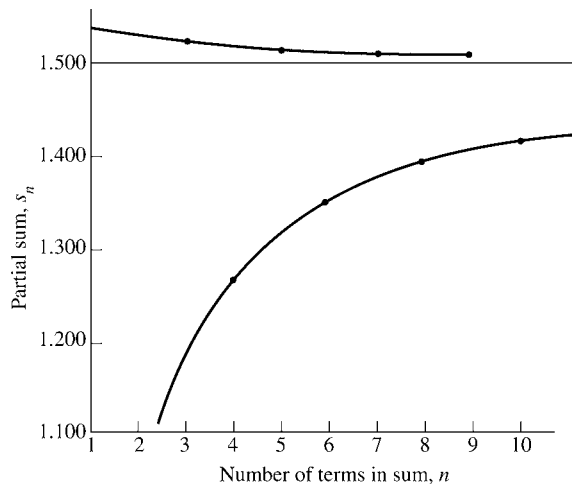

FIGURE 5.3 Alternating harmonic series — terms rearranged to give convergence to 1.5.

as **Riemann's theorem**. Obviously, conditionally convergent series must be treated with caution.

Absolutely convergent series can be multiplied without problems. This follows as a special case from the rearrangement of double series. However, conditionally convergent series cannot always be multiplied to yield convergent series, as the following example shows.

Example 5.4.1 SQUARE OF A CONDITIONALLY CONVERGENT SERIES MAY DIVERGE

The series $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{\sqrt{n}}$ converges, by the Leibniz criterion. Its square,

$$\left[\sum_n \frac{(-1)^{n-1}}{\sqrt{n}}\right]^2 = \sum_n (-1)^n \left[\frac{1}{\sqrt{1}}\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{2}}\frac{1}{\sqrt{n-2}} + \cdots + \frac{1}{\sqrt{n-1}}\frac{1}{\sqrt{1}}\right],$$

has the general term in brackets consisting of $n-1$ additive terms, each of which is greater than $\frac{1}{\sqrt{n-1}\sqrt{n-1}}$, so the product term in brackets is greater than $\frac{n-1}{n-1}$ and does not go to zero. Hence this product oscillates and therefore diverges. ■

Hence for a product of two series to converge, we have to demand as a sufficient condition that at least one of them converge absolutely. To prove this **product convergence theorem** that if $\sum_n u_n$ converges absolutely to U, $\sum_n v_n$ converges to V, then

$$\sum_n c_n, \qquad c_n = \sum_{m=0}^{n} u_m v_{n-m}$$

converges to UV, it is sufficient to show that the difference terms $D_n \equiv c_0 + c_1 + \cdots + c_{2n} - U_n V_n \to 0$ for $n \to \infty$, where U_n, V_n are the partial sums of our series. As a result, the partial sum differences

$$\begin{aligned} D_n &= u_0 v_0 + (u_0 v_1 + u_1 v_0) + \cdots + (u_0 v_{2n} + u_1 v_{2n-1} + \cdots + u_{2n} v_0) \\ &\quad - (u_0 + u_1 + \cdots + u_n)(v_0 + v_1 + \cdots + v_n) \\ &= u_0(v_{n+1} + \cdots + v_{2n}) + u_1(v_{n+1} + \cdots + v_{2n-1}) + \cdots + u_{n+1} v_{n+1} \\ &\quad + v_{n+1}(v_0 + \cdots + v_{n-1}) + \cdots + u_{2n} v_0, \end{aligned}$$

so for all sufficiently large n,

$$|D_n| < \epsilon\big(|u_0| + \cdots + |u_{n-1}|\big) + M\big(|u_{n+1}| + \cdots + |u_{2n}|\big) < \epsilon(a + M),$$

because $|v_{n+1} + v_{n+2} + \cdots + v_{n+m}| < \epsilon$ for sufficiently large n and all positive integers m as $\sum v_n$ converges, and the partial sums $V_n < B$ of $\sum_n v_n$ are bounded by M, because the sum converges. Finally we call $\sum_n |u_n| = a$, as $\sum u_n$ converges absolutely.

Two series can be multiplied, provided one of them converges absolutely. Addition and subtraction of series is also valid termwise if one series converges absolutely.

Improvement of Convergence, Rational Approximations

The series

$$\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n}, \qquad -1 < x \le 1, \tag{5.61a}$$

converges very slowly as x approaches $+1$. The **rate** of convergence may be improved substantially by multiplying both sides of Eq. (5.61a) by a polynomial and adjusting the polynomial coefficients to cancel the more slowly converging portions of the series. Consider the simplest possibility: Multiply $\ln(1+x)$ by $1 + a_1 x$:

$$(1 + a_1 x)\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n} + a_1 \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^{n+1}}{n}.$$

Combining the two series on the right, term by term, we obtain

$$\begin{aligned}(1 + a_1 x)\ln(1+x) &= x + \sum_{n=2}^{\infty} (-1)^{n-1} \left(\frac{1}{n} - \frac{a_1}{n-1} \right) x^n \\ &= x + \sum_{n=2}^{\infty} (-1)^{n-1} \frac{n(1-a_1) - 1}{n(n-1)} x^n.\end{aligned}$$

Clearly, if we take $a_1 = 1$, the n in the numerator disappears and our combined series converges as n^{-2}.

Continuing this process, we find that $(1 + 2x + x^2)\ln(1+x)$ vanishes as n^{-3} and that $(1 + 3x + 3x^2 + x^3)\ln(1+x)$ vanishes as n^{-4}. In effect we are shifting from a simple series expansion of Eq. (5.61a) to a rational fraction representation in which the function $\ln(1+x)$ is represented by the ratio of a series and a polynomial:

$$\ln(1+x) = \frac{x + \sum_{n=2}^{\infty} (-1)^n x^n / [n(n-1)]}{1+x}.$$

Such rational approximations may be both compact and accurate.

Rearrangement of Double Series

Another aspect of the rearrangement of series appears in the treatment of double series (Fig. 5.4):

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{n,m}.$$

Let us substitute

$$n = q \ge 0, \qquad m = p - q \ge 0 \qquad (q \le p).$$

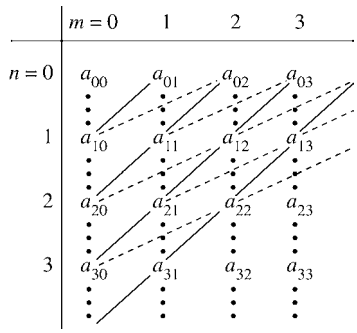

FIGURE 5.4 Double series — summation over n indicated by vertical dashed lines.

This results in the identity

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} a_{n,m} = \sum_{p=0}^{\infty}\sum_{q=0}^{p} a_{q,p-q}. \tag{5.62}$$

The summation over p and q of Eq. (5.62) is illustrated in Fig. 5.5. The substitution

$$n = s \geq 0, \qquad m = r - 2s \geq 0 \qquad \left(s \leq \frac{r}{2}\right)$$

leads to

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} a_{n,m} = \sum_{r=0}^{\infty}\sum_{s=0}^{[r/2]} a_{s,r-2s}, \tag{5.63}$$

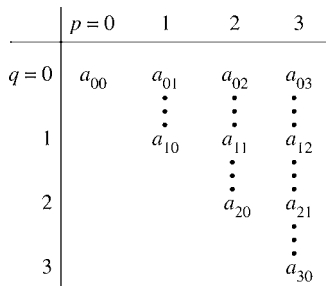

FIGURE 5.5 Double series — again, the first summation is represented by vertical dashed lines, but these vertical lines correspond to diagonals in Fig. 5.4.

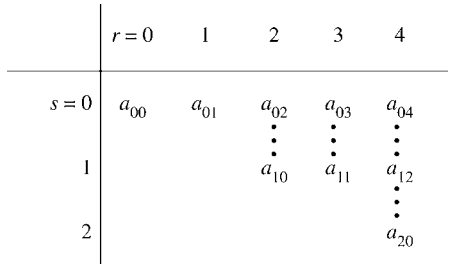

FIGURE 5.6 Double series. The summation over s corresponds to a summation along the almost-horizontal dashed lines in Fig. 5.4.

with $[r/2] = r/2$ for r even and $(r-1)/2$ for r odd. The summation over r and s of Eq. (5.63) is shown in Fig. 5.6. Equations (5.62) and (5.63) are clearly rearrangements of the array of coefficients a_{nm}, rearrangements that are valid as long as we have absolute convergence.

The combination of Eqs. (5.62) and (5.63),

$$\sum_{p=0}^{\infty}\sum_{q=0}^{p} a_{q,p-q} = \sum_{r=0}^{\infty}\sum_{s=0}^{[r/2]} a_{s,r-2s}, \tag{5.64}$$

is used in Section 12.1 in the determination of the series form of the Legendre polynomials.

Exercises

5.4.1 Given the series (derived in Section 5.6)

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4}\cdots, \qquad -1 < x \le 1,$$

show that

$$\ln\left(\frac{1+x}{1-x}\right) = 2\left(x + \frac{x^3}{3} + \frac{x^5}{5} + \cdots\right), \qquad -1 < x < 1.$$

The original series, $\ln(1+x)$, appears in an analysis of binding energy in crystals. It is $\frac{1}{2}$ the Madelung constant $(2\ln 2)$ for a chain of atoms. The second series is useful in normalizing the Legendre polynomials (Section 12.3) and in developing a second solution for Legendre's differential equation (Section 12.10).

5.4.2 Determine the values of the coefficients a_1, a_2, and a_3 that will make $(1 + a_1x + a_2x^2 + a_3x^3)\ln(1+x)$ converge as n^{-4}. Find the resulting series.

5.4.3 Show that

$$\text{(a)} \sum_{n=2}^{\infty}\big[\zeta(n) - 1\big] = 1, \qquad \text{(b)} \sum_{n=2}^{\infty}(-1)^n\big[\zeta(n) - 1\big] = \tfrac{1}{2},$$

where $\zeta(n)$ is the Riemann zeta function.

5.4.4 Write a program that will rearrange the terms of the alternating harmonic series to make the series converge to 1.5. Group your terms as indicated in Eq. (5.61). List the first 100 successive partial sums that just climb above 1.5 or just drop below 1.5, and list the new terms included in each such partial sum.

ANS.

n	1	2	3	4	5
s_n	1.5333	1.0333	1.5218	1.2718	1.5143

5.5 SERIES OF FUNCTIONS

We extend our concept of infinite series to include the possibility that each term u_n may be a function of some variable, $u_n = u_n(x)$. Numerous illustrations of such series of functions appear in Chapters 11–14. The partial sums become functions of the variable x,

$$s_n(x) = u_1(x) + u_2(x) + \cdots + u_n(x), \tag{5.65}$$

as does the series sum, defined as the limit of the partial sums:

$$\sum_{n=1}^{\infty} u_n(x) = S(x) = \lim_{n\to\infty} s_n(x). \tag{5.66}$$

So far we have concerned ourselves with the behavior of the partial sums as a function of n. Now we consider how the foregoing quantities depend on x. The key concept here is that of uniform convergence.

Uniform Convergence

If for any small $\varepsilon > 0$ there exists a number N, **independent of** x in the interval $[a, b]$ (that is, $a \le x \le b$) such that

$$|S(x) - s_n(x)| < \varepsilon, \qquad \text{for all} \quad n \ge N, \tag{5.67}$$

then the series is said to be uniformly convergent in the interval $[a, b]$. This says that for our series to be uniformly convergent, it must be possible to find a finite N so that the tail of the infinite series, $|\sum_{i=N+1}^{\infty} u_i(x)|$, will be less than an arbitrarily small ε for all x in the given interval.

This condition, Eq. (5.67), which defines uniform convergence, is illustrated in Fig. 5.7. The point is that no matter how small ε is taken to be, we can always choose n large enough so that the absolute magnitude of the difference between $S(x)$ and $s_n(x)$ is less than ε for all x, $a \le x \le b$. If this cannot be done, then $\sum u_n(x)$ is not uniformly convergent in $[a, b]$.

Example 5.5.1 NONUNIFORM CONVERGENCE

$$\sum_{n=1}^{\infty} u_n(x) = \sum_{n=1}^{\infty} \frac{x}{[(n-1)x + 1][nx + 1]}. \tag{5.68}$$

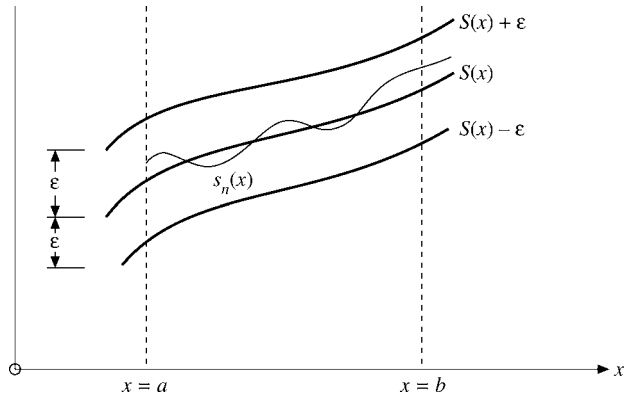

FIGURE 5.7 Uniform convergence.

The partial sum $s_n(x) = nx(nx+1)^{-1}$, as may be verified by **mathematical induction**. By inspection this expression for $s_n(x)$ holds for $n = 1, 2$. We assume it holds for n terms and then prove it holds for $n+1$ terms:

$$\begin{aligned} s_{n+1}(x) &= s_n(x) + \frac{x}{[nx+1][(n+1)x+1]} \\ &= \frac{nx}{[nx+1]} + \frac{x}{[nx+1][(n+1)x+1]} \\ &= \frac{(n+1)x}{(n+1)x+1}, \end{aligned}$$

completing the proof.

Letting n approach infinity, we obtain

$$\begin{aligned} S(0) &= \lim_{n\to\infty} s_n(0) = 0, \\ S(x \neq 0) &= \lim_{n\to\infty} s_n(x \neq 0) = 1. \end{aligned}$$

We have a discontinuity in our series limit at $x = 0$. However, $s_n(x)$ is a continuous function of $x, 0 \leq x \leq 1$, for all finite n. No matter how small ε may be, Eq. (5.67) will be violated for all sufficiently small x. Our series does not converge uniformly. ■

Weierstrass *M* (Majorant) Test

The most commonly encountered test for uniform convergence is the Weierstrass M test. If we can construct a series of numbers $\sum_1^\infty M_i$, in which $M_i \geq |u_i(x)|$ for all x in the interval $[a, b]$ and $\sum_1^\infty M_i$ is convergent, our series $u_i(x)$ will be **uniformly** convergent in $[a, b]$.

The proof of this Weierstrass M test is direct and simple. Since $\sum_i M_i$ converges, some number N exists such that for $n+1 \geq N$,

$$\sum_{i=n+1}^{\infty} M_i < \varepsilon. \tag{5.69}$$

This follows from our definition of convergence. Then, with $|u_i(x)| \leq M_i$ for all x in the interval $a \leq x \leq b$,

$$\left|\sum_{i=n+1}^{\infty} u_i(x)\right| \leq \sum_{i=n+1}^{\infty} |u_i(x)| < \varepsilon. \tag{5.70}$$

Hence

$$\left|\sum_{n=1}^{\infty} (-1)^n \frac{1}{\sqrt{n+x^2}}\right| = \left|S(x) - s_n(x)\right| < \varepsilon, \tag{5.71}$$

and by definition $\sum_{i=1}^{\infty} u_i(x)$ is uniformly convergent in $[a, b]$. Since we have specified absolute values in the statement of the Weierstrass M test, the series $\sum_{i=1}^{\infty} u_i(x)$ is also seen to be **absolutely** convergent.

Note that uniform convergence and absolute convergence are independent properties. Neither implies the other. For specific examples,

$$\sum_{n=1}^{\infty} (-1)^n \frac{1}{\sqrt{n+x^2}}, \qquad -\infty < x < \infty, \tag{5.72}$$

and

$$\sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n} = \ln(1+x), \qquad 0 \leq x \leq 1, \tag{5.73}$$

converge uniformly in the indicated intervals but do not converge absolutely. On the other hand,

$$\sum_{n=0}^{\infty} (1-x)x^n = 1, \qquad 0 \leq x < 1$$
$$= 0, \qquad x = 1, \tag{5.74}$$

converges absolutely but does not converge uniformly in $[0, 1]$.

From the definition of uniform convergence we may show that any series

$$f(x) = \sum_{n=1}^{\infty} u_n(x) \tag{5.75}$$

cannot converge uniformly in any interval that includes a discontinuity of $f(x)$ if all $u_n(x)$ are continuous.

Since the Weierstrass M test establishes both uniform and absolute convergence, it will necessarily fail for series that are uniformly but conditionally convergent.

Abel's Test

A somewhat more delicate test for uniform convergence has been given by Abel. If

$$u_n(x) = a_n f_n(x),$$

$$\sum a_n = A, \quad \text{convergent}$$

and the functions $f_n(x)$ are monotonic $[f_{n+1}(x) \le f_n(x)]$ and bounded, $0 \le f_n(x) \le M$, for all x in $[a, b]$, then $\sum_n u_n(x)$ **converges uniformly** in $[a, b]$.

This test is especially useful in analyzing power series (compare Section 5.7). Details of the proof of Abel's test and other tests for uniform convergence are given in the Additional Readings listed at the end of this chapter.

Uniformly convergent series have three particularly useful properties.

1. If the individual terms $u_n(x)$ are continuous, the series sum

$$f(x) = \sum_{n=1}^{\infty} u_n(x) \tag{5.76}$$

is also continuous.

2. If the individual terms $u_n(x)$ are continuous, the series may be integrated term by term. The sum of the integrals is equal to the integral of the sum.

$$\int_a^b f(x)\,dx = \sum_{n=1}^{\infty} \int_a^b u_n(x)\,dx. \tag{5.77}$$

3. The derivative of the series sum $f(x)$ equals the sum of the individual term derivatives:

$$\frac{d}{dx} f(x) = \sum_{n=1}^{\infty} \frac{d}{dx} u_n(x), \tag{5.78}$$

provided the following conditions are satisfied:

$$u_n(x) \text{ and } \frac{du_n(x)}{dx} \text{ are continuous in } [a, b].$$

$$\sum_{n=1}^{\infty} \frac{du_n(x)}{dx} \text{ is uniformly convergent in } [a, b].$$

Term-by-term integration of a uniformly convergent series[8] requires only continuity of the individual terms. This condition is almost always satisfied in physical applications. Term-by-term differentiation of a series is often not valid because more restrictive conditions must be satisfied. Indeed, we shall encounter Fourier series in Chapter 14 in which term-by-term differentiation of a uniformly convergent series leads to a divergent series.

[8] Term-by-term integration may also be valid in the absence of uniform convergence.

Exercises

5.5.1 Find the range of **uniform** convergence of the Dirichlet series

(a) $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^x}$, (b) $\zeta(x) = \displaystyle\sum_{n=1}^{\infty} \frac{1}{n^x}$.

ANS. (a) $0 < s \le x < \infty$.
(b) $1 < s \le x < \infty$.

5.5.2 For what range of x is the geometric series $\sum_{n=0}^{\infty} x^n$ uniformly convergent?

ANS. $-1 < -s \le x \le s < 1$.

5.5.3 For what range of positive values of x is $\sum_{n=0}^{\infty} 1/(1+x^n)$
(a) convergent? (b) uniformly convergent?

5.5.4 If the series of the coefficients $\sum a_n$ and $\sum b_n$ are absolutely convergent, show that the Fourier series

$$\sum (a_n \cos nx + b_n \sin nx)$$

is **uniformly** convergent for $-\infty < x < \infty$.

5.6 TAYLOR'S EXPANSION

This is an expansion of a function into an infinite series of powers of a variable x or into a finite series plus a remainder term. The coefficients of the successive terms of the series involve the successive derivatives of the function. We have already used Taylor's expansion in the establishment of a physical interpretation of divergence (Section 1.7) and in other sections of Chapters 1 and 2. Now we derive the Taylor expansion.

We assume that our function $f(x)$ has a continuous nth derivative[9] in the interval $a \le x \le b$. Then, integrating this nth derivative n times,

$$\int_a^x f^{(n)}(x_1)\,dx_1 = f^{(n-1)}(x_1)\Big|_a^x = f^{(n-1)}(x) - f^{(n-1)}(a),$$

$$\begin{aligned}\int_a^x dx_2 \int_a^{x_2} dx_1 f^{(n)}(x_1) &= \int_a^x dx_2 \left[f^{(n-1)}(x_2) - f^{(n-1)}(a)\right] \\ &= f^{(n-2)}(x) - f^{(n-2)}(a) - (x-a) f^{(n-1)}(a).\end{aligned} \tag{5.79}$$

Continuing, we obtain

$$\begin{aligned}\int_a^x dx_3 \int_a^{x_3} dx_2 \int_a^{x_2} dx_1 f^{(n)}(x_1) &= f^{(n-3)}(x) - f^{(n-3)}(a) - (x-a) f^{(n-2)}(a) \\ &\quad - \frac{(x-a)^2}{2!} f^{(n-1)}(a).\end{aligned} \tag{5.80}$$

[9] Taylor's expansion may be derived under slightly less restrictive conditions; compare H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics*, 3rd ed. Cambridge: Cambridge University Press (1956), Section 1.133.

Finally, on integrating for the nth time,

$$\int_a^x dx_n \cdots \int_a^{x_2} dx_1 f^{(n)}(x_1) = f(x) - f(a) - (x-a)f'(a) - \frac{(x-a)^2}{2!}f''(a) - \cdots - \frac{(x-a)^{n-1}}{(n-1)!}f^{(n-1)}(a). \tag{5.81}$$

Note that this expression is exact. No terms have been dropped, no approximations made. Now, solving for $f(x)$, we have

$$\boxed{f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}f''(a) + \cdots + \frac{(x-a)^{n-1}}{(n-1)!}f^{(n-1)}(a) + R_n.} \tag{5.82}$$

The remainder, R_n, is given by the n-fold integral

$$R_n = \int_a^x dx_n \cdots \int_a^{x_2} dx_1 f^{(n)}(x_1). \tag{5.83}$$

This remainder, Eq. (5.83), may be put into a perhaps more practical form by using the **mean value theorem** of integral calculus:

$$\int_a^x g(x)\,dx = (x-a)g(\xi), \tag{5.84}$$

with $a \le \xi \le x$. By integrating n times we get the Lagrangian form[10] of the remainder:

$$R_n = \frac{(x-a)^n}{n!}f^{(n)}(\xi). \tag{5.85}$$

With Taylor's expansion in this form we are not concerned with any questions of infinite series convergence. This series is finite, and the only questions concern the magnitude of the remainder.

When the function $f(x)$ is such that

$$\lim_{n\to\infty} R_n = 0, \tag{5.86}$$

Eq. (5.82) becomes Taylor's series:

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}f''(a) + \cdots = \sum_{n=0}^{\infty} \frac{(x-a)^n}{n!}f^{(n)}(a).^{11} \tag{5.87}$$

[10] An alternate form derived by Cauchy is

$$R_n = \frac{(x-\zeta)^{n-1}(x-a)}{(n-1)!}f^{(n)}(\zeta),$$

with $a \le \zeta \le x$.

[11] Note that $0! = 1$ (compare Section 8.1).

Our Taylor series specifies the value of a function at one point, x, in terms of the value of the function and its derivatives at a reference point a. It is an expansion in powers of the **change** in the variable, $\Delta x = x - a$ in this case. The notation may be varied at the user's convenience. With the substitution $x \to x + h$ and $a \to x$ we have an alternate form,

$$f(x+h) = \sum_{n=0}^{\infty} \frac{h^n}{n!} f^{(n)}(x).$$

When we use the **operator** $D = d/dx$, the Taylor expansion becomes

$$f(x+h) = \sum_{n=0}^{\infty} \frac{h^n D^n}{n!} f(x) = e^{hD} f(x).$$

(The transition to the exponential form anticipates Eq. (5.90), which follows.) An equivalent operator form of this Taylor expansion appears in Exercise 4.2.4. A derivation of the Taylor expansion in the context of complex variable theory appears in Section 6.5.

Maclaurin Theorem

If we expand about the origin ($a = 0$), Eq. (5.87) is known as Maclaurin's series:

$$\boxed{f(x) = f(0) + xf'(0) + \frac{x^2}{2!} f''(0) + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!} f^{(n)}(0).} \tag{5.88}$$

An immediate application of the Maclaurin series (or the Taylor series) is in the expansion of various transcendental functions into infinite (power) series.

Example 5.6.1 EXPONENTIAL FUNCTION

Let $f(x) = e^x$. Differentiating, we have

$$f^{(n)}(0) = 1 \tag{5.89}$$

for all $n, n = 1, 2, 3, \ldots$. Then, with Eq. (5.88), we have

$$\boxed{e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}.} \tag{5.90}$$

This is the series expansion of the exponential function. Some authors use this series to define the exponential function.

Although this series is clearly convergent for all x, we should check the remainder term, R_n. By Eq. (5.85) we have

$$R_n = \frac{x^n}{n!} f^{(n)}(\xi) = \frac{x^n}{n!} e^{\xi}, \qquad 0 \le |\xi| \le x. \tag{5.91}$$

Therefore

$$|R_n| \le \frac{x^n e^x}{n!} \tag{5.92}$$

and

$$\lim_{n\to\infty} R_n = 0 \tag{5.93}$$

for all **finite** values of x, which indicates that this Maclaurin expansion of e^x converges absolutely over the range $-\infty < x < \infty$. ■

Example 5.6.2 LOGARITHM

Let $f(x) = \ln(1+x)$. By differentiating, we obtain

$$\begin{aligned} f'(x) &= (1+x)^{-1}, \\ f^{(n)}(x) &= (-1)^{n-1}(n-1)!(1+x)^{-n}. \end{aligned} \tag{5.94}$$

The Maclaurin expansion (Eq. (5.88)) yields

$$\begin{aligned} \ln(1+x) &= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + R_n \\ &= \sum_{p=1}^{n} (-1)^{p-1} \frac{x^p}{p} + R_n. \end{aligned} \tag{5.95}$$

In this case our remainder is given by

$$\begin{aligned} R_n &= \frac{x^n}{n!} f^{(n)}(\xi), \qquad 0 \le \xi \le x \\ &\le \frac{x^n}{n}, \qquad 0 \le \xi \le x \le 1. \end{aligned} \tag{5.96}$$

Now, the remainder approaches zero as n is increased indefinitely, provided $0 \le x \le 1$.[12] As an infinite series,

$$\boxed{\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n}} \tag{5.97}$$

converges for $-1 < x \le 1$. The range $-1 < x < 1$ is easily established by the d'Alembert ratio test (Section 5.2). Convergence at $x = 1$ follows by the Leibniz criterion (Section 5.3). In particular, at $x = 1$ we have

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots = \sum_{n=1}^{\infty} (-1)^{n-1} n^{-1}, \tag{5.98}$$

the conditionally convergent alternating harmonic series. ■

[12] This range can easily be extended to $-1 < x \le 1$ but not to $x = -1$.

Binomial Theorem

A second, extremely important application of the Taylor and Maclaurin expansions is the derivation of the binomial theorem for negative and/or nonintegral powers.

Let $f(x) = (1+x)^m$, in which m may be negative and is not limited to integral values. Direct application of Eq. (5.88) gives

$$(1+x)^m = 1 + mx + \frac{m(m-1)}{2!}x^2 + \cdots + R_n. \tag{5.99}$$

For this function the remainder is

$$R_n = \frac{x^n}{n!}(1+\xi)^{m-n}m(m-1)\cdots(m-n+1) \tag{5.100}$$

and ξ lies between 0 and x, $0 \le \xi \le x$. Now, for $n > m$, $(1+\xi)^{m-n}$ is a maximum for $\xi = 0$. Therefore

$$R_n \le \frac{x^n}{n!}m(m-1)\cdots(m-n+1). \tag{5.101}$$

Note that the m dependent factors do not yield a zero unless m is a nonnegative integer; R_n tends to zero as $n \to \infty$ if x is restricted to the range $0 \le x < 1$. The binomial expansion therefore is shown to be

$$(1+x)^m = 1 + mx + \frac{m(m-1)}{2!}x^2 + \frac{m(m-1)(m-2)}{3!}x^3 + \cdots. \tag{5.102}$$

In other, equivalent notation,

$$\boxed{(1+x)^m = \sum_{n=0}^{\infty}\frac{m!}{n!(m-n)!}x^n = \sum_{n=0}^{\infty}\binom{m}{n}x^n.} \tag{5.103}$$

The quantity $\binom{m}{n}$, which equals $m!/[n!(m-n)!]$, is called a **binomial coefficient**. Although we have only shown that the remainder vanishes,

$$\lim_{n\to\infty} R_n = 0,$$

for $0 \le x < 1$, the series in Eq. (5.102) actually may be shown to be convergent for the extended range $-1 < x < 1$. For m an integer, $(m-n)! = \pm\infty$ if $n > m$ (Section 8.1) and the series automatically terminates at $n = m$.

Example 5.6.3 RELATIVISTIC ENERGY

The total relativistic energy of a particle of mass m and velocity v is

$$E = mc^2\left(1 - \frac{v^2}{c^2}\right)^{-1/2}. \tag{5.104}$$

Compare this expression with the classical kinetic energy, $mv^2/2$.

By Eq. (5.102) with $x = -v^2/c^2$ and $m = -1/2$ we have

$$E = mc^2\left[1 - \frac{1}{2}\left(-\frac{v^2}{c^2}\right) + \frac{(-1/2)(-3/2)}{2!}\left(-\frac{v^2}{c^2}\right)^2 + \frac{(-1/2)(-3/2)(-5/2)}{3!}\left(-\frac{v^2}{c^2}\right)^3 + \cdots\right],$$

or

$$E = mc^2 + \frac{1}{2}mv^2 + \frac{3}{8}mv^2 \cdot \frac{v^2}{c^2} + \frac{5}{16}mv^2 \cdot \left(\frac{v^2}{c^2}\right)^2 + \cdots. \tag{5.105}$$

The first term, mc^2, is identified as the rest mass energy. Then

$$E_{\text{kinetic}} = \frac{1}{2}mv^2\left[1 + \frac{3}{4}\frac{v^2}{c^2} + \frac{5}{8}\left(\frac{v^2}{c^2}\right)^2 + \cdots\right]. \tag{5.106}$$

For particle velocity $v \ll c$, the velocity of light, the expression in the brackets reduces to unity and we see that the kinetic portion of the total relativistic energy agrees with the classical result. ■

For polynomials we can generalize the binomial expansion to

$$(a_1 + a_2 + \cdots + a_m)^n = \sum \frac{n!}{n_1!n_2!\cdots n_m!} a_1^{n_1} a_2^{n_2} \cdots a_m^{n_m},$$

where the summation includes all different combinations of $n_1, n_2, \ldots, n_m$ with $\sum_{i=1}^{m} n_i = n$. Here n_i and n are all integral. This generalization finds considerable use in statistical mechanics.

Maclaurin series may sometimes appear indirectly rather than by direct use of Eq. (5.88). For instance, the most convenient way to obtain the series expansion

$$\sin^{-1} x = \sum_{n=0}^{\infty} \frac{(2n-1)!!}{(2n)!!} \cdot \frac{x^{2n+1}}{(2n+1)} = x + \frac{x^3}{6} + \frac{3x^5}{40} + \cdots, \tag{5.106a}$$

is to make use of the relation (from $\sin y = x$, get $dy/dx = 1/\sqrt{1-x^2}$)

$$\sin^{-1} x = \int_0^x \frac{dt}{(1-t^2)^{1/2}}.$$

We expand $(1-t^2)^{-1/2}$ (binomial theorem) and then integrate term by term. This term-by-term integration is discussed in Section 5.7. The result is Eq. (5.106a). Finally, we may take the limit as $x \to 1$. The series converges by Gauss' test, Exercise 5.2.5.

Taylor Expansion — More Than One Variable

If the function f has more than one independent variable, say, $f = f(x, y)$, the Taylor expansion becomes

$$\begin{aligned} f(x,y) = f(a,b) &+ (x-a)\frac{\partial f}{\partial x} + (y-b)\frac{\partial f}{\partial y} \\ &+ \frac{1}{2!}\left[(x-a)^2\frac{\partial^2 f}{\partial x^2} + 2(x-a)(y-b)\frac{\partial^2 f}{\partial x\,\partial y} + (y-b)^2\frac{\partial^2 f}{\partial y^2}\right] \\ &+ \frac{1}{3!}\left[(x-a)^3\frac{\partial^3 f}{\partial x^3} + 3(x-a)^2(y-b)\frac{\partial^3 f}{\partial x^2\,\partial y}\right. \\ &\left. + 3(x-a)(y-b)^2\frac{\partial^3 f}{\partial x\,\partial y^2} + (y-b)^3\frac{\partial^3 f}{\partial y^3}\right] + \cdots, \end{aligned} \tag{5.107}$$

with all derivatives evaluated at the point (a, b). Using $\alpha_j t = x_j - x_{j0}$, we may write the Taylor expansion for m independent variables in the symbolic form

$$f(x_1, \ldots, x_m) = \sum_{n=0}^{\infty} \frac{t^n}{n!}\left(\sum_{i=1}^{m} \alpha_i \frac{\partial}{\partial x_i}\right)^n f(x_1, \ldots, x_m)\Big|_{(x_k = x_{k0}, k=1,\ldots,m)}. \tag{5.108}$$

A convenient vector form for $m = 3$ is

$$\psi(\mathbf{r} + \mathbf{a}) = \sum_{n=0}^{\infty} \frac{1}{n!}(\mathbf{a} \cdot \nabla)^n \psi(\mathbf{r}). \tag{5.109}$$

Exercises

5.6.1 Show that

(a) $\sin x = \displaystyle\sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$,

(b) $\cos x = \displaystyle\sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}$.

In Section 6.1, e^{ix} is defined by a series expansion such that

$$e^{ix} = \cos x + i \sin x.$$

This is the basis for the polar representation of complex quantities. As a special case we find, with $x = \pi$, the intriguing relation

$$e^{i\pi} = -1.$$

5.6.2 Derive a series expansion of $\cot x$ in increasing powers of x by dividing $\cos x$ by $\sin x$. *Note*. The resultant series that starts with $1/x$ is actually a Laurent series (Section 6.5). Although the two series for $\sin x$ and $\cos x$ were valid for all x, the convergence of the series for $\cot x$ is limited by the zeros of the denominator, $\sin x$ (see Analytic Continuation in Section 6.5).

5.6.3 The Raabe test for $\sum_n (n \ln n)^{-1}$ leads to

$$\lim_{n\to\infty} n\left[\frac{(n+1)\ln(n+1)}{n\ln n} - 1\right].$$

Show that this limit is unity (which means that the Raabe test here is indeterminate).

5.6.4 Show by series expansion that

$$\frac{1}{2}\ln\frac{\eta_0+1}{\eta_0-1} = \coth^{-1}\eta_0, \qquad |\eta_0| > 1.$$

This identity may be used to obtain a second solution for Legendre's equation.

5.6.5 Show that $f(x) = x^{1/2}$ (a) has no Maclaurin expansion but (b) has a Taylor expansion about any point $x_0 \neq 0$. Find the range of convergence of the Taylor expansion about $x = x_0$.

5.6.6 Let x be an approximation for a zero of $f(x)$ and Δx be the correction. Show that by neglecting terms of order $(\Delta x)^2$,

$$\Delta x = -\frac{f(x)}{f'(x)}.$$

This is Newton's formula for finding a root. Newton's method has the virtues of illustrating series expansions and elementary calculus but is very treacherous.

5.6.7 Expand a function $\Phi(x, y, z)$ by Taylor's expansion about $(0, 0, 0)$ to $\mathcal{O}(a^3)$. Evaluate $\bar{\Phi}$, the average value of Φ, averaged over a small cube of side a centered on the origin and show that the Laplacian of Φ is a measure of deviation of Φ from $\Phi(0, 0, 0)$.

5.6.8 The ratio of two differentiable functions $f(x)$ and $g(x)$ takes on the indeterminate form $0/0$ at $x = x_0$. Using Taylor expansions prove **L'Hôpital's rule**,

$$\lim_{x\to x_0}\frac{f(x)}{g(x)} = \lim_{x\to x_0}\frac{f'(x)}{g'(x)}.$$

5.6.9 With $n > 1$, show that

$$\text{(a)}\ \frac{1}{n} - \ln\left(\frac{n}{n-1}\right) < 0, \qquad \text{(b)}\ \frac{1}{n} - \ln\left(\frac{n+1}{n}\right) > 0.$$

Use these inequalities to show that the limit defining the Euler–Mascheroni constant, Eq. (5.28), is finite.

5.6.10 Expand $(1-2tz+t^2)^{-1/2}$ in powers of t. Assume that t is small. Collect the coefficients of t^0, t^1, and t^2.

ANS. $a_0 = P_0(z) = 1$,
$a_1 = P_1(z) = z$,
$a_2 = P_2(z) = \frac{1}{2}(3z^2 - 1)$,
where $a_n = P_n(z)$, the nth Legendre polynomial.

5.6.11 Using the double factorial notation of Section 8.1, show that

$$(1+x)^{-m/2} = \sum_{n=0}^{\infty} (-1)^n \frac{(m+2n-2)!!}{2^n n!(m-2)!!} x^n,$$

for $m = 1, 2, 3, \ldots$.

5.6.12 Using binomial expansions, compare the three Doppler shift formulas:

(a) $\nu' = \nu\left(1 \mp \frac{v}{c}\right)^{-1}$ moving source;

(b) $\nu' = \nu\left(1 \pm \frac{v}{c}\right)$ moving observer;

(c) $\nu' = \nu\left(1 \pm \frac{v}{c}\right)\left(1 - \frac{v^2}{c^2}\right)^{-1/2}$ relativistic.

Note. The relativistic formula agrees with the classical formulas if terms of order v^2/c^2 can be neglected.

5.6.13 In the theory of general relativity there are various ways of relating (defining) a velocity of recession of a galaxy to its red shift, δ. Milne's model (kinematic relativity) gives

(a) $v_1 = c\delta\left(1 + \frac{1}{2}\delta\right)$,

(b) $v_2 = c\delta\left(1 + \frac{1}{2}\delta\right)(1+\delta)^{-2}$,

(c) $1 + \delta = \left[\frac{1 + v_3/c}{1 - v_3/c}\right]^{1/2}$.

1. Show that for $\delta \ll 1$ (and $v_3/c \ll 1$) all three formulas reduce to $v = c\delta$.
2. Compare the three velocities through terms of order δ^2.

Note. In special relativity (with δ replaced by z), the ratio of observed wavelength λ to emitted wavelength λ_0 is given by

$$\frac{\lambda}{\lambda_0} = 1 + z = \left(\frac{c+v}{c-v}\right)^{1/2}.$$

5.6.14 The relativistic sum w of two velocities u and v is given by

$$\frac{w}{c} = \frac{u/c + v/c}{1 + uv/c^2}.$$

If

$$\frac{v}{c} = \frac{u}{c} = 1 - \alpha,$$

where $0 \le \alpha \le 1$, find w/c in powers of α through terms in α^3.

5.6.15 The displacement x of a particle of rest mass m_0, resulting from a constant force $m_0 g$ along the x-axis, is

$$x = \frac{c^2}{g}\left\{\left[1 + \left(g\frac{t}{c}\right)^2\right]^{1/2} - 1\right\},$$

including relativistic effects. Find the displacement x as a power series in time t. Compare with the classical result,

$$x = \tfrac{1}{2} g t^2.$$

5.6.16 By use of Dirac's relativistic theory, the fine structure formula of atomic spectroscopy is given by

$$E = mc^2\left[1 + \frac{\gamma^2}{(s + n - |k|)^2}\right]^{-1/2},$$

where

$$s = \left(|k|^2 - \gamma^2\right)^{1/2}, \qquad k = \pm 1, \pm 2, \pm 3, \ldots.$$

Expand in powers of γ^2 through order γ^4 ($\gamma^2 = Ze^2/4\pi\varepsilon_0\hbar c$, with Z the atomic number). This expansion is useful in comparing the predictions of the Dirac electron theory with those of a relativistic Schrödinger electron theory. Experimental results support the Dirac theory.

5.6.17 In a head-on proton–proton collision, the ratio of the kinetic energy in the center of mass system to the incident kinetic energy is

$$R = \left[\sqrt{2mc^2\left(E_k + 2mc^2\right)} - 2mc^2\right]/E_k.$$

Find the value of this ratio of kinetic energies for

(a) $E_k \ll mc^2$ (nonrelativistic)
(b) $E_k \gg mc^2$ (extreme-relativistic).

ANS. (a) $\frac{1}{2}$, (b) 0. The latter answer is a sort of law of diminishing returns for high-energy particle accelerators (with stationary targets).

5.6.18 With binomial expansions

$$\frac{x}{1-x} = \sum_{n=1}^{\infty} x^n, \qquad \frac{x}{x-1} = \frac{1}{1-x^{-1}} = \sum_{n=0}^{\infty} x^{-n}.$$

Adding these two series yields $\sum_{n=-\infty}^{\infty} x^n = 0$.
Hopefully, we can agree that this is nonsense, but what has gone wrong?

5.6.19 (a) Planck's theory of quantized oscillators leads to an average energy

$$\langle \varepsilon \rangle = \frac{\sum_{n=1}^{\infty} n\varepsilon_0 \exp(-n\varepsilon_0/kT)}{\sum_{n=0}^{\infty} \exp(-n\varepsilon_0/kT)},$$

where ε_0 is a fixed energy. Identify the numerator and denominator as binomial expansions and show that the ratio is

$$\langle \varepsilon \rangle = \frac{\varepsilon_0}{\exp(\varepsilon_0/kT) - 1}.$$

(b) Show that the $\langle \varepsilon \rangle$ of part (a) reduces to kT, the classical result, for $kT \gg \varepsilon_0$.

5.6.20 (a) Expand by the binomial theorem and integrate term by term to obtain the Gregory series for $y = \tan^{-1} x$ (note that $\tan y = x$):

$$\tan^{-1} x = \int_0^x \frac{dt}{1+t^2} = \int_0^x \left\{1 - t^2 + t^4 - t^6 + \cdots\right\} dt$$

$$= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1}, \qquad -1 \le x \le 1.$$

(b) By comparing series expansions, show that

$$\tan^{-1} x = \frac{i}{2} \ln\left(\frac{1 - ix}{1 + ix}\right).$$

Hint. Compare Exercise 5.4.1.

5.6.21 In numerical analysis it is often convenient to approximate $d^2\psi(x)/dx^2$ by

$$\frac{d^2}{dx^2}\psi(x) \approx \frac{1}{h^2}\big[\psi(x+h) - 2\psi(x) + \psi(x-h)\big].$$

Find the error in this approximation.

ANS. Error $= \dfrac{h^2}{12}\psi^{(4)}(x)$.

5.6.22 You have a function $y(x)$ tabulated at equally spaced values of the argument

$$\begin{cases} y_n = y(x_n) \\ x_n = x + nh. \end{cases}$$

Show that the linear combination

$$\frac{1}{12h}\{-y_2 + 8y_1 - 8y_{-1} + y_{-2}\}$$

yields

$$y_0' - \frac{h^4}{30} y_0^{(5)} + \cdots.$$

Hence this linear combination yields y_0' if $(h^4/30)y_0^{(5)}$ and higher powers of h and higher derivatives of $y(x)$ are negligible.

5.6.23 In a numerical integration of a partial differential equation, the three-dimensional Laplacian is replaced by

$$\nabla^2\psi(x,y,z) \to h^{-2}\big[\psi(x+h,y,z)+\psi(x-h,y,z) + \psi(x,y+h,z)+\psi(x,y-h,z)+\psi(x,y,z+h) + \psi(x,y,z-h)-6\psi(x,y,z)\big].$$

Determine the error in this approximation. Here h is the step size, the distance between adjacent points in the x-, y-, or z-direction.

5.6.24 Using double precision, calculate e from its Maclaurin series.
Note. This simple, direct approach is the best way of calculating e to high accuracy. Sixteen terms give e to 16 significant figures. The reciprocal factorials give very rapid convergence.

5.7 POWER SERIES

The power series is a special and extremely useful type of infinite series of the form

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots = \sum_{n=0}^{\infty} a_n x^n, \tag{5.110}$$

where the coefficients a_i are constants, independent of x.[13]

Convergence

Equation (5.110) may readily be tested for convergence by either the Cauchy root test or the d'Alembert ratio test (Section 5.2). If

$$\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = R^{-1}, \tag{5.111}$$

the series converges for $-R < x < R$. This is the interval or radius of convergence. Since the root and ratio tests fail when the limit is unity, the endpoints of the interval require special attention.

For instance, if $a_n = n^{-1}$, then $R = 1$ and, from Sections 5.1, 5.2, and 5.3, the series converges for $x = -1$ but diverges for $x = +1$. If $a_n = n!$, then $R = 0$ and the series diverges for all $x \neq 0$.

Uniform and Absolute Convergence

Suppose our power series (Eq. (5.110)) has been found convergent for $-R < x < R$; then it will be uniformly and absolutely convergent in any **interior** interval, $-S \le x \le S$, where $0 < S < R$.

This may be proved directly by the Weierstrass M test (Section 5.5).

[13] Equation (5.110) may be generalized to $z = x + iy$, replacing x. The following two chapters will then yield uniform convergence, integrability, and differentiability in a region of a complex plane in place of an interval on the x-axis.

Continuity

Since each of the terms $u_n(x) = a_n x^n$ is a continuous function of x and $f(x) = \sum a_n x^n$ converges uniformly for $-S \le x \le S$, $f(x)$ must be a continuous function in the interval of uniform convergence.

This behavior is to be contrasted with the strikingly different behavior of the Fourier series (Chapter 14), in which the Fourier series is used frequently to represent discontinuous functions such as sawtooth and square waves.

Differentiation and Integration

With $u_n(x)$ continuous and $\sum a_n x^n$ uniformly convergent, we find that the differentiated series is a power series with continuous functions and the same radius of convergence as the original series. The new factors introduced by differentiation (or integration) do not affect either the root or the ratio test. Therefore our power series may be differentiated or integrated as often as desired within the interval of uniform convergence (Exercise 5.7.13).

In view of the rather severe restrictions placed on differentiation (Section 5.5), this is a remarkable and valuable result.

Uniqueness Theorem

In the preceding section, using the Maclaurin series, we expanded e^x and $\ln(1 + x)$ into infinite series. In the succeeding chapters, functions are frequently represented or perhaps defined by infinite series. We now establish that the power-series representation is unique.

If

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} a_n x^n, \qquad -R_a < x < R_a \\ &= \sum_{n=0}^{\infty} b_n x^n, \qquad -R_b < x < R_b, \end{aligned} \tag{5.112}$$

with overlapping intervals of convergence, including the origin, then

$$a_n = b_n \tag{5.113}$$

for all n; that is, we assume two (different) power-series representations and then proceed to show that the two are actually identical.

From Eq. (5.112),

$$\sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} b_n x^n, \qquad -R < x < R, \tag{5.114}$$

where R is the smaller of R_a, R_b. By setting $x = 0$ to eliminate all but the constant terms, we obtain

$$a_0 = b_0. \tag{5.115}$$

Now, exploiting the differentiability of our power series, we differentiate Eq. (5.114), getting

$$\sum_{n=1}^{\infty} n a_n x^{n-1} = \sum_{n=1}^{\infty} n b_n x^{n-1}. \tag{5.116}$$

We again set $x = 0$, to isolate the new constant terms, and find

$$a_1 = b_1. \tag{5.117}$$

By repeating this process n times, we get

$$a_n = b_n, \tag{5.118}$$

which shows that the two series coincide. Therefore our power-series representation is unique.

This will be a crucial point in Section 9.5, in which we use a power series to develop solutions of differential equations. This uniqueness of power series appears frequently in theoretical physics. The establishment of perturbation theory in quantum mechanics is one example. The power-series representation of functions is often useful in evaluating indeterminate forms, particularly when l'Hôpital's rule may be awkward to apply (Exercise 5.7.9).

Example 5.7.1 L'HÔPITAL'S RULE

Evaluate

$$\lim_{x\to 0} \frac{1-\cos x}{x^2}. \tag{5.119}$$

Replacing $\cos x$ by its Maclaurin-series expansion, we obtain

$$\frac{1-\cos x}{x^2} = \frac{1-(1-\frac{1}{2!}x^2+\frac{1}{4!}x^4-\cdots)}{x^2} = \frac{1}{2!} - \frac{x^2}{4!} + \cdots.$$

Letting $x \to 0$, we have

$$\lim_{x\to 0} \frac{1-\cos x}{x^2} = \frac{1}{2}. \tag{5.120}$$

The uniqueness of power series means that the coefficients a_n may be identified with the derivatives in a Maclaurin series. From

$$f(x) = \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(0) x^n$$

we have

$$a_n = \frac{1}{n!} f^{(n)}(0).$$

■

Inversion of Power Series

Suppose we are given a series

$$y - y_0 = a_1(x - x_0) + a_2(x - x_0)^2 + \cdots = \sum_{n=1}^{\infty} a_n (x - x_0)^n. \tag{5.121}$$

This gives $(y - y_0)$ in terms of $(x - x_0)$. However, it may be desirable to have an explicit expression for $(x - x_0)$ in terms of $(y - y_0)$. We may solve Eq. (5.121) for $x - x_0$ by inversion of our series. Assume that

$$x - x_0 = \sum_{n=1}^{\infty} b_n (y - y_0)^n, \tag{5.122}$$

with the b_n to be determined in terms of the assumed known a_n. A brute-force approach, which is perfectly adequate for the first few coefficients, is simply to substitute Eq. (5.121) into Eq. (5.122). By equating coefficients of $(x - x_0)^n$ on both sides of Eq. (5.122), since the power series is unique, we obtain

$$\begin{aligned} b_1 &= \frac{1}{a_1}, \\ b_2 &= -\frac{a_2}{a_1^3}, \\ b_3 &= \frac{1}{a_1^5}\left(2a_2^2 - a_1 a_3\right), \\ b_4 &= \frac{1}{a_1^7}\left(5a_1 a_2 a_3 - a_1^2 a_4 - 5a_2^3\right), \qquad \text{and so on.} \end{aligned} \tag{5.123}$$

Some of the higher coefficients are listed by Dwight.[14] A more general and much more elegant approach is developed by the use of complex variables in the first and second editions of *Mathematical Methods for Physicists*.

Exercises

5.7.1 The classical Langevin theory of paramagnetism leads to an expression for the magnetic polarization,

$$P(x) = c\left(\frac{\cosh x}{\sinh x} - \frac{1}{x}\right).$$

Expand $P(x)$ as a power series for small x (low fields, high temperature).

[14] H. B. Dwight, *Tables of Integrals and Other Mathematical Data*, 4th ed. New York: Macmillan (1961). (Compare Formula No. 50.)

5.7.2 The depolarizing factor L for an oblate ellipsoid in a uniform electric field parallel to the axis of rotation is

$$L = \frac{1}{\varepsilon_0}\left(1+\zeta_0^2\right)\left(1-\zeta_0 \cot^{-1}\zeta_0\right),$$

where ζ_0 defines an oblate ellipsoid in oblate spheroidal coordinates (ξ, ζ, φ). Show that

$$\lim_{\zeta_0\to\infty} L = \frac{1}{3\varepsilon_0} \quad \text{(sphere)}, \qquad \lim_{\zeta_0\to 0} L = \frac{1}{\varepsilon_0} \quad \text{(thin sheet)}.$$

5.7.3 The depolarizing factor (Exercise 5.7.2) for a prolate ellipsoid is

$$L = \frac{1}{\varepsilon_0}\left(\eta_0^2-1\right)\left(\frac{1}{2}\eta_0 \ln\frac{\eta_0+1}{\eta_0-1} - 1\right).$$

Show that

$$\lim_{\eta_0\to\infty} L = \frac{1}{3\varepsilon_0} \quad \text{(sphere)}, \qquad \lim_{\eta_0\to 0} L = 0 \quad \text{(long needle)}.$$

5.7.4 The analysis of the diffraction pattern of a circular opening involves

$$\int_0^{2\pi} \cos(c\cos\varphi)\, d\varphi.$$

Expand the integrand in a series and integrate by using

$$\int_0^{2\pi} \cos^{2n}\varphi\, d\varphi = \frac{(2n)!}{2^{2n}(n!)^2}\cdot 2\pi, \qquad \int_0^{2\pi} \cos^{2n+1}\varphi\, d\varphi = 0.$$

The result is 2π times the Bessel function $J_0(c)$.

5.7.5 Neutrons are created (by a nuclear reaction) inside a hollow sphere of radius R. The newly created neutrons are uniformly distributed over the spherical volume. Assuming that all directions are equally probable (isotropy), what is the average distance a neutron will travel before striking the surface of the sphere? Assume straight-line motion and no collisions.

(a) Show that

$$\bar{r} = \tfrac{3}{2}R\int_0^1\int_0^\pi \sqrt{1-k^2\sin^2\theta}\,k^2\, dk\sin\theta\, d\theta.$$

(b) Expand the integrand as a series and integrate to obtain

$$\bar{r} = R\left[1 - 3\sum_{n=1}^{\infty}\frac{1}{(2n-1)(2n+1)(2n+3)}\right].$$

(c) Show that the sum of this infinite series is $1/12$, giving $\bar{r} = \frac{3}{4}R$.

Hint. Show that $s_n = 1/12 - [4(2n+1)(2n+3)]^{-1}$ by mathematical induction. Then let $n\to\infty$.

5.7.6 Given that

$$\int_0^1 \frac{dx}{1+x^2} = \tan^{-1} x\Big|_0^1 = \frac{\pi}{4},$$

expand the integrand into a series and integrate term by term obtaining[15]

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \cdots + (-1)^n \frac{1}{2n+1} + \cdots,$$

which is Leibniz's formula for π. Compare the convergence of the integrand series and the integrated series at $x = 1$. See also Exercise 5.7.18.

5.7.7 Expand the incomplete factorial function

$$\gamma(n+1, x) \equiv \int_0^x e^{-t} t^n \, dt$$

in a series of powers of x. What is the range of convergence of the resulting series?

$$\textit{ANS.} \int_0^x e^{-t} t^n \, dt = x^{n+1}\left[\frac{1}{n+1} - \frac{x}{n+2} + \frac{x^2}{2!(n+3)} - \cdots \frac{(-1)^p x^p}{p!(n+p+1)} + \cdots\right].$$

5.7.8 Derive the series expansion of the incomplete beta function

$$\begin{aligned} B_x(p, q) &= \int_0^x t^{p-1}(1-t)^{q-1} \, dt \\ &= x^p \left\{\frac{1}{p} + \frac{1-q}{p+1}x + \cdots + \frac{(1-q)\cdots(n-q)}{n!(p+n)}x^n + \cdots\right\} \end{aligned}$$

for $0 \le x \le 1$, $p > 0$, and $q > 0$ (if $x = 1$).

5.7.9 Evaluate

(a) $\lim_{x\to 0}\big[\sin(\tan x) - \tan(\sin x)\big]x^{-7}$, (b) $\lim_{x\to 0} x^{-n} j_n(x), \quad n = 3,$

where $j_n(x)$ is a spherical Bessel function (Section 11.7), defined by

$$j_n(x) = (-1)^n x^n \left(\frac{1}{x}\frac{d}{dx}\right)^n \left(\frac{\sin x}{x}\right).$$

ANS. (a) $-\frac{1}{30}$, (b) $\frac{1}{1\cdot 3\cdot 5\cdots(2n+1)} \to \frac{1}{105}$ for $n = 3$.

[15] The series expansion of $\tan^{-1} x$ (upper limit 1 replaced by x) was discovered by James Gregory in 1671, 3 years before Leibniz. See Peter Beckmann's entertaining book, *A History of Pi*, 2nd ed., Boulder, CO: Golem Press (1971) and L. Berggren, J. and P. Borwein, *Pi: A Source Book*, New York: Springer (1997).

5.7.10 Neutron transport theory gives the following expression for the inverse neutron diffusion length of k:

$$\frac{a-b}{k}\tanh^{-1}\left(\frac{k}{a}\right)=1.$$

By series inversion or otherwise, determine k^2 as a series of powers of b/a. Give the first two terms of the series.

ANS. $k^2=3ab\left(1-\frac{4}{5}\frac{b}{a}\right)$.

5.7.11 Develop a series expansion of $y=\sinh^{-1}x$ (that is, $\sinh y=x$) in powers of x by

(a) inversion of the series for $\sinh y$,
(b) a direct Maclaurin expansion.

5.7.12 A function $f(z)$ is represented by a **descending** power series

$$f(z)=\sum_{n=0}^{\infty}a_nz^{-n},\qquad R\le z<\infty.$$

Show that this series expansion is unique; that is, if $f(z)=\sum_{n=0}^{\infty}b_nz^{-n}$, $R\le z<\infty$, then $a_n=b_n$ for all n.

5.7.13 A power series converges for $-R<x<R$. Show that the differentiated series and the integrated series have the same interval of convergence. (Do not bother about the endpoints $x=\pm R$.)

5.7.14 Assuming that $f(x)$ may be expanded in a power series about the origin, $f(x)=\sum_{n=0}^{\infty}a_nx^n$, with some nonzero range of convergence. Use the techniques employed in proving uniqueness of series to show that your assumed series is a Maclaurin series:

$$a_n=\frac{1}{n!}f^{(n)}(0).$$

5.7.15 The Klein–Nishina formula for the scattering of photons by electrons contains a term of the form

$$f(\varepsilon)=\frac{(1+\varepsilon)}{\varepsilon^2}\left[\frac{2+2\varepsilon}{1+2\varepsilon}-\frac{\ln(1+2\varepsilon)}{\varepsilon}\right].$$

Here $\varepsilon=h\nu/mc^2$, the ratio of the photon energy to the electron rest mass energy. Find

$$\lim_{\varepsilon\to0}f(\varepsilon).$$

ANS. $\frac{4}{3}$.

5.7.16 The behavior of a neutron losing energy by colliding elastically with nuclei of mass A is described by a parameter ξ_1,

$$\xi_1=1+\frac{(A-1)^2}{2A}\ln\frac{A-1}{A+1}.$$

An approximation, good for large A, is

$$\xi_2 = \frac{2}{A + 2/3}.$$

Expand ξ_1 and ξ_2 in powers of A^{-1}. Show that ξ_2 agrees with ξ_1 through $(A^{-1})^2$. Find the difference in the coefficients of the $(A^{-1})^3$ term.

5.7.17 Show that each of these two integrals equals Catalan's constant:

$$\text{(a)} \int_0^1 \text{arc tan}\, t \frac{dt}{t}, \qquad \text{(b)} -\int_0^1 \ln x \frac{dx}{1+x^2}.$$

Note. See $\beta(2)$ in Section 5.9 for the value of Catalan's constant.

5.7.18 Calculate π (double precision) by each of the following arc tangent expressions:

$$\pi = 16\tan^{-1}(1/5) - 4\tan^{-1}(1/239)$$
$$\pi = 24\tan^{-1}(1/8) + 8\tan^{-1}(1/57) + 4\tan^{-1}(1/239)$$
$$\pi = 48\tan^{-1}(1/18) + 32\tan^{-1}(1/57) - 20\tan^{-1}(1/239).$$

Obtain 16 significant figures. Verify the formulas using Exercise 5.6.2.
Note. These formulas have been used in some of the more accurate calculations of π.[16]

5.7.19 An analysis of the Gibbs phenomenon of Section 14.5 leads to the expression

$$\frac{2}{\pi}\int_0^\pi \frac{\sin\xi}{\xi} d\xi.$$

(a) Expand the integrand in a series and integrate term by term. Find the numerical value of this expression to four significant figures.
(b) Evaluate this expression by the Gaussian quadrature if available.

ANS. 1.178980.

5.8 ELLIPTIC INTEGRALS

Elliptic integrals are included here partly as an illustration of the use of power series and partly for their own intrinsic interest. This interest includes the occurrence of elliptic integrals in physical problems (Example 5.8.1 and Exercise 5.8.4) and applications in mathematical problems.

Example 5.8.1 PERIOD OF A SIMPLE PENDULUM

For small-amplitude oscillations, our pendulum (Fig. 5.8) has simple harmonic motion with a period $T = 2\pi(l/g)^{1/2}$. For a maximum amplitude θ_M large enough so that $\sin\theta_M \neq \theta_M$, Newton's second law of motion and Lagrange's equation (Section 17.7) lead to a nonlinear differential equation ($\sin\theta$ is a nonlinear function of θ), so we turn to a different approach.

[16] D. Shanks and J. W. Wrench, Computation of π to 100 000 decimals. *Math. Comput.* **16**: 76 (1962).

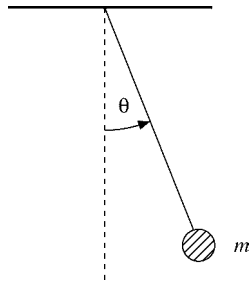

FIGURE 5.8 Simple pendulum.

The swinging mass m has a kinetic energy of $ml^2(d\theta/dt)^2/2$ and a potential energy of $-mgl\cos\theta(\theta=\pi/2$ taken for the arbitrary zero of potential energy). Since $d\theta/dt=0$ at $\theta=\theta_M$, conservation of energy gives

$$\frac{1}{2}ml^2\left(\frac{d\theta}{dt}\right)^2-mgl\cos\theta=-mgl\cos\theta_M. \tag{5.124}$$

Solving for $d\theta/dt$ we obtain

$$\frac{d\theta}{dt}=\pm\left(\frac{2g}{l}\right)^{1/2}(\cos\theta-\cos\theta_M)^{1/2}, \tag{5.125}$$

with the mass m canceling out. We take t to be zero when $\theta=0$ and $d\theta/dt>0$. An integration from $\theta=0$ to $\theta=\theta_M$ yields

$$\int_0^{\theta_M}(\cos\theta-\cos\theta_M)^{-1/2}\,d\theta=\left(\frac{2g}{l}\right)^{1/2}\int_0^t dt=\left(\frac{2g}{l}\right)^{1/2}t. \tag{5.126}$$

This is $\frac{1}{4}$ of a cycle, and therefore the time t is $\frac{1}{4}$ of the period T. We note that $\theta\le\theta_M$, and with a bit of clairvoyance we try the half-angle substitution

$$\sin\left(\frac{\theta}{2}\right)=\sin\left(\frac{\theta_M}{2}\right)\sin\varphi. \tag{5.127}$$

With this, Eq. (5.126) becomes

$$T=4\left(\frac{l}{g}\right)^{1/2}\int_0^{\pi/2}\left(1-\sin^2\left(\frac{\theta_M}{2}\right)\sin^2\varphi\right)^{-1/2}d\varphi. \tag{5.128}$$

Although not an obvious improvement over Eq. (5.126), the integral now defines the complete elliptic integral of the first kind, $K(\sin^2\theta_M/2)$. From the series expansion, the period of our pendulum may be developed as a power series — powers of $\sin\theta_M/2$:

$$T=2\pi\left(\frac{l}{g}\right)^{1/2}\left\{1+\frac{1}{4}\sin^2\frac{\theta_M}{2}+\frac{9}{64}\sin^4\frac{\theta_M}{2}+\cdots\right\}. \tag{5.129}$$

■

Definitions

Generalizing Example 5.8.1 to include the upper limit as a variable, the **elliptic integral of the first kind** is defined as

$$F(\varphi\backslash\alpha) = \int_0^{\varphi} \left(1 - \sin^2\alpha \sin^2\theta\right)^{-1/2} d\theta, \tag{5.130a}$$

or

$$F(x|m) = \int_0^{x} \left[\left(1 - t^2\right)\left(1 - mt^2\right)\right]^{-1/2} dt, \qquad 0 \le m < 1. \tag{5.130b}$$

(This is the notation of AMS-55 see footnote 4 for the reference.) For $\varphi = \pi/2$, $x = 1$, we have the **complete elliptic integral of the first kind**,

$$\begin{aligned} K(m) &= \int_0^{\pi/2} \left(1 - m\sin^2\theta\right)^{-1/2} d\theta \\ &= \int_0^{1} \left[\left(1 - t^2\right)\left(1 - mt^2\right)\right]^{-1/2} dt, \end{aligned} \tag{5.131}$$

with $m = \sin^2\alpha,\ 0 \le m < 1$.

The **elliptic integral of the second kind** is defined by

$$E(\varphi\backslash\alpha) = \int_0^{\varphi} \left(1 - \sin^2\alpha \sin^2\theta\right)^{1/2} d\theta \tag{5.132a}$$

or

$$E(x|m) = \int_0^{x} \left(\frac{1 - mt^2}{1 - t^2}\right)^{1/2} dt, \qquad 0 \le m \le 1. \tag{5.132b}$$

Again, for the case $\varphi = \pi/2,\ x = 1$, we have the **complete elliptic integral of the second kind**:

$$\begin{aligned} E(m) &= \int_0^{\pi/2} \left(1 - m\sin^2\theta\right)^{1/2} d\theta \\ &= \int_0^{1} \left(\frac{1 - mt^2}{1 - t^2}\right)^{1/2} dt, \qquad 0 \le m \le 1. \end{aligned} \tag{5.133}$$

Exercise 5.8.1 is an example of its occurrence. Figure 5.9 shows the behavior of $K(m)$ and $E(m)$. Extensive tables are available in AMS-55 (see Exercise 5.2.22 for the reference).

Series Expansion

For our range $0 \le m < 1$, the denominator of $K(m)$ may be expanded by the binomial series

$$\begin{aligned} \left(1 - m\sin^2\theta\right)^{-1/2} &= 1 + \frac{1}{2}m\sin^2\theta + \frac{3}{8}m^2\sin^4\theta + \cdots \\ &= \sum_{n=0}^{\infty} \frac{(2n-1)!!}{(2n)!!} m^n \sin^{2n}\theta. \end{aligned} \tag{5.134}$$

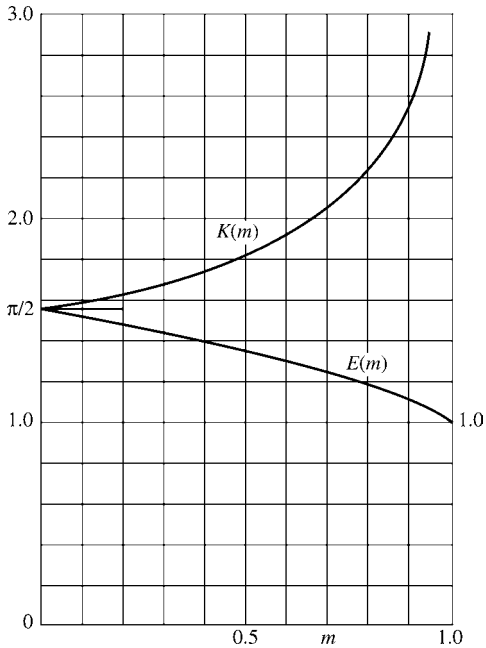

FIGURE 5.9 Complete elliptic integrals, $K(m)$ and $E(m)$.

For any closed interval $[0, m_{\max}]$, $m_{\max} < 1$, this series is uniformly convergent and may be integrated term by term. From Exercise 8.4.9,

$$\int_0^{\pi/2} \sin^{2n}\theta\, d\theta = \frac{(2n-1)!!}{(2n)!!} \cdot \frac{\pi}{2}. \tag{5.135}$$

Hence

$$K(m) = \frac{\pi}{2}\left\{1 + \sum_{n=1}^{\infty}\left[\frac{(2n-1)!!}{(2n)!!}\right]^2 m^n\right\}. \tag{5.136}$$

Similarly,

$$E(m) = \frac{\pi}{2}\left\{1 - \sum_{n=1}^{\infty}\left[\frac{(2n-1)!!}{(2n)!!}\right]^2 \frac{m^n}{2n-1}\right\} \tag{5.137}$$

(Exercise 5.8.2). In Section 13.5 these series are identified as hypergeometric functions, and we have

$$K(m) = \frac{\pi}{2}\,{}_2F_1\left(\frac{1}{2}, \frac{1}{2}; 1; m\right) \tag{5.138}$$

$$E(m) = \frac{\pi}{2}\,{}_2F_1\left(-\frac{1}{2}, \frac{1}{2}; 1; m\right). \tag{5.139}$$

Limiting Values

From the series Eqs. (5.136) and (5.137), or from the defining integrals,

$$\lim_{m\to 0} K(m) = \frac{\pi}{2}, \tag{5.140}$$

$$\lim_{m\to 0} E(m) = \frac{\pi}{2}. \tag{5.141}$$

For $m \to 1$ the series expansions are of little use. However, the integrals yield

$$\lim_{m\to 1} K(m) = \infty, \tag{5.142}$$

the integral diverging logarithmically, and

$$\lim_{m\to 1} E(m) = 1. \tag{5.143}$$

The elliptic integrals have been used extensively in the past for evaluating integrals. For instance, integrals of the form

$$I = \int_0^x R\left(t, \sqrt{a_4t^4 + a_3t^3 + a_2t^2 + a_1t^1 + a_0}\right) dt,$$

where R is a rational function of t and of the radical, may be expressed in terms of elliptic integrals. Jahnke and Emde, *Tables of Functions with Formulae and Curves*. New York: Dover (1943), Chapter 5, give pages of such transformations. With computers available for direct numerical evaluation, interest in these elliptic integral techniques has declined. However, elliptic integrals still remain of interest because of their appearance in physical problems — see Exercises 5.8.4 and 5.8.5.

For an extensive account of elliptic functions, integrals, and Jacobi theta functions, you are directed to Whittaker and Watson's treatise *A Course in Modern Analysis*, 4th ed. Cambridge, UK: Cambridge University Press (1962).

Exercises

5.8.1 The ellipse $x^2/a^2 + y^2/b^2 = 1$ may be represented parametrically by $x = a\sin\theta$, $y = b\cos\theta$. Show that the length of arc within the first quadrant is

$$a\int_0^{\pi/2} \left(1 - m\sin^2\theta\right)^{1/2} d\theta = aE(m).$$

Here $0 \le m = (a^2 - b^2)/a^2 \le 1$.

5.8.2 Derive the series expansion

$$E(m) = \frac{\pi}{2}\left\{1 - \left(\frac{1}{2}\right)^2\frac{m}{1} - \left(\frac{1\cdot 3}{2\cdot 4}\right)^2\frac{m^2}{3} - \cdots\right\}.$$

5.8.3 Show that

$$\lim_{m\to 0}\frac{(K-E)}{m} = \frac{\pi}{4}.$$

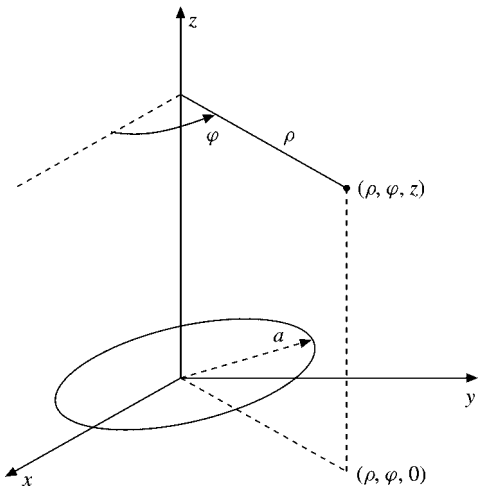

FIGURE 5.10 Circular wire loop.

5.8.4 A circular loop of wire in the xy-plane, as shown in Fig. 5.10, carries a current I. Given that the vector potential is

$$A_\varphi(\rho,\varphi,z)=\frac{a\mu_0 I}{2\pi}\int_0^\pi \frac{\cos\alpha\, d\alpha}{(a^2+\rho^2+z^2-2a\rho\cos\alpha)^{1/2}},$$

show that

$$A_\varphi(\rho,\varphi,z)=\frac{\mu_0 I}{\pi k}\left(\frac{a}{\rho}\right)^{1/2}\left[\left(1-\frac{k^2}{2}\right)K\left(k^2\right)-E\left(k^2\right)\right],$$

where

$$k^2=\frac{4a\rho}{(a+\rho)^2+z^2}.$$

Note. For extension of Exercise 5.8.4 to **B**, see Smythe, p. 270.[17]

5.8.5 An analysis of the magnetic vector potential of a circular current loop leads to the expression

$$f\left(k^2\right)=k^{-2}\left[\left(2-k^2\right)K\left(k^2\right)-2E\left(k^2\right)\right],$$

where $K(k^2)$ and $E(k^2)$ are the complete elliptic integrals of the first and second kinds. Show that for $k^2 \ll 1$ ($r \gg$ radius of loop)

$$f\left(k^2\right)\approx\frac{\pi k^2}{16}.$$

[17] W. R. Smythe, *Static and Dynamic Electricity*, 3rd ed. New York: McGraw-Hill (1969).

5.8.6 Show that

(a) $$\frac{dE(k^2)}{dk} = \frac{1}{k}(E - K),$$

(b) $$\frac{dK(k^2)}{dk} = \frac{E}{k(1-k^2)} - \frac{K}{k}.$$

Hint. For part (b) show that

$$E\left(k^2\right) = \left(1 - k^2\right) \int_0^{\pi/2} \left(1 - k\sin^2\theta\right)^{-3/2} d\theta$$

by comparing series expansions.

5.8.7 (a) Write a function subroutine that will compute $E(m)$ from the series expansion, Eq. (5.137).

(b) Test your function subroutine by using it to calculate $E(m)$ over the range $m = 0.0(0.1)0.9$ and comparing the result with the values given by AMS-55 (see Exercise 5.2.22 for the reference).

5.8.8 Repeat Exercise 5.8.7 for $K(m)$.
Note. These series for $E(m)$, Eq. (5.137), and $K(m)$, Eq. (5.136), converge only very slowly for m near 1. More rapidly converging series for $E(m)$ and $K(m)$ exist. See Dwight's Tables of Integrals:[18] No. 773.2 and 774.2. Your computer subroutine for computing E and K probably uses polynomial approximations: AMS-55, Chapter 17.

5.8.9 A simple pendulum is swinging with a maximum amplitude of θ_M. In the limit as $\theta_M \to 0$, the period is 1 s. Using the elliptic integral, $K(k^2), k = \sin(\theta_M/2)$, calculate the period T for $\theta_M = 0\ (10°)\ 90°$.
Caution. Some elliptic integral subroutines require $k = m^{1/2}$ as an input parameter, not m itself.

Check values.

θ_M	10°	50°	90°
T(sec)	1.00193	1.05033	1.18258

5.8.10 Calculate the magnetic vector potential $\mathbf{A}(\rho, \varphi, z) = \hat{\boldsymbol{\varphi}} A_\varphi(\rho, \varphi, z)$ of a circular current loop (Exercise 5.8.4) for the ranges $\rho/a = 2, 3, 4$, and $z/a = 0, 1, 2, 3, 4$.
Note. This elliptic integral calculation of the magnetic vector potential may be checked by an associated Legendre function calculation, Example 12.5.1.

Check value. For $\rho/a = 3$ and $z/a = 0$; $A_\varphi = 0.029023\mu_0 I$.

5.9 BERNOULLI NUMBERS, EULER–MACLAURIN FORMULA

The Bernoulli numbers were introduced by Jacques (James, Jacob) Bernoulli. There are several equivalent definitions, but extreme care must be taken, for some authors introduce

[18] H. B. Dwight, *Tables of Integrals and Other Mathematical Data*. New York: Macmillan (1947).

variations in numbering or in algebraic signs. One relatively simple approach is to define the Bernoulli numbers by the series[19]

$$\boxed{\frac{x}{e^x-1}=\sum_{n=0}^{\infty}\frac{B_n x^n}{n!},} \tag{5.144}$$

which converges for $|x|<2\pi$ by the ratio test substitut Eq. (5.153) (see also Example 7.1.7). By differentiating this power series repeatedly and then setting $x=0$, we obtain

$$B_n=\left[\frac{d^n}{dx^n}\left(\frac{x}{e^x-1}\right)\right]_{x=0}. \tag{5.145}$$

Specifically,

$$B_1=\frac{d}{dx}\left(\frac{x}{e^x-1}\right)\bigg|_{x=0}=\frac{1}{e^x-1}-\frac{xe^x}{(e^x-1)^2}\bigg|_{x=0}=-\frac{1}{2}, \tag{5.146}$$

as may be seen by series expansion of the denominators. Using $B_0=1$ and $B_1=-\frac{1}{2}$, it is easy to verify that the function

$$\frac{x}{e^x-1}-1+\frac{x}{2}=\sum_{n=2}^{\infty}B_n\frac{x^n}{n!}=-\frac{x}{e^{-x}-1}-1-\frac{x}{2} \tag{5.147}$$

is even in x, so all $B_{2n+1}=0$.

To derive a recursion relation for the Bernoulli numbers, we multiply

$$\begin{aligned}\frac{e^x-1}{x}\frac{x}{e^x-1}=1&=\left\{\sum_{m=0}^{\infty}\frac{x^m}{(m+1)!}\right\}\left\{1-\frac{x}{2}+\sum_{n=1}^{\infty}B_{2n}\frac{x^{2n}}{(2n)!}\right\}\\&=1+\sum_{m=1}^{\infty}x^m\left\{\frac{1}{(m+1)!}-\frac{1}{2m!}\right\}\\&\quad+\sum_{N=2}^{\infty}x^N\sum_{1\le n\le N/2}\frac{B_{2n}}{(2n)!(N-2n+1)!}.\end{aligned} \tag{5.148}$$

For $N>0$ the coefficient of x^N is zero, so Eq. (5.148) yields

$$\frac{1}{2}(N+1)-1=\sum_{1\le n\le N/2}B_{2n}\binom{N+1}{2n}=\frac{1}{2}(N-1), \tag{5.149}$$

[19]The function $x/(e^x-1)$ may be considered a **generating function** since it generates the Bernoulli numbers. Generating functions of the special functions of mathematical physics appear in Chapters 11, 12, and 13.

Table 5.1 Bernoulli Numbers

n	B_n	B_n
0	1	1.000 000 00
1	$-\frac{1}{2}$	−0.500 000 00
2	$\frac{1}{6}$	0.166 666 67
4	$-\frac{1}{30}$	−0.033 333 33
6	$\frac{1}{42}$	0.023 809 524
8	$-\frac{1}{30}$	−0.033 333 33
10	$\frac{5}{66}$	0.075 757 576

Note. Further values are given in National Bureau of Standards, *Handbook of Mathematical Functions* (AMS-55). See footnote 4 for the reference.

which is equivalent to

$$\boxed{\begin{aligned} N-\frac{1}{2} &= \sum_{n=1}^{N} B_{2n}\binom{2N+1}{2n}, \\ N-1 &= \sum_{n=1}^{N-1} B_{2n}\binom{2N}{2n}. \end{aligned}} \tag{5.150}$$

From Eq. (5.150) the Bernoulli numbers in Table 5.1 are readily obtained. If the variable x in Eq. (5.144) is replaced by $2ix$ we obtain an alternate (and equivalent) definition of B_{2n} (B_1 is set equal to $-\frac{1}{2}$ by Eq. (5.146)) by the expression

$$x\cot x = \sum_{n=0}^{\infty}(-1)^n B_{2n}\frac{(2x)^{2n}}{(2n)!}, \qquad -\pi < x < \pi. \tag{5.151}$$

Using the method of residues (Section 7.1) or working from the infinite product representation of $\sin x$ (Section 5.11), we find that

$$B_{2n} = \frac{(-1)^{n-1}2(2n)!}{(2\pi)^{2n}}\sum_{p=1}^{\infty} p^{-2n}, \qquad n = 1, 2, 3, \ldots. \tag{5.152}$$

This representation of the Bernoulli numbers was discovered by Euler. It is readily seen from Eq. (5.152) that $|B_{2n}|$ increases without limit as $n \to \infty$. Numerical values have been calculated by Glaisher.[20] Illustrating the divergent behavior of the Bernoulli numbers, we have

$$B_{20} = -5.291 \times 10^{2}$$
$$B_{200} = -3.647 \times 10^{215}.$$

[20] J. W. L. Glaisher, table of the first 250 Bernoulli's numbers (to nine figures) and their logarithms (to ten figures). *Trans. Cambridge Philos. Soc.* **12**: 390 (1871–1879).

Some authors prefer to define the Bernoulli numbers with a modified version of Eq. (5.152) by using

$$\mathcal{B}_n = \frac{2(2n)!}{(2\pi)^{2n}} \sum_{p=1}^{\infty} p^{-2n}, \tag{5.153}$$

the subscript being just half of our subscript and all signs positive. Again, when using other texts or references, you must check to see exactly how the Bernoulli numbers are defined.

The Bernoulli numbers occur frequently in number theory. The von Staudt–Clausen theorem states that

$$B_{2n} = A_n - \frac{1}{p_1} - \frac{1}{p_2} - \frac{1}{p_3} - \cdots - \frac{1}{p_k}, \tag{5.154}$$

in which A_n is an integer and $p_1, p_2, \ldots, p_k$ are prime numbers so that $p_i - 1$ is a divisor of $2n$. It may readily be verified that this holds for

$$\begin{aligned} B_6 &(A_3 = 1, \quad p = 2, 3, 7), \\ B_8 &(A_4 = 1, \quad p = 2, 3, 5), \\ B_{10} &(A_5 = 1, \quad p = 2, 3, 11), \end{aligned} \tag{5.155}$$

and other special cases.

The Bernoulli numbers appear in the summation of integral powers of the integers,

$$\sum_{j=1}^{N} j^p, \qquad p \text{ integral},$$

and in numerous series expansions of the transcendental functions, including $\tan x$, $\cot x$, $\ln|\sin x|$, $(\sin x)^{-1}$, $\ln|\cos x|$, $\ln|\tan x|$, $(\cosh x)^{-1}$, $\tanh x$, and $\coth x$. For example,

$$\tan x = x + \frac{x^3}{3} + \frac{2}{15}x^5 + \cdots + \frac{(-1)^{n-1}2^{2n}(2^{2n}-1)B_{2n}}{(2n)!}x^{2n-1} + \cdots. \tag{5.156}$$

The Bernoulli numbers are likely to come in such series expansions because of the defining equations (5.144), (5.150), and (5.151) and because of their relation to the Riemann zeta function,

$$\zeta(2n) = \sum_{p=1}^{\infty} p^{-2n}. \tag{5.157}$$

Bernoulli Polynomials

If Eq. (5.144) is generalized slightly, we have

$$\boxed{\frac{xe^{xs}}{e^x - 1} = \sum_{n=0}^{\infty} B_n(s)\frac{x^n}{n!}} \tag{5.158}$$

Table 5.2 Bernoulli Polynomials

$B_0 = 1$
$B_1 = x - \frac{1}{2}$
$B_2 = x^2 - x + \frac{1}{6}$
$B_3 = x^3 - \frac{3}{2}x^2 + \frac{1}{2}x$
$B_4 = x^4 - 2x^3 + x^2 - \frac{1}{30}$
$B_5 = x^5 - \frac{5}{2}x^4 + \frac{5}{3}x^3 - \frac{1}{6}x$
$B_6 = x^6 - 3x^5 + \frac{5}{2}x^4 - \frac{1}{2}x^2 + \frac{1}{42}$
$B_n(0) = B_n$, Bernoulli number

defining the **Bernoulli polynomials**, $B_n(s)$. The first seven Bernoulli polynomials are given in Table 5.2.

From the generating function, Eq. (5.158),

$$B_n(0) = B_n, \qquad n = 0, 1, 2, \ldots, \tag{5.159}$$

the Bernoulli polynomial evaluated at zero equals the corresponding Bernoulli number. Two particularly important properties of the Bernoulli polynomials follow from the defining relation, Eq, (5.158): a differentiation relation

$$\frac{d}{ds} B_n(s) = n B_{n-1}(s), \qquad n = 1, 2, 3, \ldots, \tag{5.160}$$

and a symmetry relation (replace $x \to -x$ in Eq. (5.158) and then set $s = 1$)

$$B_n(1) = (-1)^n B_n(0), \qquad n = 1, 2, 3, \ldots. \tag{5.161}$$

These relations are used in the development of the Euler–Maclaurin integration formula.

Euler–Maclaurin Integration Formula

One use of the Bernoulli functions is in the derivation of the Euler–Maclaurin integration formula. This formula is used in Section 8.3 for the development of an asymptotic expression for the factorial function — Stirling's series.

The technique is repeated integration by parts, using Eq. (5.160) to create new derivatives. We start with

$$\int_0^1 f(x)\,dx = \int_0^1 f(x) B_0(x)\,dx. \tag{5.162}$$

From Eq. (5.160) and Exercise 5.9.2,

$$B_1'(x) = B_0(x) = 1. \tag{5.163}$$

Substituting $B_1'(x)$ into Eq. (5.162) and integrating by parts, we obtain

$$\int_0^1 f(x)\,dx = f(1)B_1(1) - f(0)B_1(0) - \int_0^1 f'(x)B_1(x)\,dx$$
$$= \frac{1}{2}\big[f(1) + f(0)\big] - \int_0^1 f'(x)B_1(x)\,dx. \tag{5.164}$$

Again using Eq. (5.160), we have

$$B_1(x) = \frac{1}{2}B_2'(x), \tag{5.165}$$

and integrating by parts we get

$$\int_0^1 f(x)\,dx = \frac{1}{2}\big[f(1) + f(0)\big] - \frac{1}{2!}\big[f'(1)B_2(1) - f'(0)B_2(0)\big]$$
$$+ \frac{1}{2!}\int_0^1 f^{(2)}(x)B_2(x)\,dx. \tag{5.166}$$

Using the relations

$$\begin{aligned} B_{2n}(1) &= B_{2n}(0) = B_{2n}, \qquad n = 0, 1, 2, \ldots \\ B_{2n+1}(1) &= B_{2n+1}(0) = 0, \qquad n = 1, 2, 3, \ldots \end{aligned} \tag{5.167}$$

and continuing this process, we have

$$\int_0^1 f(x)\,dx = \frac{1}{2}\big[f(1) + f(0)\big] - \sum_{p=1}^{q} \frac{1}{(2p)!}B_{2p}\big[f^{(2p-1)}(1) - f^{(2p-1)}(0)\big]$$
$$+ \frac{1}{(2q)!}\int_0^1 f^{(2q)}(x)B_{2q}(x)\,dx. \tag{5.168a}$$

This is the Euler–Maclaurin integration formula. It assumes that the function $f(x)$ has the required derivatives.

The range of integration in Eq. (5.168a) may be shifted from $[0, 1]$ to $[1, 2]$ by replacing $f(x)$ by $f(x + 1)$. Adding such results up to $[n - 1, n]$, we obtain

$$\int_0^n f(x)\,dx = \frac{1}{2}f(0) + f(1) + f(2) + \cdots + f(n-1) + \frac{1}{2}f(n)$$
$$- \sum_{p=1}^{q} \frac{1}{(2p)!}B_{2p}\big[f^{(2p-1)}(n) - f^{(2p-1)}(0)\big]$$
$$+ \frac{1}{(2q)!}\int_0^1 B_{2q}(x)\sum_{\nu=0}^{n-1} f^{(2q)}(x+\nu)\,dx. \tag{5.168b}$$

The terms $\frac{1}{2}f(0) + f(1) + \cdots + \frac{1}{2}f(n)$ appear exactly as in trapezoidal integration, or quadrature. The summation over p may be interpreted as a correction to the trapezoidal approximation. Equation (5.168b) may be seen as a generalization of Eq. (5.22); it is the

Table 5.3 Riemann Zeta Function

s	$\zeta(s)$
2	1.6449340668
3	1.2020569032
4	1.0823232337
5	1.0369277551
6	1.0173430620
7	1.0083492774
8	1.0040773562
9	1.0020083928
10	1.0009945751

form used in Exercise 5.9.5 for summing positive powers of integers and in Section 8.3 for the derivation of Stirling's formula.

The Euler–Maclaurin formula is often useful in summing series by converting them to integrals.[21]

Riemann Zeta Function

This series, $\sum_{p=1}^{\infty} p^{-2n}$, was used as a comparison series for testing convergence (Section 5.2) and in Eq. (5.152) as one definition of the Bernoulli numbers, B_{2n}. It also serves to define the Riemann zeta function by

$$\boxed{\zeta(s) \equiv \sum_{n=1}^{\infty} n^{-s}, \qquad s > 1.} \tag{5.169}$$

Table 5.3 lists the values of $\zeta(s)$ for integral s, $s = 2, 3, \ldots, 10$. Closed forms for even s appear in Exercise 5.9.6. Figure 5.11 is a plot of $\zeta(s) - 1$. An integral expression for this Riemann zeta function appears in Exercise 8.2.21 as part of the development of the gamma function, and the functional relation is given in Section 14.3.

The celebrated Euler prime number product for the Riemann zeta function may be derived as

$$\zeta(s)\left(1 - 2^{-s}\right) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \cdots - \left(\frac{1}{2^s} + \frac{1}{4^s} + \frac{1}{6^s} + \cdots\right); \tag{5.170}$$

eliminating all the n^{-s}, where n is a multiple of 2. Then

$$\zeta(s)\left(1 - 2^{-s}\right)\left(1 - 3^{-s}\right) = 1 + \frac{1}{3^s} + \frac{1}{5^s} + \frac{1}{7^s} + \frac{1}{9^s} + \cdots - \left(\frac{1}{3^s} + \frac{1}{9^s} + \frac{1}{15^s} + \cdots\right); \tag{5.171}$$

[21] See R. P. Boas and C. Stutz, Estimating sums with integrals. *Am. J. Phys.* **39**: 745 (1971), for a number of examples.

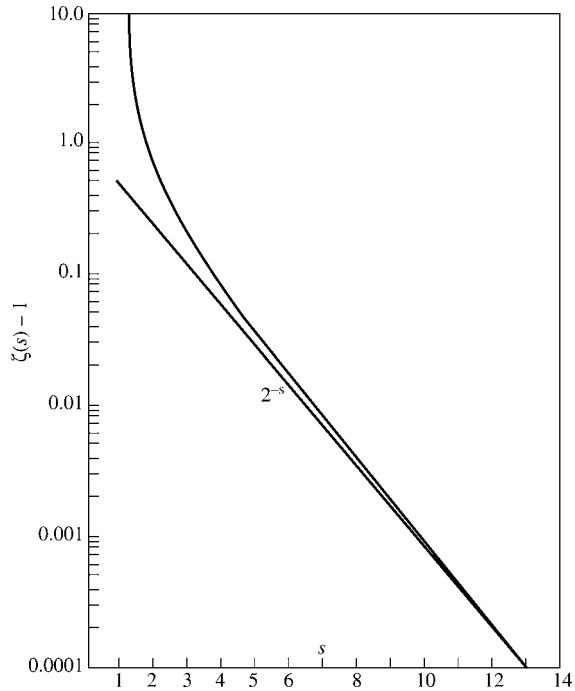

FIGURE 5.11 Riemann zeta function, $\zeta(s) - 1$ versus s.

eliminating all the remaining terms in which n is a multiple of 3. Continuing, we have $\zeta(s)(1-2^{-s})(1-3^{-s})(1-5^{-s})\cdots(1-P^{-s})$, where P is a prime number, and all terms n^{-s}, in which n is a multiple of any integer up through P, are canceled out. As $P \to \infty$,

$$\zeta(s)\left(1-2^{-s}\right)\left(1-3^{-s}\right)\cdots\left(1-P^{-s}\right) \to \zeta(s) \prod_{P(\text{prime})=2}^{\infty} \left(1-P^{-s}\right) = 1. \tag{5.172}$$

Therefore

$$\boxed{\zeta(s) = \prod_{P(\text{prime})=2}^{\infty} \left(1-P^{-s}\right)^{-1},} \tag{5.173}$$

giving $\zeta(s)$ as an infinite product.[22]

This cancellation procedure has a clear application in numerical computation. Equation (5.170) will give $\zeta(s)(1-2^{-s})$ to the same accuracy as Eq. (5.169) gives $\zeta(s)$, but

[22]This is the starting point for the extensive applications of the Riemann zeta function to analytic number theory. See H. M. Edwards, *Riemann's Zeta Function*. New York: Academic Press (1974); A. Ivić, *The Riemann Zeta Function*. New York: Wiley (1985); S. J. Patterson, *Introduction to the Theory of the Riemann Zeta Function*. Cambridge, UK: Cambridge University Press (1988).

with only half as many terms. (In either case, a correction would be made for the neglected tail of the series by the Maclaurin integral test technique — replacing the series by an integral, Section 5.2.)

Along with the Riemann zeta function, AMS-55 (Chapter 23. See Exercise 5.2.22 for the reference.) defines three other Dirichlet series related to $\zeta(s)$:

$$\eta(s) = \sum_{n=1}^{\infty} (-1)^{n-1} n^{-s} = \left(1 - 2^{1-s}\right)\zeta(s),$$

$$\lambda(s) = \sum_{n=0}^{\infty} (2n+1)^{-s} = \left(1 - 2^{-s}\right)\zeta(s),$$

and

$$\beta(s) = \sum_{n=0}^{\infty} (-1)^n (2n+1)^{-s}.$$

From the Bernoulli numbers (Exercise 5.9.6) or Fourier series (Example 14.3.3 and Exercise 14.3.13) special values are

$$\zeta(2) = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots = \frac{\pi^2}{6}$$

$$\zeta(4) = 1 + \frac{1}{2^4} + \frac{1}{3^4} + \cdots = \frac{\pi^4}{90}$$

$$\eta(2) = 1 - \frac{1}{2^2} + \frac{1}{3^2} + \cdots = \frac{\pi^2}{12}$$

$$\eta(4) = 1 - \frac{1}{2^4} + \frac{1}{3^4} + \cdots = \frac{7\pi^4}{720}$$

$$\lambda(2) = 1 + \frac{1}{3^2} + \frac{1}{5^2} + \cdots = \frac{\pi^2}{8}$$

$$\lambda(4) = 1 + \frac{1}{3^4} + \frac{1}{5^4} + \cdots = \frac{\pi^4}{96}$$

$$\beta(1) = 1 - \frac{1}{3} + \frac{1}{5} - \cdots = \frac{\pi}{4}$$

$$\beta(3) = 1 - \frac{1}{3^3} + \frac{1}{5^3} - \cdots = \frac{\pi^3}{32}.$$

Catalan's constant,

$$\beta(2) = 1 - \frac{1}{3^2} + \frac{1}{5^2} - \cdots = 0.91596559\ldots,$$

is the topic of Exercise 5.2.22.

Improvement of Convergence

If we are required to sum a convergent series $\sum_{n=1}^{\infty} a_n$ whose terms are rational functions of n, the convergence may be improved dramatically by introducing the Riemann zeta function.

Example 5.9.1 IMPROVEMENT OF CONVERGENCE

The problem is to evaluate the series $\sum_{n=1}^{\infty} 1/(1+n^2)$. Expanding $(1+n^2)^{-1} = n^{-2}(1+n^{-2})^{-1}$ by direct division, we have

$$\begin{aligned}\left(1+n^2\right)^{-1} &= n^{-2}\left(1-n^{-2}+n^{-4}-\frac{n^{-6}}{1+n^{-2}}\right)\\ &= \frac{1}{n^2}-\frac{1}{n^4}+\frac{1}{n^6}-\frac{1}{n^8+n^6}.\end{aligned}$$

Therefore

$$\sum_{n=1}^{\infty}\frac{1}{1+n^2} = \zeta(2)-\zeta(4)+\zeta(6)-\sum_{n=1}^{\infty}\frac{1}{n^8+n^6}.$$

The ζ values are tabulated and the remainder series converges as n^{-8}. Clearly, the process can be continued as desired. You make a choice between how much algebra you will do and how much arithmetic the computer will do. Other methods for improving computational effectiveness are given at the end of Sections 5.2 and 5.4. ■

Exercises

5.9.1 Show that

$$\tan x = \sum_{n=1}^{\infty}\frac{(-1)^{n-1}2^{2n}(2^{2n}-1)B_{2n}}{(2n)!}x^{2n-1}, \qquad -\frac{\pi}{2} < x < \frac{\pi}{2}.$$

Hint. $\tan x = \cot x - 2\cot 2x$.

5.9.2 Show that the first Bernoulli polynomials are

$$\begin{aligned}B_0(s) &= 1\\ B_1(s) &= s-\tfrac{1}{2}\\ B_2(s) &= s^2-s+\tfrac{1}{6}.\end{aligned}$$

Note that $B_n(0) = B_n$, the Bernoulli number.

5.9.3 Show that $B_n'(s) = nB_{n-1}(s)$, $n = 1, 2, 3, \ldots$.
Hint. Differentiate Eq. (5.158).

5.9.4 Show that

$$B_n(1) = (-1)^n B_n(0).$$

Hint. Go back to the generating function, Eq. (5.158), or Exercise 5.9.2.

5.9.5 The Euler–Maclaurin integration formula may be used for the evaluation of finite series:

$$\sum_{m=1}^{n} f(m) = \int_0^n f(x)\,dx + \frac{1}{2}f(1) + \frac{1}{2}f(n) + \frac{B_2}{2!}\big[f'(n) - f'(1)\big] + \cdots.$$

Show that

(a) $\displaystyle\sum_{m=1}^{n} m = \frac{1}{2}n(n+1).$

(b) $\displaystyle\sum_{m=1}^{n} m^2 = \frac{1}{6}n(n+1)(2n+1).$

(c) $\displaystyle\sum_{m=1}^{n} m^3 = \frac{1}{4}n^2(n+1)^2.$

(d) $\displaystyle\sum_{m=1}^{n} m^4 = \frac{1}{30}n(n+1)(2n+1)\big(3n^2+3n-1\big).$

5.9.6 From

$$B_{2n} = (-1)^{n-1}\frac{2(2n)!}{(2\pi)^{2n}}\zeta(2n),$$

show that

$$\text{(a) } \zeta(2) = \frac{\pi^2}{6} \qquad \text{(d) } \zeta(8) = \frac{\pi^8}{9450}$$

$$\text{(b) } \zeta(4) = \frac{\pi^4}{90} \qquad \text{(e) } \zeta(10) = \frac{\pi^{10}}{93,555}.$$

$$\text{(c) } \zeta(6) = \frac{\pi^6}{945}$$

5.9.7 Planck's blackbody radiation law involves the integral

$$\int_0^\infty \frac{x^3\,dx}{e^x - 1}.$$

Show that this equals $6\zeta(4)$. From Exercise 5.9.6,

$$\zeta(4) = \frac{\pi^4}{90}.$$

Hint. Make use of the gamma function, Chapter 8.

5.9.8 Prove that

$$\int_0^\infty \frac{x^n e^x\,dx}{(e^x-1)^2} = n!\zeta(n).$$

Assuming n to be real, show that each side of the equation diverges if $n = 1$. Hence the preceding equation carries the condition $n > 1$. Integrals such as this appear in the quantum theory of transport effects — thermal and electrical conductivity.

5.9.9 The Bloch–Gruneissen approximation for the resistance in a monovalent metal is

$$\rho = C\frac{T^5}{\Theta^6}\int_0^{\Theta/T} \frac{x^5\,dx}{(e^x-1)(1-e^{-x})},$$

where Θ is the Debye temperature characteristic of the metal.

(a) For $T \to \infty$, show that

$$\rho \approx \frac{C}{4}\cdot\frac{T}{\Theta^2}.$$

(b) For $T \to 0$, show that

$$\rho \approx 5!\zeta(5)C\frac{T^5}{\Theta^6}.$$

5.9.10 Show that

$$\text{(a)}\ \int_0^1 \frac{\ln(1+x)}{x}\,dx = \frac{1}{2}\zeta(2), \qquad \text{(b)}\ \lim_{a\to 1}\int_0^a \frac{\ln(1-x)}{x}\,dx = \zeta(2).$$

From Exercise 5.9.6, $\zeta(2) = \pi^2/6$. Note that the integrand in part (b) diverges for $a = 1$ but that the **integrated** series is convergent.

5.9.11 The integral

$$\int_0^1 \big[\ln(1-x)\big]^2\frac{dx}{x}$$

appears in the fourth-order correction to the magnetic moment of the electron. Show that it equals $2\zeta(3)$.
Hint. Let $1 - x = e^{-t}$.

5.9.12 Show that

$$\int_0^\infty \frac{(\ln z)^2}{1+z^2}dz = 4\left(1 - \frac{1}{3^3} + \frac{1}{5^3} - \frac{1}{7^3} + \cdots\right).$$

By contour integration (Exercise 7.1.17), this may be shown equal to $\pi^3/8$.

5.9.13 For "small" values of x,

$$\ln(x!) = -\gamma x + \sum_{n=2}^{\infty}(-1)^n\frac{\zeta(n)}{n}x^n,$$

where γ is the Euler–Mascheroni constant and $\zeta(n)$ is the Riemann zeta function. For what values of x does this series converge?

ANS. $-1 < x \le 1$.

Note that if $x = 1$, we obtain

$$\gamma = \sum_{n=2}^{\infty} (-1)^n \frac{\zeta(n)}{n},$$

a series for the Euler–Mascheroni constant. The convergence of this series is exceedingly slow. For actual computation of γ, other, indirect approaches are far superior (see Exercises 5.10.11, and 8.5.16).

5.9.14 Show that the series expansion of $\ln(x!)$ (Exercise 5.9.13) may be written as

$$\text{(a)} \quad \ln(x!) = \frac{1}{2}\ln\left(\frac{\pi x}{\sin \pi x}\right) - \gamma x - \sum_{n=1}^{\infty} \frac{\zeta(2n+1)}{2n+1} x^{2n+1},$$

$$\text{(b)} \quad \ln(x!) = \frac{1}{2}\ln\left(\frac{\pi x}{\sin \pi x}\right) - \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right) + (1-\gamma)x - \sum_{n=1}^{\infty} [\zeta(2n+1) - 1] \frac{x^{2n+1}}{2n+1}.$$

Determine the range of convergence of each of these expressions.

5.9.15 Show that Catalan's constant, $\beta(2)$, may be written as

$$\beta(2) = 2\sum_{k=1}^{\infty} (4k-3)^{-2} - \frac{\pi^2}{8}.$$

Hint. $\pi^2 = 6\zeta(2)$.

5.9.16 Derive the following expansions of the Debye functions for $n \geq 1$:

$$\int_0^x \frac{t^n\, dt}{e^t - 1} = x^n \left[\frac{1}{n} - \frac{x}{2(n+1)} + \sum_{k=1}^{\infty} \frac{B_{2k} x^{2k}}{(2k+n)(2k)!}\right], \qquad |x| < 2\pi;$$

$$\int_x^{\infty} \frac{t^n\, dt}{e^t - 1} = \sum_{k=1}^{\infty} e^{-kx} \left[\frac{x^n}{k} + \frac{n x^{n-1}}{k^2} + \frac{n(n-1)x^{n-2}}{k^3} + \cdots + \frac{n!}{k^{n+1}}\right]$$

for $x > 0$. The complete integral $(0, \infty)$ equals $n!\zeta(n+1)$, Exercise 8.2.15.

5.9.17 (a) Show that the equation $\ln 2 = \sum_{s=1}^{\infty} (-1)^{s+1} s^{-1}$ (Exercise 5.4.1) may be rewritten as

$$\ln 2 = \sum_{s=2}^{\infty} 2^{-s} \zeta(s) + \sum_{p=1}^{\infty} (2p)^{-n-1} \left[1 - \frac{1}{2p}\right]^{-1}.$$

Hint. Take the terms in pairs.

(b) Calculate $\ln 2$ to six significant figures.

5.9.18 (a) Show that the equation $\pi/4 = \sum_{n=1}^{\infty}(-1)^{n+1}(2n-1)^{-1}$ (Exercise 5.7.6) may be rewritten as

$$\frac{\pi}{4} = 1 - 2\sum_{s=1}^{\infty} 4^{-2s}\zeta(2s) - 2\sum_{p=1}^{\infty}(4p)^{-2n-2}\left[1 - \frac{1}{(4p)^2}\right]^{-1}.$$

(b) Calculate $\pi/4$ to six significant figures.

5.9.19 Write a function subprogram ZETA(N) that will calculate the Riemann zeta function for integer argument. Tabulate $\zeta(s)$ for $s = 2, 3, 4, \ldots, 20$. Check your values against Table 5.3 and AMS-55, Chapter 23. (See Exercise 5.2.22 for the reference.).
Hint. If you supply the function subprogram with the known values of $\zeta(2)$, $\zeta(3)$, and $\zeta(4)$, you avoid the more slowly converging series. Calculation time may be further shortened by using Eq. (5.170).

5.9.20 Calculate the logarithm (base 10) of $|B_{2n}|$, $n = 10, 20, \ldots, 100$.
Hint. Program $\zeta(n)$ as a function subprogram, Exercise 5.9.19.

Check values. $\log|B_{100}| = 78.45$
$\log|B_{200}| = 215.56$.

5.10 Asymptotic Series

Asymptotic series frequently occur in physics. In numerical computations they are employed for the accurate computation of a variety of functions. We consider here two types of integrals that lead to asymptotic series: first, an integral of the form

$$I_1(x) = \int_x^{\infty} e^{-u} f(u)\, du,$$

where the variable x appears as the lower limit of an integral. Second, we consider the form

$$I_2(x) = \int_0^{\infty} e^{-u} f\left(\frac{u}{x}\right) du,$$

with the function f to be expanded as a Taylor series (binomial series). Asymptotic series often occur as solutions of differential equations. An example of this appears in Section 11.6 as a solution of Bessel's equation.

Incomplete Gamma Function

The nature of an asymptotic series is perhaps best illustrated by a specific example. Suppose that we have the exponential integral function[23]

$$\mathrm{Ei}(x) = \int_{-\infty}^{x} \frac{e^u}{u} du, \tag{5.174}$$

[23]This function occurs frequently in astrophysical problems involving gas with a Maxwell–Boltzmann energy distribution.

or

$$-\mathrm{Ei}(-x) = \int_x^\infty \frac{e^{-u}}{u}\,du = E_1(x), \tag{5.175}$$

to be evaluated for large values of x. Or let us take a generalization of the incomplete factorial function (incomplete gamma function),[24]

$$I(x, p) = \int_x^\infty e^{-u}u^{-p}\,du = \Gamma(1-p, x), \tag{5.176}$$

in which x and p are positive. Again, we seek to evaluate it for large values of x.

Integrating by parts, we obtain

$$\begin{aligned} I(x, p) &= \frac{e^{-x}}{x^p} - p\int_x^\infty e^{-u}u^{-p-1}\,du \\ &= \frac{e^{-x}}{x^p} - \frac{pe^{-x}}{x^{p+1}} + p(p+1)\int_x^\infty e^{-u}u^{-p-2}\,du. \end{aligned} \tag{5.177}$$

Continuing to integrate by parts, we develop the series

$$\begin{aligned} I(x, p) = e^{-x}&\left(\frac{1}{x^p} - \frac{p}{x^{p+1}} + \frac{p(p+1)}{x^{p+2}} - \cdots + (-1)^{n-1}\frac{(p+n-2)!}{(p-1)!x^{p+n-1}}\right) \\ &+ (-1)^n\frac{(p+n-1)!}{(p-1)!}\int_x^\infty e^{-u}u^{-p-n}\,du. \end{aligned} \tag{5.178}$$

This is a remarkable series. Checking the convergence by the d'Alembert ratio test, we find

$$\lim_{n\to\infty}\frac{|u_{n+1}|}{|u_n|} = \lim_{n\to\infty}\frac{(p+n)!}{(p+n-1)!}\cdot\frac{1}{x} = \lim_{n\to\infty}\frac{p+n}{x} = \infty \tag{5.179}$$

for all finite values of x. Therefore our series as an infinite series diverges everywhere! Before discarding Eq. (5.178) as worthless, let us see how well a given partial sum approximates the incomplete factorial function, $I(x, p)$:

$$I(x, p) - s_n(x, p) = (-1)^{n+1}\frac{(p+n)!}{(p-1)!}\int_x^\infty e^{-u}u^{-p-n-1}\,du = R_n(x, p). \tag{5.180}$$

In absolute value

$$\left|I(x, p) - s_n(x, p)\right| \le \frac{(p+n)!}{(p-1)!}\int_x^\infty e^{-u}u^{-p-n-1}\,du.$$

When we substitute $u = v + x$, the integral becomes

$$\begin{aligned} \int_x^\infty e^{-u}u^{-p-n-1}\,du &= e^{-x}\int_0^\infty e^{-v}(v+x)^{-p-n-1}\,dv \\ &= \frac{e^{-x}}{x^{p+n+1}}\int_0^\infty e^{-v}\left(1+\frac{v}{x}\right)^{-p-n-1}\,dv. \end{aligned}$$

[24] See also Section 8.5.

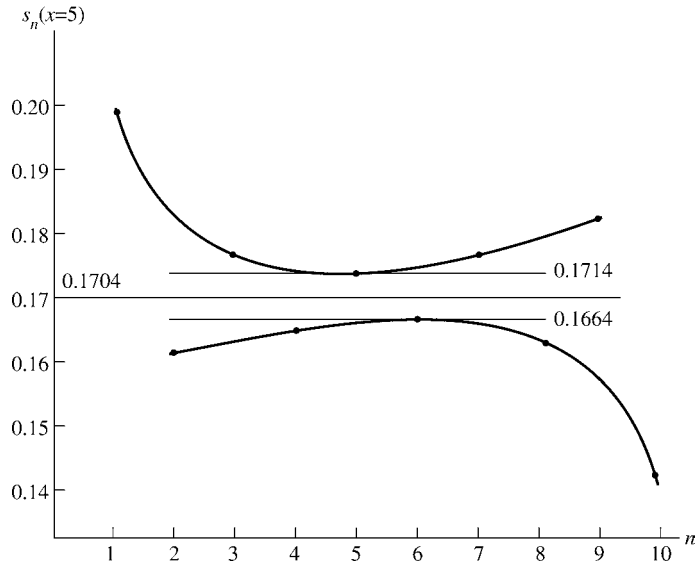

FIGURE 5.12 Partial sums of $e^x E_1(x)|_{x=5}$.

For large x the final integral approaches 1 and

$$\left|I(x,p) - s_n(x,p)\right| \approx \frac{(p+n)!}{(p-1)!} \cdot \frac{e^{-x}}{x^{p+n+1}}. \tag{5.181}$$

This means that if we take x large enough, our partial sum s_n is an arbitrarily good approximation to the function $I(x,p)$. Our divergent series (Eq. (5.178)) therefore is perfectly good for computations of partial sums. For this reason it is sometimes called a **semiconvergent** series. Note that the power of x in the denominator of the remainder $(p+n+1)$ is higher than the power of x in the last term included in $s_n(x,p)$, $(p+n)$.

Since the remainder $R_n(x,p)$ alternates in sign, the successive partial sums give alternately upper and lower bounds for $I(x,p)$. The behavior of the series (with $p=1$) as a function of the number of terms included is shown in Fig. 5.12. We have

$$\begin{aligned} e^x E_1(x) &= e^x \int_x^\infty \frac{e^{-u}}{u} du \\ &\cong s_n(x) = \frac{1}{x} - \frac{1!}{x^2} + \frac{2!}{x^2} - \frac{3!}{x^4} + \cdots + (-1)^n \frac{n!}{x^{n+1}}, \end{aligned} \tag{5.182}$$

which is evaluated at $x=5$. The optimum determination of $e^x E_1(x)$ is given by the closest approach of the upper and lower bounds, that is, between $s_4 = s_6 = 0.1664$ and $s_5 = 0.1741$ for $x=5$. Therefore

$$0.1664 \le e^x E_1(x)\big|_{x=5} \le 0.1741. \tag{5.183}$$

Actually, from tables,

$$e^x E_1(x)\big|_{x=5} = 0.1704, \tag{5.184}$$

within the limits established by our asymptotic expansion. Note that inclusion of additional terms in the series expansion beyond the optimum point literally reduces the accuracy of the representation. As x is increased, the spread between the lowest upper bound and the highest lower bound will diminish. By taking x large enough, one may compute $e^x E_1(x)$ to any desired degree of accuracy. Other properties of $E_1(x)$ are derived and discussed in Section 8.5.

Cosine and Sine Integrals

Asymptotic series may also be developed from definite integrals — if the integrand has the required behavior. As an example, the cosine and sine integrals (Section 8.5) are defined by

$$\mathrm{Ci}(x) = -\int_x^\infty \frac{\cos t}{t}\,dt, \tag{5.185}$$

$$\mathrm{si}(x) = -\int_x^\infty \frac{\sin t}{t}\,dt. \tag{5.186}$$

Combining these with regular trigonometric functions, we may define

$$\begin{aligned} f(x) &= \mathrm{Ci}(x)\sin x - \mathrm{si}(x)\cos x = \int_0^\infty \frac{\sin y}{y+x}\,dy, \\ g(x) &= -\mathrm{Ci}(x)\cos x - \mathrm{si}(x)\sin x = \int_0^\infty \frac{\cos y}{y+x}\,dy, \end{aligned} \tag{5.187}$$

with the new variable $y = t - x$. Going to complex variables, Section 6.1, we have

$$g(x) + if(x) = \int_0^\infty \frac{e^{iy}}{y+x}\,dy = \int_0^\infty \frac{ie^{-xu}}{1+iu}\,du, \tag{5.188}$$

in which $u = -iy/x$. The limits of integration, 0 to ∞, rather than 0 to $-i\infty$, may be justified by Cauchy's theorem, Section 6.3. Rationalizing the denominator and equating real part to real part and imaginary part to imaginary part, we obtain

$$g(x) = \int_0^\infty \frac{ue^{-xu}}{1+u^2}\,du, \qquad f(x) = \int_0^\infty \frac{e^{-xu}}{1+u^2}\,du. \tag{5.189}$$

For convergence of the integrals we must require that $\Re(x) > 0$.[25]

[25] $\Re(x)$ = real part of (complex) x (compare Section 6.1).

Now, to develop the asymptotic expansions, let $v = xu$ and expand the preceding factor $[1 + (v/x)^2]^{-1}$ by the binomial theorem.[26] We have

$$f(x) \approx \frac{1}{x}\int_0^\infty e^{-v} \sum_{0\le n\le N} (-1)^n \frac{v^{2n}}{x^{2n}} dv = \frac{1}{x} \sum_{0\le n\le N} (-1)^n \frac{(2n)!}{x^{2n}}, \tag{5.190}$$

$$g(x) \approx \frac{1}{x^2}\int_0^\infty e^{-v} \sum_{0\le n\le N} (-1)^n \frac{v^{2n+1}}{x^{2n}} dv = \frac{1}{x^2} \sum_{0\le n\le N} (-1)^n \frac{(2n+1)!}{x^{2n}}.$$

From Eqs. (5.187) and (5.190),

$$\begin{aligned} \mathrm{Ci}(x) &\approx \frac{\sin x}{x} \sum_{0\le n\le N} (-1)^n \frac{(2n)!}{x^{2n}} - \frac{\cos x}{x^2} \sum_{0\le n\le N} (-1)^n \frac{(2n+1)!}{x^{2n}}, \\ \mathrm{si}(x) &\approx -\frac{\cos x}{x} \sum_{0\le n\le N} (-1)^n \frac{(2n)!}{x^{2n}} - \frac{\sin x}{x^2} \sum_{0\le n\le N} (-1)^n \frac{(2n+1)!}{x^{2n}} \end{aligned} \tag{5.191}$$

are the desired asymptotic expansions.

This technique of expanding the integrand of a definite integral and integrating term by term is applied in Section 11.6 to develop an asymptotic expansion of the modified Bessel function K_ν and in Section 13.5 for expansions of the two confluent hypergeometric functions $M(a, c; x)$ and $U(a, c; x)$.

Definition of Asymptotic Series

The behavior of these series (Eqs. (5.178) and (5.191)), is consistent with the defining properties of an asymptotic series.[27] Following Poincaré, we take[28]

$$x^n R_n(x) = x^n\big[f(x) - s_n(x)\big], \tag{5.192}$$

where

$$s_n(x) = a_0 + \frac{a_1}{x} + \frac{a_2}{x^2} + \cdots + \frac{a_n}{x^n}. \tag{5.193}$$

The asymptotic expansion of $f(x)$ has the properties that

$$\lim_{x\to\infty} x^n R_n(x) = 0, \qquad \text{for fixed } n, \tag{5.194}$$

and

$$\lim_{n\to\infty} x^n R_n(x) = \infty, \qquad \text{for fixed } x.^{29} \tag{5.195}$$

[26] This step is valid for $v \le x$. The contributions from $v \ge x$ will be negligible (for large x) because of the negative exponential. It is because the binomial expansion does not converge for $v \ge x$ that our final series is asymptotic rather than convergent.

[27] It is not necessary that the asymptotic series be a power series. The required property is that the remainder $R_n(x)$ be of higher order than the last term kept — as in Eq. (5.194).

[28] Poincaré's definition allows (or neglects) exponentially decreasing functions. The refinement of Poincaré's definition is of considerable importance for the advanced theory of asymptotic expansions, particularly for extensions into the complex plane. However, for purposes of an introductory treatment and especially for numerical computation with x real and positive, Poincaré's approach is perfectly satisfactory.

See Eqs. (5.178) and (5.179) for an example of these properties. For power series, as assumed in the form of $s_n(x)$, $R_n(x) \sim x^{-n-1}$. With conditions (5.194) and (5.195) satisfied, we write

$$f(x) \approx \sum_{n=0}^{\infty} a_n x^{-n}. \tag{5.196}$$

Note the use of $\approx$ in place of $=$. The function $f(x)$ is equal to the series only in the limit as $x \to \infty$ and a finite number of terms in the series.

Asymptotic expansions of two functions may be multiplied together, and the result will be an asymptotic expansion of the product of the two functions.

The asymptotic expansion of a given function $f(t)$ may be integrated term by term (just as in a uniformly convergent series of continuous functions) from $x \le t < \infty$, and the result will be an asymptotic expansion of $\int_x^\infty f(t)\,dt$. Term-by-term differentiation, however, is valid only under very special conditions.

Some functions do not possess an asymptotic expansion; e^x is an example of such a function. However, if a function has an asymptotic expansion, it has only one. The correspondence is not one to one; many functions may have the same asymptotic expansion.

One of the most useful and powerful methods of generating asymptotic expansions, the method of steepest descents, will be developed in Section 7.3. Applications include the derivation of Stirling's formula for the (complete) factorial function (Section 8.3) and the asymptotic forms of the various Bessel functions (Section 11.6). Asymptotic series occur fairly often in mathematical physics. One of the earliest and still important approximations of quantum mechanics, the **WKB** expansion, is an asymptotic series.

Exercises

5.10.1 Stirling's formula for the logarithm of the factorial function is

$$\ln(x!) = \frac{1}{2}\ln 2\pi + \left(x + \frac{1}{2}\right)\ln x - x - \sum_{n=1}^{N} \frac{B_{2n}}{2n(2n-1)} x^{1-2n}.$$

The B_{2n} are the Bernoulli numbers (Section 5.9). Show that Stirling's formula is an **asymptotic** expansion.

5.10.2 Integrating by parts, develop asymptotic expansions of the Fresnel integrals.

$$\text{(a) } C(x) = \int_0^x \cos\frac{\pi u^2}{2}\,du, \qquad \text{(b) } s(x) = \int_0^x \sin\frac{\pi u^2}{2}\,du.$$

These integrals appear in the analysis of a knife-edge diffraction pattern.

5.10.3 Rederive the asymptotic expansions of $\mathrm{Ci}(x)$ and $\mathrm{si}(x)$ by repeated integration by parts.
Hint. $\mathrm{Ci}(x) + i\,\mathrm{si}(x) = -\int_x^\infty \frac{e^{it}}{t}\,dt$.

[29] This excludes convergent series of inverse powers of x. Some writers feel that this exclusion is artificial and unnecessary.

5.10.4 Derive the asymptotic expansion of the Gauss error function

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt$$
$$\approx 1 - \frac{e^{-x^2}}{\sqrt{\pi}x}\left(1 - \frac{1}{2x^2} + \frac{1\cdot 3}{2^2x^4} - \frac{1\cdot 3\cdot 5}{2^3x^6} + \cdots + (-1)^n \frac{(2n-1)!!}{2^n x^{2n}}\right).$$

Hint: $\operatorname{erf}(x) = 1 - \operatorname{erfc}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2}\, dt$.
Normalized so that $\operatorname{erf}(\infty) = 1$, this function plays an important role in probability theory. It may be expressed in terms of the Fresnel integrals (Exercise 5.10.2), the incomplete gamma functions (Section 8.5), and the confluent hypergeometric functions (Section 13.5).

5.10.5 The asymptotic expressions for the various Bessel functions, Section 11.6, contain the series

$$P_\nu(z) \sim 1 + \sum_{n=1}^{\infty} (-1)^n \frac{\prod_{s=1}^{2n}[4\nu^2 - (2s-1)^2]}{(2n)!(8z)^{2n}},$$

$$Q_\nu(z) \sim \sum_{n=1}^{\infty} (-1)^{n+1} \frac{\prod_{s=1}^{2n-1}[4\nu^2 - (2s-1)^2]}{(2n-1)!(8z)^{2n-1}}.$$

Show that these two series are indeed asymptotic series.

5.10.6 For $x > 1$,

$$\frac{1}{1+x} = \sum_{n=0}^{\infty} (-1)^n \frac{1}{x^{n+1}}.$$

Test this series to see if it is an asymptotic series.

5.10.7 Derive the following Bernoulli number asymptotic series for the Euler–Mascheroni constant:

$$\gamma = \sum_{s=1}^{n} s^{-1} - \ln n - \frac{1}{2n} + \sum_{k=1}^{N} \frac{B_{2k}}{(2k)n^{2k}}.$$

Hint. Apply the Euler–Maclaurin integration formula to $f(x) = x^{-1}$ over the interval $[1, n]$ for $N = 1, 2, \ldots$.

5.10.8 Develop an asymptotic series for

$$\int_0^\infty e^{-xv}\left(1 + v^2\right)^{-2} dv.$$

Take x to be real and positive.

ANS. $\dfrac{1}{x} - \dfrac{2!}{x^3} + \dfrac{4!}{x^5} - \cdots + \dfrac{(-1)^n(2n)!}{x^{2n+1}}.$

5.10.9 Calculate partial sums of $e^x E_1(x)$ for $x = 5$, 10, and 15 to exhibit the behavior shown in Fig. 5.11. Determine the width of the throat for $x = 10$ and 15, analogous to Eq. (5.183).

ANS. Throat width: $n = 10$, 0.000051
$n = 15$, 0.0000002.

5.10.10 The knife-edge diffraction pattern is described by

$$I = 0.5 I_0 \{ [C(u_0) + 0.5]^2 + [S(u_0) + 0.5]^2 \},$$

where $C(u_0)$ and $S(u_0)$ are the Fresnel integrals of Exercise 5.10.2. Here I_0 is the incident intensity and I is the diffracted intensity; u_0 is proportional to the distance away from the knife edge (measured at right angles to the incident beam). Calculate I/I_0 for u_0 varying from -1.0 to $+4.0$ in steps of 0.1. Tabulate your results and, if a plotting routine is available, plot them.

Check value. $u_0 = 1.0$, $I/I_0 = 1.259226$.

5.10.11 The Euler–Maclaurin integration formula of Section 5.9 provides a way of calculating the Euler–Mascheroni constant γ to high accuracy. Using $f(x) = 1/x$ in Eq. (5.168b) (with interval $[1, n]$) and the definition of γ (Eq. 5.28), we obtain

$$\gamma = \sum_{s=1}^{n} s^{-1} - \ln n - \frac{1}{2n} + \sum_{k=1}^{N} \frac{B_{2k}}{(2k) n^{2k}}.$$

Using double-precision arithmetic, calculate γ for $N = 1, 2, \ldots$.
Note. D. E. Knuth, Euler's constant to 1271 places. *Math. Comput.* **16**: 275 (1962). An even more precise calculation appears in Exercise 8.5.16.

ANS. For $n = 1000$, $N = 2$
$\gamma = 0.5772\ 1566\ 4901$.

5.11 INFINITE PRODUCTS

Consider a succession of positive factors $f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdots f_n (f_i > 0)$. Using capital pi ($\prod$) to indicate product, as capital sigma ($\sum$) indicates a sum, we have

$$f_1 \cdot f_2 \cdot f_3 \cdots f_n = \prod_{i=1}^{n} f_i. \tag{5.197}$$

We define p_n, a partial product, in analogy with s_n the partial sum,

$$p_n = \prod_{i=1}^{n} f_i \tag{5.198}$$

and then investigate the limit,

$$\lim_{n \to \infty} p_n = P. \tag{5.199}$$

If P is finite (but not zero), we say the infinite product is convergent. If P is infinite or zero, the infinite product is labeled divergent.

Since the product will diverge to infinity if

$$\lim_{n\to\infty} f_n > 1 \tag{5.200}$$

or to zero for

$$\lim_{n\to\infty} f_n < 1 \qquad (\text{and} > 0), \tag{5.201}$$

it is convenient to write our infinite products as

$$\prod_{n=1}^{\infty}(1+a_n).$$

The condition $a_n \to 0$ is then a necessary (but not sufficient) condition for convergence.

The infinite product may be related to an infinite series by the obvious method of taking the logarithm,

$$\ln\prod_{n=1}^{\infty}(1+a_n) = \sum_{n=1}^{\infty}\ln(1+a_n). \tag{5.202}$$

A more useful relationship is stated by the following theorem.

Convergence of Infinite Product

If $0 \le a_n < 1$, the infinite products $\prod_{n=1}^{\infty}(1+a_n)$ and $\prod_{n=1}^{\infty}(1-a_n)$ converge if $\sum_{n=1}^{\infty} a_n$ converges and diverge if $\sum_{n=1}^{\infty} a_n$ diverges.

Considering the term $1+a_n$, we see from Eq. (5.90) that

$$1 + a_n \le e^{a_n}. \tag{5.203}$$

Therefore for the partial product p_n, with s_n the partial sum of the a_i,

$$p_n \le e^{s_n}, \tag{5.204}$$

and letting $n \to \infty$,

$$\prod_{n=1}^{\infty}(1+a_n) \le \exp\sum_{n=1}^{\infty} a_n, \tag{5.205}$$

thus establishing an upper bound for the infinite product.

To develop a lower bound, we note that

$$p_n = 1 + \sum_{i=1}^{n} a_i + \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j + \cdots \ge s_n, \tag{5.206}$$

since $a_i \ge 0$. Hence

$$\prod_{n=1}^{\infty}(1+a_n) \ge \sum_{n=1}^{\infty} a_n. \tag{5.207}$$

If the infinite sum remains finite, the infinite product will also. If the infinite sum diverges, so will the infinite product.

The case of $\prod(1 - a_n)$ is complicated by the negative signs, but a proof that depends on the foregoing proof may be developed by noting that for $a_n < \frac{1}{2}$ (remember $a_n \to 0$ for convergence),

$$(1 - a_n) \le (1 + a_n)^{-1}$$

and

$$(1 - a_n) \ge (1 + 2a_n)^{-1}. \tag{5.208}$$

Sine, Cosine, and Gamma Functions

An nth-order polynomial $P_n(x)$ with n real roots may be written as a product of n factors (see Section 6.4, Gauss' fundamental theorem of algebra):

$$P_n(x) = (x - x_1)(x - x_2)\cdots(x - x_n) = \prod_{i=1}^{n}(x - x_i). \tag{5.209}$$

In much the same way we may expect that a function with an infinite number of roots may be written as an infinite product, one factor for each root. This is indeed the case for the trigonometric functions. We have two very useful infinite product representations,

$$\boxed{\sin x = x\prod_{n=1}^{\infty}\left(1 - \frac{x^2}{n^2\pi^2}\right),} \tag{5.210}$$

$$\boxed{\cos x = \prod_{n=1}^{\infty}\left[1 - \frac{4x^2}{(2n-1)^2\pi^2}\right].} \tag{5.211}$$

The most convenient and perhaps most elegant derivation of these two expressions is by the use of complex variables.[30] By our theorem of convergence, Eqs. (5.210) and (5.211) are convergent for all finite values of x. Specifically, for the infinite product for $\sin x$, $a_n = x^2/n^2\pi^2$,

$$\sum_{n=1}^{\infty} a_n = \frac{x^2}{\pi^2}\sum_{n=1}^{\infty} n^{-2} = \frac{x^2}{\pi^2}\zeta(2) = \frac{x^2}{6} \tag{5.212}$$

by Exercise 5.9.6. The series corresponding to Eq. (5.211) behaves in a similar manner.

Equation (5.210) leads to two interesting results. First, if we set $x = \pi/2$, we obtain

$$1 = \frac{\pi}{2}\prod_{n=1}^{\infty}\left[1 - \frac{1}{(2n)^2}\right] = \frac{\pi}{2}\prod_{n=1}^{\infty}\left[\frac{(2n)^2 - 1}{(2n)^2}\right]. \tag{5.213}$$

[30] See Eqs. (7.25) and (7.26).

Solving for $\pi/2$, we have

$$\frac{\pi}{2} = \prod_{n=1}^{\infty}\left[\frac{(2n)^2}{(2n-1)(2n+1)}\right] = \frac{2\cdot 2}{1\cdot 3}\cdot\frac{4\cdot 4}{3\cdot 5}\cdot\frac{6\cdot 6}{5\cdot 7}\cdots, \tag{5.214}$$

which is Wallis' famous formula for $\pi/2$.

The second result involves the gamma or factorial function (Section 8.1). One definition of the gamma function is

$$\boxed{\Gamma(x) = \left[xe^{\gamma x}\prod_{r=1}^{\infty}\left(1+\frac{x}{r}\right)e^{-x/r}\right]^{-1},} \tag{5.215}$$

where γ is the usual Euler–Mascheroni constant (compare Section 5.2). If we take the product of $\Gamma(x)$ and $\Gamma(-x)$, Eq. (5.215) leads to

$$\begin{aligned}\Gamma(x)\Gamma(-x) &= -\left[xe^{\gamma x}\prod_{r=1}^{\infty}\left(1+\frac{x}{r}\right)e^{-x/r}xe^{-\gamma x}\prod_{r=1}^{\infty}\left(1-\frac{x}{r}\right)e^{x/r}\right]^{-1}\\ &= -\frac{1}{x^2}\prod_{r=1}^{\infty}\left(1-\frac{x^2}{r^2}\right)^{-1}.\end{aligned} \tag{5.216}$$

Using Eq. (5.210) with x replaced by πx, we obtain

$$\Gamma(x)\Gamma(-x) = -\frac{\pi}{x\sin\pi x}. \tag{5.217}$$

Anticipating a recurrence relation developed in Section 8.1, we have $-x\Gamma(-x) = \Gamma(1-x)$. Equation (5.217) may be written as

$$\boxed{\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin\pi x}.} \tag{5.218}$$

This will be useful in treating the gamma function (Chapter 8).

Strictly speaking, we should check the range of x for which Eq. (5.215) is convergent. Clearly, individual factors will vanish for $x = 0, -1, -2, \ldots$. The proof that the infinite product converges for all other (finite) values of x is left as Exercise 5.11.9.

These infinite products have a variety of uses in mathematics. However, because of rather slow convergence, they are not suitable for precise numerical work in physics.

Exercises

5.11.1 Using

$$\ln\prod_{n=1}^{\infty}(1\pm a_n) = \sum_{n=1}^{\infty}\ln(1\pm a_n)$$

and the Maclaurin expansion of $\ln(1\pm a_n)$, show that the infinite product $\prod_{n=1}^{\infty}(1\pm a_n)$ converges or diverges with the infinite series $\sum_{n=1}^{\infty} a_n$.

5.11.2 An infinite product appears in the form

$$\prod_{n=1}^{\infty}\left(\frac{1+a/n}{1+b/n}\right),$$

where a and b are constants. Show that this infinite product converges only if $a = b$.

5.11.3 Show that the infinite product representations of $\sin x$ and $\cos x$ are consistent with the identity $2\sin x\cos x = \sin 2x$.

5.11.4 Determine the limit to which

$$\prod_{n=2}^{\infty}\left(1+\frac{(-1)^n}{n}\right)$$

converges.

5.11.5 Show that

$$\prod_{n=2}^{\infty}\left[1-\frac{2}{n(n+1)}\right]=\frac{1}{3}.$$

5.11.6 Prove that

$$\prod_{n=2}^{\infty}\left(1-\frac{1}{n^2}\right)=\frac{1}{2}.$$

5.11.7 Using the infinite product representations of $\sin x$, show that

$$x\cot x = 1 - 2\sum_{m,n=1}^{\infty}\left(\frac{x}{n\pi}\right)^{2m},$$

hence that the Bernoulli number

$$B_{2n} = (-1)^{n-1}\frac{2(2n)!}{(2\pi)^{2n}}\zeta(2n).$$

5.11.8 Verify the Euler identity

$$\prod_{p=1}^{\infty}\left(1+z^p\right)=\prod_{q=1}^{\infty}\left(1-z^{2q-1}\right)^{-1}, \qquad |z|<1.$$

5.11.9 Show that $\prod_{r=1}^{\infty}(1+x/r)e^{-x/r}$ converges for all finite x (except for the zeros of $1+x/r$).
Hint. Write the nth factor as $1+a_n$.

5.11.10 Calculate $\cos x$ from its infinite product representation, Eq. (5.211), using (a) 10, (b) 100, and (c) 1000 factors in the product. Calculate the absolute error. Note how slowly the partial products converge–making the infinite product quite unsuitable for precise numerical work.

ANS. For 1000 factors, $\cos\pi = -1.00051$.

Additional Readings

The topic of infinite series is treated in many texts on advanced calculus.

Bender, C. M., and S. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*. New York: McGraw-Hill (1978). Particularly recommended for methods of accelerating convergence.

Davis, H. T., *Tables of Higher Mathematical Functions*. Bloomington, IN: Principia Press (1935). Volume II contains extensive information on Bernoulli numbers and polynomials.

Dingle, R. B., *Asymptotic Expansions: Their Derivation and Interpretation*. New York: Academic Press (1973).

Galambos, J., *Representations of Real Numbers by Infinite Series*. Berlin: Springer (1976).

Gradshteyn, I. S., and I. M. Ryzhik, *Table of Integrals, Series and Products*. Corrected and enlarged 6th edition prepared by Alan Jeffrey. New York: Academic Press (2000).

Hamming, R. W., *Numerical Methods for Scientists and Engineers*. Reprinted, New York: Dover (1987).

Hansen, E., *A Table of Series and Products.* Englewood Cliffs, NJ: Prentice-Hall (1975). A tremendous compilation of series and products.

Hardy, G. H., *Divergent Series.* Oxford: Clarendon Press (1956), 2nd ed., Chelsea (1992). The standard, comprehensive work on methods of treating divergent series. Hardy includes instructive accounts of the gradual development of the concepts of convergence and divergence.

Jeffrey, A., *Handbook of Mathematical Formulas and Integrals*. San Diego: Academic Press (1995).

Knopp, K., *Theory and Application of Infinite Series.* London: Blackie and Son (2nd ed.); New York: Hafner (1971). Reprinted: A. K. Peters Classics (1997). This is a thorough, comprehensive, and authoritative work that covers infinite series and products. Proofs of almost all of the statements not proved in Chapter 5 will be found in this book.

Mangulis, V., *Handbook of Series for Scientists and Engineers*. New York: Academic Press (1965). A most convenient and useful collection of series. Includes algebraic functions, Fourier series, and series of the special functions: Bessel, Legendre, and so on.

Olver, F. W. J., *Asymptotics and Special Functions*. New York: Academic Press (1974). A detailed, readable development of asymptotic theory. Considerable attention is paid to error bounds for use in computation.

Rainville, E. D., *Infinite Series*. New York: Macmillan (1967). A readable and useful account of series constants and functions.

Sokolnikoff, I. S., and R. M. Redheffer, *Mathematics of Physics and Modern Engineering*, 2nd ed. New York: McGraw-Hill (1966). A long Chapter 2 (101 pages) presents infinite series in a thorough but very readable form. Extensions to the solutions of differential equations, to complex series, and to Fourier series are included.

This page intentionally left blank

CHAPTER 6

FUNCTIONS OF A COMPLEX VARIABLE I

ANALYTIC PROPERTIES, MAPPING

The imaginary numbers are a wonderful flight of God's spirit;
they are almost an amphibian between being and not being.

GOTTFRIED WILHELM VON LEIBNIZ, 1702

We turn now to a study of functions of a complex variable. In this area we develop some of the most powerful and widely useful tools in all of analysis. To indicate, at least partly, why complex variables are important, we mention briefly several areas of application.

1. For many pairs of functions u and v, both u and v satisfy Laplace's equation,

$$\nabla^2\psi = \frac{\partial^2\psi(x,y)}{\partial x^2} + \frac{\partial^2\psi(x,y)}{\partial y^2} = 0.$$

Hence either u or v may be used to describe a two-dimensional electrostatic potential. The other function, which gives a family of curves orthogonal to those of the first function, may then be used to describe the electric field $\mathbf{E}$. A similar situation holds for the hydrodynamics of an ideal fluid in irrotational motion. The function u might describe the velocity potential, whereas the function v would then be the stream function.

In many cases in which the functions u and v are unknown, mapping or transforming in the complex plane permits us to create a coordinate system tailored to the particular problem.

2. In Chapter 9 we shall see that the second-order differential equations of interest in physics may be solved by power series. The same power series may be used in the complex plane to replace x by the complex variable z. The dependence of the solution $f(z)$ at a given z_0 on the behavior of $f(z)$ elsewhere gives us greater insight into the behavior of our

solution and a powerful tool (analytic continuation) for extending the region in which the solution is valid.

3. The change of a parameter k from real to imaginary, $k \to ik$, transforms the Helmholtz equation into the diffusion equation. The same change transforms the Helmholtz equation solutions (Bessel and spherical Bessel functions) into the diffusion equation solutions (modified Bessel and modified spherical Bessel functions).

4. Integrals in the complex plane have a wide variety of useful applications:

- Evaluating definite integrals;
- Inverting power series;
- Forming infinite products;
- Obtaining solutions of differential equations for large values of the variable (asymptotic solutions);
- Investigating the stability of potentially oscillatory systems;
- Inverting integral transforms.

5. Many physical quantities that were originally real become complex as a simple physical theory is made more general. The real index of refraction of light becomes a complex quantity when absorption is included. The real energy associated with an energy level becomes complex when the finite lifetime of the level is considered.

6.1 COMPLEX ALGEBRA

A complex number is nothing more than an ordered pair of two real numbers, (a, b). Similarly, a complex variable is an ordered pair of two real variables,[1]

$$z \equiv (x, y). \tag{6.1}$$

The ordering is significant. In general (a, b) is not equal to (b, a) and (x, y) is not equal to (y, x). As usual, we continue writing a real number $(x, 0)$ simply as x, and we call $i \equiv (0, 1)$ the imaginary unit.

All our complex variable analysis can be developed in terms of ordered pairs of numbers (a, b), variables (x, y), and functions $(u(x, y), v(x, y))$.

We now define **addition** of complex numbers in terms of their Cartesian components as

$$\boxed{z_1 + z_2 = (x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2),} \tag{6.2a}$$

that is, two-dimensional vector addition. In Chapter 1 the points in the xy-plane are identified with the two-dimensional displacement vector $\mathbf{r} = \hat{\mathbf{x}}x + \hat{\mathbf{y}}y$. As a result, two-dimensional vector analogs can be developed for much of our complex analysis. Exercise 6.1.2 is one simple example; Cauchy's theorem, Section 6.3, is another.

Multiplication of complex numbers is defined as

$$\boxed{z_1 z_2 = (x_1, y_1) \cdot (x_2, y_2) = (x_1 x_2 - y_1 y_2, x_1 y_2 + x_2 y_1).} \tag{6.2b}$$

[1] This is precisely how a computer does complex arithmetic.

Using Eq. (6.2b) we verify that $i^2 = (0, 1) \cdot (0, 1) = (-1, 0) = -1$, so we can also identify $i = \sqrt{-1}$, as usual and further rewrite Eq. (6.1) as

$$\boxed{z = (x, y) = (x, 0) + (0, y) = x + (0, 1) \cdot (y, 0) = x + iy.} \tag{6.2c}$$

Clearly, the i is not necessary here but it is convenient. It serves to keep pairs in order — somewhat like the unit vectors of Chapter 1.[2]

Permanence of Algebraic Form

All our elementary functions, e^z, $\sin z$, and so on, can be extended into the complex plane (compare Exercise 6.1.9). For instance, they can be defined by power-series expansions, such as

$$e^z = 1 + \frac{z}{1!} + \frac{z^2}{2!} + \cdots = \sum_{n=0}^{\infty} \frac{z^n}{n!} \tag{6.3}$$

for the exponential. Such definitions agree with the real variable definitions along the real x-axis and extend the corresponding real functions into the complex plane. This result is often called **permanence of the algebraic form**.

It is convenient to employ a graphical representation of the complex variable. By plotting x — the real part of z — as the abscissa and y — the imaginary part of z — as the ordinate, we have the complex plane, or Argand plane, shown in Fig. 6.1. If we assign specific values to x and y, then z corresponds to a point (x, y) in the plane. In terms of the ordering mentioned before, it is obvious that the point (x, y) does not coincide with the point (y, x) except for the special case of $x = y$. Further, from Fig. 6.1 we may write

$$x = r\cos\theta, \qquad y = r\sin\theta \tag{6.4a}$$

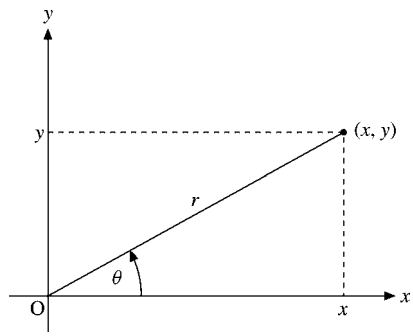

FIGURE 6.1 Complex plane — Argand diagram.

[2]The algebra of complex numbers, (a, b), is isomorphic with that of matrices of the form

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

(compare Exercise 3.2.4).

and

$$z = r(\cos\theta + i\sin\theta). \tag{6.4b}$$

Using a result that is suggested (but not rigorously proved)[3] by Section 5.6 and Exercise 5.6.1, we have the useful polar representation

$$\boxed{z = r(\cos\theta + i\sin\theta) = re^{i\theta}.} \tag{6.4c}$$

In order to prove this identity, we use $i^3 = -i,\ i^4 = 1, \ldots$ in the Taylor expansion of the exponential and trigonometric functions and separate even and odd powers in

$$\begin{aligned} e^{i\theta} &= \sum_{n=0}^{\infty} \frac{(i\theta)^n}{n!} = \sum_{\nu=0}^{\infty} \frac{(i\theta)^{2\nu}}{(2\nu)!} + \sum_{\nu=0}^{\infty} \frac{(i\theta)^{2\nu+1}}{(2\nu+1)!} \\ &= \sum_{\nu=0}^{\infty} (-1)^{\nu} \frac{\theta^{2\nu}}{(2\nu)!} + i\sum_{\nu=0}^{\infty} (-1)^{\nu} \frac{\theta^{2\nu+1}}{(2\nu+1)!} = \cos\theta + i\sin\theta. \end{aligned}$$

For the special values $\theta = \pi/2$ and $\theta = \pi$, we obtain

$$e^{i\pi/2} = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2} = i, \qquad e^{i\pi} = \cos(\pi) = -1,$$

intriguing connections between e, i, and π. Moreover, the exponential function $e^{i\theta}$ is periodic with period 2π, just like $\sin\theta$ and $\cos\theta$.

In this representation r is called the **modulus** or **magnitude** of z ($r = |z| = (x^2 + y^2)^{1/2}$) and the angle θ ($= \tan^{-1}(y/x)$) is labeled the argument or **phase** of z. (Note that the arctan function $\tan^{-1}(y/x)$ has infinitely many branches.)

The choice of polar representation, Eq. (6.4c), or Cartesian representation, Eqs. (6.1) and (6.2c), is a matter of convenience. Addition and subtraction of complex variables are easier in the Cartesian representation, Eq. (6.2a). Multiplication, division, powers, and roots are easier to handle in polar form, Eq. (6.4c).

Analytically or graphically, using the vector analogy, we may show that the modulus of the sum of two complex numbers is no greater than the sum of the moduli and no less than the difference, Exercise 6.1.3,

$$|z_1| - |z_2| \le |z_1 + z_2| \le |z_1| + |z_2|. \tag{6.5}$$

Because of the vector analogy, these are called the **triangle** inequalities.

Using the polar form, Eq. (6.4c), we find that the magnitude of a product is the product of the magnitudes:

$$|z_1 \cdot z_2| = |z_1| \cdot |z_2|. \tag{6.6}$$

Also,

$$\arg(z_1 \cdot z_2) = \arg z_1 + \arg z_2. \tag{6.7}$$

[3] Strictly speaking, Chapter 5 was limited to real variables. The development of power-series expansions for complex functions is taken up in Section 6.5 (Laurent expansion).

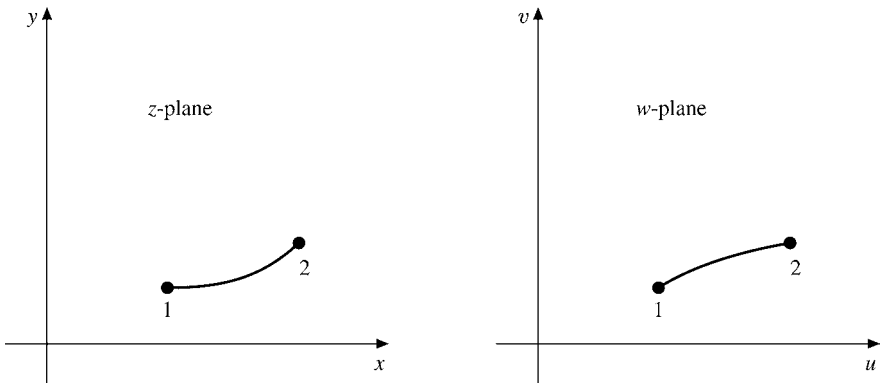

FIGURE 6.2 The function $w(z) = u(x, y) + iv(x, y)$ maps points in the xy-plane into points in the uv-plane.

From our complex variable z complex functions $f(z)$ or $w(z)$ may be constructed. These complex functions may then be resolved into real and imaginary parts,

$$w(z) = u(x, y) + iv(x, y), \tag{6.8}$$

in which the separate functions $u(x, y)$ and $v(x, y)$ are pure real. For example, if $f(z) = z^2$, we have

$$f(z) = (x + iy)^2 = \left(x^2 - y^2\right) + i2xy.$$

The **real part** of a function $f(z)$ will be labeled $\Re f(z)$, whereas the **imaginary part** will be labeled $\Im f(z)$. In Eq. (6.8)

$$\Re w(z) = \mathrm{Re}(w) = u(x, y), \qquad \Im w(z) = \mathrm{Im}(w) = v(x, y).$$

The relationship between the independent variable z and the dependent variable w is perhaps best pictured as a mapping operation. A given $z = x + iy$ means a given point in the z-plane. The complex value of $w(z)$ is then a point in the w-plane. Points in the z-plane map into points in the w-plane and curves in the z-plane map into curves in the w-plane, as indicated in Fig. 6.2.

Complex Conjugation

In all these steps, complex number, variable, and function, the operation of replacing i by $-i$ is called "taking the complex conjugate." The complex conjugate of z is denoted by z^*, where[4]

$$\boxed{z^* = x - iy.} \tag{6.9}$$

[4]The complex conjugate is often denoted by $\bar{z}$ in the mathematical literature.

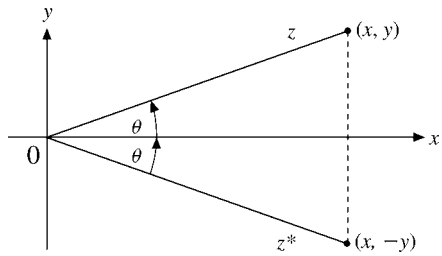

FIGURE 6.3 Complex conjugate points.

The complex variable z and its complex conjugate z^* are mirror images of each other reflected in the x-axis, that is, inversion of the y-axis (compare Fig. 6.3). The product zz^* leads to

$$\boxed{zz^* = (x + iy)(x - iy) = x^2 + y^2 = r^2.} \tag{6.10}$$

Hence

$$(zz^*)^{1/2} = |z|,$$

the **magnitude** of z.

Functions of a Complex Variable

All the elementary functions of real variables may be extended into the complex plane—replacing the real variable x by the complex variable z. This is an example of the analytic continuation mentioned in Section 6.5. The extremely important relation of Eq. (6.4c) is an illustration. Moving into the complex plane opens up new opportunities for analysis.

Example 6.1.1 DE MOIVRE'S FORMULA

If Eq. (6.4c) (setting $r = 1$) is raised to the nth power, we have

$$e^{in\theta} = (\cos\theta + i\sin\theta)^n. \tag{6.11}$$

Expanding the exponential now with argument $n\theta$, we obtain

$$\cos n\theta + i\sin n\theta = (\cos\theta + i\sin\theta)^n. \tag{6.12}$$

De Moivre's formula is generated if the right-hand side of Eq. (6.12) is expanded by the binomial theorem; we obtain $\cos n\theta$ as a series of powers of $\cos\theta$ and $\sin\theta$, Exercise 6.1.6. ■

Numerous other examples of relations among the exponential, hyperbolic, and trigonometric functions in the complex plane appear in the exercises.

Occasionally there are complications. The logarithm of a complex variable may be expanded using the polar representation

$$\ln z = \ln re^{i\theta} = \ln r + i\theta. \tag{6.13a}$$

This is not complete. To the phase angle, θ, we may add any integral multiple of 2π without changing z. Hence Eq. (6.13a) should read

$$\ln z = \ln r e^{i(\theta+2n\pi)} = \ln r + i(\theta + 2n\pi). \tag{6.13b}$$

The parameter n may be any integer. This means that $\ln z$ is a **multivalued** function having an infinite number of values for a single pair of real values r and θ. To avoid ambiguity, the simplest choice is $n = 0$ and limitation of the phase to an interval of length 2π, such as $(-\pi, \pi)$.[5] The line in the z-plane that is not crossed, the negative real axis in this case, is labeled a **cut line** or **branch cut**. The value of $\ln z$ with $n = 0$ is called the **principal value** of $\ln z$. Further discussion of these functions, including the logarithm, appears in Section 6.7.

Exercises

6.1.1 (a) Find the reciprocal of $x + iy$, working entirely in the Cartesian representation.
(b) Repeat part (a), working in polar form but expressing the final result in Cartesian form.

6.1.2 The complex quantities $a = u + iv$ and $b = x + iy$ may also be represented as two-dimensional vectors $\mathbf{a} = \hat{\mathbf{x}}u + \hat{\mathbf{y}}v$, $\mathbf{b} = \hat{\mathbf{x}}x + \hat{\mathbf{y}}y$. Show that

$$a^* b = \mathbf{a} \cdot \mathbf{b} + i\hat{\mathbf{z}} \cdot \mathbf{a} \times \mathbf{b}.$$

6.1.3 Prove algebraically that for complex numbers,

$$|z_1| - |z_2| \le |z_1 + z_2| \le |z_1| + |z_2|.$$

Interpret this result in terms of two-dimensional vectors. Prove that

$$|z - 1| < \left|\sqrt{z^2 - 1}\right| < |z + 1|, \qquad \text{for} \quad \Re(z) > 0.$$

6.1.4 We may define a complex conjugation operator K such that $Kz = z^*$. Show that K is not a linear operator.

6.1.5 Show that complex numbers have square roots and that the square roots are contained in the complex plane. What are the square roots of i?

6.1.6 Show that

(a) $\cos n\theta = \cos^n \theta - \binom{n}{2} \cos^{n-2}\theta \sin^2\theta + \binom{n}{4} \cos^{n-4}\theta \sin^4\theta - \cdots$.
(b) $\sin n\theta = \binom{n}{1} \cos^{n-1}\theta \sin\theta - \binom{n}{3} \cos^{n-3}\theta \sin^3\theta + \cdots$.

Note. The quantities $\binom{n}{m}$ are binomial coefficients: $\binom{n}{m} = n!/[(n-m)!m!]$.

6.1.7 Prove that

(a) $$\sum_{n=0}^{N-1} \cos nx = \frac{\sin(Nx/2)}{\sin x/2} \cos(N-1)\frac{x}{2},$$

[5] There is no standard choice of phase; the appropriate phase depends on each problem.

(b) $$\sum_{n=0}^{N-1} \sin nx = \frac{\sin(Nx/2)}{\sin x/2} \sin(N-1)\frac{x}{2}.$$

These series occur in the analysis of the multiple-slit diffraction pattern. Another application is the analysis of the Gibbs phenomenon, Section 14.5.
Hint. Parts (a) and (b) may be combined to form a geometric series (compare Section 5.1).

6.1.8 For $-1 < p < 1$ prove that

(a) $$\sum_{n=0}^{\infty} p^n \cos nx = \frac{1 - p\cos x}{1 - 2p\cos x + p^2},$$

(b) $$\sum_{n=0}^{\infty} p^n \sin nx = \frac{p\sin x}{1 - 2p\cos x + p^2}.$$

These series occur in the theory of the Fabry–Perot interferometer.

6.1.9 Assume that the trigonometric functions and the hyperbolic functions are defined for complex argument by the appropriate power series

$$\sin z = \sum_{n=1,\text{odd}}^{\infty} (-1)^{(n-1)/2} \frac{z^n}{n!} = \sum_{s=0}^{\infty} (-1)^s \frac{z^{2s+1}}{(2s+1)!},$$

$$\cos z = \sum_{n=0,\text{even}}^{\infty} (-1)^{n/2} \frac{z^n}{n!} = \sum_{s=0}^{\infty} (-1)^s \frac{z^{2s}}{(2s)!},$$

$$\sinh z = \sum_{n=1,\text{odd}}^{\infty} \frac{z^n}{n!} = \sum_{s=0}^{\infty} \frac{z^{2s+1}}{(2s+1)!},$$

$$\cosh z = \sum_{n=0,\text{even}}^{\infty} \frac{z^n}{n!} = \sum_{s=0}^{\infty} \frac{z^{2s}}{(2s)!}.$$

(a) Show that

$$i\sin z = \sinh iz, \qquad \sin iz = i\sinh z,$$
$$\cos z = \cosh iz, \qquad \cos iz = \cosh z.$$

(b) Verify that familiar functional relations such as

$$\cosh z = \tfrac{e^z + e^{-z}}{2},$$
$$\sin(z_1 + z_2) = \sin z_1 \cos z_2 + \sin z_2 \cos z_1,$$

still hold in the complex plane.

6.1.10 Using the identities

$$\cos z = \frac{e^{iz} + e^{-iz}}{2}, \qquad \sin z = \frac{e^{iz} - e^{-iz}}{2i},$$

established from comparison of power series, show that

(a) $\sin(x + iy) = \sin x \cosh y + i \cos x \sinh y,$

$\cos(x + iy) = \cos x \cosh y - i \sin x \sinh y,$

(b) $|\sin z|^2 = \sin^2 x + \sinh^2 y, \qquad |\cos z|^2 = \cos^2 x + \sinh^2 y.$

This demonstrates that we may have $|\sin z|, |\cos z| > 1$ in the complex plane.

6.1.11 From the identities in Exercises 6.1.9 and 6.1.10 show that

(a) $\sinh(x + iy) = \sinh x \cos y + i \cosh x \sin y,$

$\cosh(x + iy) = \cosh x \cos y + i \sinh x \sin y,$

(b) $|\sinh z|^2 = \sinh^2 x + \sin^2 y, \qquad |\cosh z|^2 = \cosh^2 x + \sin^2 y.$

6.1.12 Prove that

(a) $|\sin z| \geq |\sin x|$ (b) $|\cos z| \geq |\cos x|.$

6.1.13 Show that the exponential function e^z is periodic with a pure imaginary period of $2\pi i$.

6.1.14 Show that

(a) $\tanh \frac{z}{2} = \frac{\sinh x + i \sin y}{\cosh x + \cos y}, \qquad$ (b) $\coth \frac{z}{2} = \frac{\sinh x - i \sin y}{\cosh x - \cos y}.$

6.1.15 Find all the zeros of

(a) $\sin z,$ (b) $\cos z,$ (c) $\sinh z,$ (d) $\cosh z.$

6.1.16 Show that

(a) $\sin^{-1} z = -i \ln\left(iz \pm \sqrt{1 - z^2}\right),$

(b) $\cos^{-1} z = -i \ln\left(z \pm \sqrt{z^2 - 1}\right),$

(c) $\tan^{-1} z = \frac{i}{2} \ln\left(\frac{i + z}{i - z}\right),$

(d) $\sinh^{-1} z = \ln\left(z + \sqrt{z^2 + 1}\right),$

(e) $\cosh^{-1} z = \ln\left(z + \sqrt{z^2 - 1}\right),$

(f) $\tanh^{-1} z = \frac{1}{2} \ln\left(\frac{1 + z}{1 - z}\right).$

Hint. 1. Express the trigonometric and hyperbolic functions in terms of exponentials. 2. Solve for the exponential and then for the exponent.

6.1.17 In the quantum theory of the photoionization we encounter the identity

$$\left(\frac{ia - 1}{ia + 1}\right)^{ib} = \exp\left(-2b \cot^{-1} a\right),$$

in which a and b are real. Verify this identity.

6.1.18 A plane wave of light of angular frequency ω is represented by

$$e^{i\omega(t-nx/c)}.$$

In a certain substance the simple real index of refraction n is replaced by the complex quantity $n - ik$. What is the effect of k on the wave? What does k correspond to physically? The generalization of a quantity from real to complex form occurs frequently in physics. Examples range from the complex Young's modulus of viscoelastic materials to the complex (optical) potential of the "cloudy crystal ball" model of the atomic nucleus.

6.1.19 We see that for the angular momentum components defined in Exercise 2.5.14,

$$L_x - iL_y \neq (L_x + iL_y)^*.$$

Explain why this occurs.

6.1.20 Show that the **phase** of $f(z) = u + iv$ is equal to the imaginary part of the logarithm of $f(z)$. Exercise 8.2.13 depends on this result.

6.1.21 (a) Show that $e^{\ln z}$ always equals z.
(b) Show that $\ln e^z$ does not always equal z.

6.1.22 The infinite product representations of Section 5.11 hold when the real variable x is replaced by the complex variable z. From this, develop infinite product representations for
(a) $\sinh z$, (b) $\cosh z$.

6.1.23 The equation of motion of a mass m **relative to a rotating coordinate system** is

$$m\frac{d^2\mathbf{r}}{dt^2} = \mathbf{F} - m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) - 2m\left(\boldsymbol{\omega} \times \frac{d\mathbf{r}}{dt}\right) - m\left(\frac{d\boldsymbol{\omega}}{dt} \times \mathbf{r}\right).$$

Consider the case $\mathbf{F} = 0$, $\mathbf{r} = \hat{\mathbf{x}}x + \hat{\mathbf{y}}y$, and $\boldsymbol{\omega} = \omega\hat{\mathbf{z}}$, with ω constant. Show that the replacement of $\mathbf{r} = \hat{\mathbf{x}}x + \hat{\mathbf{y}}y$ by $z = x + iy$ leads to

$$\frac{d^2z}{dt^2} + i2\omega\frac{dz}{dt} - \omega^2 z = 0.$$

Note. This ODE may be solved by the substitution $z = fe^{-i\omega t}$.

6.1.24 Using the complex arithmetic available in FORTRAN, write a program that will calculate the complex exponential e^z from its series expansion (definition). Calculate e^z for $z = e^{in\pi/6}$, $n = 0, 1, 2, \ldots, 12$. Tabulate the phase angle ($\theta = n\pi/6$), $\Re z$, $\Im z$, $\Re(e^z)$, $\Im(e^z)$, $|e^z|$, and the phase of e^z.

Check value. $n = 5, \theta = 2.61799, \Re(z) = -0.86602,$
$\Im z = 0.50000, \Re(e^z) = 0.36913, \Im(e^z) = 0.20166,$
$|e^z| = 0.42062,$ phase$(e^z) = 0.50000.$

6.1.25 Using the complex arithmetic available in FORTRAN, calculate and tabulate $\Re(\sinh z)$, $\Im(\sinh z)$, $|\sinh z|$, and phase $(\sinh z)$ for $x = 0.0(0.1)1.0$ and $y = 0.0(0.1)1.0$.

Hint. Beware of dividing by zero when calculating an angle as an arc tangent.

Check value. $z = 0.2 + 0.1i$, $\Re(\sinh z) = 0.20033$, $\Im(\sinh z) = 0.10184$, $|\sinh z| = 0.22473$, $\text{phase}(\sinh z) = 0.47030$.

6.1.26 Repeat Exercise 6.1.25 for $\cosh z$.

6.2 CAUCHY–RIEMANN CONDITIONS

Having established complex functions of a complex variable, we now proceed to differentiate them. The derivative of $f(z)$, like that of a real function, is defined by

$$\lim_{\delta z \to 0} \frac{f(z+\delta z) - f(z)}{z + \delta z - z} = \lim_{\delta z \to 0} \frac{\delta f(z)}{\delta z} = \frac{df}{dz} = f'(z), \tag{6.14}$$

provided that the limit is **independent** of the particular approach to the point z. For real variables we require that the right-hand limit ($x \to x_0$ from above) and the left-hand limit ($x \to x_0$ from below) be equal for the derivative $df(x)/dx$ to exist at $x = x_0$. Now, with z (or z_0) some point in a plane, our requirement that the limit be independent of the direction of approach is very restrictive.

Consider increments δx and δy of the variables x and y, respectively. Then

$$\delta z = \delta x + i\delta y. \tag{6.15}$$

Also,

$$\delta f = \delta u + i\delta v, \tag{6.16}$$

so that

$$\frac{\delta f}{\delta z} = \frac{\delta u + i\delta v}{\delta x + i\delta y}. \tag{6.17}$$

Let us take the limit indicated by Eq. (6.14) by two different approaches, as shown in Fig. 6.4. First, with $\delta y = 0$, we let $\delta x \to 0$. Equation (6.14) yields

$$\lim_{\delta z \to 0} \frac{\delta f}{\delta z} = \lim_{\delta x \to 0} \left(\frac{\delta u}{\delta x} + i \frac{\delta v}{\delta x} \right) = \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x}, \tag{6.18}$$

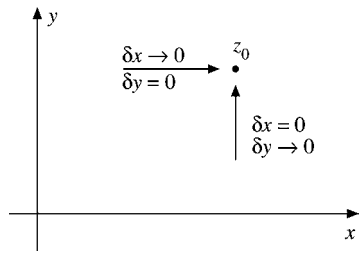

FIGURE 6.4 Alternate approaches to z_0.

assuming the partial derivatives exist. For a second approach, we set $\delta x = 0$ and then let $\delta y \to 0$. This leads to

$$\lim_{\delta z \to 0} \frac{\delta f}{\delta z} = \lim_{\delta y \to 0} \left(-i \frac{\delta u}{\delta y} + \frac{\delta v}{\delta y} \right) = -i \frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}. \tag{6.19}$$

If we are to have a derivative df/dz, Eqs. (6.18) and (6.19) must be identical. Equating real parts to real parts and imaginary parts to imaginary parts (like components of vectors), we obtain

$$\boxed{\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.} \tag{6.20}$$

These are the famous **Cauchy–Riemann** conditions. They were discovered by Cauchy and used extensively by Riemann in his theory of analytic functions. These Cauchy–Riemann conditions are necessary for the existence of a derivative of $f(z)$; that is, if df/dz exists, the Cauchy–Riemann conditions must hold.

Conversely, if the Cauchy–Riemann conditions are satisfied and the partial derivatives of $u(x, y)$ and $v(x, y)$ are continuous, the derivative df/dz exists. This may be shown by writing

$$\delta f = \left(\frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x} \right) \delta x + \left(\frac{\partial u}{\partial y} + i \frac{\partial v}{\partial y} \right) \delta y. \tag{6.21}$$

The justification for this expression depends on the continuity of the partial derivatives of u and v. Dividing by δz, we have

$$\begin{aligned} \frac{\delta f}{\delta z} &= \frac{(\partial u/\partial x + i(\partial v/\partial x))\delta x + (\partial u/\partial y + i(\partial v/\partial y))\delta y}{\delta x + i\delta y} \\ &= \frac{(\partial u/\partial x + i(\partial v/\partial x)) + (\partial u/\partial y + i(\partial v/\partial y))\delta y/\delta x}{1 + i(\delta y/\delta x)}. \end{aligned} \tag{6.22}$$

If $\delta f/\delta z$ is to have a unique value, the dependence on $\delta y/\delta x$ must be eliminated. Applying the Cauchy–Riemann conditions to the y derivatives, we obtain

$$\frac{\partial u}{\partial y} + i \frac{\partial v}{\partial y} = -\frac{\partial v}{\partial x} + i \frac{\partial u}{\partial x}. \tag{6.23}$$

Substituting Eq. (6.23) into Eq. (6.22), we may cancel out the $\delta y/\delta x$ dependence and

$$\frac{\delta f}{\delta z} = \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x}, \tag{6.24}$$

which shows that $\lim \delta f/\delta z$ is independent of the direction of approach in the complex plane as long as the partial derivatives are continuous. Thus, $\frac{df}{dz}$ exists and f is analytic at z.

It is worthwhile noting that the Cauchy–Riemann conditions guarantee that the curves $u = c_1$ will be orthogonal to the curves $v = c_2$ (compare Section 2.1). This is fundamental in application to potential problems in a variety of areas of physics. If $u = c_1$ is a line of

electric force, then $v = c_2$ is an equipotential line (surface), and vice versa. To see this, let us write the Cauchy–Riemann conditions as a product of ratios of partial derivatives,

$$\frac{u_x}{u_y} \cdot \frac{v_x}{v_y} = -1, \tag{6.25}$$

with the abbreviations

$$\frac{\partial u}{\partial x} \equiv u_x, \qquad \frac{\partial u}{\partial y} \equiv u_y, \qquad \frac{\partial v}{\partial x} \equiv v_x, \qquad \frac{\partial v}{\partial y} \equiv v_y.$$

Now recall the geometric meaning of $-u_x/u_y$ as the slope of the tangent of each curve $u(x, y) =$ const. and similarly for $v(x, y) =$ const. This means that the $u =$ const. and $v =$ const. curves are mutually orthogonal at each intersection. Alternatively,

$$u_x\,dx + u_y\,dy = 0 = v_y\,dx - v_x\,dy$$

says that, if (dx, dy) is tangent to the u-curve, then the orthogonal $(-dy, dx)$ is tangent to the v-curve at the intersection point, $z = (x, y)$. Or equivalently, $u_x v_x + u_y v_y = 0$ implies that the **gradient vectors** (u_x, u_y) and (v_x, v_y) **are perpendicular**. A further implication for potential theory is developed in Exercise 6.2.1.

Analytic Functions

Finally, if $f(z)$ is differentiable at $z = z_0$ and in some small region around z_0, we say that $f(z)$ is **analytic**[6] at $z = z_0$. If $f(z)$ is analytic everywhere in the (finite) complex plane, we call it an **entire** function. Our theory of complex variables here is one of analytic functions of a complex variable, which points up the crucial importance of the Cauchy–Riemann conditions. The concept of analyticity carried on in advanced theories of modern physics plays a crucial role in dispersion theory (of elementary particles). If $f'(z)$ does not exist at $z = z_0$, then z_0 is labeled a singular point and consideration of it is postponed until Section 6.6.

To illustrate the Cauchy–Riemann conditions, consider two very simple examples.

Example 6.2.1 z^2 Is Analytic

Let $f(z) = z^2$. Then the real part $u(x, y) = x^2 - y^2$ and the imaginary part $v(x, y) = 2xy$. Following Eq. (6.20),

$$\frac{\partial u}{\partial x} = 2x = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -2y = -\frac{\partial v}{\partial x}.$$

We see that $f(z) = z^2$ satisfies the Cauchy–Riemann conditions throughout the complex plane. Since the partial derivatives are clearly continuous, we conclude that $f(z) = z^2$ is analytic. ■

[6] Some writers use the term **holomorphic** or **regular**.

Example 6.2.2 z^* Is Not Analytic

Let $f(z) = z^*$. Now $u = x$ and $v = -y$. Applying the Cauchy–Riemann conditions, we obtain

$$\frac{\partial u}{\partial x} = 1 \neq \frac{\partial v}{\partial y} = -1.$$

The Cauchy–Riemann conditions are not satisfied and $f(z) = z^*$ is not an analytic function of z. It is interesting to note that $f(z) = z^*$ is continuous, thus providing an example of a function that is everywhere continuous but nowhere differentiable in the complex plane.

The derivative of a real function of a real variable is essentially a local characteristic, in that it provides information about the function only in a local neighborhood — for instance, as a truncated Taylor expansion. The existence of a derivative of a function of a complex variable has much more far-reaching implications. The real and imaginary parts of our analytic function must separately satisfy Laplace's equation. This is Exercise 6.2.1. Further, our analytic function is guaranteed derivatives of all orders, Section 6.4. In this sense the derivative not only governs the local behavior of the complex function, but controls the distant behavior as well. ■

Exercises

6.2.1 The functions $u(x, y)$ and $v(x, y)$ are the real and imaginary parts, respectively, of an analytic function $w(z)$.

(a) Assuming that the required derivatives exist, show that

$$\nabla^2 u = \nabla^2 v = 0.$$

Solutions of Laplace's equation such as $u(x, y)$ and $v(x, y)$ are called **harmonic** functions.

(b) Show that

$$\frac{\partial u}{\partial x}\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\frac{\partial v}{\partial y} = 0,$$

and give a geometric interpretation.
Hint. The technique of Section 1.6 allows you to construct vectors normal to the curves $u(x, y) = c_i$ and $v(x, y) = c_j$.

6.2.2 Show whether or not the function $f(z) = \Re(z) = x$ is analytic.

6.2.3 Having shown that the real part $u(x, y)$ and the imaginary part $v(x, y)$ of an analytic function $w(z)$ each satisfy Laplace's equation, show that $u(x, y)$ and $v(x, y)$ **cannot both have either a maximum or a minimum** in the interior of any region in which $w(z)$ is analytic. (They can have saddle points only.)

6.2.4 Let $A = \partial^2 w/\partial x^2$, $B = \partial^2 w/\partial x \partial y$, $C = \partial^2 w/\partial y^2$. From the calculus of functions of two variables, $w(x, y)$, we have a **saddle point** if

$$B^2 - AC > 0.$$

With $f(z) = u(x, y) + iv(x, y)$, apply the Cauchy–Riemann conditions and show that **neither** $u(x, y)$ nor $v(x, y)$ **has a maximum or a minimum** in a finite region of the complex plane. (See also Section 7.3.)

6.2.5 Find the analytic function

$$w(z) = u(x, y) + iv(x, y)$$

if (a) $u(x, y) = x^3 - 3xy^2$, (b) $v(x, y) = e^{-y} \sin x$.

6.2.6 If there is some common region in which $w_1 = u(x, y) + iv(x, y)$ and $w_2 = w_1^* = u(x, y) - iv(x, y)$ are both analytic, prove that $u(x, y)$ and $v(x, y)$ are constants.

6.2.7 The function $f(z) = u(x, y) + iv(x, y)$ is analytic. Show that $f^*(z^*)$ is also analytic.

6.2.8 Using $f(re^{i\theta}) = R(r, \theta)e^{i\Phi(r,\theta)}$, in which $R(r, \theta)$ and $\Phi(r, \theta)$ are differentiable real functions of r and θ, show that the Cauchy–Riemann conditions in polar coordinates become

$$\text{(a)}\ \frac{\partial R}{\partial r} = \frac{R}{r}\frac{\partial \Theta}{\partial \theta}, \qquad \text{(b)}\ \frac{1}{r}\frac{\partial R}{\partial \theta} = -R\frac{\partial \Theta}{\partial r}.$$

Hint. Set up the derivative first with δz radial and then with δz tangential.

6.2.9 As an extension of Exercise 6.2.8 show that $\Theta(r, \theta)$ satisfies Laplace's equation in polar coordinates. Equation (2.35) (without the final term and set to zero) is the Laplacian in polar coordinates.

6.2.10 Two-dimensional irrotational fluid flow is described by the complex potential $f(z) = u(x, y) + iv(x, y)$. We label the real part, $u(x, y)$, the velocity potential and the imaginary part, $v(x, y)$, the stream function. The fluid velocity $\mathbf{V}$ is given by $\mathbf{V} = \nabla u$. If $f(z)$ is analytic,

(a) Show that $df/dz = V_x - iV_y$;
(b) Show that $\nabla \cdot \mathbf{V} = 0$ (no sources or sinks);
(c) Show that $\nabla \times \mathbf{V} = 0$ (irrotational, nonturbulent flow).

6.2.11 A proof of the Schwarz inequality (Section 10.4) involves minimizing an expression,

$$f = \psi_{aa} + \lambda\psi_{ab} + \lambda^*\psi_{ab}^* + \lambda\lambda^*\psi_{bb} \geq 0.$$

The ψ are integrals of products of functions; ψ_{aa} and ψ_{bb} are real, ψ_{ab} is complex and λ is a complex parameter.

(a) Differentiate the preceding expression with respect to λ^*, treating λ as an independent parameter, independent of λ^*. Show that setting the derivative $\partial f/\partial \lambda^*$ equal to zero yields

$$\lambda = -\frac{\psi_{ab}^*}{\overline{\psi_{bb}}}.$$

(b) Show that $\partial f/\partial\lambda = 0$ leads to the same result.

(c) Let $\lambda = x + iy, \lambda^* = x - iy$. Set the x and y derivatives equal to zero and show that again

$$\lambda = -\frac{\psi_{ab}^*}{\psi_{bb}}.$$

This independence of λ and λ^* appears again in Section 17.7.

6.2.12 The function $f(z)$ is analytic. Show that the derivative of $f(z)$ with respect to z^* does not exist unless $f(z)$ is a constant.
Hint. Use the chain rule and take $x = (z + z^*)/2$, $y = (z - z^*)/2i$.
Note. This result emphasizes that our analytic function $f(z)$ is not just a complex function of two real variables x and y. It is a function of the complex variable $x + iy$.

6.3 CAUCHY'S INTEGRAL THEOREM

Contour Integrals

With differentiation under control, we turn to integration. The integral of a complex variable over a contour in the complex plane may be defined in close analogy to the (Riemann) integral of a real function integrated along the real x-axis.

We divide the contour from z_0 to z_0' into n intervals by picking $n - 1$ intermediate points $z_1, z_2, \ldots$ on the contour (Fig. 6.5). Consider the sum

$$S_n = \sum_{j=1}^{n} f(\zeta_j)(z_j - z_{j-1}), \tag{6.26}$$

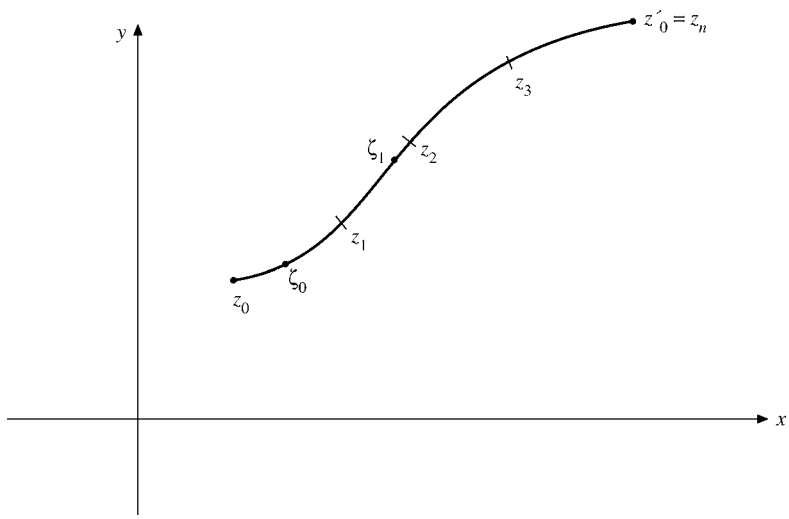

FIGURE 6.5 Integration path.

where ζ_j is a point on the curve between z_j and z_{j-1}. Now let $n \to \infty$ with

$$|z_j - z_{j-1}| \to 0$$

for all j. If the $\lim_{n\to\infty} S_n$ exists and is independent of the details of choosing the points z_j and ζ_j, then

$$\lim_{n\to\infty} \sum_{j=1}^{n} f(\zeta_j)(z_j - z_{j-1}) = \int_{z_0}^{z_0'} f(z)\,dz. \tag{6.27}$$

The right-hand side of Eq. (6.27) is called the contour integral of $f(z)$ (along the specified contour C from $z = z_0$ to $z = z_0'$).

The preceding development of the contour integral is closely analogous to the Riemann integral of a real function of a real variable. As an alternative, the contour integral may be defined by

$$\begin{aligned}\int_{z_1}^{z_2} f(z)dz &= \int_{x_1,y_1}^{x_2,y_2} \big[u(x,y) + iv(x,y)\big][dx + idy] \\ &= \int_{x_1,y_1}^{x_2,y_2} \big[u(x,y)\,dx - v(x,y)\,dy\big] + i\int_{x_1,y_1}^{x_2,y_2} \big[v(x,y)\,dx + u(x,y)\,dy\big]\end{aligned}$$

with the path joining (x_1, y_1) and (x_2, y_2) specified. This reduces the complex integral to the complex sum of real integrals. It is somewhat analogous to the replacement of a vector integral by the vector sum of scalar integrals, Section 1.10.

An important example is the contour integral $\int_C z^n\,dz$, where C is a circle of radius $r > 0$ around the origin $z = 0$ in the positive mathematical sense (counterclockwise). In polar coordinates of Eq. (6.4c) we parameterize the circle as $z = re^{i\theta}$ and $dz = ire^{i\theta}d\theta$. For $n \neq -1$, n an integer, we then obtain

$$\begin{aligned}\frac{1}{2\pi i}\int_C z^n\,dz &= \frac{r^{n+1}}{2\pi}\int_0^{2\pi} \exp\big[i(n+1)\theta\big]\,d\theta \\ &= \big[2\pi i(n+1)\big]^{-1} r^{n+1}\big[e^{i(n+1)\theta}\big]\Big|_0^{2\pi} = 0\end{aligned} \tag{6.27a}$$

because 2π is a period of $e^{i(n+1)\theta}$, while for $n = -1$

$$\frac{1}{2\pi i}\int_C \frac{dz}{z} = \frac{1}{2\pi}\int_0^{2\pi} d\theta = 1, \tag{6.27b}$$

again independent of r.

Alternatively, we can **integrate around a rectangle** with the corners z_1, z_2, z_3, z_4 to obtain for $n \neq -1$

$$\int z^n\,dz = \frac{z^{n+1}}{n+1}\bigg|_{z_1}^{z_2} + \frac{z^{n+1}}{n+1}\bigg|_{z_2}^{z_3} + \frac{z^{n+1}}{n+1}\bigg|_{z_3}^{z_4} + \frac{z^{n+1}}{n+1}\bigg|_{z_4}^{z_1} = 0,$$

because each corner point appears once as an upper and a lower limit that cancel. For $n = -1$ the corresponding real parts of the logarithms cancel similarly, but their imaginary parts involve the increasing arguments of the points from z_1 to z_4 and, when we come back to the first corner z_1, its argument has increased by 2π due to the multivaluedness of the

logarithm, so $2\pi i$ is left over as the value of the integral. Thus, **the value of the integral involving a multivalued function must be that which is reached in a continuous fashion on the path being taken**. These integrals are examples of Cauchy's integral theorem, which we consider in the next section.

Stokes' Theorem Proof

Cauchy's integral theorem is the first of two basic theorems in the theory of the behavior of functions of a complex variable. First, we offer a proof under relatively restrictive conditions — conditions that are intolerable to the mathematician developing a beautiful abstract theory but that are usually satisfied in physical problems.

If a function $f(z)$ is analytic, that is, if its partial derivatives are continuous throughout some **simply connected region** R,[7] for every closed path C (Fig. 6.6) in R, and if it is single-valued (assumed for simplicity here), the line integral of $f(z)$ around C is zero, or

$$\int_C f(z)\,dz = \oint_C f(z)\,dz = 0. \tag{6.27c}$$

Recall that in Section 1.13 such a function $f(z)$, identified as a force, was labeled conservative. The symbol $\oint$ is used to emphasize that the path is closed. Note that the interior of the simply connected region bounded by a contour is that region lying to the left when moving in the direction implied by the contour; as a rule, a simply connected region is bounded by a single closed curve.

In this form the Cauchy integral theorem may be proved by direct application of Stokes' theorem (Section 1.12). With $f(z) = u(x, y) + iv(x, y)$ and $dz = dx + i\,dy$,

$$\begin{aligned}\oint_C f(z)\,dz &= \oint_C (u + iv)(dx + i\,dy)\\ &= \oint_C (u\,dx - v\,dy) + i\oint (v\,dx + u\,dy).\end{aligned} \tag{6.28}$$

These two line integrals may be converted to surface integrals by Stokes' theorem, a procedure that is justified if the partial derivatives are continuous within C. In applying Stokes' theorem, note that the final two integrals of Eq. (6.28) are real. Using

$$\mathbf{V} = \hat{\mathbf{x}}V_x + \hat{\mathbf{y}}V_y,$$

Stokes' theorem says that

$$\oint_C (V_x\,dx + V_y\,dy) = \int \left(\frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y}\right) dx\,dy. \tag{6.29}$$

For the first integral in the last part of Eq. (6.28) let $u = V_x$ and $v = -V_y$.[8] Then

[7] Any closed simple curve (one that does not intersect itself) inside a simply connected region or domain may be contracted to a single point that still belongs to the region. If a region is not simply connected, it is called multiply connected. As an example of a multiply connected region, consider the z-plane with the interior of the unit circle **excluded**.

[8] In the proof of Stokes' theorem, Section 1.12, V_x and V_y are any two functions (with continuous partial derivatives).

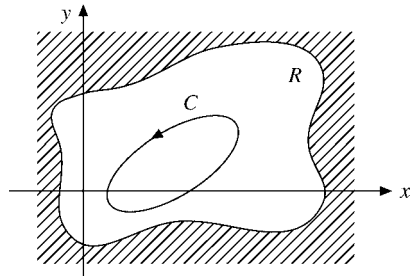

FIGURE 6.6 A closed contour C within a simply connected region R.

$$\oint_C (u\,dx - v\,dy) = \oint_C (V_x\,dx + V_y\,dy)$$
$$= \int \left(\frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y}\right) dx\,dy = -\int \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right) dx\,dy. \quad (6.30)$$

For the second integral on the right side of Eq. (6.28) we let $u = V_y$ and $v = V_x$. Using Stokes' theorem again, we obtain

$$\oint (v\,dx + u\,dy) = \int \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) dx\,dy. \quad (6.31)$$

On application of the Cauchy–Riemann conditions, which must hold, since $f(z)$ is assumed analytic, each integrand vanishes and

$$\oint f(z)\,dz = -\int \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right) dx\,dy + i \int \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) dx\,dy = 0. \quad (6.32)$$

Cauchy–Goursat Proof

This completes the proof of Cauchy's integral theorem. However, the proof is marred from a theoretical point of view by the need for continuity of the first partial derivatives. Actually, as shown by Goursat, this condition is not necessary. An outline of the Goursat proof is as follows. We subdivide the region inside the contour C into a network of small squares, as indicated in Fig. 6.7. Then

$$\oint_C f(z)\,dz = \sum_j \oint_{C_j} f(z)\,dz, \quad (6.33)$$

all integrals along interior lines canceling out. To estimate the $\oint_{C_j} f(z)\,dz$, we construct the function

$$\delta_j(z, z_j) = \frac{f(z) - f(z_j)}{z - z_j} - \left.\frac{df(z)}{dz}\right|_{z=z_j}, \quad (6.34)$$

with z_j an interior point of the jth subregion. Note that $[f(z) - f(z_j)]/(z - z_j)$ is an approximation to the derivative at $z = z_j$. Equivalently, we may note that if $f(z)$ had

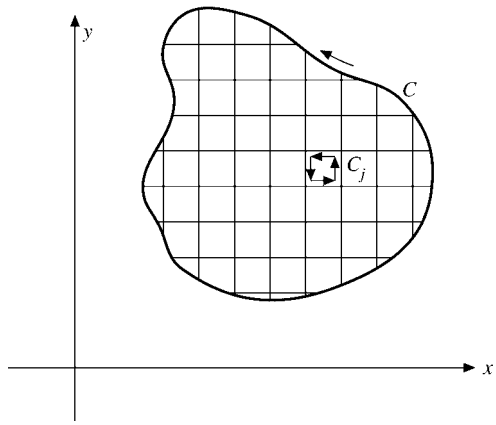

FIGURE 6.7 Cauchy–Goursat contours.

a Taylor expansion (which we have not yet proved), then $\delta_j(z, z_j)$ would be of order $z - z_j$, approaching zero as the network was made finer. But since $f'(z_j)$ exists, that is, is finite, we may make

$$|\delta_j(z, z_j)| < \varepsilon, \tag{6.35}$$

where ε is an arbitrarily chosen small positive quantity. Solving Eq. (6.34) for $f(z)$ and integrating around C_j, we obtain

$$\oint_{C_j} f(z)\,dz = \oint_{C_j} (z - z_j)\delta_j(z, z_j)\,dz, \tag{6.36}$$

the integrals of the other terms vanishing.[9] When Eqs. (6.35) and (6.36) are combined, one shows that

$$\left| \sum_j \oint_{C_j} f(z)\,dz \right| < A\varepsilon, \tag{6.37}$$

where A is a term of the order of the area of the enclosed region. Since ε is arbitrary, we let $\varepsilon \to 0$ and conclude that if a function $f(z)$ is analytic on and within a closed path C,

$$\oint_C f(z)\,dz = 0. \tag{6.38}$$

Details of the proof of this significantly more general and more powerful form can be found in Churchill in the Additional Readings. Actually we can still prove the theorem for $f(z)$ analytic within the interior of C and only continuous on C.

The consequence of the Cauchy integral theorem is that for analytic functions the line integral is a function only of its endpoints, independent of the path of integration,

$$\int_{z_1}^{z_2} f(z)\,dz = F(z_2) - F(z_1) = -\int_{z_2}^{z_1} f(z)\,dz, \tag{6.39}$$

again exactly like the case of a conservative force, Section 1.13.

[9] $\oint dz$ and $\oint z\,dz = 0$ by Eq. (6.27a).

Multiply Connected Regions

The original statement of Cauchy's integral theorem demanded a simply connected region. This restriction may be relaxed by the creation of a barrier, a contour line. The purpose of the following contour-line construction is to permit, within a multiply connected region, the identification of curves that can be shrunk to a point within the region, that is, the construction of a subregion that is simply connected.

Consider the multiply connected region of Fig. 6.8, in which $f(z)$ is not defined for the interior, R'. Cauchy's integral theorem is not valid for the contour C, as shown, but we can construct a contour C' for which the theorem holds. We draw a line from the interior forbidden region, R', to the forbidden region exterior to R and then run a new contour, C', as shown in Fig. 6.9.

The new contour, C', through *ABDEFGA* never crosses the contour line that literally converts R into a simply connected region. The three-dimensional analog of this technique was used in Section 1.14 to prove Gauss' law. By Eq. (6.39),

$$\int_G^A f(z)\,dz = -\int_E^D f(z)\,dz, \tag{6.40}$$

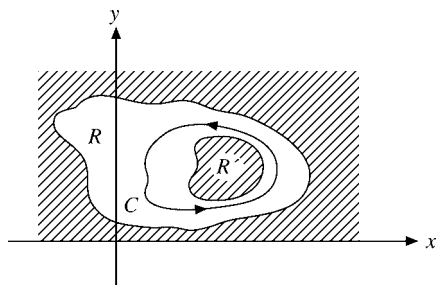

FIGURE 6.8 A closed contour C in a multiply connected region.

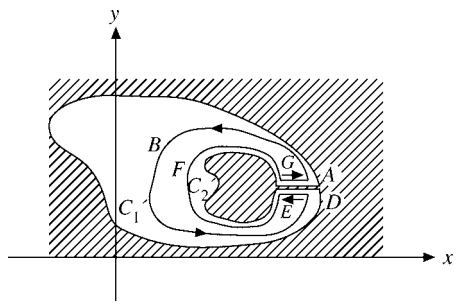

FIGURE 6.9 Conversion of a multiply connected region into a simply connected region.

with $f(z)$ having been continuous across the contour line and line segments DE and GA arbitrarily close together. Then

$$\oint_{C'} f(z)\,dz = \int_{ABD} f(z)\,dz + \int_{EFG} f(z)\,dz = 0 \tag{6.41}$$

by Cauchy's integral theorem, with region R now simply connected. Applying Eq. (6.39) once again with $ABD \to C_1'$ and $EFG \to -C_2'$, we obtain

$$\oint_{C_1'} f(z)\,dz = \oint_{C_2'} f(z)\,dz, \tag{6.42}$$

in which C_1' and C_2' are both traversed in the same (counterclockwise, that is, positive) direction.

Let us emphasize that the contour line here is a matter of mathematical convenience, to permit the application of Cauchy's integral theorem. Since $f(z)$ is analytic in the annular region, it is necessarily single-valued and continuous across any such contour line.

Exercises

6.3.1 Show that $\int_{z_1}^{z_2} f(z)\,dz = -\int_{z_2}^{z_1} f(z)\,dz$.

6.3.2 Prove that

$$\left| \int_C f(z)\,dz \right| \le |f|_{\max} \cdot L,$$

where $|f|_{\max}$ is the maximum value of $|f(z)|$ along the contour C and L is the length of the contour.

6.3.3 Verify that

$$\int_{0,0}^{1,1} z^*\,dz$$

depends on the path by evaluating the integral for the two paths shown in Fig. 6.10. Recall that $f(z) = z^*$ is not an analytic function of z and that Cauchy's integral theorem therefore does not apply.

6.3.4 Show that

$$\oint_C \frac{dz}{z^2 + z} = 0,$$

in which the contour C is a circle defined by $|z| = R > 1$.

Hint. Direct use of the Cauchy integral theorem is illegal. Why? The integral may be evaluated by transforming to polar coordinates and using tables. This yields 0 for $R > 1$ and $2\pi i$ for $R < 1$.

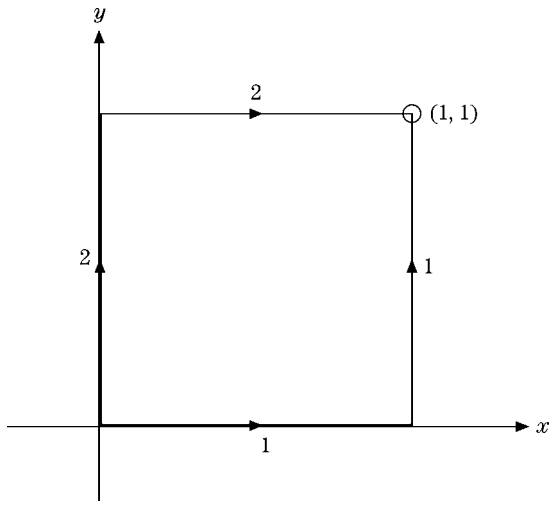

FIGURE 6.10 Contour.

6.4 CAUCHY'S INTEGRAL FORMULA

As in the preceding section, we consider a function $f(z)$ that is analytic on a closed contour C and within the interior region bounded by C. We seek to prove that

$$\boxed{\frac{1}{2\pi i}\oint_C \frac{f(z)}{z-z_0}dz = f(z_0),} \tag{6.43}$$

in which z_0 is any point in the interior region bounded by C. This is the second of the two basic theorems mentioned in Section 6.3. Note that since z is on the contour C while z_0 is in the interior, $z - z_0 \neq 0$ and the integral Eq. (6.43) is well defined. Although $f(z)$ is assumed analytic, the integrand is $f(z)/(z - z_0)$ and is not analytic at $z = z_0$ unless $f(z_0) = 0$. If the contour is deformed as shown in Fig. 6.11 (or Fig. 6.9, Section 6.3), Cauchy's integral theorem applies. By Eq. (6.42),

$$\oint_C \frac{f(z)}{z-z_0}dz - \oint_{C_2} \frac{f(z)}{z-z_0}dz = 0, \tag{6.44}$$

where C is the original outer contour and C_2 is the circle surrounding the point z_0 traversed in a **counterclockwise** direction. Let $z = z_0 + re^{i\theta}$, using the polar representation because of the circular shape of the path around z_0. Here r is small and will eventually be made to approach zero. We have (with $dz = ire^{ir\theta}d\theta$ from Eq. (6.27a))

$$\oint_{C_2} \frac{f(z)}{z-z_0}dz = \oint_{C_2} \frac{f(z_0 + re^{i\theta})}{re^{i\theta}} rie^{i\theta}\, d\theta.$$

Taking the limit as $r \to 0$, we obtain

$$\oint_{C_2} \frac{f(z)}{z-z_0}dz = if(z_0)\int_{C_2} d\theta = 2\pi i f(z_0), \tag{6.45}$$

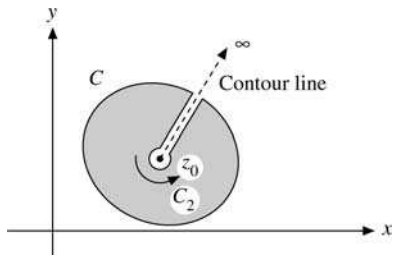

FIGURE 6.11 Exclusion of a singular point.

since $f(z)$ is analytic and therefore continuous at $z = z_0$. This proves the Cauchy integral formula.

Here is a remarkable result. The value of an analytic function $f(z)$ is given at an interior point $z = z_0$ once the values on the boundary C are specified. This is closely analogous to a two-dimensional form of Gauss' law (Section 1.14) in which the magnitude of an interior line charge would be given in terms of the cylindrical surface integral of the electric field **E**.

A further analogy is the determination of a function in real space by an integral of the function and the corresponding Green's function (and their derivatives) over the bounding surface. Kirchhoff diffraction theory is an example of this.

It has been emphasized that z_0 is an interior point. What happens if z_0 is exterior to C? In this case the entire integrand is analytic on and within C. Cauchy's integral theorem, Section 6.3, applies and the integral vanishes. We have

$$\frac{1}{2\pi i}\oint_C \frac{f(z)\,dz}{z - z_0} = \begin{cases} f(z_0), & z_0 \quad \text{interior} \\ 0, & z_0 \quad \text{exterior.} \end{cases}$$

Derivatives

Cauchy's integral formula may be used to obtain an expression for the derivative of $f(z)$. From Eq. (6.43), with $f(z)$ analytic,

$$\frac{f(z_0 + \delta z_0) - f(z_0)}{\delta z_0} = \frac{1}{2\pi i \delta z_0}\left(\oint \frac{f(z)}{z - z_0 - \delta z_0}dz - \oint \frac{f(z)}{z - z_0}dz\right).$$

Then, by definition of derivative (Eq. (6.14)),

$$\begin{aligned} f'(z_0) &= \lim_{\delta z_0 \to 0} \frac{1}{2\pi i \delta z_0}\oint \frac{\delta z_0 f(z)}{(z - z_0 - \delta z_0)(z - z_0)}dz \\ &= \frac{1}{2\pi i}\oint \frac{f(z)}{(z - z_0)^2}dz. \end{aligned} \tag{6.46}$$

This result could have been obtained by differentiating Eq. (6.43) under the integral sign with respect to z_0. This formal, or turning-the-crank, approach is valid, but the justification for it is contained in the preceding analysis.

This technique for constructing derivatives may be repeated. We write $f'(z_0 + \delta z_0)$ and $f'(z_0)$, using Eq. (6.46). Subtracting, dividing by δz_0, and finally taking the limit as $\delta z_0 \to 0$, we have

$$f^{(2)}(z_0) = \frac{2}{2\pi i} \oint \frac{f(z)\,dz}{(z - z_0)^3}.$$

Note that $f^{(2)}(z_0)$ is independent of the direction of δz_0, as it must be. Continuing, we get[10]

$$\boxed{f^{(n)}(z_0) = \frac{n!}{2\pi i} \oint \frac{f(z)\,dz}{(z - z_0)^{n+1}};} \tag{6.47}$$

that is, the requirement that $f(z)$ be analytic guarantees not only a first derivative but derivatives of **all** orders as well! The derivatives of $f(z)$ are automatically analytic. Notice that this statement assumes the Goursat version of the Cauchy integral theorem. This is also why Goursat's contribution is so significant in the development of the theory of complex variables.

Morera's Theorem

A further application of Cauchy's integral formula is in the proof of Morera's **theorem**, which is the converse of Cauchy's integral theorem. The theorem states the following:

> *If a function $f(z)$ is continuous in a simply connected region R and $\oint_C f(z)\,dz = 0$ for every closed contour C within R, then $f(z)$ is analytic throughout R.*

Let us integrate $f(z)$ from z_1 to z_2. Since every closed-path integral of $f(z)$ vanishes, the integral is independent of path and depends only on its endpoints. We label the result of the integration $F(z)$, with

$$F(z_2) - F(z_1) = \int_{z_1}^{z_2} f(z)\,dz. \tag{6.48}$$

As an identity,

$$\frac{F(z_2) - F(z_1)}{z_2 - z_1} - f(z_1) = \frac{\int_{z_1}^{z_2} [f(t) - f(z_1)]\,dt}{z_2 - z_1}, \tag{6.49}$$

using t as another complex variable. Now we take the limit as $z_2 \to z_1$:

$$\lim_{z_2 \to z_1} \frac{\int_{z_1}^{z_2} [f(t) - f(z_1)]\,dt}{z_2 - z_1} = 0, \tag{6.50}$$

[10]This expression is the starting point for defining derivatives of **fractional order**. See A. Erdelyi (ed.), *Tables of Integral Transforms*, Vol. 2. New York: McGraw-Hill (1954). For recent applications to mathematical analysis, see T. J. Osler, An integral analogue of Taylor's series and its use in computing Fourier transforms. *Math. Comput.* **26**: 449 (1972), and references therein.

since $f(t)$ is continuous.[11] Therefore

$$\lim_{z_2 \to z_1} \frac{F(z_2) - F(z_1)}{z_2 - z_1} = F'(z)\Big|_{z=z_1} = f(z_1) \tag{6.51}$$

by definition of derivative (Eq. (6.14)). We have proved that $F'(z)$ at $z = z_1$ exists and equals $f(z_1)$. Since z_1 is any point in R, we see that $F(z)$ is analytic. Then by Cauchy's integral formula (compare Eq. (6.47)), $F'(z) = f(z)$ is also analytic, proving Morera's theorem.

Drawing once more on our electrostatic analog, we might use $f(z)$ to represent the electrostatic field **E**. If the net charge within every closed region in R is zero (Gauss' law), the charge density is everywhere zero in R. Alternatively, in terms of the analysis of Section 1.13, $f(z)$ represents a conservative force (by definition of conservative), and then we find that it is always possible to express it as the derivative of a potential function $F(z)$.

An important application of Cauchy's integral formula is the following **Cauchy inequality**. If $f(z) = \sum a_n z^n$ is analytic and bounded, $|f(z)| \le M$ on a circle of radius r about the origin, then

$$|a_n| r^n \le M \qquad \text{(Cauchy's inequality)} \tag{6.52}$$

gives upper bounds for the coefficients of its Taylor expansion. To prove Eq. (6.52) let us define $M(r) = \max_{|z|=r} |f(z)|$ and use the Cauchy integral for a_n:

$$|a_n| = \frac{1}{2\pi} \left| \int_{|z|=r} \frac{f(z)}{z^{n+1}} dz \right| \le M(r) \frac{2\pi r}{2\pi r^{n+1}}.$$

An immediate consequence of the inequality (6.52) is **Liouville's theorem**: If $f(z)$ is analytic and bounded in the entire complex plane it is a constant. In fact, if $|f(z)| \le M$ for all z, then Cauchy's inequality (6.52) gives $|a_n| \le M r^{-n} \to 0$ as $r \to \infty$ for $n > 0$. Hence $f(z) = a_0$.

Conversely, the slightest deviation of an analytic function from a constant value implies that there must be at least one singularity somewhere in the infinite complex plane. Apart from the trivial constant functions, then, singularities are a fact of life, and we must learn to live with them. But we shall do more than that. We shall next expand a function in a Laurent series at a singularity, and we shall use singularities to develop the powerful and useful calculus of residues in Chapter 7.

A famous application of Liouville's theorem yields the **fundamental theorem of algebra** (due to C. F. Gauss), which says that any polynomial $P(z) = \sum_{\nu=0}^{n} a_\nu z^\nu$ with $n > 0$ and $a_n \neq 0$ has n roots. To prove this, suppose $P(z)$ has no zero. Then $1/P(z)$ is analytic and bounded as $|z| \to \infty$. Hence $P(z)$ is a constant by Liouville's theorem, q.e.a. Thus, $P(z)$ has at least one root that we can divide out. Then we repeat the process for the resulting polynomial of degree $n - 1$. This leads to the conclusion that $P(z)$ has exactly n roots.

[11] We quote the mean value theorem of calculus here.

Exercises

6.4.1 Show that

$$\oint_C (z - z_0)^n \, dz = \begin{cases} 2\pi i, & n = -1, \\ 0, & n \neq -1, \end{cases}$$

where the contour C encircles the point $z = z_0$ in a positive (counterclockwise) sense. The exponent n is an integer. See also Eq. (6.27a). The calculus of residues, Chapter 7, is based on this result.

6.4.2 Show that

$$\frac{1}{2\pi i} \oint z^{m-n-1} \, dz, \qquad m \text{ and } n \text{ integers}$$

(with the contour encircling the origin once counterclockwise) is a representation of the Kronecker δ_{mn}.

6.4.3 Solve Exercise 6.3.4 by separating the integrand into partial fractions and then applying Cauchy's integral theorem for multiply connected regions.
Note. Partial fractions are explained in Section 15.8 in connection with Laplace transforms.

6.4.4 Evaluate

$$\oint_C \frac{dz}{z^2 - 1},$$

where C is the circle $|z| = 2$.

6.4.5 Assuming that $f(z)$ is analytic on and within a closed contour C and that the point z_0 is within C, show that

$$\oint_C \frac{f'(z)}{z - z_0} \, dz = \oint_C \frac{f(z)}{(z - z_0)^2} dz.$$

6.4.6 You know that $f(z)$ is analytic on and within a closed contour C. You suspect that the nth derivative $f^{(n)}(z_0)$ is given by

$$f^{(n)}(z_0) = \frac{n!}{2\pi i} \oint_C \frac{f(z)}{(z - z_0)^{n+1}} dz.$$

Using mathematical induction, prove that this expression is correct.

6.4.7 (a) A function $f(z)$ is analytic within a closed contour C (and continuous on C). If $f(z) \neq 0$ within C and $|f(z)| \leq M$ on C, show that

$$\frac{1}{|f(z)|} \leq M$$

for all points within C.
Hint. Consider $w(z) = 1/f(z)$.

(b) If $f(z) = 0$ within the contour C, show that the foregoing result does not hold and that it is possible to have $|f(z)| = 0$ at one or more points in the interior with $|f(z)| > 0$ over the entire bounding contour. Cite a specific example of an analytic function that behaves this way.

6.4.8 Using the Cauchy integral formula for the nth derivative, convert the following Rodrigues formulas into the corresponding so-called Schlaefli integrals.

(a) Legendre:

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}\left(x^2-1\right)^n.$$

ANS. $\dfrac{(-1)^n}{2^n}\cdot\dfrac{1}{2\pi i}\oint\dfrac{(1-z^2)^n}{(z-x)^{n+1}}dz.$

(b) Hermite:

$$H_n(x) = (-1)^n e^{x^2}\frac{d^n}{dx^n}e^{-x^2}.$$

(c) Laguerre:

$$L_n(x) = \frac{e^x}{n!}\frac{d^n}{dx^n}\left(x^n e^{-x}\right).$$

Note. From the Schlaefli integral representations one can develop generating functions for these special functions. Compare Sections 12.4, 13.1, and 13.2.

6.5 LAURENT EXPANSION

Taylor Expansion

The Cauchy integral formula of the preceding section opens up the way for another derivation of Taylor's series (Section 5.6), but this time for functions of a complex variable. Suppose we are trying to expand $f(z)$ about $z = z_0$ and we have $z = z_1$ as the nearest point on the Argand diagram for which $f(z)$ is not analytic. We construct a circle C centered at $z = z_0$ with radius less than $|z_1 - z_0|$ (Fig. 6.12). Since z_1 was assumed to be the nearest point at which $f(z)$ was not analytic, $f(z)$ is necessarily analytic on and within C.

From Eq. (6.43), the Cauchy integral formula,

$$\begin{aligned} f(z) &= \frac{1}{2\pi i}\oint_C \frac{f(z')\,dz'}{z'-z} \\ &= \frac{1}{2\pi i}\oint_C \frac{f(z')\,dz'}{(z'-z_0)-(z-z_0)} \\ &= \frac{1}{2\pi i}\oint_C \frac{f(z')\,dz'}{(z'-z_0)[1-(z-z_0)/(z'-z_0)]}. \end{aligned} \tag{6.53}$$

Here z' is a point on the contour C and z is any point interior to C. It is not legal yet to expand the denominator of the integrand in Eq. (6.53) by the binomial theorem, for we have not yet proved the binomial theorem for complex variables. Instead, we note the identity

$$\frac{1}{1-t} = 1 + t + t^2 + t^3 + \cdots = \sum_{n=0}^{\infty} t^n, \tag{6.54}$$

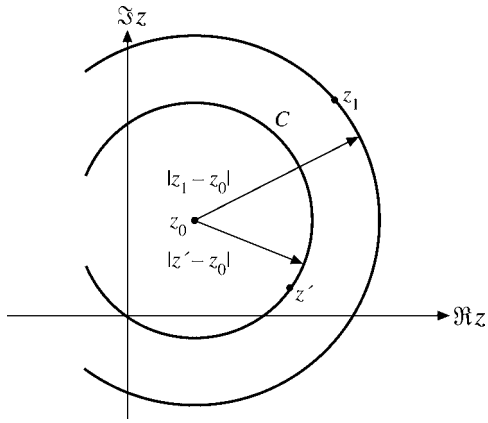

FIGURE 6.12 Circular domain for Taylor expansion.

which may easily be verified by multiplying both sides by $1 - t$. The infinite series, following the methods of Section 5.2, is convergent for $|t| < 1$.

Now, for a point z interior to C, $|z - z_0| < |z' - z_0|$, and, using Eq. (6.54), Eq. (6.53) becomes

$$f(z) = \frac{1}{2\pi i} \oint_C \sum_{n=0}^{\infty} \frac{(z - z_0)^n f(z')\, dz'}{(z' - z_0)^{n+1}}. \tag{6.55}$$

Interchanging the order of integration and summation (valid because Eq. (6.54) is uniformly convergent for $|t| < 1$), we obtain

$$f(z) = \frac{1}{2\pi i} \sum_{n=0}^{\infty} (z - z_0)^n \oint_C \frac{f(z')\, dz'}{(z' - z_0)^{n+1}}. \tag{6.56}$$

Referring to Eq. (6.47), we get

$$f(z) = \sum_{n=0}^{\infty} (z - z_0)^n \frac{f^{(n)}(z_0)}{n!}, \tag{6.57}$$

which is our desired Taylor expansion. Note that it is based only on the assumption that $f(z)$ is analytic for $|z - z_0| < |z_1 - z_0|$. Just as for real variable power series (Section 5.7), this expansion is unique for a given z_0.

From the Taylor expansion for $f(z)$ a binomial theorem may be derived (Exercise 6.5.2).

Schwarz Reflection Principle

From the binomial expansion of $g(z) = (z - x_0)^n$ for integral n it is easy to see that the complex conjugate of the function g is the function of the complex conjugate for real x_0:

$$g^*(z) = \left[(z - x_0)^n\right]^* = (z^* - x_0)^n = g(z^*). \tag{6.58}$$

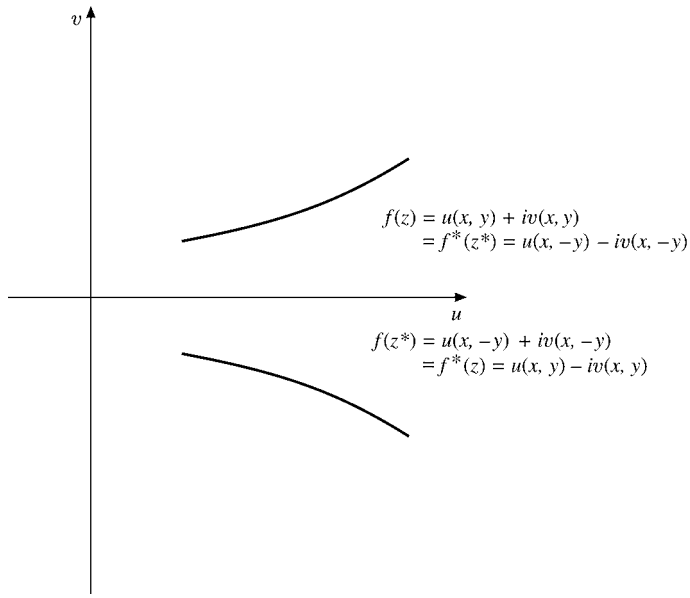

FIGURE 6.13 Schwarz reflection.

This leads us to the Schwarz reflection principle:

> *If a function* $f(z)$ *is* (1) *analytic over some region including the real axis and* (2) *real when* z *is real, then*
>
> $$f^*(z) = f(z^*). \tag{6.59}$$

(See Fig. 6.13.)

Expanding $f(z)$ about some (nonsingular) point x_0 on the real axis,

$$f(z) = \sum_{n=0}^{\infty} (z - x_0)^n \frac{f^{(n)}(x_0)}{n!} \tag{6.60}$$

by Eq. (6.56). Since $f(z)$ is analytic at $z = x_0$, this Taylor expansion exists. Since $f(z)$ is real when z is real, $f^{(n)}(x_0)$ must be real for all n. Then when we use Eq. (6.58), Eq. (6.59), the Schwarz reflection principle, follows immediately. Exercise 6.5.6 is another form of this principle. This completes the proof within a circle of convergence. Analytic continuation then permits extending this result to the entire region of analyticity.

Analytic Continuation

It is natural to think of the values $f(z)$ of an analytic function f as a single entity, which is usually defined in some restricted region S_1 of the complex plane, for example, by a Taylor series (see Fig. 6.14). Then f is analytic inside the **circle of convergence** C_1, whose radius is given by the distance r_1 from the center of C_1 to the **nearest singularity** of f at z_1 (in Fig. 6.14). A singularity is any point where f is not analytic. If we choose a point inside C_1

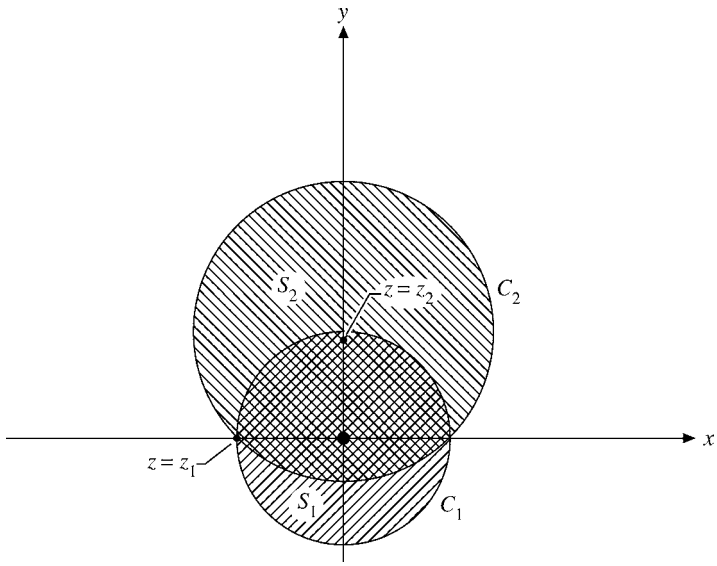

FIGURE 6.14 Analytic continuation.

that is farther than r_1 from the singularity z_1 and make a Taylor expansion of f about it (z_2 in Fig. 6.14), then the circle of convergence, C_2 will usually extend beyond the first circle, C_1. In the overlap region of both circles, C_1, C_2, the function f is uniquely defined. In the region of the circle C_2 that extends beyond C_1, $f(z)$ is uniquely defined by the Taylor series about the center of C_2 and is analytic there, although the Taylor series about the center of C_1 is no longer convergent there. After Weierstrass this process is called **analytic continuation**. It defines the analytic functions in terms of its original definition (in C_1, say) and all its continuations.

A specific example is the function

$$f(z) = \frac{1}{1+z}, \tag{6.61}$$

which has a (simple) pole at $z = -1$ and is analytic elsewhere. The geometric series expansion

$$\frac{1}{1+z} = 1 - z + z^2 + \cdots = \sum_{n=0}^{\infty} (-z)^n \tag{6.62}$$

converges for $|z| < 1$, that is, inside the circle C_1 in Fig. 6.14.

Suppose we expand $f(z)$ about $z = i$, so

$$\begin{aligned} f(z) = \frac{1}{1+z} &= \frac{1}{1+i+(z-i)} = \frac{1}{(1+i)(1+(z-i)/(1+i))} \\ &= \left[1 - \frac{z-i}{1+i} + \frac{(z-i)^2}{(1+i)^2} - \cdots\right] \frac{1}{1+i} \end{aligned} \tag{6.63}$$

converges for $|z - i| < |1 + i| = \sqrt{2}$. Our circle of convergence is C_2 in Fig. 6.14. Now

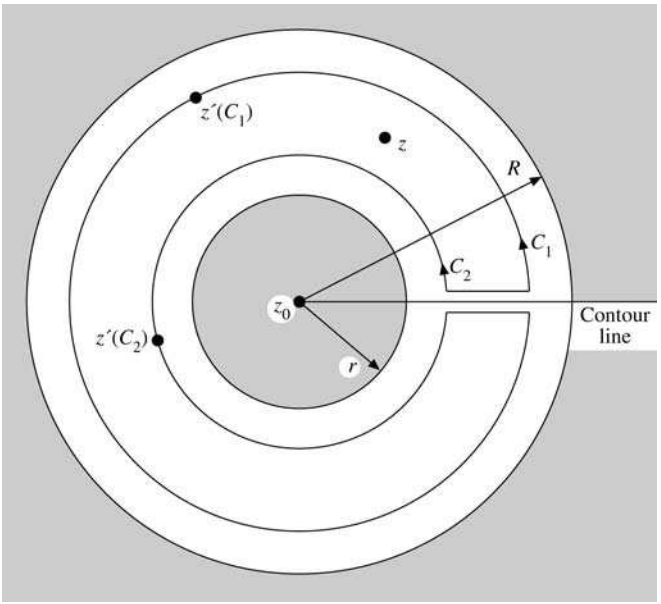

FIGURE 6.15 $|z' - z_0|_{C_1} > |z - z_0|$; $|z' - z_0|_{C_2} < |z - z_0|$.

$f(z)$ is defined by the expansion (6.63) in S_2, which overlaps S_1 and extends further out in the complex plane.[12] This extension is an analytic continuation, and when we have only isolated singular points to contend with, the function can be extended indefinitely. Equations (6.61), (6.62), and (6.63) are three different representations of the same function. Each representation has its own domain of convergence. Equation (6.62) is a Maclaurin series. Equation (6.63) is a Taylor expansion about $z = i$ and from the following paragraphs Eq. (6.61) is seen to be a one-term Laurent series.

Analytic continuation may take many forms, and the series expansion just considered is not necessarily the most convenient technique. As an alternate technique we shall use a functional relation in Section 8.1 to extend the factorial function around the isolated singular points $z = -n$, $n = 1, 2, 3, \ldots$. As another example, the hypergeometric equation is satisfied by the hypergeometric function defined by the series, Eq. (13.115), for $|z| < 1$. The integral representation given in Exercise 13.4.7 permits a continuation into the complex plane.

[12] One of the most powerful and beautiful results of the more abstract theory of functions of a complex variable is that if two analytic functions coincide in any region, such as the overlap of S_1 and S_2, or coincide on any line segment, they are the same function, in the sense that they will coincide everywhere as long as they are both well defined. In this case the agreement of the expansions (Eqs. (6.62) and (6.63)) over the region common to S_1 and S_2 would establish the identity of the functions these expansions represent. Then Eq. (6.63) would represent an analytic continuation or extension of $f(z)$ into regions not covered by Eq. (6.62). We could equally well say that $f(z) = 1/(1 + z)$ is itself an analytic continuation of either of the series given by Eqs. (6.62) and (6.63).

Laurent Series

We frequently encounter functions that are analytic and single-valued in an annular region, say, of inner radius r and outer radius R, as shown in Fig. 6.15. Drawing an imaginary contour line to convert our region into a simply connected region, we apply Cauchy's integral formula, and for two circles C_2 and C_1 centered at $z = z_0$ and with radii r_2 and r_1, respectively, where $r < r_2 < r_1 < R$, we have[13]

$$f(z) = \frac{1}{2\pi i}\oint_{C_1} \frac{f(z')\,dz'}{z'-z} - \frac{1}{2\pi i}\oint_{C_2} \frac{f(z')\,dz'}{z'-z}. \tag{6.64}$$

Note that in Eq. (6.64) an explicit minus sign has been introduced so that the contour C_2 (like C_1) is to be traversed in the positive (counterclockwise) sense. The treatment of Eq. (6.64) now proceeds exactly like that of Eq. (6.53) in the development of the Taylor series. Each denominator is written as $(z'-z_0)-(z-z_0)$ and expanded by the binomial theorem, which now follows from the Taylor series (Eq. (6.57)).

Noting that for C_1, $|z'-z_0| > |z-z_0|$ while for C_2, $|z'-z_0| < |z-z_0|$, we find

$$\begin{aligned} f(z) = {} & \frac{1}{2\pi i}\sum_{n=0}^{\infty}(z-z_0)^n \oint_{C_1} \frac{f(z')\,dz'}{(z'-z_0)^{n+1}} \\ & + \frac{1}{2\pi i}\sum_{n=1}^{\infty}(z-z_0)^{-n} \oint_{C_2} (z'-z_0)^{n-1} f(z')\,dz'. \end{aligned} \tag{6.65}$$

The minus sign of Eq. (6.64) has been absorbed by the binomial expansion. Labeling the first series S_1 and the second S_2 we have

$$S_1 = \frac{1}{2\pi i}\sum_{n=0}^{\infty}(z-z_0)^n \oint_{C_1} \frac{f(z')\,dz'}{(z'-z_0)^{n+1}}, \tag{6.66}$$

which is the regular Taylor expansion, convergent for $|z-z_0| < |z'-z_0| = r_1$, that is, for all z **interior** to the larger circle, C_1. For the second series in Eq. (6.65) we have

$$S_2 = \frac{1}{2\pi i}\sum_{n=1}^{\infty}(z-z_0)^{-n} \oint_{C_2} (z'-z_0)^{n-1} f(z')\,dz', \tag{6.67}$$

convergent for $|z-z_0| > |z'-z_0| = r_2$, that is, for all z **exterior** to the smaller circle, C_2. Remember, C_2 now goes counterclockwise.

These two series are combined into one series[14] (a Laurent series) by

$$\boxed{f(z) = \sum_{n=-\infty}^{\infty} a_n (z-z_0)^n,} \tag{6.68}$$

[13] We may take r_2 arbitrarily close to r and r_1 arbitrarily close to R, maximizing the area enclosed between C_1 and C_2.

[14] Replace n by $-n$ in S_2 and add.

where

$$\boxed{a_n = \frac{1}{2\pi i}\oint_C \frac{f(z')\,dz'}{(z'-z_0)^{n+1}}.} \tag{6.69}$$

Since, in Eq. (6.69), convergence of a binomial expansion is no longer a problem, C may be any contour within the annular region $r < |z - z_0| < R$ encircling z_0 once in a counterclockwise sense. If we assume that such an annular region of convergence does exist, then Eq. (6.68) is the Laurent series, or Laurent expansion, of $f(z)$.

The use of the contour line (Fig. 6.15) is convenient in converting the annular region into a simply connected region. Since our function is analytic in this annular region (and single-valued), the contour line is not essential and, indeed, does not appear in the final result, Eq. (6.69).

Laurent series coefficients need not come from evaluation of contour integrals (which may be very intractable). Other techniques, such as ordinary series expansions, may provide the coefficients.

Numerous examples of Laurent series appear in Chapter 7. We limit ourselves here to one simple example to illustrate the application of Eq. (6.68).

Example 6.5.1 LAURENT EXPANSION

Let $f(z) = [z(z-1)]^{-1}$. If we choose $z_0 = 0$, then $r = 0$ and $R = 1$, $f(z)$ diverging at $z = 1$. A partial fraction expansion yields the Laurent series

$$\frac{1}{z(z-1)} = -\frac{1}{1-z} - \frac{1}{z} = -\frac{1}{z} - 1 - z - z^2 - z^3 - \cdots = -\sum_{n=-1}^{\infty} z^n. \tag{6.70}$$

From Eqs. (6.70), (6.68), and (6.69) we then have

$$a_n = \frac{1}{2\pi i}\oint \frac{dz'}{(z')^{n+2}(z'-1)} = \begin{cases} -1 & \text{for } n \geq -1, \\ 0 & \text{for } n < -1. \end{cases} \tag{6.71}$$

The integrals in Eq. (6.71) can also be directly evaluated by substituting the geometric-series expansion of $(1 - z')^{-1}$ used already in Eq. (6.70) for $(1 - z)^{-1}$:

$$a_n = \frac{-1}{2\pi i}\oint \sum_{m=0}^{\infty} (z')^m \frac{dz'}{(z')^{n+2}}. \tag{6.72}$$

Upon interchanging the order of summation and integration (uniformly convergent series), we have

$$a_n = -\frac{1}{2\pi i}\sum_{m=0}^{\infty}\oint \frac{dz'}{(z')^{n+2-m}}. \tag{6.73}$$

If we employ the polar form, as in Eq. (6.47) (or compare Exercise 6.4.1),

$$a_n = -\frac{1}{2\pi i}\sum_{m=0}^{\infty}\oint \frac{rie^{i\theta}\,d\theta}{r^{n+2-m}e^{i(n+2-m)\theta}}$$

$$= -\frac{1}{2\pi i}\cdot 2\pi i\sum_{m=0}^{\infty}\delta_{n+2-m,1}, \tag{6.74}$$

which agrees with Eq. (6.71). ■

The Laurent series differs from the Taylor series by the obvious feature of negative powers of $(z - z_0)$. For this reason the Laurent series will always diverge at least at $z = z_0$ and perhaps as far out as some distance r (Fig. 6.15).

Exercises

6.5.1 Develop the Taylor expansion of $\ln(1 + z)$.

ANS. $\displaystyle\sum_{n=1}^{\infty}(-1)^{n-1}\frac{z^n}{n}$.

6.5.2 Derive the binomial expansion

$$(1+z)^m = 1 + mz + \frac{m(m-1)}{1\cdot 2}z^2 + \cdots = \sum_{n=0}^{\infty}\binom{m}{n}z^n$$

for m any real number. The expansion is convergent for $|z| < 1$. Why?

6.5.3 A function $f(z)$ is analytic on and within the unit circle. Also, $|f(z)| < 1$ for $|z| \le 1$ and $f(0) = 0$. Show that $|f(z)| < |z|$ for $|z| \le 1$.
Hint. One approach is to show that $f(z)/z$ is analytic and then to express $[f(z_0)/z_0]^n$ by the Cauchy integral formula. Finally, consider absolute magnitudes and take the nth root. This exercise is sometimes called Schwarz's theorem.

6.5.4 If $f(z)$ is a real function of the complex variable $z = x + iy$, that is, if $f(x) = f^*(x)$, and the Laurent expansion about the origin, $f(z) = \sum a_n z^n$, has $a_n = 0$ for $n < -N$, show that all of the coefficients a_n are real.
Hint. Show that $z^N f(z)$ is analytic (via Morera's theorem, Section 6.4).

6.5.5 A function $f(z) = u(x, y) + iv(x, y)$ satisfies the conditions for the Schwarz reflection principle. Show that
(a) u is an even function of y. (b) v is an odd function of y.

6.5.6 A function $f(z)$ can be expanded in a Laurent series about the origin with the coefficients a_n real. Show that the complex conjugate of this function of z is the same function of the complex conjugate of z; that is,

$$f^*(z) = f(z^*).$$

Verify this explicitly for

(a) $f(z) = z^n$, n an integer, (b) $f(z) = \sin z$.
If $f(z) = iz$ ($a_1 = i$), show that the foregoing statement does not hold.

6.5.7 The function $f(z)$ is analytic in a domain that includes the real axis. When z is real ($z = x$), $f(x)$ is pure imaginary.

(a) Show that

$$f(z^*) = -\big[f(z)\big]^*.$$

(b) For the specific case $f(z) = iz$, develop the Cartesian forms of $f(z)$, $f(z^*)$, and $f^*(z)$. Do not quote the general result of part (a).

6.5.8 Develop the first three nonzero terms of the Laurent expansion of

$$f(z) = \left(e^z - 1\right)^{-1}$$

about the origin. Notice the resemblance to the Bernoulli number–generating function, Eq. (5.144) of Section 5.9.

6.5.9 Prove that the Laurent expansion of a given function about a given point is unique; that is, if

$$f(z) = \sum_{n=-N}^{\infty} a_n (z - z_0)^n = \sum_{n=-N}^{\infty} b_n (z - z_0)^n,$$

show that $a_n = b_n$ for all n.
Hint. Use the Cauchy integral formula.

6.5.10 (a) Develop a Laurent expansion of $f(z) = [z(z-1)]^{-1}$ about the point $z = 1$ valid for small values of $|z - 1|$. Specify the exact range over which your expansion holds. This is an analytic continuation of Eq. (6.70).
(b) Determine the Laurent expansion of $f(z)$ about $z = 1$ but for $|z - 1|$ large.
Hint. Partial fraction this function and use the geometric series.

6.5.11 (a) Given $f_1(z) = \int_0^\infty e^{-zt}\,dt$ (with t real), show that the domain in which $f_1(z)$ exists (and is analytic) is $\Re(z) > 0$.
(b) Show that $f_2(z) = 1/z$ equals $f_1(z)$ over $\Re(z) > 0$ and is therefore an analytic continuation of $f_1(z)$ over the entire z-plane except for $z = 0$.
(c) Expand $1/z$ about the point $z = i$. You will have $f_3(z) = \sum_{n=0}^{\infty} a_n (z - i)^n$. What is the domain of $f_3(z)$?

$$\text{ANS. } \frac{1}{z} = -i \sum_{n=0}^{\infty} i^n (z - i)^n, \qquad |z - i| < 1.$$

6.6 SINGULARITIES

The Laurent expansion represents a generalization of the Taylor series in the presence of singularities. We define the point z_0 as an **isolated singular point** of the function $f(z)$ if $f(z)$ is not analytic at $z = z_0$ but is analytic at all neighboring points.

Poles

In the Laurent expansion of $f(z)$ about z_0,

$$f(z) = \sum_{m=-\infty}^{\infty} a_m (z - z_0)^m, \tag{6.75}$$

if $a_m = 0$ for $m < -n < 0$ and $a_{-n} \neq 0$, we say that z_0 is a pole of order n. For instance, if $n = 1$, that is, if $a_{-1}/(z - z_0)$ is the first nonvanishing term in the Laurent series, we have a pole of order 1, often called a **simple** pole.

If, on the other hand, the summation continues to $m = -\infty$, then z_0 is a pole of infinite order and is called an **essential singularity**. These essential singularities have many pathological features. For instance, we can show that in any small neighborhood of an essential singularity of $f(z)$ the function $f(z)$ comes arbitrarily close to any (and therefore every) preselected complex quantity w_0.[15] Here, the entire w-plane is mapped by f into the neighborhood of the point z_0. One point of fundamental difference between a pole of finite order n and an essential singularity is that by multiplying $f(z)$ by $(z - z_0)^n$, $f(z)(z - z_0)^n$ is no longer singular at z_0. This obviously cannot be done for an essential singularity.

The behavior of $f(z)$ as $z \to \infty$ is defined in terms of the behavior of $f(1/t)$ as $t \to 0$. Consider the function

$$\sin z = \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{(2n+1)!}. \tag{6.76}$$

As $z \to \infty$, we replace the z by $1/t$ to obtain

$$\sin\left(\frac{1}{t}\right) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)! t^{2n+1}}. \tag{6.77}$$

From the definition, $\sin z$ has an essential singularity at infinity. This result could be anticipated from Exercise 6.1.9 since

$$\sin z = \sin iy = i \sinh y, \qquad \text{when } x = 0,$$

which approaches infinity exponentially as $y \to \infty$. Thus, although the absolute value of $\sin x$ for real x is equal to or less than unity, the absolute value of $\sin z$ is not bounded.

A function that is analytic throughout the finite complex plane **except** for isolated poles is called **meromorphic**, such as ratios of two polynomials or $\tan z$, $\cot z$. Examples are also **entire** functions that have no singularities in the finite complex plane, such as $\exp(z)$, $\sin z$, $\cos z$ (see Sections 5.9, 5.11).

[15] This theorem is due to Picard. A proof is given by E. C. Titchmarsh, *The Theory of Functions*, 2nd ed. New York: Oxford University Press (1939).

Branch Points

There is another sort of singularity that will be important in Chapter 7. Consider

$$f(z) = z^a,$$

in which a is not an integer.[16] As z moves around the unit circle from e^0 to $e^{2\pi i}$,

$$f(z) \to e^{2\pi a i} \neq e^{0\cdot a} = 1,$$

for nonintegral a. We have a branch point at the origin and another at infinity. If we set $z = 1/t$, a similar analysis of $f(z)$ for $t \to 0$ shows that $t = 0$; that is, $z = \infty$ is also a branch point. The points e^{0i} and $e^{2\pi i}$ in the z-plane coincide, but these **coincident points lead to different values** of $f(z)$; that is, $f(z)$ is a **multivalued function**. The problem is resolved by constructing a **cut line joining both branch points** so that $f(z)$ will be uniquely specified for a given point in the z-plane. For z^a, the cut line can go out at any angle. Note that the point at infinity must be included here; that is, the cut line may join finite branch points via the point at infinity. The next example is a case in point. If $a = p/q$ is a rational number, then q is called the order of the branch point, because one needs to go around the branch point q times before coming back to the starting point. If a is irrational, then the order of the branch point is infinite, just as for the logarithm.

Note that a function with a branch point and a required cut line will not be continuous across the cut line. Often there will be a phase difference on opposite sides of this cut line. Hence line integrals on opposite sides of this branch point cut line will not generally cancel each other. Numerous examples of this case appear in the exercises.

The contour line used to convert a multiply connected region into a simply connected region (Section 6.3) is completely different. Our function is continuous across that contour line, and no phase difference exists.

Example 6.6.1 BRANCH POINTS OF ORDER 2

Consider the function

$$f(z) = \left(z^2 - 1\right)^{1/2} = (z+1)^{1/2}(z-1)^{1/2}. \tag{6.78}$$

The first factor on the right-hand side, $(z+1)^{1/2}$, has a branch point at $z = -1$. The second factor has a branch point at $z = +1$. At infinity $f(z)$ has a simple pole. This is best seen by substituting $z = 1/t$ and making a binomial expansion at $t = 0$:

$$\left(z^2 - 1\right)^{1/2} = \frac{1}{t}\left(1 - t^2\right)^{1/2} = \frac{1}{t}\sum_{n=0}^{\infty}\binom{1/2}{n}(-1)^n t^{2n} = \frac{1}{t} - \frac{1}{2}t - \frac{1}{8}t^3 + \cdots.$$

The cut line has to connect both branch points, so it is not possible to encircle either branch point completely. To check on the possibility of taking the line segment joining $z = +1$ and

[16] $z = 0$ is a singular point, for z^a has only a finite number of derivatives, whereas an analytic function is guaranteed an infinite number of derivatives (Section 6.4). The problem is that $f(z)$ is not single-valued as we encircle the origin. The Cauchy integral formula may not be applied.

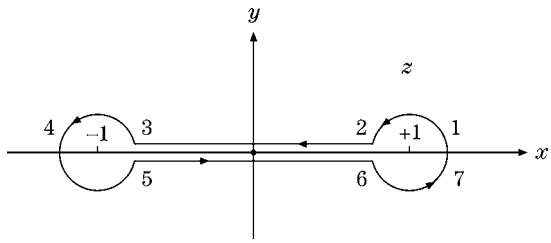

FIGURE 6.16 Branch cut and phases of Table 6.1.

Table 6.1 Phase Angle

Point	θ	φ	$\frac{\theta+\varphi}{2}$
1	0	0	0
2	0	π	$\frac{\pi}{2}$
3	0	π	$\frac{\pi}{2}$
4	π	π	π
5	2π	π	$\frac{3\pi}{2}$
6	2π	π	$\frac{3\pi}{2}$
7	2π	2π	2π

$z=-1$ as a cut line, let us follow the phases of these two factors as we move along the contour shown in Fig. 6.16.

For convenience in following the changes of phase let $z+1=re^{i\theta}$ and $z-1=\rho e^{i\varphi}$. Then the phase of $f(z)$ is $(\theta+\varphi)/2$. We start at point 1, where both $z+1$ and $z-1$ have a phase of zero. Moving from point 1 to point 2, φ, the phase of $z-1=\rho e^{i\varphi}$, increases by π. ($z-1$ becomes negative.) φ then stays constant until the circle is completed, moving from 6 to 7. θ, the phase of $z+1=re^{i\theta}$, shows a similar behavior, increasing by 2π as we move from 3 to 5. The phase of the function $f(z)=(z+1)^{1/2}(z-1)^{1/2}=r^{1/2}\rho^{1/2}e^{i(\theta+\varphi)/2}$ is $(\theta+\varphi)/2$. This is tabulated in the final column of Table 6.1.

Two features emerge:

1. The phase at points 5 and 6 is not the same as the phase at points 2 and 3. This behavior can be expected at a branch cut.

2. The phase at point 7 exceeds that at point 1 by 2π, and the function $f(z)=(z^2-1)^{1/2}$ is therefore **single-valued** for the contour shown, encircling **both** branch points.

If we take the x-axis, $-1\le x\le 1$, as a cut line, $f(z)$ is uniquely specified. Alternatively, the positive x-axis for $x>1$ and the negative x-axis for $x<-1$ may be taken as cut lines. The branch points cannot be encircled, and the function remains single-valued. These two cut lines are, in fact, one branch cut from -1 to $+1$ via the point at infinity. ■

Generalizing from this example, we have that the phase of a function

$$f(z)=f_1(z)\cdot f_2(z)\cdot f_3(z)\cdots$$

is the algebraic sum of the phase of its individual factors:

$$\arg f(z)=\arg f_1(z)+\arg f_2(z)+\arg f_3(z)+\cdots .$$

The phase of an individual factor may be taken as the arctangent of the ratio of its imaginary part to its real part (choosing the appropriate branch of the arctan function $\tan^{-1} y/x$, which has infinitely many branches),

$$\arg f_i(z) = \tan^{-1}\left(\frac{v_i}{u_i}\right).$$

For the case of a factor of the form

$$f_i(z) = (z - z_0),$$

the phase corresponds to the phase angle of a two-dimensional vector from $+z_0$ to z, the phase increasing by 2π as the point $+z_0$ is encircled. Conversely, the traversal of any closed loop not encircling z_0 does not change the phase of $z - z_0$.

Exercises

6.6.1 The function $f(z)$ expanded in a Laurent series exhibits a pole of order m at $z = z_0$. Show that the coefficient of $(z - z_0)^{-1}$, a_{-1}, is given by

$$a_{-1} = \frac{1}{(m-1)!}\frac{d^{m-1}}{dz^{m-1}}\Big[(z - z_0)^m f(z)\Big]_{z=z_0},$$

with

$$a_{-1} = \Big[(z - z_0) f(z)\Big]_{z=z_0},$$

when the pole is a simple pole ($m = 1$). These equations for a_{-1} are extremely useful in determining the residue to be used in the residue theorem of Section 7.1.
Hint. The technique that was so successful in proving the uniqueness of power series, Section 5.7, will work here also.

6.6.2 A function $f(z)$ can be represented by

$$f(z) = \frac{f_1(z)}{f_2(z)},$$

in which $f_1(z)$ and $f_2(z)$ are analytic. The denominator, $f_2(z)$, vanishes at $z = z_0$, showing that $f(z)$ has a pole at $z = z_0$. However, $f_1(z_0) \neq 0$, $f_2'(z_0) \neq 0$. Show that a_{-1}, the coefficient of $(z - z_0)^{-1}$ in a Laurent expansion of $f(z)$ at $z = z_0$, is given by

$$a_{-1} = \frac{f_1(z_0)}{f_2'(z_0)}.$$

(This result leads to the Heaviside expansion theorem, Exercise 15.12.11.)

6.6.3 In analogy with Example 6.6.1, consider in detail the phase of each factor and the resultant overall phase of $f(z) = (z^2 + 1)^{1/2}$ following a contour similar to that of Fig. 6.16 but encircling the new branch points.

6.6.4 The Legendre function of the second kind, $Q_\nu(z)$, has branch points at $z = \pm 1$. The branch points are joined by a cut line along the real (x) axis.

(a) Show that $Q_0(z) = \frac{1}{2}\ln((z+1)/(z-1))$ is single-valued (with the real axis $-1 \le x \le 1$ taken as a cut line).

(b) For real argument x and $|x| < 1$ it is convenient to take

$$Q_0(x) = \frac{1}{2}\ln\frac{1+x}{1-x}.$$

Show that

$$Q_0(x) = \frac{1}{2}\big[Q_0(x+i0) + Q_0(x-i0)\big].$$

Here $x + i0$ indicates that z approaches the real axis from above, and $x - i0$ indicates an approach from below.

6.6.5 As an example of an essential singularity, consider $e^{1/z}$ as z approaches zero. For any complex number z_0, $z_0 \neq 0$, show that

$$e^{1/z} = z_0$$

has an infinite number of solutions.

6.7 MAPPING

In the preceding sections we have defined analytic functions and developed some of their main features. Here we introduce some of the more geometric aspects of functions of complex variables, aspects that will be useful in visualizing the integral operations in Chapter 7 and that are valuable in their own right in solving Laplace's equation in two-dimensional systems.

In ordinary analytic geometry we may take $y = f(x)$ and then plot y versus x. Our problem here is more complicated, for z is a function of two variables, x and y. We use the notation

$$w = f(z) = u(x, y) + iv(x, y). \tag{6.79}$$

Then for a point in the z-plane (specific values for x and y) there may correspond specific values for $u(x, y)$ and $v(x, y)$ that then yield a point in the w-plane. As points in the z-plane transform, or are mapped into points in the w-plane, lines or areas in the z-plane will be mapped into lines or areas in the w-plane. Our immediate purpose is to see how lines and areas map from the z-plane to the w-plane for a number of simple functions.

Translation

$$w = z + z_0. \tag{6.80}$$

The function w is equal to the variable z plus a constant, $z_0 = x_0 + iy_0$. By Eqs. (6.1) and (6.79),

$$u = x + x_0, \qquad v = y + y_0, \tag{6.81}$$

representing a pure translation of the coordinate axes, as shown in Fig. 6.17.

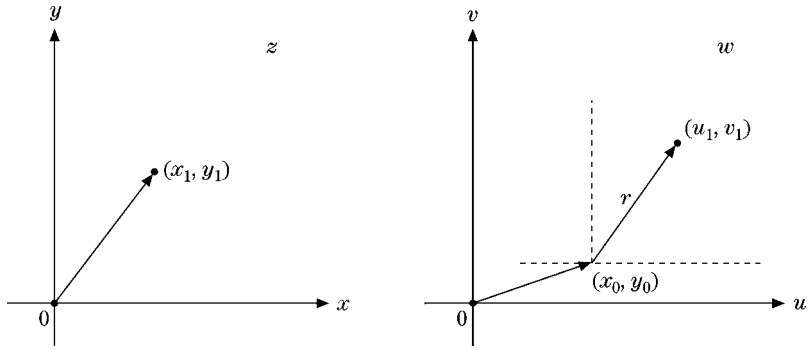

FIGURE 6.17 Translation.

Rotation

$$w = zz_0. \tag{6.82}$$

Here it is convenient to return to the polar representation, using

$$w = \rho e^{i\varphi}, \qquad z = re^{i\theta}, \qquad \text{and} \qquad z_0 = r_0 e^{i\theta_0}, \tag{6.83}$$

then

$$\rho e^{i\varphi} = rr_0 e^{i(\theta+\theta_0)}, \tag{6.84}$$

or

$$\rho = rr_0, \qquad \varphi = \theta + \theta_0. \tag{6.85}$$

Two things have occurred. First, the modulus r has been modified, either expanded or contracted, by the factor r_0. Second, the argument θ has been increased by the additive constant θ_0 (Fig. 6.18). This represents a rotation of the complex variable through an angle θ_0. For the special case of $z_0 = i$, we have a pure rotation through $\pi/2$ radians.

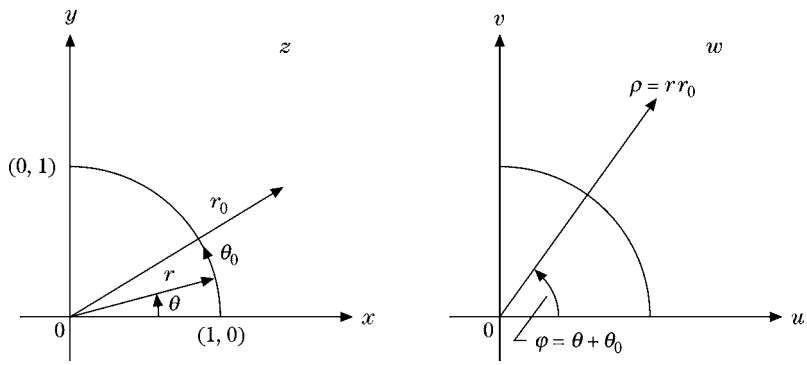

FIGURE 6.18 Rotation.

Inversion

$$w = \frac{1}{z}. \tag{6.86}$$

Again, using the polar form, we have

$$\rho e^{i\varphi} = \frac{1}{re^{i\theta}} = \frac{1}{r}e^{-i\theta}, \tag{6.87}$$

which shows that

$$\rho = \frac{1}{r}, \qquad \varphi = -\theta. \tag{6.88}$$

The first part of Eq. (6.87) shows that inversion clearly. The interior of the unit circle is mapped onto the exterior and vice versa (Fig. 6.19). In addition, the second part of Eq. (6.87) shows that the polar angle is reversed in sign. Equation (6.88) therefore also involves a reflection of the y-axis, exactly like the complex conjugate equation.

To see how curves in the z-plane transform into the w-plane, we return to the Cartesian form:

$$u + iv = \frac{1}{x + iy}. \tag{6.89}$$

Rationalizing the right-hand side by multiplying numerator and denominator by z^* and then equating the real parts and the imaginary parts, we have

$$\begin{aligned} u &= \frac{x}{x^2 + y^2}, & x &= \frac{u}{u^2 + v^2}, \\ v &= -\frac{y}{x^2 + y^2}, & y &= -\frac{v}{u^2 + v^2}. \end{aligned} \tag{6.90}$$

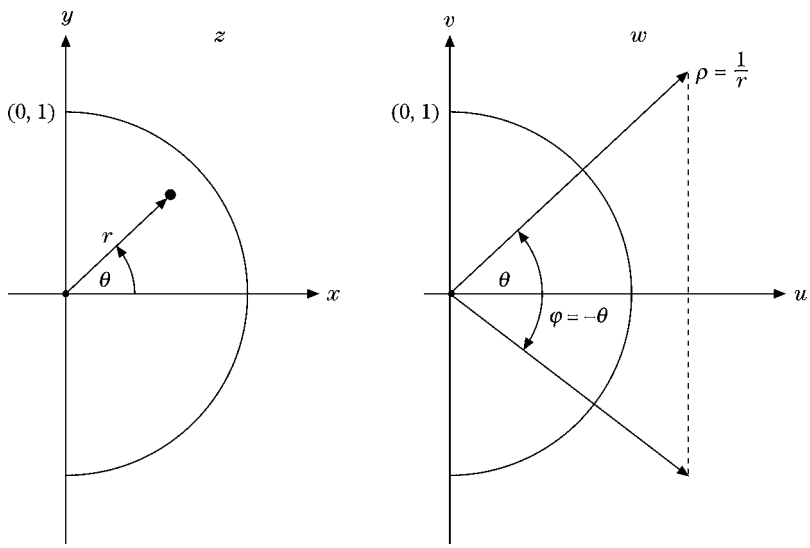

FIGURE 6.19 Inversion.

A circle centered at the origin in the z-plane has the form

$$x^2 + y^2 = r^2 \tag{6.91}$$

and by Eqs. (6.90) transforms into

$$\frac{u^2}{(u^2+v^2)^2} + \frac{v^2}{(u^2+v^2)^2} = r^2. \tag{6.92}$$

Simplifying Eq. (6.92), we obtain

$$u^2 + v^2 = \frac{1}{r^2} = \rho^2, \tag{6.93}$$

which describes a circle in the w-plane also centered at the origin.

The horizontal line $y = c_1$ transforms into

$$\frac{-v}{u^2+v^2} = c_1, \tag{6.94}$$

or

$$u^2 + \left(v + \frac{1}{2c_1}\right)^2 = \frac{1}{(2c_1)^2}, \tag{6.95}$$

which describes a circle in the w-plane of radius $(1/2c_1)$ and centered at $u = 0$, $v = -\frac{1}{2c_1}$ (Fig. 6.20).

We pick up the other three possibilities, $x = \pm c_1$, $y = -c_1$, by rotating the xy-axes. In general, any straight line or circle in the z-plane will transform into a straight line or a circle in the w-plane (compare Exercise 6.7.1).

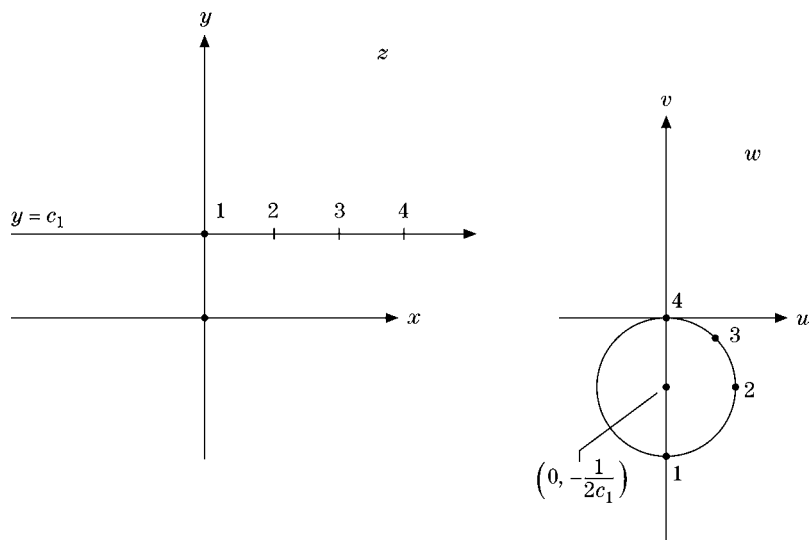

FIGURE 6.20 Inversion, line ↔ circle.

Branch Points and Multivalent Functions

The three transformations just discussed have all involved one-to-one correspondence of points in the z-plane to points in the w-plane. Now to illustrate the variety of transformations that are possible and the problems that can arise, we introduce first a two-to-one correspondence and then a many-to-one correspondence. Finally, we take up the inverses of these two transformations.

Consider first the transformation

$$w = z^2, \tag{6.96}$$

which leads to

$$\rho = r^2, \qquad \varphi = 2\theta. \tag{6.97}$$

Clearly, our transformation is nonlinear, for the modulus is squared, but the significant feature of Eq. (6.96) is that the phase angle or argument is doubled. This means that the

- first quadrant of z, $0 \le \theta < \dfrac{\pi}{2}$, $\rightarrow$ upper half-plane of w, $0 \le \varphi < \pi$,
- upper half-plane of z, $0 \le \theta < \pi$, $\rightarrow$ whole plane of w, $0 \le \varphi < 2\pi$.

The lower half-plane of z maps into the already covered entire plane of w, thus covering the w-plane a second time. This is our two-to-one correspondence, that is, two distinct points in the z-plane, z_0 and $z_0 e^{i\pi} = -z_0$, corresponding to the single point $w = z_0^2$.

In Cartesian representation,

$$u + iv = (x + iy)^2 = x^2 - y^2 + i2xy, \tag{6.98}$$

leading to

$$u = x^2 - y^2, \qquad v = 2xy. \tag{6.99}$$

Hence the lines $u = c_1, v = c_2$ in the w-plane correspond to $x^2 - y^2 = c_1, 2xy = c_2$, rectangular (and orthogonal) hyperbolas in the z-plane (Fig. 6.21). To every point on the hyperbola $x^2 - y^2 = c_1$ in the right half-plane, $x > 0$, one point on the line $u = c_1$ corresponds, and vice versa. However, every point on the line $u = c_1$ also corresponds to a point on the hyperbola $x^2 - y^2 = c_1$ in the left half-plane, $x < 0$, as already explained.

It will be shown in Section 6.8 that if lines in the w-plane are orthogonal, the corresponding lines in the z-plane are also orthogonal, as long as the transformation is analytic. Since $u = c_1$ and $v = c_2$ are constructed perpendicular to each other, the corresponding hyperbolas in the z-plane are orthogonal. We have constructed a new orthogonal system of hyperbolic lines (or surfaces if we add an axis perpendicular to x and y). Exercise 2.1.3 was an analysis of this system. It might be noted that if the hyperbolic lines are electric or magnetic lines of force, then we have a quadrupole lens useful in focusing beams of high-energy particles.

The inverse of the fourth transformation (Eq. (6.96)) is

$$w = z^{1/2}. \tag{6.100}$$

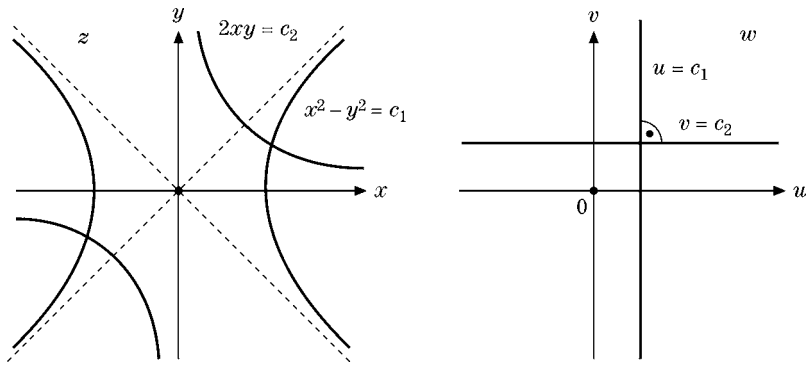

FIGURE 6.21 Mapping — hyperbolic coordinates.

From the relation

$$\rho e^{i\varphi} = r^{1/2} e^{i\theta/2} \tag{6.101}$$

and

$$2\varphi = \theta, \tag{6.102}$$

we now have two points in the w-plane (arguments φ and $\varphi + \pi$) corresponding to one point in the z-plane (except for the point $z = 0$). Or, to put it another way, θ and $\theta + 2\pi$ correspond to φ and $\varphi + \pi$, two distinct points in the w-plane. This is the complex variable analog of the simple real variable equation $y^2 = x$, in which two values of y, plus and minus, correspond to each value of x.

The important point here is that we can make the function w of Eq. (6.100) a single-valued function instead of a double-valued function if we agree to restrict θ to a range such as $0 \le \theta < 2\pi$. This may be done by agreeing never to cross the line $\theta = 0$ in the z-plane (Fig. 6.22). Such a line of demarcation is called a **cut line** or **branch cut**. Note that branch points occur in pairs.

The **cut line joins the two branch point singularities**, here at 0 and ∞ (for the latter, transform $z = 1/t$ for $t \to 0$). Any line from $z = 0$ to infinity would serve equally well. The purpose of the cut line is to restrict the argument of z. The points z and $z\exp(2\pi i)$ coincide in the z-plane but yield different points w and $-w = w\exp(\pi i)$ in the w-plane. Hence in the absence of a cut line, the function $w = z^{1/2}$ is ambiguous. Alternatively, since the function $w = z^{1/2}$ is double-valued, we can also glue two sheets of the complex z-plane together along the branch cut so that $\arg(z)$ increases beyond 2π along the branch cut and continues from 4π on the second sheet to reach the same function values for z as for $ze^{-4\pi i}$, that is, the start on the first sheet again. This construction is called the **Riemann surface** of $w = z^{1/2}$. We shall encounter branch points and cut lines (branch cuts) frequently in Chapter 7.

The transformation

$$w = e^z \tag{6.103}$$

leads to

$$\rho e^{i\varphi} = e^{x+iy}, \tag{6.104}$$

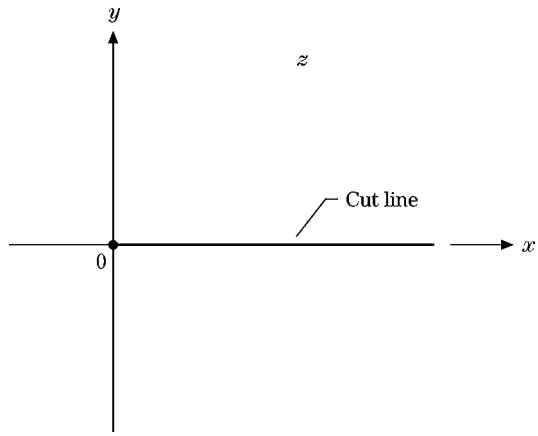

FIGURE 6.22 A cut line.

or

$$\rho = e^x, \qquad \varphi = y. \tag{6.105}$$

If y ranges from $0 \le y < 2\pi$ (or $-\pi < y \le \pi$), then φ covers the same range. But this is the whole w-plane. In other words, a horizontal strip in the z-plane of width 2π maps into the entire w-plane. Further, any point $x + i(y + 2n\pi)$, in which n is any integer, maps into the same point (by Eq. (6.104)) in the w-plane. We have a many-(infinitely many)-to-one correspondence.

Finally, as the inverse of the fifth transformation (Eq. (6.103)), we have

$$w = \ln z. \tag{6.106}$$

By expanding it, we obtain

$$u + iv = \ln re^{i\theta} = \ln r + i\theta. \tag{6.107}$$

For a given point z_0 in the z-plane the argument θ is unspecified within an integral multiple of 2π. This means that

$$v = \theta + 2n\pi, \tag{6.108}$$

and, as in the exponential transformation, we have an infinitely many-to-one correspondence.

Equation (6.108) has a nice physical representation. If we go around the unit circle in the z-plane, $r = 1$, and by Eq. (6.107), $u = \ln r = 0$; but $v = \theta$, and θ is steadily increasing and continues to increase as θ continues past 2π.

The cut line joins the branch point at the origin with infinity. As θ increases past 2π we glue a new sheet of the complex z-plane along the cut line, etc. Going around the unit circle in the z-plane is like the advance of a screw as it is rotated or the ascent of a person walking up a spiral staircase (Fig. 6.23), which is the **Riemann surface** of $w = \ln z$.

As in the preceding example, we can also make the correspondence unique (and Eq. (6.106) unambiguous) by restricting θ to a range such as $0 \le \theta < 2\pi$ by taking the

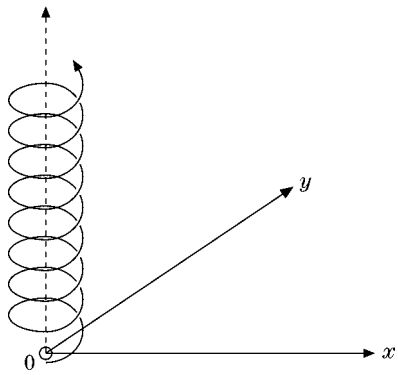

FIGURE 6.23 This is the Riemann surface for $\ln z$, a multivalued function.

line $\theta = 0$ (positive real axis) as a cut line. This is equivalent to taking one and only one complete turn of the spiral staircase.

The concept of mapping is a very broad and useful one in mathematics. Our mapping from a complex z-plane to a complex w-plane is a simple generalization of one definition of function: a mapping of x (from one set) into y in a second set. A more sophisticated form of mapping appears in Section 1.15 where we use the Dirac delta function $\delta(x-a)$ to map a function $f(x)$ into its value at the point a. Then in Chapter 15 integral transforms are used to map one function $f(x)$ in x-space into a second (related) function $F(t)$ in t-space.

Exercises

6.7.1 How do circles centered on the origin in the z-plane transform for

(a) $w_1(z) = z + \dfrac{1}{z}$, (b) $w_2(z) = z - \dfrac{1}{z}$, for $z \neq 0$?

What happens when $|z| \to 1$?

6.7.2 What part of the z-plane corresponds to the interior of the unit circle in the w-plane if

(a) $w = \dfrac{z-1}{z+1}$, (b) $w = \dfrac{z-i}{z+i}$?

6.7.3 Discuss the transformations

(a) $w(z) = \sin z$, (c) $w(z) = \sinh z$,

(b) $w(z) = \cos z$, (d) $w(z) = \cosh z$.

Show how the lines $x = c_1$, $y = c_2$ map into the w-plane. Note that the last three transformations can be obtained from the first one by appropriate translation and/or rotation.

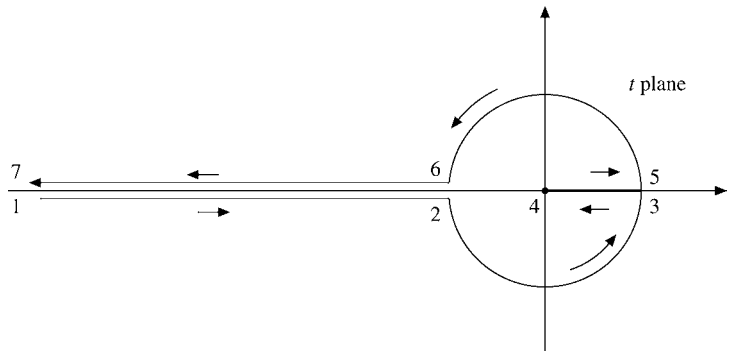

FIGURE 6.24 Bessel function integration contour.

6.7.4 Show that the function

$$w(z) = \left(z^2 - 1\right)^{1/2}$$

is single-valued if we take $-1 \le x \le 1$, $y = 0$ as a cut line.

6.7.5 Show that negative numbers have logarithms in the complex plane. In particular, find $\ln(-1)$.

ANS. $\ln(-1) = i\pi$.

6.7.6 An integral representation of the Bessel function follows the contour in the t-plane shown in Fig. 6.24. Map this contour into the θ-plane with $t = e^{\theta}$. Many additional examples of mapping are given in Chapters 11, 12, and 13.

6.7.7 For noninteger m, show that the binomial expansion of Exercise 6.5.2 holds only for a suitably defined branch of the function $(1 + z)^m$. Show how the z-plane is cut. Explain why $|z| < 1$ may be taken as the circle of convergence for the expansion of this branch, in light of the cut you have chosen.

6.7.8 The Taylor expansion of Exercises 6.5.2 and 6.7.7 is **not** suitable for branches other than the one suitably defined branch of the function $(1 + z)^m$ for noninteger m. [Note that other branches cannot have the same Taylor expansion since they must be distinguishable.] Using the same branch cut of the earlier exercises for all other branches, find the corresponding Taylor expansions, detailing the phase assignments and Taylor coefficients.

6.8 CONFORMAL MAPPING

In Section 6.7 hyperbolas were mapped into straight lines and straight lines were mapped into circles. Yet in all these transformations one feature stayed constant. This constancy was a result of the fact that all the transformations of Section 6.7 were analytic.

As long as $w = f(z)$ is an analytic function, we have

$$\frac{df}{dz} = \frac{dw}{dz} = \lim_{\Delta z \to 0} \frac{\Delta w}{\Delta z}. \tag{6.109}$$

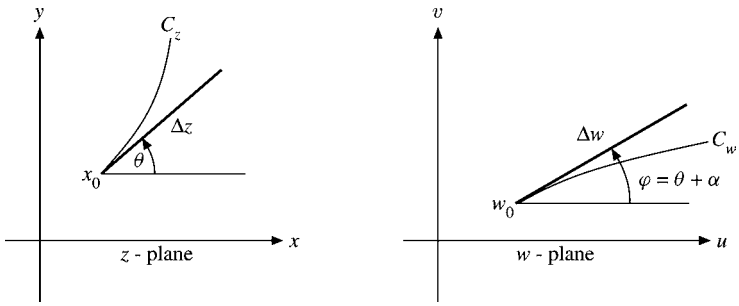

FIGURE 6.25 Conformal mapping — preservation of angles.

Assuming that this equation is in polar form, we may equate modulus to modulus and argument to argument. For the latter (assuming that $df/dz \neq 0$),

$$\arg \lim_{\Delta z \to 0} \frac{\Delta w}{\Delta z} = \lim_{\Delta z \to 0} \arg \frac{\Delta w}{\Delta z}$$

$$= \lim_{\Delta z \to 0} \arg \Delta w - \lim_{\Delta z \to 0} \arg \Delta z = \arg \frac{df}{dz} = \alpha, \tag{6.110}$$

where α, the argument of the derivative, may depend on z but is a constant for a fixed z, independent of the direction of approach. To see the significance of this, consider two curves C_z in the z-plane and the corresponding curve C_w in the w-plane (Fig. 6.25). The increment Δz is shown at an angle of θ relative to the real (x) axis, whereas the corresponding increment Δw forms an angle of φ with the real (u) axis. From Eq. (6.110),

$$\varphi = \theta + \alpha, \tag{6.111}$$

or any line in the z-plane is rotated through an angle α in the w-plane as long as w is an analytic transformation and the derivative is not zero.[17]

Since this result holds for any line through z_0, it will hold for a pair of lines. Then for the angle between these two lines,

$$\varphi_2 - \varphi_1 = (\theta_2 + \alpha) - (\theta_1 + \alpha) = \theta_2 - \theta_1, \tag{6.112}$$

which shows that the included angle is preserved under an analytic transformation. Such angle-preserving transformations are called **conformal**. The rotation angle α will, in general, depend on z. In addition $|f'(z)|$ will usually be a function of z.

Historically, these conformal transformations have been of great importance to scientists and engineers in solving Laplace's equation for problems of electrostatics, hydrodynamics, heat flow, and so on. Unfortunately, the conformal transformation approach, however elegant, is limited to problems that can be reduced to two dimensions. The method is often beautiful if there is a high degree of symmetry present but often impossible if the symmetry is broken or absent. Because of these limitations and primarily because electronic computers offer a useful alternative (iterative solution of the partial differential equation), the details and applications of conformal mappings are omitted.

[17] If $df/dz = 0$, its argument or phase is undefined and the (analytic) transformation will not necessarily preserve angles.

Exercises

6.8.1 Expand $w(x)$ in a Taylor series about the point $z = z_0$, where $f'(z_0) = 0$. (Angles are not preserved.) Show that if the first $n - 1$ derivatives vanish but $f^{(n)}(z_0) \neq 0$, then angles in the z-plane with vertices at $z = z_0$ appear in the w-plane multiplied by n.

6.8.2 Develop the transformations that create each of the four cylindrical coordinate systems:

(a) Circular cylindrical: $x = \rho\cos\varphi$, $y = \rho\sin\varphi$.

(b) Elliptic cylindrical: $x = a\cosh u\cos v$, $y = a\sinh u\sin v$.

(c) Parabolic cylindrical: $x = \xi\eta$, $y = \frac{1}{2}(\eta^2 - \xi^2)$.

(d) Bipolar: $x = \dfrac{a\sinh\eta}{\cosh\eta - \cos\xi}$, $y = \dfrac{a\sin\xi}{\cosh\eta - \cos\xi}$.

Note. These transformations are not necessarily analytic.

6.8.3 In the transformation

$$e^z = \frac{a - w}{a + w},$$

how do the coordinate lines in the z-plane transform? What coordinate system have you constructed?

Additional Readings

Ahlfors, L. V., *Complex Analysis*, 3rd ed. New York: McGraw-Hill (1979). This text is detailed, thorough, rigorous, and extensive.

Churchill, R. V., J. W. Brown, and R. F. Verkey, *Complex Variables and Applications*, 5th ed. New York: McGraw-Hill (1989). This is an excellent text for both the beginning and advanced student. It is readable and quite complete. A detailed proof of the Cauchy–Goursat theorem is given in Chapter 5.

Greenleaf, F. P., *Introduction to Complex Variables*. Philadelphia: Saunders (1972). This very readable book has detailed, careful explanations.

Kurala, A., *Applied Functions of a Complex Variable*. New York: Wiley (Interscience) (1972). An intermediate-level text designed for scientists and engineers. Includes many physical applications.

Levinson, N., and R. M. Redheffer, *Complex Variables*. San Francisco: Holden-Day (1970). This text is written for scientists and engineers who are interested in applications.

Morse, P. M., and H. Feshbach, *Methods of Theoretical Physics*. New York: McGraw-Hill (1953). Chapter 4 is a presentation of portions of the theory of functions of a complex variable of interest to theoretical physicists.

Remmert, R., *Theory of Complex Functions*. New York: Springer (1991).

Sokolnikoff, I. S., and R. M. Redheffer, *Mathematics of Physics and Modern Engineering*, 2nd ed. New York: McGraw-Hill (1966). Chapter 7 covers complex variables.

Spiegel, M. R., *Complex Variables*. New York: McGraw-Hill (1985). An excellent summary of the theory of complex variables for scientists.

Titchmarsh, E. C., *The Theory of Functions*, 2nd ed. New York: Oxford University Press (1958). A classic.

Watson, G. N., *Complex Integration and Cauchy's Theorem*. New York: Hafner (orig. 1917, reprinted 1960). A short work containing a rigorous development of the Cauchy integral theorem and integral formula. Applications to the calculus of residues are included. *Cambridge Tracts in Mathematics, and Mathematical Physics*, No. 15.

Other references are given at the end of Chapter 15.

CHAPTER 7

FUNCTIONS OF A COMPLEX VARIABLE II

In this chapter we return to the analysis that started with the Cauchy–Riemann conditions in Chapter 6 and develop the residue theorem, with major applications to the evaluation of definite and principal part integrals of interest to scientists and asymptotic expansion of integrals by the method of steepest descent. We also develop further specific analytic functions, such as pole expansions of meromorphic functions and product expansions of entire functions. Dispersion relations are included because they represent an important application of complex variable methods for physicists.

7.1 CALCULUS OF RESIDUES

Residue Theorem

If the Laurent expansion of a function $f(z) = \sum_{n=-\infty}^{\infty} a_n (z - z_0)^n$ is integrated term by term by using a closed contour that encircles one isolated singular point z_0 once in a counterclockwise sense, we obtain (Exercise 6.4.1)

$$a_n \oint (z - z_0)^n \, dz = a_n \frac{(z - z_0)^{n+1}}{n+1} \bigg|_{z_1}^{z_1} = 0, \qquad n \neq -1. \tag{7.1}$$

However, if $n = -1$,

$$a_{-1} \oint (z - z_0)^{-1} \, dz = a_{-1} \oint \frac{i r e^{i\theta} \, d\theta}{r e^{i\theta}} = 2\pi i a_{-1}. \tag{7.2}$$

Summarizing Eqs. (7.1) and (7.2), we have

$$\frac{1}{2\pi i} \oint f(z) \, dz = a_{-1}. \tag{7.3}$$

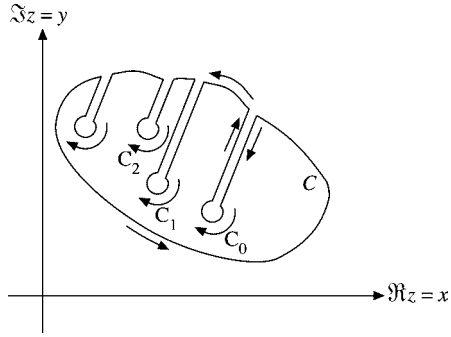

FIGURE 7.1 Excluding isolated singularities.

The constant a_{-1}, the coefficient of $(z - z_0)^{-1}$ in the Laurent expansion, is called the residue of $f(z)$ at $z = z_0$.

A set of isolated singularities can be handled by deforming our contour as shown in Fig. 7.1. Cauchy's integral theorem (Section 6.3) leads to

$$\oint_C f(z)\,dz + \oint_{C_0} f(z)\,dz + \oint_{C_1} f(z)\,dz + \oint_{C_2} f(z)\,dz + \cdots = 0. \tag{7.4}$$

The circular integral around any given singular point is given by Eq. (7.3),

$$\oint_{C_i} f(z)\,dz = -2\pi i a_{-1,z_i}, \tag{7.5}$$

assuming a Laurent expansion about the singular point $z = z_i$. The negative sign comes from the clockwise integration, as shown in Fig. 7.1. Combining Eqs. (7.4) and (7.5), we have

$$\boxed{\oint_C f(z)\,dz = 2\pi i (a_{-1z_0} + a_{-1z_1} + a_{-1z_2} + \cdots)}$$

$$= 2\pi i \times (\text{sum of enclosed residues}). \tag{7.6}$$

This is the **residue theorem**. The problem of evaluating one or more contour integrals is replaced by the algebraic problem of computing residues at the enclosed singular points.

We first use this residue theorem to develop the concept of the Cauchy principal value. Then in the remainder of this section we apply the residue theorem to a wide variety of definite integrals of mathematical and physical interest.

Using the transformation $z = 1/w$ for w approaching 0, we can find the nature of a singularity at z going to ∞ and the residue of a function $f(z)$ with just isolated singularities and no branch points. In such cases we know that

$$\sum \{\text{residues in the finite } z\text{-plane}\} + \{\text{residue at } z \to \infty\} = 0.$$

Cauchy Principal Value

Occasionally an isolated pole will be directly on the contour of integration, causing the integral to diverge. Let us illustrate a physical case.

Example 7.1.1 FORCED CLASSICAL OSCILLATOR

The inhomogeneous differential equation for a classical, undamped, driven harmonic oscillator,

$$\ddot{x}(t) + \omega_0^2 x(t) = f(t), \tag{7.7}$$

may be solved by representing the driving force $f(t) = \int \delta(t' - t) f(t')\, dt'$ as a superposition of impulses by analogy with an extended charge distribution in electrostatics.[1] If we solve first the simpler differential equation

$$\ddot{G} + \omega_0^2 G = \delta(t - t') \tag{7.8}$$

for $G(t, t')$, which is independent of the driving term f (model dependent), then $x(t) = \int G(t, t') f(t')\, dt'$ solves the original problem. First, we verify this by substituting the integrals for $x(t)$ and its time derivatives into the differential equation for $x(t)$ using the differential equation for G. Then we look for $G(t, t') = \int \tilde{G}(\omega) e^{i\omega t} \frac{d\omega}{2\pi}$ in terms of an integral weighted by $\tilde{G}$, which is suggested by a similar integral form for $\delta(t - t') = \int e^{i\omega(t-t')} \frac{d\omega}{2\pi}$ (see Eq. (1.193c) in Section 1.15).

Upon substituting G and $\ddot{G}$ into the differential equation for G, we obtain

$$\int \left[\left(\omega_0^2 - \omega^2\right) \tilde{G} - e^{-i\omega t'} \right] e^{i\omega t}\, d\omega = 0. \tag{7.9}$$

Because this integral is zero for all t, the expression in brackets must vanish for all ω. This relation is no longer a differential equation but an algebraic relation that we can solve for $\tilde{G}$:

$$\tilde{G}(\omega) = \frac{e^{-i\omega t'}}{\omega_0^2 - \omega^2} = \frac{e^{-i\omega t'}}{2\omega_0(\omega + \omega_0)} - \frac{e^{-i\omega t'}}{2\omega_0(\omega - \omega_0)}. \tag{7.10}$$

Substituting $\tilde{G}$ into the integral for G yields

$$G(t, t') = \frac{1}{4\pi\omega_0} \int_{-\infty}^{\infty} \left[\frac{e^{i\omega(t-t')}}{\omega + \omega_0} - \frac{e^{i\omega(t-t')}}{\omega - \omega_0} \right] d\omega. \tag{7.11}$$

Here, the dependence of G on $t - t'$ in the exponential is consistent with the same dependence of $\delta(t - t')$, its driving term. Now, the problem is that this integral diverges because the integrand blows up at $\omega = \pm\omega_0$, since the integration goes right through the first-order poles. To explain why this happens, we note that the δ-function driving term for G includes all frequencies with the same amplitude. Next, we see that the equation for $\tilde{G}$ at $t' = 0$ has its driving term equal to unity for all frequencies ω, including the resonant ω_0.

[1] Adapted from A. Yu. Grosberg, priv. comm.

We know from physics that forcing an oscillator at resonance leads to an indefinitely growing amplitude when there is no friction. With friction, the amplitude remains finite, even at resonance. This suggests including a small friction term in the differential equations for $x(t)$ and G.

With a small friction term $\eta\dot{G}$, $\eta > 0$, in the differential equation for $G(t, t')$ (and $\eta\dot{x}$ for $x(t)$), we can still solve the algebraic equation

$$\left(\omega_0^2 - \omega^2 + i\eta\omega\right)\tilde{G} = e^{-i\omega t'} \tag{7.12}$$

for $\tilde{G}$ with friction. The solution is

$$\tilde{G} = \frac{e^{-i\omega t'}}{\omega_0^2 - \omega^2 + i\eta\omega} = \frac{e^{-i\omega t'}}{2\Omega}\left(\frac{1}{\omega - \omega_-} - \frac{1}{\omega - \omega_+}\right), \tag{7.13}$$

$$\omega_\pm = \pm\Omega + \frac{i\eta}{2}, \qquad \Omega = \omega_0\sqrt{1 - \left(\frac{\eta}{2\omega_0}\right)^2}. \tag{7.14}$$

For small friction, $0 < \eta \ll \omega_0$, Ω is nearly equal to ω_0 and real, whereas $\omega_\pm$ each pick up a small imaginary part. This means that the integration of the integral for G,

$$G(t, t') = \frac{1}{4\pi\Omega}\int_{-\infty}^{\infty}\left[\frac{e^{i\omega(t-t')}}{\omega - \omega_-} - \frac{e^{i\omega(t-t')}}{\omega - \omega_+}\right]d\omega, \tag{7.15}$$

no longer encounters a pole and remains finite. ■

This treatment of an integral with a pole moves the pole off the contour and then considers the limiting behavior as it is brought back, as in Example 7.1.1 for $\eta \to 0$. This example also suggests treating ω as a complex variable in case the singularity is a first-order pole, deforming the integration path to avoid the singularity, which is equivalent to adding a small imaginary part to the pole position, and evaluating the integral by means of the residue theorem.

Therefore, if the integration path of an integral $\int \frac{dz}{z-x_0}$ for real x_0 goes right through the pole x_0, we may deform the contour to include or exclude the residue, as desired, by including a semicircular detour of **infinitesimal radius**. This is shown in Fig. 7.2. The integration over the semicircle then gives, with $z - x_0 = \delta e^{i\varphi}$, $dz = i\,\delta e^{i\varphi}\,d\varphi$ (see Eq. (6.27a)),

$$\int \frac{dz}{z - x_0} = i\int_{\pi}^{2\pi} d\varphi = i\pi, \text{ i.e., } \pi i a_{-1} \qquad \text{if counterclockwise,}$$

$$\int \frac{dz}{z - x_0} = i\int_{\pi}^{0} d\varphi = -i\pi, \text{ i.e., } -\pi i a_{-1} \qquad \text{if clockwise.}$$

This contribution, $+$ or $-$, appears on the left-hand side of Eq. (7.6). If our detour were clockwise, the residue would not be enclosed and there would be no corresponding term on the right-hand side of Eq. (7.6).

However, if our detour were counterclockwise, this residue would be enclosed by the contour C and a term $2\pi i a_{-1}$ would appear on the right-hand side of Eq. (7.6).

The net result for either clockwise or counterclockwise detour is that a simple pole on the contour is counted as one-half of what it would be if it were within the contour. This corresponds to taking the Cauchy principal value.

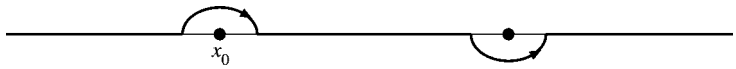

FIGURE 7.2 Bypassing singular points.

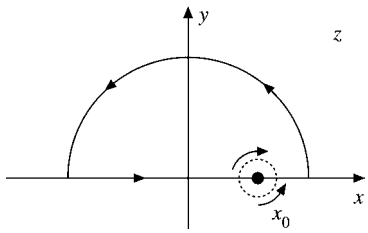

FIGURE 7.3 Closing the contour with an infinite-radius semicircle.

For instance, let us suppose that $f(z)$ with a simple pole at $z = x_0$ is integrated over the entire real axis. The contour is closed with an infinite semicircle in the upper half-plane (Fig. 7.3). Then

$$\oint f(z)\,dz = \int_{-\infty}^{x_0-\delta} f(x)\,dx + \int_{C_{x_0}} f(z)\,dz + \int_{x_0+\delta}^{\infty} f(x)\,dx + \int_{C} \text{infinite semicircle}$$
$$= 2\pi i \sum \textbf{enclosed} \text{ residues.} \tag{7.16}$$

If the small semicircle C_{x_0}, includes x_0 (by going below the x-axis, counterclockwise), x_0 is enclosed, and its contribution appears **twice**—as $\pi i a_{-1}$ in $\int_{C_{x_0}}$ and as $2\pi i a_{-1}$ in the term $2\pi i \sum$ enclosed residues—for a net contribution of $\pi i a_{-1}$. If the upper small semicircle is selected, x_0 is excluded. The only contribution is from the **clockwise** integration over C_{x_0}, which yields $-\pi i a_{-1}$. Moving this to the extreme right of Eq. (7.16), we have $+\pi i a_{-1}$, as before.

The integrals along the x-axis may be combined and the semicircle radius permitted to approach zero. We therefore define

$$\lim_{\delta\to 0}\left\{\int_{-\infty}^{x_0-\delta} f(x)\,dx + \int_{x_0+\delta}^{\infty} f(x)\,dx\right\} = P\int_{-\infty}^{\infty} f(x)\,dx. \tag{7.17}$$

P indicates the Cauchy **principal value** and represents the preceding limiting process. Note that the Cauchy principal value is a balancing (or canceling) process. In the vicinity of our singularity at $z = x_0$,

$$f(x) \approx \frac{a_{-1}}{x - x_0}. \tag{7.18}$$

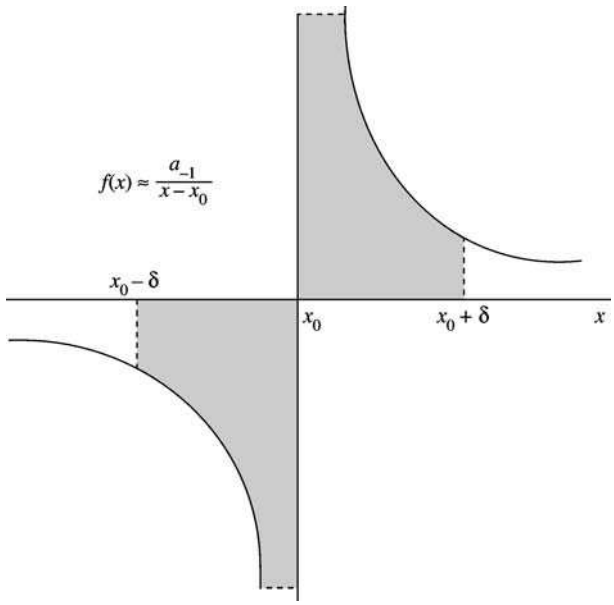

FIGURE 7.4 Cancellation at a simple pole.

This is odd, relative to x_0. The symmetric or even interval (relative to x_0) provides cancellation of the shaded areas, Fig. 7.4. The contribution of the singularity is in the integration about the semicircle.

In general, if a function $f(x)$ has a singularity x_0 somewhere inside the interval $a \leq x_0 \leq b$ and is integrable over every portion of this interval that does not contain the point x_0, then we **define**

$$\int_a^b f(x)\,dx = \lim_{\delta_1 \to 0} \int_a^{x_0 - \delta_1} f(x)\,dx + \lim_{\delta_2 \to 0} \int_{x_0 + \delta_2}^b f(x)\,dx,$$

when the limit exists as $\delta_j \to 0$ **independently**, else the integral is said to diverge. If this limit does not exist but the limit $\delta_1 = \delta_2 = \delta \to 0$ exists, it is defined to be the principal value of the integral.

This same limiting technique is applicable to the integration limits $\pm\infty$. We define

$$\int_{-\infty}^{\infty} f(x)\,dx = \lim_{a \to -\infty, b \to \infty} \int_a^b f(x)\,dx, \tag{7.19a}$$

if the integral exists with a, b approaching their limits independently, else the integral diverges. In case the integral diverges but

$$\lim_{a \to \infty} \int_{-a}^{a} f(x)\,dx = P \int_{-\infty}^{\infty} f(x)\,dx \tag{7.19b}$$

exist, it is defined as its principal value.

Pole Expansion of Meromorphic Functions

Analytic functions $f(z)$ that have only isolated poles as singularities are called **meromorphic**. Examples are $\cot z$ [from $\frac{d}{dz}\ln\sin z$ in Eq. (5.210)] and ratios of polynomials. For simplicity we assume that these poles at finite $z = a_n$ with $0 < |a_1| < |a_2| < \cdots$ are all simple with residues b_n. Then an expansion of $f(z)$ in terms of $b_n(z - a_n)^{-1}$ depends in a systematic way on all singularities of $f(z)$, in contrast to the Taylor expansion about an arbitrarily chosen analytic point z_0 of $f(z)$ or the Laurent expansion about one of the singular points of $f(z)$.

Let us consider a series of concentric circles C_n about the origin so that C_n includes $a_1, a_2, \ldots, a_n$ but no other poles, its radius $R_n \to \infty$ as $n \to \infty$. To guarantee convergence we assume that $|f(z)| < \varepsilon R_n$ for any small positive constant ε and all z on C_n. Then the series

$$\boxed{f(z) = f(0) + \sum_{n=1}^{\infty} b_n\left\{(z - a_n)^{-1} + a_n^{-1}\right\}} \tag{7.20}$$

converges to $f(z)$. To prove this **theorem** (due to Mittag–Leffler) we use the residue theorem to evaluate the contour integral for z inside C_n:

$$\begin{aligned} I_n &= \frac{1}{2\pi i}\int_{C_n} \frac{f(w)}{w(w - z)}dw \\ &= \sum_{m=1}^{n} \frac{b_m}{a_m(a_m - z)} + \frac{f(z) - f(0)}{z}. \end{aligned} \tag{7.21}$$

On C_n we have, for $n \to \infty$,

$$|I_n| \le 2\pi R_n \frac{\max_{w \text{ on } C_n} |f(w)|}{2\pi R_n(R_n - |z|)} < \frac{\varepsilon R_n}{R_n - |z|} \to \varepsilon$$

for $R_n \gg |z|$. Using $I_n \to 0$ in Eq. (7.21) proves Eq. (7.20).

If $|f(z)| < \varepsilon R_n^{p+1}$, then we evaluate similarly the integral

$$I_n = \frac{1}{2\pi i}\int \frac{f(w)}{w^{p+1}(w - z)}dw \to 0 \qquad \text{as } n \to \infty$$

and obtain the analogous pole expansion

$$f(z) = f(0) + zf'(0) + \cdots + \frac{z^p f^{(p)}(0)}{p!} + \sum_{n=1}^{\infty} \frac{b_n z^{p+1}/a_n^{p+1}}{z - a_n}. \tag{7.22}$$

Note that the convergence of the series in Eqs. (7.20) and (7.22) is implied by the bound of $|f(z)|$ for $|z| \to \infty$.

Product Expansion of Entire Functions

A function $f(z)$ that is analytic for all finite z is called an **entire** function. The logarithmic derivative f'/f is a meromorphic function with a pole expansion.

If $f(z)$ has a simple zero at $z = a_n$, then $f(z) = (z - a_n)g(z)$ with analytic $g(z)$ and $g(a_n) \neq 0$. Hence the logarithmic derivative

$$\frac{f'(z)}{f(z)} = (z - a_n)^{-1} + \frac{g'(z)}{g(z)} \tag{7.23}$$

has a simple pole at $z = a_n$ with residue 1, and g'/g is analytic there. If f'/f satisfies the conditions that lead to the pole expansion in Eq. (7.20), then

$$\frac{f'(z)}{f(z)} = \frac{f'(0)}{f(0)} + \sum_{n=1}^{\infty}\left[\frac{1}{a_n} + \frac{1}{z - a_n}\right] \tag{7.24}$$

holds. Integrating Eq. (7.24) yields

$$\begin{aligned}\int_0^z \frac{f'(z)}{f(z)}dz &= \ln f(z) - \ln f(0) \\ &= \frac{zf'(0)}{f(0)} + \sum_{n=1}^{\infty}\left\{\ln(z - a_n) - \ln(-a_n) + \frac{z}{a_n}\right\},\end{aligned}$$

and exponentiating we obtain the product expansion

$$\boxed{f(z) = f(0)\exp\left(\frac{zf'(0)}{f(0)}\right)\prod_1^{\infty}\left(1 - \frac{z}{a_n}\right)e^{z/a_n}.} \tag{7.25}$$

Examples are the product expansions (see Chapter 5) for

$$\boxed{\begin{aligned}\sin z &= z\prod_{\substack{n=-\infty \\ n\neq 0}}^{\infty}\left(1 - \frac{z}{n\pi}\right)e^{z/n\pi} = z\prod_{n=1}^{\infty}\left(1 - \frac{z^2}{n^2\pi^2}\right), \\ \cos z &= \prod_{n=1}^{\infty}\left\{1 - \frac{z^2}{(n - 1/2)^2\pi^2}\right\}.\end{aligned}} \tag{7.26}$$

Another example is the product expansion of the gamma function, which will be discussed in Chapter 8.

As a consequence of Eq. (7.23) the contour integral of the logarithmic derivative may be used to count the number N_f of zeros (including their multiplicities) of the function $f(z)$ inside the contour C:

$$\frac{1}{2\pi i}\int_C \frac{f'(z)}{f(z)}dz = N_f. \tag{7.27}$$

Moreover, using

$$\int \frac{f'(z)}{f(z)}dz = \ln f(z) = \ln|f(z)| + i \arg f(z), \tag{7.28}$$

we see that the real part in Eq. (7.28) does not change as z moves once around the contour, while the corresponding change in arg f must be

$$\Delta_C \arg(f) = 2\pi N_f. \tag{7.29}$$

This leads to **Rouché's theorem**: *If $f(z)$ and $g(z)$ are analytic inside and on a closed contour C and $|g(z)| < |f(z)|$ on C then $f(z)$ and $f(z) + g(z)$ have the same number of zeros inside C.*

To show this we use

$$2\pi N_{f+g} = \Delta_C \arg(f+g) = \Delta_C \arg(f) + \Delta_C \arg\left(1 + \frac{g}{f}\right).$$

Since $|g| < |f|$ on C, the point $w = 1 + g(z)/f(z)$ is always an interior point of the circle in the w-plane with center at 1 and radius 1. Hence $\arg(1 + g/f)$ must return to its original value when z moves around C (it does not circle the origin); it cannot decrease or increase by a multiple of 2π so that $\Delta_C \arg(1 + g/f) = 0$.

Rouché's theorem may be used for an alternative proof of the fundamental theorem of algebra: A polynomial $\sum_{m=0}^{n} a_m z^m$ with $a_n \neq 0$ has n zeros. We define $f(z) = a_n z^n$. Then f has an n-fold zero at the origin and no other zeros. Let $g(z) = \sum_{m=0}^{n-1} a_m z^m$. We apply Rouché's theorem to a circle C with center at the origin and radius $R > 1$. On C, $|f(z)| = |a_n|R^n$ and

$$|g(z)| \leq |a_0| + |a_1|R + \cdots + |a_{n-1}|R^{n-1} \leq \left(\sum_{m=0}^{n-1} |a_m|\right) R^{n-1}.$$

Hence $|g(z)| < |f(z)|$ for z on C, provided $R > (\sum_{m=0}^{n-1} |a_m|)/|a_n|$. For all sufficiently large circles C therefore, $f + g = \sum_{m=0}^{n} a_m z^m$ has n zeros inside C according to Rouché's theorem.

Evaluation of Definite Integrals

Definite integrals appear repeatedly in problems of mathematical physics as well as in pure mathematics. Three moderately general techniques are useful in evaluating definite integrals: (1) contour integration, (2) conversion to gamma or beta functions (Chapter 8), and (3) numerical quadrature. Other approaches include series expansion with term-by-term integration and integral transforms. As will be seen subsequently, the method of contour integration is perhaps the most versatile of these methods, since it is applicable to a wide variety of integrals.

Definite Integrals: $\int_0^{2\pi} f(\sin\theta, \cos\theta)\, d\theta$

The calculus of residues is useful in evaluating a wide variety of definite integrals in both physical and purely mathematical problems. We consider, first, integrals of the form

$$I = \int_0^{2\pi} f(\sin\theta, \cos\theta)\, d\theta, \tag{7.30}$$

where f is finite for all values of θ. We also require f to be a rational function of $\sin\theta$ and $\cos\theta$ so that it will be single-valued. Let

$$z = e^{i\theta}, \qquad dz = ie^{i\theta}\, d\theta.$$

From this,

$$d\theta = -i\frac{dz}{z}, \qquad \sin\theta = \frac{z - z^{-1}}{2i}, \qquad \cos\theta = \frac{z + z^{-1}}{2}. \tag{7.31}$$

Our integral becomes

$$I = -i \oint f\left(\frac{z - z^{-1}}{2i}, \frac{z + z^{-1}}{2}\right)\frac{dz}{z}, \tag{7.32}$$

with the path of integration the unit circle. By the residue theorem, Eq. (7.16),

$$I = (-i)2\pi i \sum \text{residues within the unit circle.} \tag{7.33}$$

Note that we are after the residues of f/z. Illustrations of integrals of this type are provided by Exercises 7.1.7–7.1.10.

Example 7.1.2 INTEGRAL OF COS IN DENOMINATOR

Our problem is to evaluate the definite integral

$$I = \int_0^{2\pi} \frac{d\theta}{1 + \varepsilon\cos\theta}, \qquad |\varepsilon| < 1.$$

By Eq. (7.32) this becomes

$$\begin{aligned} I &= -i \oint_{\text{unit circle}} \frac{dz}{z[1 + (\varepsilon/2)(z + z^{-1})]} \\ &= -i\frac{2}{\varepsilon} \oint \frac{dz}{z^2 + (2/\varepsilon)z + 1}. \end{aligned}$$

The denominator has roots

$$z_- = -\frac{1}{\varepsilon} - \frac{1}{\varepsilon}\sqrt{1 - \varepsilon^2} \qquad \text{and} \qquad z_+ = -\frac{1}{\varepsilon} + \frac{1}{\varepsilon}\sqrt{1 - \varepsilon^2},$$

where z_+ is within the unit circle and z_- is outside. Then by Eq. (7.33) and Exercise 6.6.1,

$$I = -i\frac{2}{\varepsilon} \cdot 2\pi i \left.\frac{1}{2z + 2/\varepsilon}\right|_{z=-1/\varepsilon+(1/\varepsilon)\sqrt{1-\varepsilon^2}}.$$

We obtain

$$\int_0^{2\pi} \frac{d\theta}{1+\varepsilon\cos\theta} = \frac{2\pi}{\sqrt{1-\varepsilon^2}}, \qquad |\varepsilon| < 1. \quad \blacksquare$$

Evaluation of Definite Integrals: $\int_{-\infty}^{\infty} f(x)\,dx$

Suppose that our definite integral has the form

$$I = \int_{-\infty}^{\infty} f(x)\,dx \tag{7.34}$$

and satisfies the two conditions:

- $f(z)$ is analytic in the upper half-plane except for a finite number of poles. (It will be assumed that there are no poles on the real axis. If poles are present on the real axis, they may be included or excluded as discussed earlier in this section.)
- $f(z)$ vanishes as strongly[2] as $1/z^2$ for $|z| \to \infty,\ 0 \le \arg z \le \pi$.

With these conditions, we may take as a contour of integration the real axis and a semicircle in the upper half-plane, as shown in Fig. 7.5. We let the radius R of the semicircle become infinitely large. Then

$$\begin{aligned}\oint f(z)\,dz &= \lim_{R\to\infty}\int_{-R}^{R} f(x)\,dx + \lim_{R\to\infty}\int_0^{\pi} f\left(Re^{i\theta}\right) iRe^{i\theta}\,d\theta \\ &= 2\pi i \sum \text{ residues (upper half-plane).}\end{aligned} \tag{7.35}$$

From the second condition the second integral (over the semicircle) vanishes and

$$\int_{-\infty}^{\infty} f(x)\,dx = 2\pi i \sum \text{ residues (upper half-plane).} \tag{7.36}$$

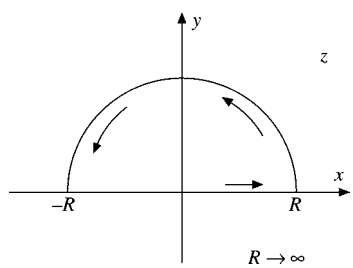

FIGURE 7.5 Half-circle contour.

[2]We could use $f(z)$ vanishes faster than $1/z$, and we wish to have $f(z)$ single-valued.

Example 7.1.3 INTEGRAL OF MEROMORPHIC FUNCTION

Evaluate

$$I = \int_{-\infty}^{\infty} \frac{dx}{1+x^2}. \tag{7.37}$$

From Eq. (7.36),

$$\int_{-\infty}^{\infty} \frac{dx}{1+x^2} = 2\pi i \sum \text{ residues (upper half-plane).}$$

Here and in every other similar problem we have the question: Where are the poles? Rewriting the integrand as

$$\frac{1}{z^2+1} = \frac{1}{z+i} \cdot \frac{1}{z-i}, \tag{7.38}$$

we see that there are simple poles (order 1) at $z = i$ and $z = -i$.

A simple pole at $z = z_0$ indicates (and is indicated by) a Laurent expansion of the form

$$f(z) = \frac{a_{-1}}{z-z_0} + a_0 + \sum_{n=1}^{\infty} a_n (z-z_0)^n. \tag{7.39}$$

The residue a_{-1} is easily isolated as (Exercise 6.6.1)

$$a_{-1} = (z-z_0) f(z)|_{z=z_0}. \tag{7.40}$$

Using Eq. (7.40), we find that the residue at $z = i$ is $1/2i$, whereas that at $z = -i$ is $-1/2i$.

Then

$$\int_{-\infty}^{\infty} \frac{dx}{1+x^2} = 2\pi i \cdot \frac{1}{2i} = \pi. \tag{7.41}$$

Here we have used $a_{-1} = 1/2i$ for the residue of the one included pole at $z = i$. Note that it is possible to use the lower semicircle and that this choice will lead to the same result, $I = \pi$. A somewhat more delicate problem is provided by the next example. ■

Evaluation of Definite Integrals: $\int_{-\infty}^{\infty} f(x) e^{iax}\, dx$

Consider the definite integral

$$I = \int_{-\infty}^{\infty} f(x) e^{iax}\, dx, \tag{7.42}$$

with a real and positive. (This is a Fourier transform, Chapter 15.) We assume the two conditions:

- $f(z)$ is analytic in the upper half-plane except for a finite number of poles.

- $\lim_{|z|\to\infty} f(z) = 0, \qquad 0 \le \arg z \le \pi.$ (7.43)

Note that this is a less restrictive condition than the second condition imposed on $f(z)$ for integrating $\int_{-\infty}^{\infty} f(x)\,dx$ previously.

We employ the contour shown in Fig. 7.5. The application of the calculus of residues is the same as the one just considered, but here we have to work harder to show that the integral over the (infinite) semicircle goes to zero. This integral becomes

$$I_R = \int_0^{\pi} f\left(Re^{i\theta}\right) e^{iaR\cos\theta - aR\sin\theta} iRe^{i\theta}\,d\theta. \tag{7.44}$$

Let R be so large that $|f(z)| = |f(Re^{i\theta})| < \varepsilon$. Then

$$|I_R| \le \varepsilon R \int_0^{\pi} e^{-aR\sin\theta}\,d\theta = 2\varepsilon R \int_0^{\pi/2} e^{-aR\sin\theta}\,d\theta. \tag{7.45}$$

In the range $[0, \pi/2]$,

$$\frac{2}{\pi}\theta \le \sin\theta.$$

Therefore (Fig. 7.6)

$$|I_R| \le 2\varepsilon R \int_0^{\pi/2} e^{-2aR\theta/\pi}\,d\theta. \tag{7.46}$$

Now, integrating by inspection, we obtain

$$|I_R| \le 2\varepsilon R \frac{1 - e^{-aR}}{2aR/\pi}.$$

Finally,

$$\lim_{R\to\infty} |I_R| \le \frac{\pi}{a}\varepsilon. \tag{7.47}$$

From Eq. (7.43), $\varepsilon \to 0$ as $R \to \infty$ and

$$\lim_{R\to\infty} |I_R| = 0. \tag{7.48}$$

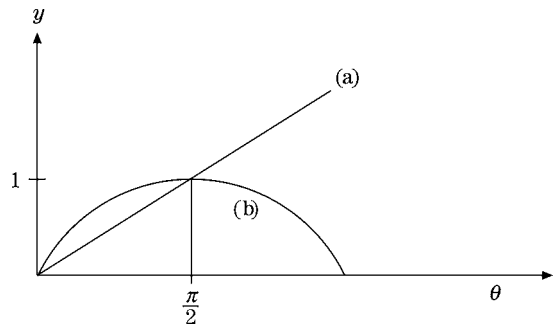

FIGURE 7.6 (a) $y = (2/\pi)\theta$, (b) $y = \sin\theta$.

This useful result is sometimes called **Jordan's lemma**. With it, we are prepared to tackle Fourier integrals of the form shown in Eq. (7.42).

Using the contour shown in Fig. 7.5, we have

$$\int_{-\infty}^{\infty} f(x)e^{iax}\,dx + \lim_{R\to\infty} I_R = 2\pi i \sum \text{residues (upper half-plane)}.$$

Since the integral over the upper semicircle I_R vanishes as $R \to \infty$ (Jordan's lemma),

$$\int_{-\infty}^{\infty} f(x)e^{iax}\,dx = 2\pi i \sum \text{residues (upper half-plane)} \qquad (a > 0). \tag{7.49}$$

Example 7.1.4 SIMPLE POLE ON CONTOUR OF INTEGRATION

The problem is to evaluate

$$I = \int_0^{\infty} \frac{\sin x}{x} dx. \tag{7.50}$$

This may be taken as the imaginary part[3] of

$$I_2 = P\int_{-\infty}^{\infty} \frac{e^{iz}dz}{z}. \tag{7.51}$$

Now the only pole is a simple pole at $z = 0$ and the residue there by Eq. (7.40) is $a_{-1} = 1$. We choose the contour shown in Fig. 7.7 (1) to avoid the pole, (2) to include the real axis, and (3) to yield a vanishingly small integrand for $z = iy$, $y \to \infty$. Note that in this case a large (infinite) semicircle in the lower half-plane would be disastrous. We have

$$\oint \frac{e^{iz}\,dz}{z} = \int_{-R}^{-r} e^{ix}\frac{dx}{x} + \int_{C_1} \frac{e^{iz}\,dz}{z} + \int_r^R \frac{e^{ix}\,dx}{x} + \int_{C_2} \frac{e^{iz}\,dz}{z} = 0, \tag{7.52}$$

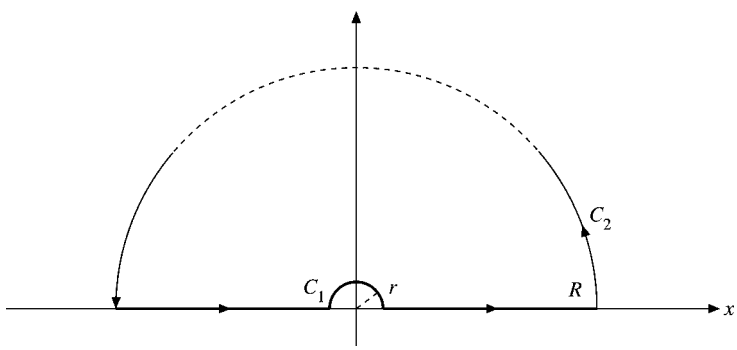

FIGURE 7.7 Singularity on contour.

[3]One can use $\int[(e^{iz} - e^{-iz})/2iz]\,dz$, but then two different contours will be needed for the two exponentials (compare Example 7.1.5).

the final zero coming from the residue theorem (Eq. (7.6)). By Jordan's lemma

$$\int_{C_2} \frac{e^{iz}\,dz}{z} = 0, \tag{7.53}$$

and

$$\oint \frac{e^{iz}\,dz}{z} = \int_{C_1} \frac{e^{iz}\,dz}{z} + P\int_{-\infty}^{\infty} \frac{e^{ix}\,dx}{x} = 0. \tag{7.54}$$

The integral over the small semicircle yields $(-)\pi i$ times the residue of 1, and minus, as a result of going clockwise. Taking the imaginary part,[4] we have

$$\int_{-\infty}^{\infty} \frac{\sin x}{x} dx = \pi \tag{7.55}$$

or

$$\int_{0}^{\infty} \frac{\sin x}{x} dx = \frac{\pi}{2}. \tag{7.56}$$

The contour of Fig. 7.7, although convenient, is not at all unique. Another choice of contour for evaluating Eq. (7.50) is presented as Exercise 7.1.15. ■

Example 7.1.5 QUANTUM MECHANICAL SCATTERING

The quantum mechanical analysis of scattering leads to the function

$$I(\sigma) = \int_{-\infty}^{\infty} \frac{x \sin x\,dx}{x^2 - \sigma^2}, \tag{7.57}$$

where σ is real and positive. This integral is divergent and therefore ambiguous. From the physical conditions of the problem there is a further requirement: $I(\sigma)$ is to have the form $e^{i\sigma}$ so that it will represent an outgoing scattered wave.

Using

$$\sin z = \frac{1}{i}\sinh iz = \frac{1}{2i}e^{iz} - \frac{1}{2i}e^{-iz}, \tag{7.58}$$

we write Eq. (7.57) in the complex plane as

$$I(\sigma) = I_1 + I_2, \tag{7.59}$$

with

$$I_1 = \frac{1}{2i}\int_{-\infty}^{\infty} \frac{ze^{iz}}{z^2 - \sigma^2} dz,$$

$$I_2 = -\frac{1}{2i}\int_{-\infty}^{\infty} \frac{ze^{-iz}}{z^2 - \sigma^2} dz. \tag{7.60}$$

[4] Alternatively, we may combine the integrals of Eq. (7.52) as

$$\int_{-R}^{-r} e^{ix}\frac{dx}{x} + \int_{r}^{R} e^{ix}\frac{dx}{x} = \int_{r}^{R} \left(e^{ix} - e^{-ix}\right)\frac{dx}{x} = 2i\int_{r}^{R} \frac{\sin x}{x} dx.$$

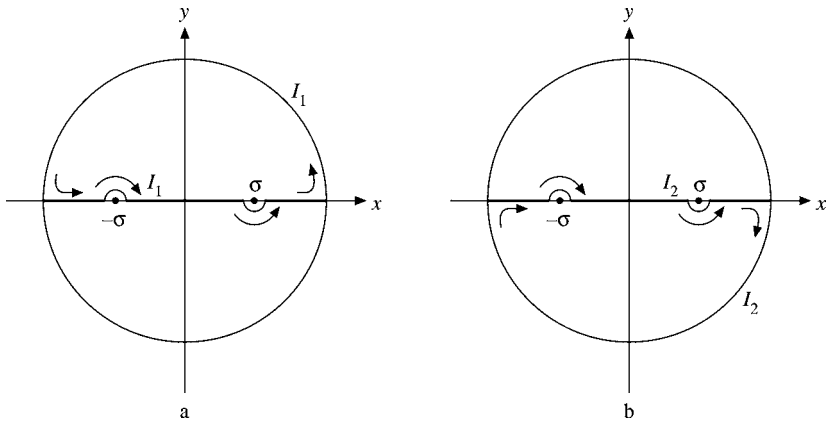

FIGURE 7.8 Contours.

Integral I_1 is similar to Example 7.1.4 and, as in that case, we may complete the contour by an infinite semicircle in the upper half-plane, as shown in Fig. 7.8a. For I_2 the exponential is negative and we complete the contour by an infinite semicircle in the lower half-plane, as shown in Fig. 7.8b. As in Example 7.1.4, neither semicircle contributes anything to the integral — Jordan's lemma.

There is still the problem of locating the poles and evaluating the residues. We find poles at $z = +\sigma$ and $z = -\sigma$ **on the contour of integration**. The residues are (Exercises 6.6.1 and 7.1.1)

	$z=\sigma$	$z=-\sigma$
I_1	$\frac{e^{i\sigma}}{2}$	$\frac{e^{-i\sigma}}{2}$
I_2	$\frac{e^{-i\sigma}}{2}$	$\frac{e^{i\sigma}}{2}$

Detouring around the poles, as shown in Fig. 7.8 (it matters little whether we go above or below), we find that the residue theorem leads to

$$PI_1 - \pi i\left(\frac{1}{2i}\right)\frac{e^{-i\sigma}}{2} + \pi i\left(\frac{1}{2i}\right)\frac{e^{i\sigma}}{2} = 2\pi i\left(\frac{1}{2i}\right)\frac{e^{i\sigma}}{2}, \tag{7.61}$$

for we have enclosed the singularity at $z = \sigma$ but excluded the one at $z = -\sigma$. In similar fashion, but noting that the contour for I_2 is clockwise,

$$PI_2 - \pi i\left(\frac{-1}{2i}\right)\frac{e^{i\sigma}}{2} + \pi i\left(\frac{-1}{2i}\right)\frac{e^{-i\sigma}}{2} = -2\pi i\left(\frac{-1}{2i}\right)\frac{e^{i\sigma}}{2}. \tag{7.62}$$

Adding Eqs. (7.61) and (7.62), we have

$$PI(\sigma) = PI_1 + PI_2 = \frac{\pi}{2}\left(e^{i\sigma} + e^{-i\sigma}\right) = \pi\cosh i\sigma = \pi\cos\sigma. \tag{7.63}$$

This is a perfectly good evaluation of Eq. (7.57), but unfortunately the cosine dependence is appropriate for a standing wave and not for the outgoing scattered wave as specified.

To obtain the desired form, we try a different technique (compare Example 7.1.1). Instead of dodging around the singular points, let us move them off the real axis. Specifically, let $\sigma \to \sigma + i\gamma$, $-\sigma \to -\sigma - i\gamma$, where γ is positive but small and will eventually be made to approach zero; that is, for I_1 we include one pole and for I_2 the other one,

$$I_+(\sigma) = \lim_{\gamma \to 0} I(\sigma + i\gamma). \tag{7.64}$$

With this simple substitution, the first integral I_1 becomes

$$I_1(\sigma + i\gamma) = 2\pi i \left(\frac{1}{2i}\right) \frac{e^{i(\sigma + i\gamma)}}{2} \tag{7.65}$$

by direct application of the residue theorem. Also,

$$I_2(\sigma + i\gamma) = -2\pi i \left(\frac{-1}{2i}\right) \frac{e^{i(\sigma + i\gamma)}}{2}. \tag{7.66}$$

Adding Eqs. (7.65) and (7.66) and then letting $\gamma \to 0$, we obtain

$$\begin{aligned} I_+(\sigma) &= \lim_{\gamma \to 0} \left[I_1(\sigma + i\gamma) + I_2(\sigma + i\gamma) \right] \\ &= \lim_{\gamma \to 0} \pi e^{i(\sigma + i\gamma)} = \pi e^{i\sigma}, \end{aligned} \tag{7.67}$$

a result that does fit the boundary conditions of our scattering problem.

It is interesting to note that the substitution $\sigma \to \sigma - i\gamma$ would have led to

$$I_-(\sigma) = \pi e^{-i\sigma}, \tag{7.68}$$

which could represent an incoming wave. Our earlier result (Eq. (7.63)) is seen to be the arithmetic average of Eqs. (7.67) and (7.68). This average is the Cauchy principal value of the integral. Note that we have these possibilities (Eqs. (7.63), (7.67), and (7.68)) because our integral is not uniquely defined until we specify the particular limiting process (or average) to be used. ■

Evaluation of Definite Integrals: Exponential Forms

With exponential or hyperbolic functions present in the integrand, life gets somewhat more complicated than before. Instead of a general overall prescription, the contour must be chosen to fit the specific integral. These cases are also opportunities to illustrate the versatility and power of contour integration.

As an example, we consider an integral that will be quite useful in developing a relation between $\Gamma(1 + z)$ and $\Gamma(1 - z)$. Notice how the periodicity along the imaginary axis is exploited.

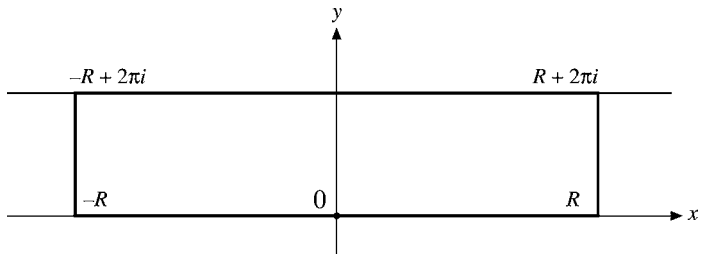

FIGURE 7.9 Rectangular contour.

Example 7.1.6 FACTORIAL FUNCTION

We wish to evaluate

$$I = \int_{-\infty}^{\infty} \frac{e^{ax}}{1+e^x} dx, \qquad 0 < a < 1. \tag{7.69}$$

The limits on a are sufficient (but not necessary) to prevent the integral from diverging as $x \to \pm\infty$. This integral (Eq. (7.69)) may be handled by replacing the real variable x by the complex variable z and integrating around the contour shown in Fig. 7.9. If we take the limit as $R \to \infty$, the real axis, of course, leads to the integral we want. The return path along $y = 2\pi$ is chosen to leave the denominator of the integral invariant, at the same time introducing a constant factor $e^{i2\pi a}$ in the numerator. We have, in the complex plane,

$$\begin{aligned} \oint \frac{e^{az}}{1+e^z} dz &= \lim_{R\to\infty} \left(\int_{-R}^{R} \frac{e^{ax}}{1+e^x} dx - e^{i2\pi a} \int_{-R}^{R} \frac{e^{ax}}{1+e^x} dx \right) \\ &= \left(1 - e^{i2\pi a}\right) \int_{-\infty}^{\infty} \frac{e^{ax}}{1+e^x} dx. \end{aligned} \tag{7.70}$$

In addition there are two vertical sections ($0 \le y \le 2\pi$), which vanish (exponentially) as $R \to \infty$.

Now where are the poles and what are the residues? We have a pole when

$$e^z = e^x e^{iy} = -1. \tag{7.71}$$

Equation (7.71) is satisfied at $z = 0 + i\pi$. By a Laurent expansion[5] in powers of $(z - i\pi)$ the pole is seen to be a simple pole with a residue of $-e^{i\pi a}$. Then, applying the residue theorem,

$$\left(1 - e^{i2\pi a}\right) \int_{-\infty}^{\infty} \frac{e^{ax}}{1+e^x} dx = 2\pi i \left(-e^{i\pi a}\right). \tag{7.72}$$

This quickly reduces to

$$\int_{-\infty}^{\infty} \frac{e^{ax}}{1+e^x} dx = \frac{\pi}{\sin a\pi}, \qquad 0 < a < 1. \tag{7.73}$$

[5] $1 + e^z = 1 + e^{z-i\pi} e^{i\pi} = 1 - e^{z-i\pi} = -(z - i\pi)(1 + \frac{z-i\pi}{2!} + \frac{(z-i\pi)^2}{3!} + \cdots)$.

Using the beta function (Section 8.4), we can show the integral to be equal to the product $\Gamma(a)\Gamma(1-a)$. This results in the interesting and useful factorial function relation

$$\Gamma(a+1)\Gamma(1-a) = \frac{\pi a}{\sin \pi a}. \tag{7.74}$$

Although Eq. (7.73) holds for real $a, 0 < a < 1$, Eq. (7.74) may be extended by analytic continuation to all values of a, real and complex, excluding only real integral values. ■

As a final example of contour integrals of exponential functions, we consider Bernoulli numbers again.

Example 7.1.7 BERNOULLI NUMBERS

In Section 5.9 the Bernoulli numbers were defined by the expansion

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} \frac{B_n}{n!} x^n. \tag{7.75}$$

Replacing x with z (analytic continuation), we have a Taylor series (compare Eq. (6.47)) with

$$B_n = \frac{n!}{2\pi i} \oint_{C_0} \frac{z}{e^z - 1} \frac{dz}{z^{n+1}}, \tag{7.76}$$

where the contour C_0 is around the origin counterclockwise with $|z| < 2\pi$ to avoid the poles at $2\pi i n$.

For $n = 0$ we have a simple pole at $z = 0$ with a residue of $+1$. Hence by Eq. (7.25),

$$B_0 = \frac{0!}{2\pi i} \cdot 2\pi i (1) = 1. \tag{7.77}$$

For $n = 1$ the singularity at $z = 0$ becomes a second-order pole. The residue may be shown to be $-\frac{1}{2}$ by series expansion of the exponential, followed by a binomial expansion. This results in

$$B_1 = \frac{1!}{2\pi i} \cdot 2\pi i \left(-\frac{1}{2}\right) = -\frac{1}{2}. \tag{7.78}$$

For $n \geq 2$ this procedure becomes rather tedious, and we resort to a different means of evaluating Eq. (7.76). The contour is deformed, as shown in Fig. 7.10.

The new contour C still encircles the origin, as required, but now it also encircles (in a negative direction) an infinite series of singular points along the imaginary axis at $z = \pm p 2\pi i$, $p = 1, 2, 3, \ldots$. The integration back and forth along the x-axis cancels out, and for $R \to \infty$ the integration over the infinite circle yields zero. Remember that $n \geq 2$. Therefore

$$\oint_{C_0} \frac{z}{e^z - 1} \frac{dz}{z^{n+1}} = -2\pi i \sum_{p=1}^{\infty} \text{residues} \qquad (z = \pm p 2\pi i). \tag{7.79}$$

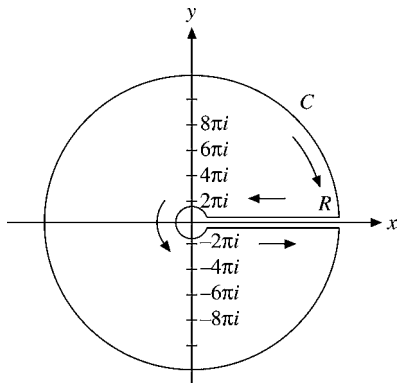

FIGURE 7.10 Contour of integration for Bernoulli numbers.

At $z = p2\pi i$ we have a simple pole with a residue $(p2\pi i)^{-n}$. When n is odd, the residue from $z = p2\pi i$ exactly cancels that from $z = -p2\pi i$ and $B_n = 0, n = 3, 5, 7$, and so on. For n even the residues add, giving

$$B_n = \frac{n!}{2\pi i}(-2\pi i)2\sum_{p=1}^{\infty}\frac{1}{p^n(2\pi i)^n}$$

$$= -\frac{(-1)^{n/2}2n!}{(2\pi)^n}\sum_{p=1}^{\infty}p^{-n} = -\frac{(-1)^{n/2}2n!}{(2\pi)^n}\zeta(n) \qquad (n \text{ even}), \tag{7.80}$$

where $\zeta(n)$ is the Riemann zeta function introduced in Section 5.9. Equation (7.80) corresponds to Eq. (5.152) of Section 5.9. ■

Exercises

7.1.1 Determine the nature of the singularities of each of the following functions and evaluate the residues $(a > 0)$.

(a) $\dfrac{1}{z^2 + a^2}$.

(b) $\dfrac{1}{(z^2 + a^2)^2}$.

(c) $\dfrac{z^2}{(z^2 + a^2)^2}$.

(d) $\dfrac{\sin 1/z}{z^2 + a^2}$.

(e) $\dfrac{ze^{+iz}}{z^2 + a^2}$.

(f) $\dfrac{ze^{+iz}}{z^2 - a^2}$.

(g) $\dfrac{e^{+iz}}{z^2 - a^2}$.

(h) $\dfrac{z^{-k}}{z + 1}$, $\quad 0 < k < 1$.

Hint. For the point at infinity, use the transformation $w = 1/z$ for $|z| \to 0$. For the residue, transform $f(z)\,dz$ into $g(w)\,dw$ and look at the behavior of $g(w)$.

7.1.2 Locate the singularities and evaluate the residues of each of the following functions.

(a) $z^{-n}(e^z - 1)^{-1}, \qquad z \neq 0,$

(b) $\dfrac{z^2 e^z}{1 + e^{2z}}.$

(c) Find a closed-form expression (that is, not a sum) for the sum of the finite-plane singularities.

(d) Using the result in part (c), what is the residue at $|z| \to \infty$?

Hint. See Section 5.9 for expressions involving Bernoulli numbers. Note that Eq. (5.144) cannot be used to investigate the singularity at $z \to \infty$, since this series is only valid for $|z| < 2\pi$.

7.1.3 The statement that the integral halfway around a singular point is equal to one-half the integral all the way around was limited to simple poles. Show, by a specific example, that

$$\int_{\text{Semicircle}} f(z)\,dz = \frac{1}{2}\oint_{\text{Circle}} f(z)\,dz$$

does not necessarily hold if the integral encircles a pole of higher order.
Hint. Try $f(z) = z^{-2}$.

7.1.4 A function $f(z)$ is analytic along the real axis except for a third-order pole at $z = x_0$. The Laurent expansion about $z = x_0$ has the form

$$f(z) = \frac{a_{-3}}{(z - x_0)^3} + \frac{a_{-1}}{z - x_0} + g(z),$$

with $g(z)$ analytic at $z = x_0$. Show that the Cauchy principal value technique is applicable, in the sense that

(a) $\lim_{\delta \to 0}\left\{\int_{-\infty}^{x_0 - \delta} f(x)\,dx + \int_{x_0 + \delta}^{\infty} f(x)\,dx\right\}$ is finite.

(b) $\int_{C_{x_0}} f(z)\,dz = \pm i\pi a_{-1},$
where C_{x_0} denotes a **small semicircle** about $z = x_0$.

7.1.5 The unit step function is defined as (compare Exercise 1.15.13)

$$u(s - a) = \begin{cases} 0, & s < a \\ 1, & s > a. \end{cases}$$

Show that $u(s)$ has the integral representations

(a) $u(s) = \lim_{\varepsilon \to 0^+} \dfrac{1}{2\pi i}\displaystyle\int_{-\infty}^{\infty} \frac{e^{ixs}}{x - i\varepsilon} dx,$

(b) $u(s) = \dfrac{1}{2} + \dfrac{1}{2\pi i} P \displaystyle\int_{-\infty}^{\infty} \frac{e^{ixs}}{x} dx.$

Note. The parameter s is real.

7.1.6 Most of the special functions of mathematical physics may be generated (defined) by a generating function of the form

$$g(t, x) = \sum_n f_n(x) t^n.$$

Given the following integral representations, derive the corresponding generating function:

(a) Bessel:

$$J_n(x) = \frac{1}{2\pi i} \oint e^{(x/2)(t-1/t)} t^{-n-1}\, dt.$$

(b) Modified Bessel:

$$I_n(x) = \frac{1}{2\pi i} \oint e^{(x/2)(t+1/t)} t^{-n-1}\, dt.$$

(c) Legendre:

$$P_n(x) = \frac{1}{2\pi i} \oint \left(1 - 2tx + t^2\right)^{-1/2} t^{-n-1}\, dt.$$

(d) Hermite:

$$H_n(x) = \frac{n!}{2\pi i} \oint e^{-t^2+2tx} t^{-n-1}\, dt.$$

(e) Laguerre:

$$L_n(x) = \frac{1}{2\pi i} \oint \frac{e^{-xt/(1-t)}}{(1-t)t^{n+1}} dt.$$

(f) Chebyshev:

$$T_n(x) = \frac{1}{4\pi i} \oint \frac{(1-t^2)t^{-n-1}}{(1-2tx+t^2)} dt.$$

Each of the contours encircles the origin and no other singular points.

7.1.7 Generalizing Example 7.1.2, show that

$$\int_0^{2\pi} \frac{d\theta}{a \pm b\cos\theta} = \int_0^{2\pi} \frac{d\theta}{a \pm b\sin\theta} = \frac{2\pi}{(a^2-b^2)^{1/2}}, \qquad \text{for } a > |b|.$$

What happens if $|b| > |a|$?

7.1.8 Show that

$$\int_0^{\pi} \frac{d\theta}{(a+\cos\theta)^2} = \frac{\pi a}{(a^2-1)^{3/2}}, \qquad a > 1.$$

7.1.9 Show that

$$\int_0^{2\pi} \frac{d\theta}{1-2t\cos\theta+t^2} = \frac{2\pi}{1-t^2}, \qquad \text{for } |t|<1.$$

What happens if $|t|>1$? What happens if $|t|=1$?

7.1.10 With the calculus of residues show that

$$\int_0^{\pi} \cos^{2n}\theta\, d\theta = \pi\frac{(2n)!}{2^{2n}(n!)^2} = \pi\frac{(2n-1)!!}{(2n)!!}, \qquad n=0,1,2,\ldots.$$

(The double factorial notation is defined in Section 8.1.)
Hint. $\cos\theta = \frac{1}{2}(e^{i\theta}+e^{-i\theta}) = \frac{1}{2}(z+z^{-1}), \qquad |z|=1.$

7.1.11 Evaluate

$$\int_{-\infty}^{\infty} \frac{\cos bx-\cos ax}{x^2}dx, \qquad a>b>0.$$

ANS. $\pi(a-b)$.

7.1.12 Prove that

$$\int_{-\infty}^{\infty} \frac{\sin^2 x}{x^2}\,dx = \frac{\pi}{2}.$$

Hint. $\sin^2 x = \frac{1}{2}(1-\cos 2x)$.

7.1.13 A quantum mechanical calculation of a transition probability leads to the function $f(t,\omega) = 2(1-\cos\omega t)/\omega^2$. Show that

$$\int_{-\infty}^{\infty} f(t,\omega)d\omega = 2\pi t.$$

7.1.14 Show that $(a>0)$

(a) $$\int_{-\infty}^{\infty} \frac{\cos x}{x^2+a^2}dx = \frac{\pi}{a}e^{-a}.$$

How is the right side modified if $\cos x$ is replaced by $\cos kx$?

(b) $$\int_{-\infty}^{\infty} \frac{x\sin x}{x^2+a^2}dx = \pi e^{-a}.$$

How is the right side modified if $\sin x$ is replaced by $\sin kx$?

These integrals may also be interpreted as Fourier cosine and sine transforms—Chapter 15.

7.1.15 Use the contour shown (Fig. 7.11) with $R\to\infty$ to prove that

$$\int_{-\infty}^{\infty} \frac{\sin x}{x}dx = \pi.$$

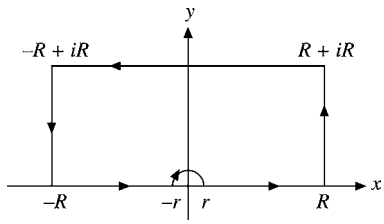

FIGURE 7.11 Large square contour.

7.1.16 In the quantum theory of atomic collisions we encounter the integral

$$I = \int_{-\infty}^{\infty} \frac{\sin t}{t} e^{ipt}\, dt,$$

in which p is real. Show that

$$\begin{aligned} I &= 0, \qquad |p| > 1 \\ I &= \pi, \qquad |p| < 1. \end{aligned}$$

What happens if $p = \pm 1$?

7.1.17 Evaluate

$$\int_0^{\infty} \frac{(\ln x)^2}{1+x^2} dx$$

(a) by appropriate series expansion of the integrand to obtain

$$4\sum_{n=0}^{\infty} (-1)^n (2n+1)^{-3},$$

(b) and by contour integration to obtain

$$\frac{\pi^3}{8}.$$

Hint. $x \to z = e^t$. Try the contour shown in Fig. 7.12, letting $R \to \infty$.

7.1.18 Show that

$$\int_0^{\infty} \frac{x^a}{(x+1)^2} dx = \frac{\pi a}{\sin \pi a},$$

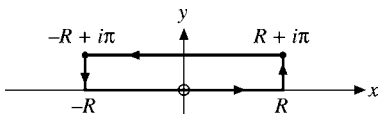

FIGURE 7.12 Small square contour.

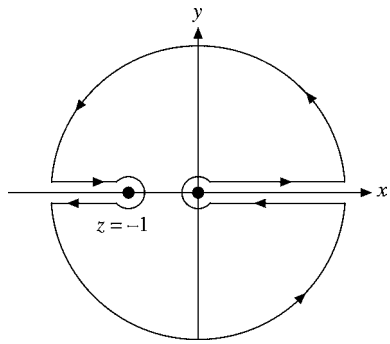

FIGURE 7.13 Contour avoiding branch point and pole.

where $-1 < a < 1$. Here is still another way of deriving Eq. (7.74).
Hint. Use the contour shown in Fig. 7.13, noting that $z = 0$ is a branch point and the positive x-axis is a cut line. Note also the comments on phases following Example 6.6.1.

7.1.19 Show that

$$\int_0^\infty \frac{x^{-a}}{x+1}dx = \frac{\pi}{\sin a\pi},$$

where $0 < a < 1$. This opens up another way of deriving the factorial function relation given by Eq. (7.74).
Hint. You have a branch point and you will need a cut line. Recall that $z^{-a} = w$ in polar form is

$$\left[re^{i(\theta+2\pi n)}\right]^{-a} = \rho e^{i\varphi},$$

which leads to $-a\theta - 2an\pi = \varphi$. You must restrict n to zero (or any other single integer) in order that φ may be uniquely specified. Try the contour shown in Fig. 7.14.

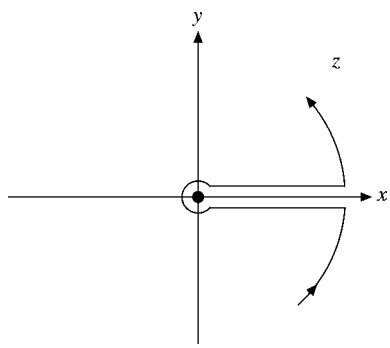

FIGURE 7.14 Alternative contour avoiding branch point.

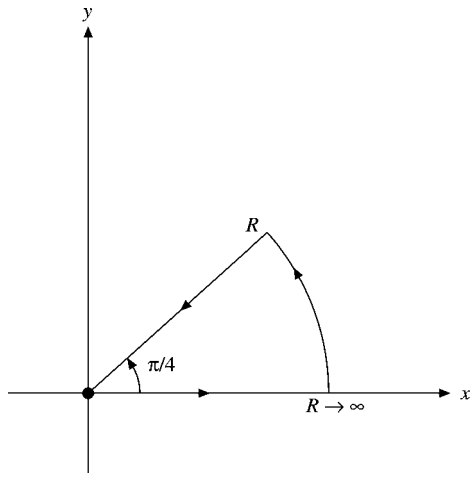

FIGURE 7.15 Angle contour.

7.1.20 Show that

$$\int_0^\infty \frac{dx}{(x^2+a^2)^2} = \frac{\pi}{4a^3}, \qquad a > 0.$$

7.1.21 Evaluate

$$\int_{-\infty}^{\infty} \frac{x^2}{1+x^4} dx.$$

ANS. $\pi/\sqrt{2}$.

7.1.22 Show that

$$\int_0^\infty \cos\left(t^2\right) dt = \int_0^\infty \sin\left(t^2\right) dt = \frac{\sqrt{\pi}}{2\sqrt{2}}.$$

Hint. Try the contour shown in Fig. 7.15.
Note. These are the Fresnel integrals for the special case of infinity as the upper limit. For the general case of a varying upper limit, asymptotic expansions of the Fresnel integrals are the topic of Exercise 5.10.2. Spherical Bessel expansions are the subject of Exercise 11.7.13.

7.1.23 Several of the Bromwich integrals, Section 15.12, involve a portion that may be approximated by

$$I(y) = \int_{a-iy}^{a+iy} \frac{e^{zt}}{z^{1/2}} dz.$$

Here a and t are positive and finite. Show that

$$\lim_{y\to\infty} I(y) = 0.$$

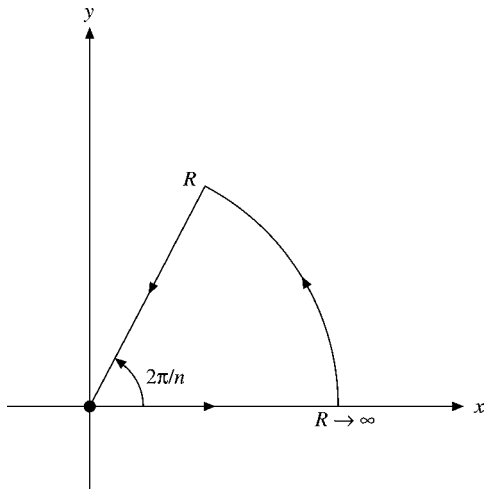

FIGURE 7.16 Sector contour.

7.1.24 Show that

$$\int_0^\infty \frac{1}{1+x^n}dx = \frac{\pi/n}{\sin(\pi/n)}.$$

Hint. Try the contour shown in Fig. 7.16.

7.1.25 (a) Show that

$$f(z) = z^4 - 2\cos 2\theta z^2 + 1$$

has zeros at $e^{i\theta}$, $e^{-i\theta}$, $-e^{i\theta}$, and $-e^{-i\theta}$.

(b) Show that

$$\int_{-\infty}^\infty \frac{dx}{x^4 - 2\cos 2\theta x^2 + 1} = \frac{\pi}{2\sin\theta} = \frac{\pi}{2^{1/2}(1-\cos 2\theta)^{1/2}}.$$

Exercise 7.1.24 ($n = 4$) is a special case of this result.

7.1.26 Show that

$$\int_{-\infty}^\infty \frac{x^2\,dx}{x^4 - 2\cos 2\theta x^2 + 1} = \frac{\pi}{2\sin\theta} = \frac{\pi}{2^{1/2}(1-\cos 2\theta)^{1/2}}.$$

Exercise 7.1.21 is a special case of this result.

7.1.27 Apply the techniques of Example 7.1.5 to the evaluation of the improper integral

$$I = \int_{-\infty}^\infty \frac{dx}{x^2 - \sigma^2}.$$

(a) Let $\sigma \to \sigma + i\gamma$.
(b) Let $\sigma \to \sigma - i\gamma$.
(c) Take the Cauchy principal value.

7.1.28 The integral in Exercise 7.1.17 may be transformed into

$$\int_0^\infty e^{-y}\frac{y^2}{1+e^{-2y}}dy = \frac{\pi^3}{16}.$$

Evaluate this integral by the Gauss–Laguerre quadrature and compare your result with $\pi^3/16$.

ANS. Integral = 1.93775 (10 points).

7.2 DISPERSION RELATIONS

The concept of dispersion relations entered physics with the work of Kronig and Kramers in optics. The name **dispersion** comes from optical dispersion, a result of the dependence of the index of refraction on wavelength, or angular frequency. The index of refraction n may have a real part determined by the phase velocity and a (negative) imaginary part determined by the absorption — see Eq. (7.94). Kronig and Kramers showed in 1926–1927 that the real part of (n^2-1) could be expressed as an integral of the imaginary part. Generalizing this, we shall apply the label **dispersion relations** to any pair of equations giving the real part of a function as an integral of its imaginary part and the imaginary part as an integral of its real part — Eqs. (7.86a) and (7.86b), which follow. The existence of such integral relations might be suspected as an integral analog of the Cauchy–Riemann differential relations, Section 6.2.

The applications in modern physics are widespread. For instance, the real part of the function might describe the forward scattering of a gamma ray in a nuclear Coulomb field (a dispersive process). Then the imaginary part would describe the electron–positron pair production in that same Coulomb field (the absorptive process). As will be seen later, the dispersion relations may be taken as a consequence of causality and therefore are independent of the details of the particular interaction.

We consider a complex function $f(z)$ that is analytic in the upper half-plane and on the real axis. We also require that

$$\lim_{|z|\to\infty} |f(z)| = 0, \qquad 0 \le \arg z \le \pi, \tag{7.81}$$

in order that the integral over an infinite semicircle will vanish. The point of these conditions is that we may express $f(z)$ by the Cauchy integral formula, Eq. (6.43),

$$f(z_0) = \frac{1}{2\pi i}\oint \frac{f(z)}{z-z_0}dz. \tag{7.82}$$

The integral over the upper semicircle[6] vanishes and we have

$$f(z_0) = \frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{f(x)}{x-z_0}dx. \tag{7.83}$$

The integral over the contour shown in Fig. 7.17 has become an integral along the x-axis.

Equation (7.83) assumes that z_0 is in the upper half-plane — interior to the closed contour. If z_0 were in the lower half-plane, the integral would yield zero by the Cauchy integral

[6] The use of a semicircle to close the path of integration is convenient, not mandatory. Other paths are possible.

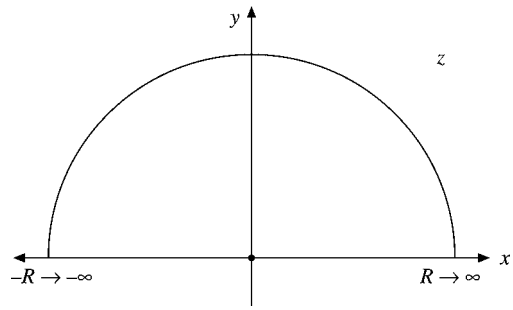

FIGURE 7.17 Semicircle contour.

theorem, Section 6.3. Now, either letting z_0 approach the real axis from above $(z_0 - x_0)$ or placing it on the real axis and taking an average of Eq. (7.83) and zero, we find that Eq. (7.83) becomes

$$f(x_0) = \frac{1}{\pi i} P \int_{-\infty}^{\infty} \frac{f(x)}{x - x_0} dx, \tag{7.84}$$

where P indicates the Cauchy principal value. Splitting Eq. (7.84) into real and imaginary parts[7] yields

$$\begin{aligned} f(x_0) &= u(x_0) + i v(x_0) \\ &= \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{v(x)}{x - x_0} dx - \frac{i}{\pi} P \int_{-\infty}^{\infty} \frac{u(x)}{x - x_0} dx. \end{aligned} \tag{7.85}$$

Finally, equating real part to real part and imaginary part to imaginary part, we obtain

$$\boxed{\begin{aligned} u(x_0) &= \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{v(x)}{x - x_0} dx \qquad (7.86a) \\ v(x_0) &= -\frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{u(x)}{x - x_0} dx. \qquad (7.86b) \end{aligned}}$$

These are the dispersion relations. The real part of our complex function is expressed as an integral over the imaginary part. The imaginary part is expressed as an integral over the real part. The real and imaginary parts are **Hilbert transforms** of each other. Note that these relations are meaningful only when $f(x)$ is a complex function of the real variable x. Compare Exercise 7.2.1.

From a physical point of view $u(x)$ and/or $v(x)$ represent some physical measurements. Then $f(z) = u(z) + i v(z)$ is an analytic continuation over the upper half-plane, with the value on the real axis serving as a boundary condition.

[7] The second argument, $y = 0$, is dropped: $u(x_0, 0) \to u(x_0)$.

Symmetry Relations

On occasion $f(x)$ will satisfy a symmetry relation and the integral from $-\infty$ to $+\infty$ may be replaced by an integral over positive values only. This is of considerable physical importance because the variable x might represent a frequency and only zero and positive frequencies are available for physical measurements. Suppose[8]

$$f(-x) = f^*(x). \tag{7.87}$$

Then

$$u(-x) + iv(-x) = u(x) - iv(x). \tag{7.88}$$

The real part of $f(x)$ is even and the imaginary part is odd.[9] In quantum mechanical scattering problems these relations (Eq. (7.88)) are called crossing conditions. To exploit these **crossing conditions**, we rewrite Eq. (7.86a) as

$$u(x_0) = \frac{1}{\pi} P \int_{-\infty}^{0} \frac{v(x)}{x - x_0} dx + \frac{1}{\pi} P \int_{0}^{\infty} \frac{v(x)}{x - x_0} dx. \tag{7.89}$$

Letting $x \to -x$ in the first integral on the right-hand side of Eq. (7.89) and substituting $v(-x) = -v(x)$ from Eq. (7.88), we obtain

$$\begin{aligned} u(x_0) &= \frac{1}{\pi} P \int_{0}^{\infty} v(x) \left\{ \frac{1}{x + x_0} + \frac{1}{x - x_0} \right\} dx \\ &= \frac{2}{\pi} P \int_{0}^{\infty} \frac{x v(x)}{x^2 - x_0^2} dx. \end{aligned} \tag{7.90}$$

Similarly,

$$v(x_0) = -\frac{2}{\pi} P \int_{0}^{\infty} \frac{x_0 u(x)}{x^2 - x_0^2} dx. \tag{7.91}$$

The original Kronig–Kramers optical dispersion relations were in this form. The asymptotic behavior ($x_0 \to \infty$) of Eqs. (7.90) and (7.91) lead to quantum mechanical **sum rules**, Exercise 7.2.4.

Optical Dispersion

The function $\exp[i(kx - \omega t)]$ describes an electromagnetic wave moving along the x-axis in the positive direction with velocity $v = \omega/k$; ω is the angular frequency, k the wave number or propagation vector, and $n = ck/\omega$ the index of refraction. From Maxwell's

[8] This is not just a happy coincidence. It ensures that the Fourier transform of $f(x)$ will be real. In turn, Eq. (7.87) is a consequence of obtaining $f(x)$ as the Fourier transform of a real function.

[9] $u(x, 0) = u(-x, 0)$, $v(x, 0) = -v(-x, 0)$. Compare these symmetry conditions with those that follow from the Schwarz reflection principle, Section 6.5.

equations, electric permittivity ε, and Ohm's law with conductivity σ, the propagation vector k for a dielectric becomes[10]

$$k^2 = \varepsilon \frac{\omega^2}{c^2}\left(1 + i\frac{4\pi\sigma}{\omega\varepsilon}\right) \tag{7.92}$$

(with μ, the magnetic permeability, taken to be unity). The presence of the conductivity (which means absorption) gives rise to an imaginary part. The propagation vector k (and therefore the index of refraction n) have become complex.

Conversely, the (positive) imaginary part implies absorption. For poor conductivity ($4\pi\sigma/\omega\varepsilon \ll 1$) a binomial expansion yields

$$k = \sqrt{\varepsilon}\frac{\omega}{c} + i\frac{2\pi\sigma}{c\sqrt{\varepsilon}}$$

and

$$e^{i(kx-\omega t)} = e^{i\omega(x\sqrt{\varepsilon}/c - t)}e^{-2\pi\sigma x/c\sqrt{\varepsilon}},$$

an attenuated wave.

Returning to the general expression for k^2, Eq. (7.92), we find that the index of refraction becomes

$$n^2 = \frac{c^2k^2}{\omega^2} = \varepsilon + i\frac{4\pi\sigma}{\omega}. \tag{7.93}$$

We take n^2 to be a function of the **complex** variable ω (with ε and σ depending on ω). However, n^2 does not vanish as $\omega \to \infty$ but instead approaches unity. So to satisfy the condition, Eq. (7.81), one works with $f(\omega) = n^2(\omega) - 1$. The original Kronig–Kramers optical dispersion relations were in the form of

$$\begin{aligned}
\Re\big[n^2(\omega_0) - 1\big] &= \frac{2}{\pi}P\int_0^\infty \frac{\omega\Im[n^2(\omega) - 1]}{\omega^2 - \omega_0^2}d\omega, \\
\Im\big[n^2(\omega_0) - 1\big] &= -\frac{2}{\pi}P\int_0^\infty \frac{\omega_0\Re[n^2(\omega) - 1]}{\omega^2 - \omega_0^2}d\omega.
\end{aligned} \tag{7.94}$$

Knowledge of the absorption coefficient at all frequencies specifies the real part of the index of refraction, and vice versa.

The Parseval Relation

When the functions $u(x)$ and $v(x)$ are Hilbert transforms of each other (given by Eqs. (7.86)) and each is square integrable,[11] the two functions are related by

$$\int_{-\infty}^{\infty} |u(x)|^2\,dx = \int_{-\infty}^{\infty} |v(x)|^2\,dx. \tag{7.95}$$

[10] See J. D. Jackson, *Classical Electrodynamics*, 3rd ed. New York: Wiley (1999), Sections 7.7 and 7.10. Equation (7.92) is in Gaussian units.

[11] This means that $\int_{-\infty}^{\infty} |u(x)|^2\,dx$ and $\int_{-\infty}^{\infty} |v(x)|^2\,dx$ are finite.

This is the Parseval relation.

To derive Eq. (7.95), we start with

$$\int_{-\infty}^{\infty} |u(x)|^2\,dx = \int_{-\infty}^{\infty} \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{v(s)\,ds}{s-x}\,\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{v(t)\,dt}{t-x}\,dx,$$

using Eq. (7.86a) twice. Integrating first with respect to x, we have

$$\int_{-\infty}^{\infty} |u(x)|^2\,dx = \int_{-\infty}^{\infty} v(s)\,ds \int_{-\infty}^{\infty} v(t)\,dt \int_{-\infty}^{\infty} \frac{dx}{\pi^2\,(s-x)(t-x)}. \tag{7.96}$$

From Exercise 7.2.8, the x integration yields a delta function:

$$\frac{1}{\pi^2}\int_{-\infty}^{\infty} \frac{dx}{(s-x)(t-x)} = \delta(s-t).$$

We have

$$\int_{-\infty}^{\infty} |u(x)|^2\,dx = \int_{-\infty}^{\infty} v(t)\,dt \int_{-\infty}^{\infty} v(s)\delta(s-t)\,ds. \tag{7.97}$$

Then the s integration is carried out by inspection, using the defining property of the delta function:

$$\int_{-\infty}^{\infty} v(s)\delta(s-t)\,ds = v(t). \tag{7.98}$$

Substituting Eq. (7.98) into Eq. (7.97), we have Eq. (7.95), the Parseval relation. Again, in terms of optics, the presence of refraction over some frequency range ($n \neq 1$) implies the existence of absorption, and vice versa.

Causality

The real significance of dispersion relations in physics is that they are a direct consequence of assuming that the particular physical system obeys causality. Causality is awkward to define precisely, but the general meaning is that the effect cannot precede the cause. A scattered wave cannot be emitted by the scattering center before the incident wave has arrived. For linear systems the most general relation between an input function G (the cause) and an output function H (the effect) may be written as

$$H(t) = \int_{-\infty}^{\infty} F(t-t')G(t')\,dt'. \tag{7.99}$$

Causality is imposed by requiring that

$$F(t-t') = 0 \qquad \text{for } t-t' < 0.$$

Equation (7.99) gives the time dependence. The frequency dependence is obtained by taking Fourier transforms. By the Fourier convolution theorem, Section 15.5,

$$h(\omega) = f(\omega)g(\omega),$$

where $f(\omega)$ is the Fourier transform of $F(t)$, and so on. Conversely, $F(t)$ is the Fourier transform of $f(\omega)$.

The connection with the dispersion relations is provided by the Titchmarsh **theorem**.[12] This states that if $f(\omega)$ is square integrable over the real ω-axis, then any one of the following three statements implies the other two.

1. The Fourier transform of $f(\omega)$ is zero for $t < 0$: Eq. (7.99).
2. Replacing ω by z, the function $f(z)$ is analytic in the complex z-plane for $y > 0$ and approaches $f(x)$ almost everywhere as $y \to 0$. Further,
$$\int_{-\infty}^{\infty} |f(x+iy)|^2 dx < K \qquad \text{for } y > 0;$$
that is, the integral is bounded.
3. The real and imaginary parts of $f(z)$ are Hilbert transforms of each other: Eqs. (7.86a) and (7.86b).

The assumption that the relationship between the input and the output of our linear system is causal (Eq. (7.99)) means that the first statement is satisfied. If $f(\omega)$ is square integrable, then the Titchmarsh theorem has the third statement as a consequence and we have dispersion relations.

Exercises

7.2.1 The function $f(z)$ satisfies the conditions for the dispersion relations. In addition, $f(z) = f^*(z^*)$, the Schwarz reflection principle, Section 6.5. Show that $f(z)$ is identically zero.

7.2.2 For $f(z)$ such that we may replace the closed contour of the Cauchy integral formula by an integral over the real axis we have
$$f(x_0) = \frac{1}{2\pi i}\left\{\int_{-\infty}^{x_0-\delta} \frac{f(x)}{x-x_0}dx + \int_{x_0+\delta}^{\infty} \frac{f(x)}{x-x_0}dx\right\} + \frac{1}{2\pi i}\int_{C_{x_0}} \frac{f(x)}{x-x_0}dx.$$
Here C_{x_0} designates a small semicircle about x_0 in the lower half-plane. Show that this reduces to
$$f(x_0) = \frac{1}{\pi i} P\int_{-\infty}^{\infty} \frac{f(x)}{x-x_0}dx,$$
which is Eq. (7.84).

7.2.3 (a) For $f(z) = e^{iz}$, Eq. (7.81) does not hold at the endpoints, $\arg z = 0, \pi$. Show, with the help of Jordan's lemma, Section 7.1, that Eq. (7.82) still holds.
(b) For $f(z) = e^{iz}$ verify the dispersion relations, Eq. (7.89) or Eqs. (7.90) and (7.91), by direct integration.

7.2.4 With $f(x) = u(x) + iv(x)$ and $f(x) = f^*(-x)$, show that as $x_0 \to \infty$,

[12] Refer to E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, 2nd ed. New York: Oxford University Press (1937). For a more informal discussion of the Titchmarsh theorem and further details on causality see J. Hilgevoord, *Dispersion Relations and Causal Description*. Amsterdam: North-Holland (1962).

(a) $u(x_0) \sim -\dfrac{2}{\pi x_0^2}\displaystyle\int_0^\infty x v(x)\,dx,$

(b) $v(x_0) \sim \dfrac{2}{\pi x_0}\displaystyle\int_0^\infty u(x)\,dx.$

In quantum mechanics relations of this form are often called **sum rules**.

7.2.5 (a) Given the integral equation

$$\frac{1}{1+x_0^2} = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{u(x)}{x - x_0} dx,$$

use Hilbert transforms to determine $u(x_0)$.

(b) Verify that the integral equation of part (a) is satisfied.

(c) From $f(z)|_{y=0} = u(x) + i v(x)$, replace x by z and determine $f(z)$. Verify that the conditions for the Hilbert transforms are satisfied.

(d) Are the crossing conditions satisfied?

ANS. (a) $u(x_0) = \dfrac{x_0}{1+x_0^2}$, (c) $f(z) = (z+i)^{-1}$.

7.2.6 (a) If the real part of the complex index of refraction (squared) is constant (no optical dispersion), show that the imaginary part is zero (no absorption).

(b) Conversely, if there is absorption, show that there must be dispersion. In other words, if the imaginary part of $n^2 - 1$ is not zero, show that the real part of $n^2 - 1$ is not constant.

7.2.7 Given $u(x) = x/(x^2+1)$ and $v(x) = -1/(x^2+1)$, show by direct evaluation of each integral that

$$\int_{-\infty}^{\infty} |u(x)|^2 dx = \int_{-\infty}^{\infty} |v(x)|^2 dx = \int_{-\infty}^{\infty} |u(x) + i v(x)|^2 dx.$$

ANS. π.

7.2.8 Take $u(x) = \delta(x)$, a delta function, and **assume** that the Hilbert transform equations hold.

(a) Show that

$$\delta(w) = \frac{1}{\pi^2} \int_{-\infty}^{\infty} \frac{dy}{y(y-w)}.$$

(b) With changes of variables $w = s - t$ and $x = s - y$, transform the δ representation of part (a) into

$$\delta(s-t) = \frac{1}{\pi^2} \int_{-\infty}^{\infty} \frac{dx}{(x-s)(s-t)}.$$

Note. The δ function is discussed in Section 1.15.

7.2.9 Show that

$$\delta(x) = \frac{1}{\pi^2} \int_{-\infty}^{\infty} \frac{dt}{t(t-x)}$$

is a valid representation of the delta function in the sense that

$$\int_{-\infty}^{\infty} f(x)\delta(x)\,dx = f(0).$$

Assume that $f(x)$ satisfies the condition for the existence of a Hilbert transform.
Hint. Apply Eq. (7.84) twice.

7.3 METHOD OF STEEPEST DESCENTS

Analytic Landscape

In analyzing problems in mathematical physics, one often finds it desirable to know the behavior of a function for large values of the variable or some parameter s, that is, the asymptotic behavior of the function. Specific examples are furnished by the gamma function (Chapter 8) and various Bessel functions (Chapter 11). All these analytic functions are defined by integrals

$$I(s) = \int_C F(z, s)\,dz, \tag{7.100}$$

where F is analytic in z and depends on a real parameter s. We write $F(z)$ whenever possible.

So far we have evaluated such definite integrals of analytic functions along the real axis by deforming the path C to C' in the complex plane, so $|F|$ becomes small for all z on C'. This method succeeds as long as only isolated poles occur in the area between C and C'. The poles are taken into account by applying the residue theorem of Section 7.1. The residues give a measure of the simple poles, where $|F| \to \infty$, which usually dominate and determine the value of the integral.

The behavior of the integral in Eq. (7.100) clearly depends on the absolute value $|F|$ of the integrand. Moreover, the contours of $|F|$ often become more pronounced as s becomes large. Let us focus on a plot of $|F(x+iy)|^2 = U^2(x, y) + V^2(x, y)$, rather than the real part $\Re F = U$ and the imaginary part $\Im F = V$ separately. Such a plot of $|F|^2$ over the complex plane is called the **analytic landscape**, after Jensen, who, in 1912, proved that it has **only saddle points and troughs but no peaks**. Moreover, the troughs reach down all the way to the complex plane. In the absence of (simple) poles, **saddle points** are next in line to **dominate the integral** in Eq. (7.100). Hence the name **saddle point method**. At a saddle point the real (or imaginary) part U of F has a local maximum, which implies that

$$\frac{\partial U}{\partial x} = \frac{\partial U}{\partial y} = 0,$$

and therefore by the use of the Cauchy–Riemann conditions of Section 6.2,

$$\frac{\partial V}{\partial x} = \frac{\partial V}{\partial y} = 0,$$

so V has a minimum, or vice versa, and $F'(z) = 0$. Jensen's theorem prevents U and V from having either a maximum or a minimum. See Fig. 7.18 for a typical shape (and Exercises 6.2.3 and 6.2.4). **Our strategy will be to choose the path C so that it runs over the saddle point, which gives the dominant contribution, and in the valleys elsewhere.** If there are several saddle points, we treat each alike, and their contributions will add to $I(s \to \infty)$.

To prove that there are no peaks, assume there is one at z_0. That is, $|F(z_0)|^2 > |F(z)|^2$ for all z of a neighborhood $|z - z_0| \le r$. If

$$F(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$$

is the Taylor expansion at z_0, the mean value $m(F)$ on the circle $z = z_0 + r\exp(i\varphi)$ becomes

$$\begin{aligned} m(F) &\equiv \frac{1}{2\pi}\int_0^{2\pi} \left|F\left(z_0 + re^{i\varphi}\right)\right|^2 d\varphi \\ &= \frac{1}{2\pi}\int_0^{2\pi} \sum_{m,n=0}^{\infty} a_m^* a_n r^{m+n} e^{i(n-m)\varphi}\, d\varphi \\ &= \sum_{n=0}^{\infty} |a_n|^2 r^{2n} \ge |a_0|^2 = \left|F(z_0)\right|^2, \end{aligned} \tag{7.101}$$

using orthogonality, $\frac{1}{2\pi}\int_0^{2\pi} \exp i(n-m)\varphi\, d\varphi = \delta_{nm}$. Since $m(F)$ is the mean value of $|F|^2$ on the circle of radius r, there must be a point z_1 on it so that $|F(z_1)|^2 \ge m(F) \ge |F(z_0)|^2$, which contradicts our assumption. Hence there can be no such peak.

Next, let us assume there is a minimum at z_0 so that $0 < |F(z_0)|^2 < |F(z)|^2$ for all z of a neighborhood of z_0. In other words, the dip in the valley does not go down to the complex plane. Then $|F(z)|^2 > 0$ and, since $1/F(z)$ is analytic there, it has a Taylor expansion and z_0 would be a peak of $1/|F(z)|^2$, which is impossible. This proves Jensen's theorem. We now turn our attention back to the integral in Eq. (7.100).

Saddle Point Method

Since each saddle point z_0 necessarily lies above the complex plane, that is, $|F(z_0)|^2 > 0$, we write F in exponential form, $e^{f(z,s)}$, in its vicinity without loss of generality. Note that having no zero in the complex plane is a characteristic property of the exponential function. Moreover, any saddle point with $F(z) = 0$ becomes a trough of $|F(z)|^2$ because $|F(z)|^2 \ge 0$. A case in point is the function z^2 at $z = 0$, where $d(z^2)/dz = 2z = 0$. Here $z^2 = (x + iy)^2 = x^2 - y^2 + 2ixy$, and $2xy$ has a saddle point at $z = 0$, and so has $x^2 - y^2$, but $|z|^4$ has a trough there.

At z_0 the tangential plane is horizontal; that is, $\frac{\partial F}{\partial z}|_{z=z_0} = 0$, or equivalently $\frac{\partial f}{\partial z}|_{z=z_0} = 0$. **This condition locates the saddle point.** Our next goal is to determine the **direction of steepest descent.** At z_0, f has a power series

$$f(z) = f(z_0) + \frac{1}{2} f''(z_0)(z - z_0)^2 + \cdots, \tag{7.102}$$

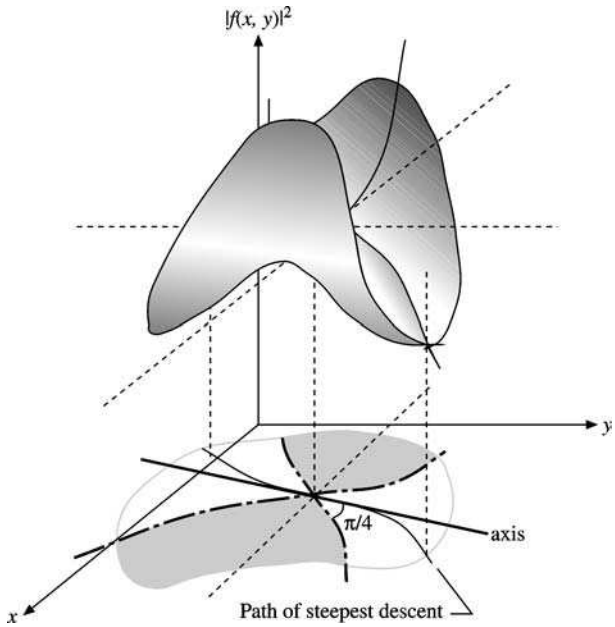

FIGURE 7.18 A saddle point.

or

$$f(z) = f(z_0) + \frac{1}{2}\big(f''(z_0) + \varepsilon\big)(z - z_0)^2, \tag{7.103}$$

upon collecting all higher powers in the (small) ε. Let us take $f''(z_0) \neq 0$ for simplicity. Then

$$f''(z_0)(z - z_0)^2 = -t^2, \qquad t \text{ real}, \tag{7.104}$$

defines a line through z_0 (saddle point **axis** in Fig. 7.18). At z_0, $t = 0$. Along the axis $\Im f''(z_0)(z - z_0)^2$ is zero and $v = \Im f(z) \approx \Im f(z_0)$ is constant if ε in Eq. (7.103) is neglected. Equation (7.104) can also be expressed in terms of angles,

$$\arg(z - z_0) = \frac{\pi}{2} - \frac{1}{2}\arg f''(z_0) = \text{constant}. \tag{7.105}$$

Since $|F(z)|^2 = \exp(2\Re f)$ varies monotonically with $\Re f$, $|F(z)|^2 \approx \exp(-t^2)$ falls off exponentially from its maximum at $t = 0$ along this axis. Hence the name **steepest descent**. The line through z_0 defined by

$$f''(z_0)(z - z_0)^2 = +t^2 \tag{7.106}$$

is orthogonal to this axis (**dashed** in Fig. 7.18), which is evident from its angle,

$$\arg(z - z_0) = -\frac{1}{2}\arg f''(z_0) = \text{constant}, \tag{7.107}$$

when compared with Eq. (7.105). Here $|F(z)|^2$ grows exponentially.

The curves $\Re f(z) = \Re f(z_0)$ go through z_0, so $\Re[(f''(z_0) + \varepsilon)(z - z_0)^2] = 0$, or $(f''(z_0) + \varepsilon)(z - z_0)^2 = it$ for real t. Expressing this in angles as

$$\arg(z - z_0) = \frac{\pi}{4} - \frac{1}{2}\arg\big(f''(z_0) + \varepsilon\big), \qquad t > 0, \tag{7.108a}$$

$$\arg(z - z_0) = -\frac{\pi}{4} - \frac{1}{2}\arg\big(f''(z_0) + \varepsilon\big), \qquad t < 0, \tag{7.108b}$$

and comparing with Eqs. (7.105) and (7.107) we note that these curves (**dot-dashed** in Fig. 7.18) divide the saddle point region into four sectors, two with $\Re f(z) > \Re f(z_0)$ (hence $|F(z)| > |F(z_0)|$), shown shaded in Fig. 7.18, and two with $\Re f(z) < \Re f(z_0)$ (hence $|F(z)| < |F(z_0)|$). They are at $\pm\frac{\pi}{4}$ angles from the axis. Thus, the integration path has to avoid the shaded areas, where $|F|$ rises. If a path is chosen to run up the slopes above the saddle point, the large imaginary part of $f(z)$ leads to rapid oscillations of $F(z) = e^{f(z)}$ and cancelling contributions to the integral.

So far, our **treatment has been general**, except for $f''(z_0) \neq 0$, which can be relaxed. Now we are ready to **specialize the integrand** F further in order to tie up the path selection with the asymptotic behavior as $s \to \infty$.

We assume that s appears linearly in the exponent, that is, we replace $\exp f(z, s) \to \exp(sf(z))$. This dependence on s ensures that the saddle point contribution at z_0 grows with $s \to \infty$ providing steep slopes, as is the case in most applications in physics. In order to account for the region far away from the saddle point that is not influenced by s, we include another analytic function, $g(z)$, which varies slowly near the saddle point and is independent of s.

Altogether, then, **our integral has the more appropriate and specific form**

$$I(s) = \int_C g(z)e^{sf(z)}\,dz. \tag{7.109}$$

The path of steepest descent is the saddle point axis when we neglect the higher-order terms, ε, in Eq. (7.103). With ε, the path of steepest descent is the curve close to the axis within the unshaded sectors, where $v = \Im f(z)$ is strictly constant, while $\Im f(z)$ is only approximately constant on the axis. We approximate $I(s)$ by the integral along the piece of the axis inside the patch in Fig. 7.18, where (compare with Eq. (7.104))

$$z = z_0 + xe^{i\alpha}, \qquad \alpha = \frac{\pi}{2} - \frac{1}{2}\arg f''(z_0), \qquad a \le x \le b. \tag{7.110}$$

We find

$$I(s) \approx e^{i\alpha}\int_a^b g\big(z_0 + xe^{i\alpha}\big)\exp\big[sf\big(z_0 + xe^{i\alpha}\big)\big]\,dx, \tag{7.111a}$$

and the omitted part is small and can be estimated because $\Re(f(z) - f(z_0))$ has an upper negative bound, $-R$ say, that depends on the size of the saddle point patch in Fig. 7.18 (that is, the values of a, b in Eq. (7.110)) that we choose. In Eq. (7.111) we use the power expansions

$$\begin{aligned} f\big(z_0 + xe^{i\alpha}\big) &= f(z_0) + \frac{1}{2}f''(z_0)e^{2i\alpha}x^2 + \cdots, \\ g\big(z_0 + xe^{i\alpha}\big) &= g(z_0) + g'(z_0)e^{i\alpha}x + \cdots, \end{aligned} \tag{7.111b}$$

and recall from Eq. (7.110) that

$$\frac{1}{2}f''(z_0)e^{2i\alpha} = -\frac{1}{2}\left|f''(z_0)\right| < 0.$$

We find for the leading term for $s \to \infty$:

$$I(s) = g(z_0)e^{sf(z_0)+i\alpha}\int_a^b e^{-\frac{1}{2}s|f''(z_0)|x^2}\,dx. \tag{7.112}$$

Since the integrand in Eq. (7.112) is essentially zero when x departs appreciably from the origin, we let $b \to \infty$ and $a \to -\infty$. The small error involved is straightforward to estimate. Noting that the remaining integral is just a Gauss error integral,

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}a^2x^2}\,dx = \frac{1}{a}\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,dx = \frac{\sqrt{2\pi}}{a},$$

we finally obtain

$$\boxed{I(s) = \frac{\sqrt{2\pi}\,g(z_0)e^{sf(z_0)}e^{i\alpha}}{|sf''(z_0)|^{1/2}},} \tag{7.113}$$

where the phase α was introduced in Eqs. (7.110) and (7.105).

A note of warning: We assumed that the only significant contribution to the integral came from the immediate vicinity of the saddle point(s) $z = z_0$. This condition must be checked for each new problem (Exercise 7.3.5).

Example 7.3.1 Asymptotic Form of the Hankel Function $H_\nu^{(1)}(s)$

In Section 11.4 it is shown that the Hankel functions, which satisfy Bessel's equation, may be defined by

$$H_\nu^{(1)}(s) = \frac{1}{\pi i}\int_{C_1,0}^{\infty e^{i\pi}} e^{(s/2)(z-1/z)}\frac{dz}{z^{\nu+1}}, \tag{7.114}$$

$$H_\nu^{(2)}(s) = \frac{1}{\pi i}\int_{C_2,\infty e^{-i\pi}}^{0} e^{(s/2)(z-1/z)}\frac{dz}{z^{\nu+1}}. \tag{7.115}$$

The contour C_1 is the curve in the upper half-plane of Fig. 7.19. The contour C_2 is in the lower half-plane. We apply the method of steepest descents to the first Hankel function, $H_\nu^{(1)}(s)$, which is conveniently in the form specified by Eq. (7.109), with $f(z)$ given by

$$f(z) = \frac{1}{2}\left(z - \frac{1}{z}\right). \tag{7.116}$$

By differentiating, we obtain

$$f'(z) = \frac{1}{2} + \frac{1}{2z^2}. \tag{7.117}$$

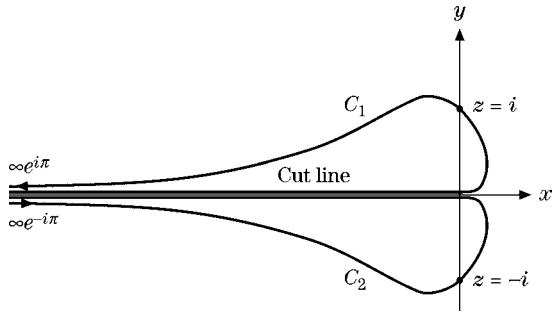

FIGURE 7.19 Hankel function contours.

Setting $f'(z) = 0$, we obtain

$$z = i, -i. \tag{7.118}$$

Hence there are saddle points at $z = +i$ and $z = -i$. At $z = i$, $f''(i) = -i$, or $\arg f''(i) = -\pi/2$, so the saddle point direction is given by Eq. (7.110) as $\alpha = \frac{\pi}{2} + \frac{\pi}{4} = \frac{3}{4}\pi$. For the integral for $H_\nu^{(1)}(s)$ we must choose the contour through the point $z = +i$ so that it starts at the origin, moves out tangentially to the positive real axis, and then moves around through the saddle point at $z = +i$ in the direction given by the angle $\alpha = 3\pi/4$ and then on out to minus infinity, asymptotic with the negative real axis. The path of steepest ascent, which we must avoid, has the phase $-\frac{1}{2}\arg f''(i) = \frac{\pi}{4}$, according to Eq. (7.107), and is orthogonal to the axis, our path of steepest descent.

Direct substitution into Eq. (7.113) with $\alpha = 3\pi/4$ now yields

$$\begin{aligned} H_\nu^{(1)}(s) &= \frac{1}{\pi i}\frac{\sqrt{2\pi i^{-\nu-1}}e^{(s/2)(i-1/i)}e^{3\pi i/4}}{|(s/2)(-2/i^3)|^{1/2}} \\ &= \sqrt{\frac{2}{\pi s}}e^{(i\pi/2)(-\nu-2)}e^{is}e^{i(3\pi/4)}. \end{aligned} \tag{7.119}$$

By combining terms, we obtain

$$H_\nu^{(1)}(s) \approx \sqrt{\frac{2}{\pi s}}e^{i(s-\nu(\pi/2)-\pi/4)} \tag{7.120}$$

as the leading term of the asymptotic expansion of the Hankel function $H_\nu^{(1)}(s)$. Additional terms, if desired, may be picked up from the power series of f and g in Eq. (7.111b). The other Hankel function can be treated similarly using the saddle point at $z = -i$. ■

Example 7.3.2 ASYMPTOTIC FORM OF THE FACTORIAL FUNCTION $\Gamma(1 + s)$

In many physical problems, particularly in the field of statistical mechanics, it is desirable to have an accurate approximation of the gamma or factorial function of very large

numbers. As developed in Section 8.1, the factorial function may be defined by the Euler integral

$$\Gamma(1+s) = \int_0^\infty \rho^s e^{-\rho}\, d\rho = s^{s+1} \int_0^\infty e^{s(\ln z - z)}\, dz. \tag{7.121}$$

Here we have made the substitution $\rho = zs$ in order to convert the integral to the form required by Eq. (7.109). As before, we assume that s is real and positive, from which it follows that the integrand vanishes at the limits 0 and ∞. By differentiating the z-dependence appearing in the exponent, we obtain

$$\frac{df(z)}{dz} = \frac{d}{dz}(\ln z - z) = \frac{1}{z} - 1, \qquad f''(z) = -\frac{1}{z^2}, \tag{7.122}$$

which shows that the point $z = 1$ is a saddle point and $\arg f''(1) = \arg(-1) = \pi$. According to Eq. (7.109) we let

$$z - 1 = xe^{i\alpha}, \qquad \alpha = \frac{\pi}{2} - \frac{1}{2}\arg f''(1) = \frac{\pi}{2} - \frac{\pi}{2} = 0, \tag{7.123}$$

with x small, to describe the contour in the vicinity of the saddle point. From this we see that the direction of steepest descent is along the real axis, a conclusion that we could have reached more or less intuitively.

Direct substitution into Eq. (7.113) with $\alpha = 0$ now gives

$$\Gamma(1+s) \approx \frac{\sqrt{2\pi}s^{s+1}e^{-s}}{|s(-1^{-2})|^{1/2}}. \tag{7.124}$$

Thus the first term in the asymptotic expansion of the factorial function is

$$\Gamma(1+s) \approx \sqrt{2\pi s}s^s e^{-s}. \tag{7.125}$$

This result is the first term in Stirling's expansion of the factorial function. The method of steepest descent is probably the easiest way of obtaining this first term. If more terms in the expansion are desired, then the method of Section 8.3 is preferable. ■

In the foregoing example the calculation was carried out by assuming s to be real. This assumption is not necessary. We may show (Exercise 7.3.6) that Eq. (7.125) also holds when s is replaced by the complex variable w, provided only that the real part of w be required to be large and positive.

Asymptotic limits of integral representations of functions are extremely important in many approximations and applications in physics:

$$\int_C g(z)e^{sf(z)}\, dz \sim \frac{\sqrt{2\pi}\, g(z_0)e^{sf(z_0)}e^{i\alpha}}{\sqrt{|sf''(z_0)|}}, \qquad f'(z_0) = 0.$$

The saddle point method is one method of choice for deriving them and belongs in the toolkit of every physicist and engineer.

Exercises

7.3.1 Using the method of steepest descents, evaluate the second Hankel function, given by

$$H_\nu^{(2)}(s) = \frac{1}{\pi i}\int_{-\infty C_2}^{0} e^{(s/2)(z-1/z)}\frac{dz}{z^{\nu+1}},$$

with contour C_2 as shown in Fig. 7.19.

ANS. $H_\nu^{(2)}(s) \approx \sqrt{\dfrac{2}{\pi s}}e^{-i(s-\pi/4-\nu\pi/2)}$.

7.3.2 Find the steepest path and leading asymptotic expansion for the Fresnel integrals $\int_0^s \cos x^2\,dx$, $\int_0^s \sin x^2\,dx$.
Hint. Use $\int_0^1 e^{isz^2}\,dz$.

7.3.3 (a) In applying the method of steepest descent to the Hankel function $H_\nu^{(1)}(s)$, show that

$$\Re\big[f(z)\big] < \Re\big[f(z_0)\big] = 0$$

for z on the contour C_1 but away from the point $z = z_0 = i$.

(b) Show that

$$\Re\big[f(z)\big] > 0 \qquad \text{for} \quad 0 < r < 1, \qquad \begin{cases} \dfrac{\pi}{2} < \theta \le \pi \\ -\pi \le \theta < \dfrac{\pi}{2} \end{cases}$$

and

$$\Re\big[f(z)\big] < 0 \qquad \text{for} \quad r > 1, \qquad -\frac{\pi}{2} < \theta < \frac{\pi}{2}$$

(Fig. 7.20). This is why C_1 may not be deformed to pass through the second saddle point, $z = -i$. Compare with and verify the dot-dashed lines in Fig. 7.18 for this case.

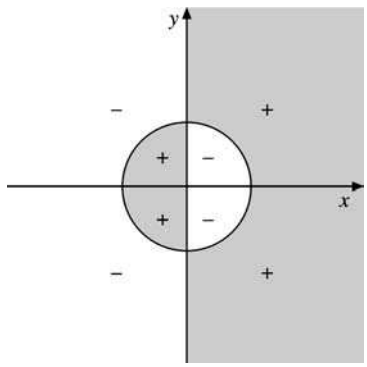

FIGURE 7.20

7.3.4 Determine the asymptotic dependence of the modified Bessel functions $I_\nu(x)$, given

$$I_\nu(x) = \frac{1}{2\pi i}\int_C e^{(x/2)(t+1/t)}\frac{dt}{t^{\nu+1}}.$$

The contour starts and ends at $t = -\infty$, encircling the origin in a positive sense. There are two saddle points. Only the one at $z = +1$ contributes significantly to the asymptotic form.

7.3.5 Determine the asymptotic dependence of the modified Bessel function of the second kind, $K_\nu(x)$, by using

$$K_\nu(x) = \frac{1}{2}\int_0^\infty e^{(-x/2)(s+1/s)}\frac{ds}{s^{1-\nu}}.$$

7.3.6 Show that Stirling's formula,

$$\Gamma(1+s) \approx \sqrt{2\pi s}s^s e^{-s},$$

holds for complex values of s (with $\Re(s)$ large and positive).
Hint. This involves assigning a phase to s and then demanding that $\Im[sf(z)] =$ constant in the vicinity of the saddle point.

7.3.7 Assume $H_\nu^{(1)}(s)$ to have a negative power-series expansion of the form

$$H_\nu^{(1)}(s) = \sqrt{\frac{2}{\pi s}}e^{i(s-\nu(\pi/2)-\pi/4)}\sum_{n=0}^{\infty}a_{-n}s^{-n},$$

with the coefficient of the summation obtained by the method of steepest descent. Substitute into Bessel's equation and show that you reproduce the asymptotic series for $H_\nu^{(1)}(s)$ given in Section 11.6.

Additional Readings

Nussenzveig, H. M., *Causality and Dispersion Relations*, Mathematics in Science and Engineering Series, Vol. 95. New York: Academic Press (1972). This is an advanced text covering causality and dispersion relations in the first chapter and then moving on to develop the implications in a variety of areas of theoretical physics.

Wyld, H. W., *Mathematical Methods for Physics*. Reading, MA: Benjamin/Cummings (1976), Perseus Books (1999). This is a relatively advanced text that contains an extensive discussion of the dispersion relations.

This page intentionally left blank

CHAPTER 8

THE GAMMA FUNCTION (FACTORIAL FUNCTION)

The gamma function appears occasionally in physical problems such as the normalization of Coulomb wave functions and the computation of probabilities in statistical mechanics. In general, however, it has less direct physical application and interpretation than, say, the Legendre and Bessel functions of Chapters 11 and 12. Rather, its importance stems from its usefulness in developing other functions that have direct physical application. The gamma function, therefore, is included here.

8.1 DEFINITIONS, SIMPLE PROPERTIES

At least three different, convenient definitions of the gamma function are in common use. Our first task is to state these definitions, to develop some simple, direct consequences, and to show the equivalence of the three forms.

Infinite Limit (Euler)

The first definition, named after Euler, is

$$\Gamma(z) \equiv \lim_{n\to\infty} \frac{1\cdot 2\cdot 3\cdots n}{z(z+1)(z+2)\cdots(z+n)} n^z, \qquad z \neq 0, -1, -2, -3, \ldots. \tag{8.1}$$

This definition of $\Gamma(z)$ is useful in developing the Weierstrass infinite-product form of $\Gamma(z)$, Eq. (8.16), and in obtaining the derivative of $\ln\Gamma(z)$ (Section 8.2). Here and else-

where in this chapter z may be either real or complex. Replacing z with $z+1$, we have

$$\begin{aligned}\Gamma(z+1) &= \lim_{n\to\infty} \frac{1\cdot 2\cdot 3\cdots n}{(z+1)(z+2)(z+3)\cdots(z+n+1)} n^{z+1} \\ &= \lim_{n\to\infty} \frac{nz}{z+n+1} \cdot \frac{1\cdot 2\cdot 3\cdots n}{z(z+1)(z+2)\cdots(z+n)} n^{z} \\ &= z\Gamma(z). \end{aligned} \tag{8.2}$$

This is the basic functional relation for the gamma function. It should be noted that it is a **difference** equation. It has been shown that the gamma function is one of a general class of functions that do not satisfy any differential equation with rational coefficients. Specifically, the gamma function is one of the very few functions of mathematical physics that does not satisfy either the hypergeometric differential equation (Section 13.4) or the confluent hypergeometric equation (Section 13.5).

Also, from the definition,

$$\Gamma(1) = \lim_{n\to\infty} \frac{1\cdot 2\cdot 3\cdots n}{1\cdot 2\cdot 3\cdots n(n+1)} n = 1. \tag{8.3}$$

Now, application of Eq. (8.2) gives

$$\begin{aligned}&\Gamma(2) = 1, \\ &\Gamma(3) = 2\Gamma(2) = 2, \ldots \\ &\Gamma(n) = 1\cdot 2\cdot 3\cdots(n-1) = (n-1)!. \end{aligned} \tag{8.4}$$

Definite Integral (Euler)

A second definition, also frequently called the Euler integral, is

$$\boxed{\Gamma(z) \equiv \int_0^\infty e^{-t} t^{z-1}\, dt, \qquad \Re(z) > 0.} \tag{8.5}$$

The restriction on z is necessary to avoid divergence of the integral. When the gamma function does appear in physical problems, it is often in this form or some variation, such as

$$\Gamma(z) = 2\int_0^\infty e^{-t^2} t^{2z-1}\, dt, \qquad \Re(z) > 0. \tag{8.6}$$

$$\Gamma(z) = \int_0^1 \left[\ln\left(\frac{1}{t}\right)\right]^{z-1} dt, \qquad \Re(z) > 0. \tag{8.7}$$

When $z = \frac{1}{2}$, Eq. (8.6) is just the Gauss error integral, and we have the interesting result

$$\Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi}. \tag{8.8}$$

Generalizations of Eq. (8.6), the Gaussian integrals, are considered in Exercise 8.1.11. This definite integral form of $\Gamma(z)$, Eq. (8.5), leads to the beta function, Section 8.4.

To show the equivalence of these two definitions, Eqs. (8.1) and (8.5), consider the function of two variables

$$F(z,n)=\int_0^n\left(1-\frac{t}{n}\right)^n t^{z-1}\,dt,\qquad \Re(z)>0, \tag{8.9}$$

with n a positive integer.[1] Since

$$\lim_{n\to\infty}\left(1-\frac{t}{n}\right)^n\equiv e^{-t}, \tag{8.10}$$

from the definition of the exponential

$$\lim_{n\to\infty}F(z,n)=F(z,\infty)=\int_0^\infty e^{-t}t^{z-1}\,dt\equiv\Gamma(z) \tag{8.11}$$

by Eq. (8.5).

Returning to $F(z,n)$, we evaluate it in successive integrations by parts. For convenience let $u=t/n$. Then

$$F(z,n)=n^z\int_0^1(1-u)^n u^{z-1}\,du. \tag{8.12}$$

Integrating by parts, we obtain

$$\frac{F(z,n)}{n^z}=(1-u)^n\frac{u^z}{z}\bigg|_0^1+\frac{n}{z}\int_0^1(1-u)^{n-1}u^z\,du. \tag{8.13}$$

Repeating this with the integrated part vanishing at both endpoints each time, we finally get

$$\begin{aligned}F(z,n)&=n^z\frac{n(n-1)\cdots1}{z(z+1)\cdots(z+n-1)}\int_0^1 u^{z+n-1}\,du\\&=\frac{1\cdot2\cdot3\cdots n}{z(z+1)(z+2)\cdots(z+n)}n^z.\end{aligned} \tag{8.14}$$

This is identical with the expression on the right side of Eq. (8.1). Hence

$$\lim_{n\to\infty}F(z,n)=F(z,\infty)\equiv\Gamma(z), \tag{8.15}$$

by Eq. (8.1), completing the proof.

Infinite Product (Weierstrass)

The third definition (Weierstrass' form) is

$$\boxed{\frac{1}{\Gamma(z)}\equiv ze^{\gamma z}\prod_{n=1}^{\infty}\left(1+\frac{z}{n}\right)e^{-z/n},} \tag{8.16}$$

[1] The form of $F(z,n)$ is suggested by the beta function (compare Eq. (8.60)).

where γ is the Euler–Mascheroni constant,

$$\gamma = 0.5772156619\ldots. \tag{8.17}$$

This infinite-product form may be used to develop the reflection identity, Eq. (8.23), and applied in the exercises, such as Exercise 8.1.17. This form can be derived from the original definition (Eq. (8.1)) by rewriting it as

$$\Gamma(z) = \lim_{n\to\infty} \frac{1\cdot 2\cdot 3\cdots n}{z(z+1)\cdots(z+n)} n^z = \lim_{n\to\infty} \frac{1}{z}\prod_{m=1}^{n}\left(1+\frac{z}{m}\right)^{-1} n^z. \tag{8.18}$$

Inverting Eq. (8.18) and using

$$n^{-z} = e^{(-\ln n)z}, \tag{8.19}$$

we obtain

$$\frac{1}{\Gamma(z)} = z\lim_{n\to\infty} e^{(-\ln n)z}\prod_{m=1}^{n}\left(1+\frac{z}{m}\right). \tag{8.20}$$

Multiplying and dividing by

$$\exp\left[\left(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}\right)z\right] = \prod_{m=1}^{n} e^{z/m}, \tag{8.21}$$

we get

$$\begin{aligned}\frac{1}{\Gamma(z)} = z&\left\{\lim_{n\to\infty}\exp\left[\left(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}-\ln n\right)z\right]\right\}\\ &\times\left[\lim_{n\to\infty}\prod_{m=1}^{n}\left(1+\frac{z}{m}\right)e^{-z/m}\right].\end{aligned} \tag{8.22}$$

As shown in Section 5.2, the parenthesis in the exponent approaches a limit, namely γ, the Euler–Mascheroni constant. Hence Eq. (8.16) follows.

It was shown in Section 5.11 that the Weierstrass infinite-product definition of $\Gamma(z)$ led directly to an important identity,

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin z\pi}. \tag{8.23}$$

Alternatively, we can start from the product of Euler integrals,

$$\begin{aligned}\Gamma(z+1)\Gamma(1-z) &= \int_0^\infty s^z e^{-s}\,ds\int_0^\infty t^{-z}e^{-t}\,dt\\ &= \int_0^\infty v^z\frac{dv}{(v+1)^2}\int_0^\infty e^{-u}u\,du = \frac{\pi z}{\sin\pi z},\end{aligned}$$

transforming from the variables s, t to $u = s+t$, $v = s/t$, as suggested by combining the exponentials and the powers in the integrands. The Jacobian is

$$J = -\begin{vmatrix} 1 & 1\\ \frac{1}{t} & -\frac{s}{t^2}\end{vmatrix} = \frac{s+t}{t^2} = \frac{(v+1)^2}{u},$$

where $(v+1)t = u$. The integral $\int_0^\infty e^{-u}u\,du = 1$, while that over v may be derived by contour integration, giving $\frac{\pi z}{\sin \pi z}$.

This identity may also be derived by contour integration (Example 7.1.6 and Exercises 7.1.18 and 7.1.19) and the beta function, Section 8.4. Setting $z = \frac{1}{2}$ in Eq. (8.23), we obtain

$$\Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi} \tag{8.24a}$$

(taking the positive square root), in agreement with Eq. (8.8).

Similarly one can establish **Legendre's duplication formula**,

$$\Gamma(1+z)\Gamma\left(z+\tfrac{1}{2}\right) = 2^{-2z}\sqrt{\pi}\,\Gamma(2z+1). \tag{8.24b}$$

The Weierstrass definition shows immediately that $\Gamma(z)$ has simple poles at $z = 0, -1, -2, -3, \ldots$ and that $[\Gamma(z)]^{-1}$ has no poles in the finite complex plane, which means that $\Gamma(z)$ has no zeros. This behavior may also be seen in Eq. (8.23), in which we note that $\pi/(\sin \pi z)$ is never equal to zero.

Actually the infinite-product definition of $\Gamma(z)$ may be derived from the Weierstrass factorization theorem with the specification that $[\Gamma(z)]^{-1}$ have simple zeros at $z = 0, -1, -2, -3, \ldots$. The Euler–Mascheroni constant is fixed by requiring $\Gamma(1) = 1$. See also the products expansions of entire functions in Section 7.1.

In probability theory the gamma distribution (probability density) is given by

$$f(x) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x > 0 \\ 0, & x \le 0. \end{cases} \tag{8.24c}$$

The constant $[\beta^\alpha \Gamma(\alpha)]^{-1}$ is chosen so that the total (integrated) probability will be unity. For $x \to E$, kinetic energy, $\alpha \to \frac{3}{2}$, and $\beta \to kT$, Eq. (8.24c) yields the classical Maxwell–Boltzmann statistics.

Factorial Notation

So far this discussion has been presented in terms of the classical notation. As pointed out by Jeffreys and others, the -1 of the $z-1$ exponent in our second definition (Eq. (8.5)) is a continual nuisance. Accordingly, Eq. (8.5) is sometimes rewritten as

$$\int_0^\infty e^{-t} t^z\,dt \equiv z!, \qquad \Re(z) > -1, \tag{8.25}$$

to **define** a factorial function $z!$. Occasionally we may still encounter Gauss' notation, $\prod(z)$, for the factorial function:

$$\prod(z) = z! = \Gamma(z+1). \tag{8.26}$$

The Γ notation is due to Legendre. The factorial function of Eq. (8.25) is related to the gamma function by

$$\Gamma(z) = (z-1)! \qquad \text{or} \qquad \Gamma(z+1) = z!. \tag{8.27}$$

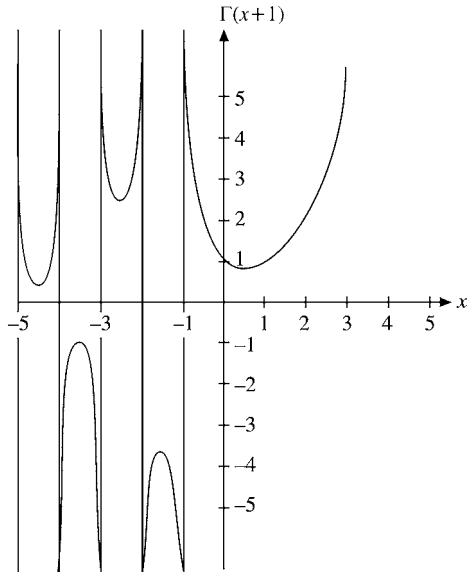

FIGURE 8.1 The factorial function — extension to negative arguments.

If $z = n$, a positive integer (Eq. (8.4)) shows that

$$z! = n! = 1 \cdot 2 \cdot 3 \cdots n, \tag{8.28}$$

the familiar factorial. However, it should be noted that since $z!$ is now defined by Eq. (8.25) (or equivalently by Eq. (8.27)) the factorial function is no longer limited to positive integral values of the argument (Fig. 8.1). The difference relation (Eq. (8.2)) becomes

$$(z-1)! = \frac{z!}{z}. \tag{8.29}$$

This shows immediately that

$$0! = 1 \tag{8.30}$$

and

$$n! = \pm\infty \qquad \text{for } n, \text{ a } \textbf{negative} \text{ integer.} \tag{8.31}$$

In terms of the factorial, Eq. (8.23) becomes

$$z!(-z)! = \frac{\pi z}{\sin \pi z}. \tag{8.32}$$

By restricting ourselves to the real values of the argument, we find that $\Gamma(x+1)$ defines the curves shown in Figs. 8.1 and 8.2. The minimum of the curve is

$$\Gamma(x+1) = x! = (0.46163\ldots)! = 0.88560\ldots. \tag{8.33a}$$

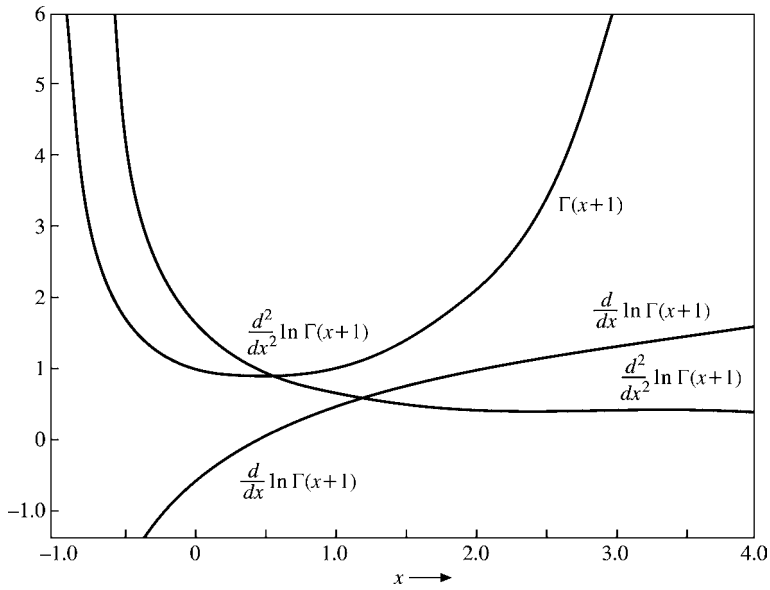

FIGURE 8.2 The factorial function and the first two derivatives of $\ln(\Gamma(x+1))$.

Double Factorial Notation

In many problems of mathematical physics, particularly in connection with Legendre polynomials (Chapter 12), we encounter products of the odd positive integers and products of the even positive integers. For convenience these are given special labels as double factorials:

$$\begin{aligned} 1\cdot 3\cdot 5\cdots(2n+1) &= (2n+1)!! \\ 2\cdot 4\cdot 6\cdots(2n) &= (2n)!!. \end{aligned} \tag{8.33b}$$

Clearly, these are related to the regular factorial functions by

$$(2n)!! = 2^n n! \qquad \text{and} \qquad (2n+1)!! = \frac{(2n+1)!}{2^n n!}. \tag{8.33c}$$

We also define $(-1)!! = 1$, a special case that does not follow from Eq. (8.33c).

Integral Representation

An integral representation that is useful in developing asymptotic series for the Bessel functions is

$$\int_C e^{-z} z^{\nu}\, dz = \left(e^{2\pi i\nu} - 1\right)\Gamma(\nu+1), \tag{8.34}$$

where C is the contour shown in Fig. 8.3. This contour integral representation is only useful when ν is not an integer, $z = 0$ then being a **branch point**. Equation (8.34) may be

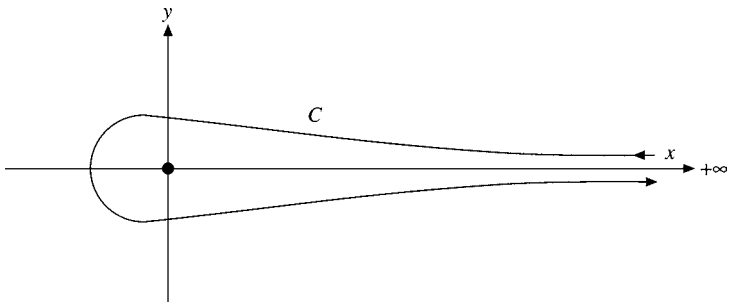

FIGURE 8.3 Factorial function contour.

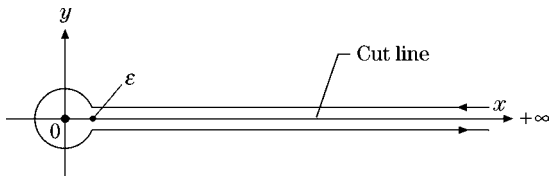

FIGURE 8.4 The contour of Fig. 8.3 deformed.

readily verified for $\nu > -1$ by deforming the contour as shown in Fig. 8.4. The integral from ∞ into the origin yields $-(\nu!)$, placing the phase of z at 0. The integral out to ∞ (in the fourth quadrant) then yields $e^{2\pi i\nu}\nu!$, the phase of z having increased to 2π. Since the circle around the origin contributes nothing when $\nu > -1$, Eq. (8.34) follows.

It is often convenient to cast this result into a more symmetrical form:

$$\int_C e^{-z}(-z)^{\nu}\,dz = 2i\Gamma(\nu+1)\sin(\nu\pi). \tag{8.35}$$

This analysis establishes Eqs. (8.34) and (8.35) for $\nu > -1$. It is relatively simple to extend the range to include all nonintegral ν. First, we note that the integral exists for $\nu < -1$ as long as we stay away from the origin. Second, integrating by parts we find that Eq. (8.35) yields the familiar difference relation (Eq. (8.29)). If we take the difference relation to define the factorial function of $\nu < -1$, then Eqs. (8.34) and (8.35) are verified for all ν (except negative integers).

Exercises

8.1.1 Derive the recurrence relations

$$\Gamma(z+1) = z\Gamma(z)$$

from the Euler integral (Eq. (8.5)),

$$\Gamma(z) = \int_0^{\infty} e^{-t}t^{z-1}\,dt.$$

8.1.2 In a power-series solution for the Legendre functions of the second kind we encounter the expression

$$\frac{(n+1)(n+2)(n+3)\cdots(n+2s-1)(n+2s)}{2\cdot 4\cdot 6\cdot 8\cdots(2s-2)(2s)\cdot(2n+3)(2n+5)(2n+7)\cdots(2n+2s+1)},$$

in which s is a positive integer. Rewrite this expression in terms of factorials.

8.1.3 Show that, as $s-n\to$ negative integer,

$$\frac{(s-n)!}{(2s-2n)!}\to\frac{(-1)^{n-s}(2n-2s)!}{(n-s)!}.$$

Here s and n are integers with $s<n$. This result can be used to avoid negative factorials, such as in the series representations of the spherical Neumann functions and the Legendre functions of the second kind.

8.1.4 Show that $\Gamma(z)$ may be written

$$\Gamma(z)=2\int_0^\infty e^{-t^2}t^{2z-1}\,dt,\qquad \Re(z)>0,$$

$$\Gamma(z)=\int_0^1\left[\ln\left(\frac{1}{t}\right)\right]^{z-1}dt,\qquad \Re(z)>0.$$

8.1.5 In a Maxwellian distribution the fraction of particles with speed between v and $v+dv$ is

$$\frac{dN}{N}=4\pi\left(\frac{m}{2\pi kT}\right)^{3/2}\exp\left(-\frac{mv^2}{2kT}\right)v^2\,dv,$$

N being the total number of particles. The average or expectation value of v^n is defined as $\langle v^n\rangle=N^{-1}\int v^n\,dN$. Show that

$$\langle v^n\rangle=\left(\frac{2kT}{m}\right)^{n/2}\frac{\Gamma\left(\frac{n+3}{2}\right)}{\Gamma(3/2)}.$$

8.1.6 By transforming the integral into a gamma function, show that

$$-\int_0^1 x^k\ln x\,dx=\frac{1}{(k+1)^2},\qquad k>-1.$$

8.1.7 Show that

$$\int_0^\infty e^{-x^4}\,dx=\Gamma\left(\frac{5}{4}\right).$$

8.1.8 Show that

$$\lim_{x\to 0}\frac{(ax-1)!}{(x-1)!}=\frac{1}{a}.$$

8.1.9 Locate the poles of $\Gamma(z)$. Show that they are simple poles and determine the residues.

8.1.10 Show that the equation $x!=k$, $k\neq 0$, has an infinite number of real roots.

8.1.11 Show that

(a) $$\int_0^\infty x^{2s+1}\exp\left(-ax^2\right)dx = \frac{s!}{2a^{s+1}}.$$

(b) $$\int_0^\infty x^{2s}\exp\left(-ax^2\right)dx = \frac{(s-\frac{1}{2})!}{2a^{s+1/2}} = \frac{(2s-1)!!}{2^{s+1}a^s}\sqrt{\frac{\pi}{a}}.$$

These Gaussian integrals are of major importance in statistical mechanics.

8.1.12 (a) Develop recurrence relations for $(2n)!!$ and for $(2n+1)!!$.
(b) Use these recurrence relations to calculate (or to define) $0!!$ and $(-1)!!$.

ANS. $0!! = 1, \quad (-1)!! = 1.$

8.1.13 For s a nonnegative integer, show that

$$(-2s-1)!! = \frac{(-1)^s}{(2s-1)!!} = \frac{(-1)^s 2^s s!}{(2s)!}.$$

8.1.14 Express the coefficient of the nth term of the expansion of $(1+x)^{1/2}$

(a) in terms of factorials of integers,
(b) in terms of the double factorial (!!) functions.

$$\text{ANS. } a_n = (-1)^{n+1}\frac{(2n-3)!}{2^{2n-2}n!(n-2)!} = (-1)^{n+1}\frac{(2n-3)!!}{(2n)!!}, \qquad n = 2, 3, \ldots.$$

8.1.15 Express the coefficient of the nth term of the expansion of $(1+x)^{-1/2}$

(a) in terms of the factorials of integers,
(b) in terms of the double factorial (!!) functions.

$$\text{ANS. } a_n = (-1)^n\frac{(2n)!}{2^{2n}(n!)^2} = (-1)^n\frac{(2n-1)!!}{(2n)!!}, \qquad n = 1, 2, 3, \ldots.$$

8.1.16 The Legendre polynomial may be written as

$$\begin{aligned} P_n(\cos\theta) = 2\frac{(2n-1)!!}{(2n)!!}\Bigg\{&\cos n\theta + \frac{1}{1}\cdot\frac{n}{2n-1}\cos(n-2)\theta \\ &+ \frac{1\cdot 3}{1\cdot 2}\frac{n(n-1)}{(2n-1)(2n-3)}\cos(n-4)\theta \\ &+ \frac{1\cdot 3\cdot 5}{1\cdot 2\cdot 3}\frac{n(n-1)(n-2)}{(2n-1)(2n-3)(2n-5)}\cos(n-6)\theta + \cdots\Bigg\}. \end{aligned}$$

Let $n = 2s+1$. Then

$$P_n(\cos\theta) = P_{2s+1}(\cos\theta) = \sum_{m=0}^{s} a_m\cos(2m+1)\theta.$$

Find a_m in terms of factorials and double factorials.

8.1.17 (a) Show that

$$\Gamma\left(\tfrac{1}{2}-n\right)\Gamma\left(\tfrac{1}{2}+n\right)=(-1)^n\pi,$$

where n is an integer.

(b) Express $\Gamma(\frac{1}{2}+n)$ and $\Gamma(\frac{1}{2}-n)$ separately in terms of $\pi^{1/2}$ and a !! function.

$$\text{ANS. } \Gamma(\tfrac{1}{2}+n)=\frac{(2n-1)!!}{2^n}\pi^{1/2}.$$

8.1.18 From one of the definitions of the factorial or gamma function, show that

$$|(ix)!|^2=\frac{\pi x}{\sinh \pi x}\prod_{n=0}^{\infty}\left[1+\frac{x^2}{(n+1)^2}\right].$$

8.1.19 Prove that

$$\left|\Gamma(\alpha+i\beta)\right|=\left|\Gamma(\alpha)\right|\prod_{n=0}^{\infty}\left[1+\frac{\beta^2}{(\alpha+n)^2}\right]^{-1/2}.$$

This equation has been useful in calculations of beta decay theory.

8.1.20 Show that

$$\left|(n+ib)!\right|=\left(\frac{\pi b}{\sinh \pi b}\right)^{1/2}\prod_{s=1}^{n}\left(s^2+b^2\right)^{1/2}$$

for n, a positive integer.

8.1.21 Show that

$$|x!|\geq\left|(x+iy)!\right|$$

for all x. The variables x and y are real.

8.1.22 Show that

$$\left|\Gamma\left(\frac{1}{2}+iy\right)\right|^2=\frac{\pi}{\cosh \pi y}.$$

8.1.23 The probability density associated with the normal distribution of statistics is given by

$$f(x)=\frac{1}{\sigma(2\pi)^{1/2}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],$$

with $(-\infty,\infty)$ for the range of x. Show that

(a) the mean value of x, $\langle x\rangle$ is equal to μ,

(b) the standard deviation $(\langle x^2\rangle-\langle x\rangle^2)^{1/2}$ is given by σ.

8.1.24 From the gamma distribution

$$f(x)=\begin{cases}\dfrac{1}{\beta^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}, & x>0,\\ 0, & x\leq 0,\end{cases}$$

show that

(a) $\langle x\rangle$ (mean) $=\alpha\beta$, (b) σ^2 (variance) $\equiv\langle x^2\rangle-\langle x\rangle^2=\alpha\beta^2$.

8.1.25 The wave function of a particle scattered by a Coulomb potential is $\psi(r,\theta)$. At the origin the wave function becomes

$$\psi(0) = e^{-\pi\gamma/2}\Gamma(1+i\gamma),$$

where $\gamma = Z_1 Z_2 e^2/\hbar v$. Show that

$$\left|\psi(0)\right|^2 = \frac{2\pi\gamma}{e^{2\pi\gamma}-1}.$$

8.1.26 Derive the contour integral representation of Eq. (8.34),

$$2i\nu!\sin\nu\pi = \int_C e^{-z}(-z)^\nu\,dz.$$

8.1.27 Write a function subprogram *FACT*(N) (fixed-point independent variable) that will calculate $N!$. Include provision for rejection and appropriate error message if N is negative.
Note. For small integer N, direct multiplication is simplest. For large N, Eq. (8.55), Stirling's series would be appropriate.

8.1.28 (a) Write a function subprogram to calculate the double factorial ratio $(2N-1)!!/(2N)!!$. Include provision for $N=0$ and for rejection and an error message if N is negative. Calculate and tabulate this ratio for $N = 1(1)100$.
(b) Check your function subprogram calculation of $199!!/200!!$ against the value obtained from Stirling's series (Section 8.3).

$$\text{ANS. } \frac{199!!}{200!!} = 0.056348.$$

8.1.29 Using either the FORTRAN-supplied GAMMA or a library-supplied subroutine for $x!$ or $\Gamma(x)$, determine the value of x for which $\Gamma(x)$ is a minimum $(1 \le x \le 2)$ and this minimum value of $\Gamma(x)$. Notice that although the minimum value of $\Gamma(x)$ may be obtained to about six significant figures (single precision), the corresponding value of x is much less accurate. Why this relatively low accuracy?

8.1.30 The factorial function expressed in integral form can be evaluated by the Gauss–Laguerre quadrature. For a 10-point formula the resultant $x!$ is theoretically exact for x an integer, 0 up through 19. What happens if x is not an integer? Use the Gauss–Laguerre quadrature to evaluate $x!$, $x = 0.0(0.1)2.0$. Tabulate the absolute error as a function of x.

Check value. $x!_{\text{exact}} - x!_{\text{quadrature}} = 0.00034 \quad \text{for} \quad x = 1.3.$

8.2 DIGAMMA AND POLYGAMMA FUNCTIONS

Digamma Functions

As may be noted from the three definitions in Section 8.1, it is inconvenient to deal with the derivatives of the gamma or factorial function directly. Instead, it is customary to take

the natural logarithm of the factorial function (Eq. (8.1)), convert the product to a sum, and then differentiate; that is,

$$\Gamma(z+1) = z\Gamma(z) = \lim_{n\to\infty} \frac{n!}{(z+1)(z+2)\cdots(z+n)} n^z \tag{8.36}$$

and

$$\begin{aligned}\ln\Gamma(z+1) = \lim_{n\to\infty} \big[&\ln(n!) + z\ln n - \ln(z+1) \\ &- \ln(z+2) - \cdots - \ln(z+n)\big],\end{aligned} \tag{8.37}$$

in which the logarithm of the limit is equal to the limit of the logarithm. Differentiating with respect to z, we obtain

$$\frac{d}{dz}\ln\Gamma(z+1) \equiv \psi(z+1) = \lim_{n\to\infty}\left(\ln n - \frac{1}{z+1} - \frac{1}{z+2} - \cdots - \frac{1}{z+n}\right), \tag{8.38}$$

which defines $\psi(z+1)$, the digamma function. From the definition of the Euler–Mascheroni constant,[2] Eq. (8.38) may be rewritten as

$$\begin{aligned}\psi(z+1) &= -\gamma - \sum_{n=1}^{\infty}\left(\frac{1}{z+n} - \frac{1}{n}\right) \\ &= -\gamma + \sum_{n=1}^{\infty} \frac{z}{n(n+z)}.\end{aligned} \tag{8.39}$$

One application of Eq. (8.39) is in the derivation of the series form of the Neumann function (Section 11.3). Clearly,

$$\psi(1) = -\gamma = -0.577\,215\,664\,901\ldots.^{3} \tag{8.40}$$

Another, perhaps more useful, expression for $\psi(z)$ is derived in Section 8.3.

Polygamma Function

The digamma function may be differentiated repeatedly, giving rise to the polygamma function:

$$\begin{aligned}\psi^{(m)}(z+1) &\equiv \frac{d^{m+1}}{dz^{m+1}}\ln(z!) \\ &= (-1)^{m+1} m! \sum_{n=1}^{\infty} \frac{1}{(z+n)^{m+1}}, \qquad m = 1, 2, 3, \ldots.\end{aligned} \tag{8.41}$$

[2] Compare Sections 5.2 and 5.9. We add and substract $\sum_{s=1}^{n} s^{-1}$.

[3] γ has been computed to 1271 places by D. E. Knuth, *Math. Comput.* **16**: 275 (1962), and to 3566 decimal places by D. W. Sweeney, *ibid.* **17**: 170 (1963). It may be of interest that the fraction 228/395 gives γ accurate to six places.

A plot of $\psi(x+1)$ and $\psi'(x+1)$ is included in Fig. 8.2. Since the series in Eq. (8.41) defines the Riemann zeta function[4] (with $z=0$),

$$\zeta(m) \equiv \sum_{n=1}^{\infty} \frac{1}{n^m}, \tag{8.42}$$

we have

$$\psi^{(m)}(1) = (-1)^{m+1} m! \zeta(m+1), \qquad m = 1, 2, 3, \ldots. \tag{8.43}$$

The values of the polygamma functions of positive integral argument, $\psi^{(m)}(n+1)$, may be calculated by using Exercise 8.2.6.

In terms of the perhaps more common Γ notation,

$$\frac{d^{n+1}}{dz^{n+1}} \ln \Gamma(z) = \frac{d^n}{dz^n} \psi(z) = \psi^{(n)}(z). \tag{8.44a}$$

Maclaurin Expansion, Computation

It is now possible to write a Maclaurin expansion for $\ln \Gamma(z+1)$:

$$\ln \Gamma(z+1) = \sum_{n=1}^{\infty} \frac{z^n}{n!} \psi^{(n-1)}(1) = -\gamma z + \sum_{n=2}^{\infty} (-1)^n \frac{z^n}{n} \zeta(n) \tag{8.44b}$$

convergent for $|z| < 1$; for $z = x$, the range is $-1 < x \leq 1$. Alternate forms of this series appear in Exercise 5.9.14. Equation (8.44b) is a possible means of computing $\Gamma(z+1)$ for real or complex z, but Stirling's series (Section 8.3) is usually better, and in addition, an excellent table of values of the gamma function for complex arguments based on the use of Stirling's series and the recurrence relation (Eq. (8.29)) is now available.[5]

Series Summation

The digamma and polygamma functions may also be used in summing series. If the general term of the series has the form of a rational fraction (with the highest power of the index in the numerator at least two less than the highest power of the index in the denominator), it may be transformed by the method of partial fractions (compare Section 15.8). The infinite series may then be expressed as a finite sum of digamma and polygamma functions. The usefulness of this method depends on the availability of tables of digamma and polygamma functions. Such tables and examples of series summation are given in AMS-55, Chapter 6 (see Additional Readings for the reference).

[4] See Section 5.9. For $z \neq 0$ this series may be used to define a generalized zeta function.

[5] *Table of the Gamma Function for Complex Arguments*, Applied Mathematics Series No. 34. Washington, DC: National Bureau of Standards (1954).

Example 8.2.1 CATALAN'S CONSTANT

Catalan's constant, Exercise 5.2.22, or $\beta(2)$ of Section 5.9 is given by

$$K = \beta(2) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)^2}. \tag{8.44c}$$

Grouping the positive and negative terms separately and starting with unit index (to match the form of $\psi^{(1)}$, Eq. (8.41)), we obtain

$$K = 1 + \sum_{n=1}^{\infty} \frac{1}{(4n+1)^2} - \frac{1}{9} - \sum_{n=1}^{\infty} \frac{1}{(4n+3)^2}.$$

Now, quoting Eq. (8.41), we get

$$K = \tfrac{8}{9} + \tfrac{1}{16}\psi^{(1)}\left(1+\tfrac{1}{4}\right) - \tfrac{1}{16}\psi^{(1)}\left(1+\tfrac{3}{4}\right). \tag{8.44d}$$

Using the values of $\psi^{(1)}$ from Table 6.1 of AMS-55 (see Additional Readings for the reference), we obtain

$$K = 0.91596559\ldots.$$

Compare this calculation of Catalan's constant with the calculations of Chapter 5, either direct summation or a modification using Riemann zeta function values. ■

Exercises

8.2.1 Verify that the following two forms of the digamma function,

$$\psi(x+1) = \sum_{r=1}^{x} \frac{1}{r} - \gamma$$

and

$$\psi(x+1) = \sum_{r=1}^{\infty} \frac{x}{r(r+x)} - \gamma,$$

are equal to each other (for x a positive integer).

8.2.2 Show that $\psi(z+1)$ has the series expansion

$$\psi(z+1) = -\gamma + \sum_{n=2}^{\infty} (-1)^n \zeta(n) z^{n-1}.$$

8.2.3 For a power-series expansion of $\ln(z!)$, AMS-55 (see Additional Readings for reference) lists

$$\ln(z!) = -\ln(1+z) + z(1-\gamma) + \sum_{n=2}^{\infty} (-1)^n \frac{[\zeta(n)-1]z^n}{n}.$$

(a) Show that this agrees with Eq. (8.44b) for $|z| < 1$.
(b) What is the range of convergence of this new expression?

8.2.4 Show that

$$\frac{1}{2}\ln\left(\frac{\pi z}{\sin \pi z}\right) = \sum_{n=1}^{\infty} \frac{\zeta(2n)}{2n} z^{2n}, \qquad |z| < 1.$$

Hint. Try Eq. (8.32).

8.2.5 Write out a Weierstrass infinite-product definition of $\ln(z!)$. Without differentiating, show that this leads directly to the Maclaurin expansion of $\ln(z!)$, Eq. (8.44b).

8.2.6 Derive the difference relation for the polygamma function

$$\psi^{(m)}(z+2) = \psi^{(m)}(z+1) + (-1)^m \frac{m!}{(z+1)^{m+1}}, \qquad m = 0, 1, 2, \ldots.$$

8.2.7 Show that if

$$\Gamma(x + iy) = u + iv,$$

then

$$\Gamma(x - iy) = u - iv.$$

This is a special case of the Schwarz reflection principle, Section 6.5.

8.2.8 The Pochhammer symbol $(a)_n$ is defined as

$$(a)_n = a(a+1)\cdots(a+n-1), \qquad (a)_0 = 1$$

(for integral n).

(a) Express $(a)_n$ in terms of factorials.
(b) Find $(d/da)(a)_n$ in terms of $(a)_n$ and digamma functions.

$$\text{ANS. } \frac{d}{da}(a)_n = (a)_n\big[\psi(a+n) - \psi(a)\big].$$

(c) Show that

$$(a)_{n+k} = (a+n)_k \cdot (a)_n.$$

8.2.9 Verify the following special values of the ψ form of the di- and polygamma functions:

$$\psi(1) = -\gamma, \qquad \psi^{(1)}(1) = \zeta(2), \qquad \psi^{(2)}(1) = -2\zeta(3).$$

8.2.10 Derive the polygamma function recurrence relation

$$\psi^{(m)}(1+z) = \psi^{(m)}(z) + (-1)^m m!/z^{m+1}, \qquad m = 0, 1, 2, \ldots.$$

8.2.11 Verify

(a) $\displaystyle\int_0^\infty e^{-r} \ln r \, dr = -\gamma.$

(b) $\displaystyle\int_0^\infty re^{-r}\ln r\,dr = 1-\gamma.$

(c) $\displaystyle\int_0^\infty r^n e^{-r}\ln r\,dr = (n-1)! + n\int_0^\infty r^{n-1}e^{-r}\ln r\,dr, \qquad n = 1, 2, 3, \ldots.$

Hint. These may be verified by integration by parts, three parts, or differentiating the integral form of $n!$ with respect to n.

8.2.12 Dirac relativistic wave functions for hydrogen involve factors such as $[2(1-\alpha^2Z^2)^{1/2}]!$ where α, the fine structure constant, is $\frac{1}{137}$ and Z is the atomic number. Expand $[2(1-\alpha^2Z^2)^{1/2}]!$ in a series of powers of α^2Z^2.

8.2.13 The quantum mechanical description of a particle in a Coulomb field requires a knowledge of the phase of the complex factorial function. Determine the phase of $(1+ib)!$ for small b.

8.2.14 The total energy radiated by a blackbody is given by

$$u = \frac{8\pi k^4T^4}{c^3h^3}\int_0^\infty \frac{x^3}{e^x-1}dx.$$

Show that the integral in this expression is equal to $3!\zeta(4)$.
$[\zeta(4) = \pi^4/90 = 1.0823\ldots]$ The final result is the Stefan–Boltzmann law.

8.2.15 As a generalization of the result in Exercise 8.2.14, show that

$$\int_0^\infty \frac{x^s\,dx}{e^x-1} = s!\zeta(s+1), \qquad \Re(s) > 0.$$

8.2.16 The neutrino energy density (Fermi distribution) in the early history of the universe is given by

$$\rho_\nu = \frac{4\pi}{h^3}\int_0^\infty \frac{x^3}{\exp(x/kT)+1}dx.$$

Show that

$$\rho_\nu = \frac{7\pi^5}{30h^3}(kT)^4.$$

8.2.17 Prove that

$$\int_0^\infty \frac{x^s\,dx}{e^x+1} = s!\left(1-2^{-s}\right)\zeta(s+1), \qquad \Re(s) > 0.$$

Exercises 8.2.15 and 8.2.17 actually constitute Mellin integral transforms (compare Section 15.1).

8.2.18 Prove that

$$\psi^{(n)}(z) = (-1)^{n+1}\int_0^\infty \frac{t^ne^{-zt}}{1-e^{-t}}dt, \qquad \Re(z) > 0.$$

8.2.19 Using di- and polygamma functions, sum the series

$$\text{(a)} \sum_{n=1}^{\infty} \frac{1}{n(n+1)}, \qquad \text{(b)} \sum_{n=2}^{\infty} \frac{1}{n^2-1}.$$

Note. You can use Exercise 8.2.6 to calculate the needed digamma functions.

8.2.20 Show that

$$\sum_{n=1}^{\infty} \frac{1}{(n+a)(n+b)} = \frac{1}{(b-a)}\big\{\psi(1+b) - \psi(1+a)\big\},$$

where $a \neq b$ and neither a nor b is a negative integer. It is of some interest to compare this summation with the corresponding integral,

$$\int_1^{\infty} \frac{dx}{(x+a)(x+b)} = \frac{1}{b-a}\big\{\ln(1+b) - \ln(1+a)\big\}.$$

The relation between $\psi(x)$ and $\ln x$ is made explicit in Eq. (8.51) in the next section.

8.2.21 Verify the contour integral representation of $\zeta(s)$,

$$\zeta(s) = -\frac{(-s)!}{2\pi i} \int_C \frac{(-z)^{s-1}}{e^z - 1} dz.$$

The contour C is the same as that for Eq. (8.35). The points $z = \pm 2n\pi i,\ n = 1, 2, 3, \ldots,$ are all excluded.

8.2.22 Show that $\zeta(s)$ is analytic in the entire finite complex plane except at $s = 1$, where it has a simple pole with a residue of $+1$.
Hint. The contour integral representation will be useful.

8.2.23 Using the complex variable capability of FORTRAN calculate $\Re(1+ib)!$, $\Im(1+ib)!$, $|(1+ib)!|$ and phase $(1+ib)!$ for $b = 0.0(0.1)1.0$. Plot the phase of $(1+ib)!$ versus b.
Hint. Exercise 8.2.3 offers a convenient approach. You will need to calculate $\zeta(n)$.

8.3 STIRLING'S SERIES

For computation of $\ln(z!)$ for very large z (statistical mechanics) and for numerical computations at nonintegral values of z, a series expansion of $\ln(z!)$ in negative powers of z is desirable. Perhaps the most elegant way of deriving such an expansion is by the method of steepest descents (Section 7.3). The following method, starting with a numerical integration formula, does not require knowledge of contour integration and is particularly direct.

Derivation from Euler–Maclaurin Integration Formula

The Euler–Maclaurin formula for evaluating a definite integral[6] is

$$\int_0^n f(x)\,dx = \tfrac{1}{2}f(0) + f(1) + f(2) + \cdots + \tfrac{1}{2}f(n) - b_2\big[f'(n) - f'(0)\big] - b_4\big[f'''(n) - f'''(0)\big] - \cdots, \tag{8.45}$$

in which the b_{2n} are related to the Bernoulli numbers B_{2n} (compare Section 5.9) by

$$(2n)!b_{2n} = B_{2n}, \tag{8.46}$$

$$\begin{aligned} B_0 &= 1, & B_6 &= \tfrac{1}{42}, \\ B_2 &= \tfrac{1}{6}, & B_8 &= -\tfrac{1}{30}, \\ B_4 &= -\tfrac{1}{30}, & B_{10} &= \tfrac{5}{66}, \qquad \text{and so on.} \end{aligned} \tag{8.47}$$

By applying Eq. (8.45) to the definite integral

$$\int_0^\infty \frac{dx}{(z+x)^2} = \frac{1}{z} \tag{8.48}$$

(for z not on the negative real axis), we obtain

$$\frac{1}{z} = \frac{1}{2z^2} + \psi^{(1)}(z+1) - \frac{2!b_2}{z^3} - \frac{4!b_4}{z^5} - \cdots. \tag{8.49}$$

This is the reason for using Eq. (8.48). The Euler–Maclaurin evaluation yields $\psi^{(1)}(z+1)$, which is $d^2 \ln \Gamma(z+1)/dz^2$.

Using Eq. (8.46) and solving for $\psi^{(1)}(z+1)$, we have

$$\begin{aligned} \psi^{(1)}(z+1) = \frac{d}{dz}\psi(z+1) &= \frac{1}{z} - \frac{1}{2z^2} + \frac{B_2}{z^3} + \frac{B_4}{z^5} + \cdots \\ &= \frac{1}{z} - \frac{1}{2z^2} + \sum_{n=1}^{\infty} \frac{B_{2n}}{z^{2n+1}}. \end{aligned} \tag{8.50}$$

Since the Bernoulli numbers diverge strongly, this series does not converge. It is a semi-convergent, or asymptotic, series, useful if one retains a small enough number of terms (compare Section 5.10).

Integrating once, we get the digamma function

$$\begin{aligned} \psi(z+1) &= C_1 + \ln z + \frac{1}{2z} - \frac{B_2}{2z^2} - \frac{B_4}{4z^4} - \cdots \\ &= C_1 + \ln z + \frac{1}{2z} - \sum_{n=1}^{\infty} \frac{B_{2n}}{2nz^{2n}}. \end{aligned} \tag{8.51}$$

Integrating Eq. (8.51) with respect to z from $z-1$ to z and then letting z approach infinity, C_1, the constant of integration, may be shown to vanish. This gives us a second expression for the digamma function, often more useful than Eq. (8.38) or (8.44b).

[6] This is obtained by repeated integration by parts, Section 5.9.

Stirling's Series

The indefinite integral of the digamma function (Eq. (8.51)) is

$$\ln\Gamma(z+1) = C_2 + \left(z+\frac{1}{2}\right)\ln z - z + \frac{B_2}{2z} + \cdots + \frac{B_{2n}}{2n(2n-1)z^{2n-1}} + \cdots, \tag{8.52}$$

in which C_2 is another constant of integration. To fix C_2 we find it convenient to use the doubling, or Legendre duplication, formula derived in Section 8.4,

$$\Gamma(z+1)\Gamma\left(z+\tfrac{1}{2}\right) = 2^{-2z}\pi^{1/2}\Gamma(2z+1). \tag{8.53}$$

This may be proved directly when z is a positive integer by writing $\Gamma(2z+1)$ as a product of even terms times a product of odd terms and extracting a factor of 2 from each term (Exercise 8.3.5). Substituting Eq. (8.52) into the logarithm of the doubling formula, we find that C_2 is

$$C_2 = \tfrac{1}{2}\ln 2\pi, \tag{8.54}$$

giving

$$\boxed{\ln\Gamma(z+1) = \frac{1}{2}\ln 2\pi + \left(z+\frac{1}{2}\right)\ln z - z + \frac{1}{12z} - \frac{1}{360z^3} + \frac{1}{1260z^5} - \cdots.} \tag{8.55}$$

This is Stirling's series, an asymptotic expansion. The absolute value of the error is less than the absolute value of the first term omitted.

The constants of integration C_1 and C_2 may also be evaluated by comparison with the first term of the series expansion obtained by the method of "steepest descent." This is carried out in Section 7.3.

To help convey a feeling of the remarkable precision of Stirling's series for $\Gamma(s+1)$, the ratio of the first term of Stirling's approximation to $\Gamma(s+1)$ is plotted in Fig. 8.5. A tabulation gives the ratio of the first term in the expansion to $\Gamma(s+1)$ and the ratio of the first two terms in the expansion to $\Gamma(s+1)$ (Table 8.1). The derivation of these forms is Exercise 8.3.1.

Exercises

8.3.1 Rewrite Stirling's series to give $\Gamma(z+1)$ instead of $\ln\Gamma(z+1)$.

$$\text{ANS. } \Gamma(z+1) = \sqrt{2\pi}z^{z+1/2}e^{-z}\left(1 + \frac{1}{12z} + \frac{1}{288z^2} - \frac{139}{51{,}840z^3} + \cdots\right).$$

8.3.2 Use Stirling's formula to estimate 52!, the number of possible rearrangements of cards in a standard deck of playing cards.

8.3.3 By integrating Eq. (8.51) from $z-1$ to z and then letting $z \to \infty$, evaluate the constant C_1 in the asymptotic series for the digamma function $\psi(z)$.

8.3.4 Show that the constant C_2 in Stirling's formula equals $\frac{1}{2}\ln 2\pi$ by using the logarithm of the doubling formula.

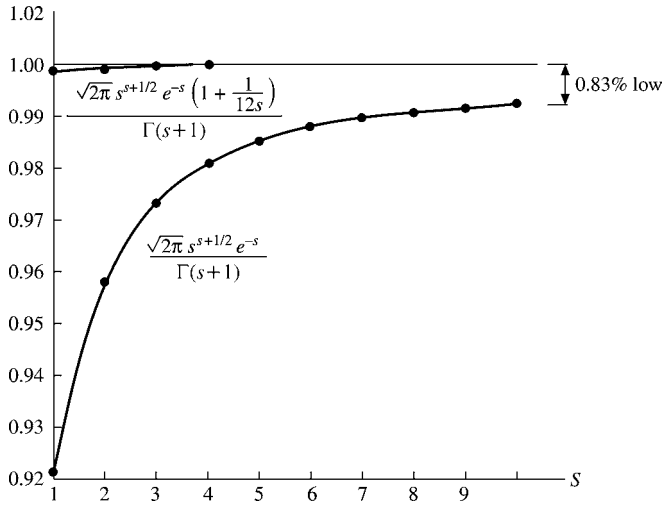

FIGURE 8.5 Accuracy of Stirling's formula.

Table 8.1

s	$\frac{1}{\Gamma(s+1)}\sqrt{2\pi}s^{s+1/2}e^{-s}$	$\frac{1}{\Gamma(s+1)}\sqrt{2\pi}s^{s+1/2}e^{-s}\left(1+\frac{1}{12s}\right)$
1	0.92213	0.99898
2	0.95950	0.99949
3	0.97270	0.99972
4	0.97942	0.99983
5	0.98349	0.99988
6	0.98621	0.99992
7	0.98817	0.99994
8	0.98964	0.99995
9	0.99078	0.99996
10	0.99170	0.99998

8.3.5 By direct expansion, verify the doubling formula for $z = n + \frac{1}{2}$; n is an integer.

8.3.6 Without using Stirling's series show that

$$\text{(a) } \ln(n!) < \int_1^{n+1} \ln x\, dx, \qquad \text{(b) } \ln(n!) > \int_1^{n} \ln x\, dx;\ n \text{ is an integer} \geq 2.$$

Notice that the arithmetic mean of these two integrals gives a good approximation for Stirling's series.

8.3.7 Test for convergence

$$\sum_{p=0}^{\infty}\left[\frac{(p-\frac{1}{2})!}{p!}\right]^2 \times \frac{2p+1}{2p+2} = \pi \sum_{p=0}^{\infty} \frac{(2p-1)!!(2p+1)!!}{(2p)!!(2p+2)!!}.$$

This series arises in an attempt to describe the magnetic field created by and enclosed by a current loop.

8.3.8 Show that

$$\lim_{x\to\infty} x^{b-a}\frac{(x+a)!}{(x+b)!} = 1.$$

8.3.9 Show that

$$\lim_{n\to\infty} \frac{(2n-1)!!}{(2n)!!} n^{1/2} = \pi^{-1/2}.$$

8.3.10 Calculate the binomial coefficient $\binom{2n}{n}$ to six significant figures for $n = 10$, 20, and 30. Check your values by

(a) a Stirling series approximation through terms in n^{-1},
(b) a double precision calculation.

ANS. $\binom{20}{10} = 1.84756 \times 10^5$, $\binom{40}{20} = 1.37846 \times 10^{11}$,
$\binom{60}{30} = 1.18264 \times 10^{17}$.

8.3.11 Write a program (or subprogram) that will calculate $\log_{10}(x!)$ directly from Stirling's series. Assume that $x \geq 10$. (Smaller values could be calculated via the factorial recurrence relation.) Tabulate $\log_{10}(x!)$ versus x for $x = 10(10)300$. Check your results against AMS-55 (see Additional Readings for this reference) or by direct multiplication (for $n = 10$, 20, and 30).

Check value. $\log_{10}(100!) = 157.97$.

8.3.12 Using the complex arithmetic capability of FORTRAN, write a subroutine that will calculate $\ln(z!)$ for complex z based on Stirling's series. Include a test and an appropriate error message if z is too close to a negative real integer. Check your subroutine against alternate calculations for z real, z pure imaginary, and $z = 1 + ib$ (Exercise 8.2.23).

Check values. $|(i0.5)!| = 0.82618$
phase $(i0.5)! = -0.24406$.

8.4 THE BETA FUNCTION

Using the integral definition (Eq. (8.25)), we write the product of two factorials as the product of two integrals. To facilitate a change in variables, we take the integrals over a finite range:

$$m!n! = \lim_{a^2\to\infty} \int_0^{a^2} e^{-u}u^m\,du \int_0^{a^2} e^{-v}v^n\,dv, \qquad \begin{matrix}\Re(m) > -1,\\ \Re(n) > -1.\end{matrix} \tag{8.56a}$$

Replacing u with x^2 and v with y^2, we obtain

$$m!n! = \lim_{a\to\infty} 4\int_0^{a} e^{-x^2}x^{2m+1}\,dx \int_0^{a} e^{-y^2}y^{2n+1}\,dy. \tag{8.56b}$$

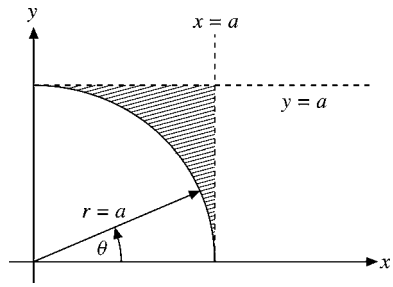

FIGURE 8.6 Transformation from Cartesian to polar coordinates.

Transforming to polar coordinates gives us

$$m!n! = \lim_{a\to\infty} 4\int_0^a e^{-r^2} r^{2m+2n+3}\,dr \int_0^{\pi/2} \cos^{2m+1}\theta \sin^{2n+1}\theta\,d\theta$$

$$= (m+n+1)!2\int_0^{\pi/2} \cos^{2m+1}\theta \sin^{2n+1}\theta\,d\theta. \tag{8.57}$$

Here the Cartesian area element $dx\,dy$ has been replaced by $r\,dr\,d\theta$ (Fig. 8.6). The last equality in Eq. (8.57) follows from Exercise 8.1.11.

The definite integral, together with the factor 2, has been named the beta function:

$$B(m+1, n+1) \equiv 2\int_0^{\pi/2} \cos^{2m+1}\theta \sin^{2n+1}\theta\,d\theta$$

$$= \frac{m!n!}{(m+n+1)!}. \tag{8.58a}$$

Equivalently, in terms of the gamma function and noting its symmetry,

$$B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}, \qquad B(q,p) = B(p,q). \tag{8.58b}$$

The only reason for choosing $m+1$ and $n+1$, rather than m and n, as the arguments of B is to be in agreement with the conventional, historical beta function.

Definite Integrals, Alternate Forms

The beta function is useful in the evaluation of a wide variety of definite integrals. The substitution $t = \cos^2\theta$ converts Eq. (8.58a) to[7]

$$B(m+1, n+1) = \frac{m!n!}{(m+n+1)!} = \int_0^1 t^m (1-t)^n\,dt. \tag{8.59a}$$

[7] The Laplace transform convolution theorem provides an alternate derivation of Eq. (8.58a), compare Exercise 15.11.2.

Replacing t by x^2, we obtain

$$\frac{m!n!}{2(m+n+1)!} = \int_0^1 x^{2m+1}\left(1-x^2\right)^n dx. \tag{8.59b}$$

The substitution $t = u/(1+u)$ in Eq. (8.59a) yields still another useful form,

$$\frac{m!n!}{(m+n+1)!} = \int_0^\infty \frac{u^m}{(1+u)^{m+n+2}} du. \tag{8.60}$$

The beta function as a definite integral is useful in establishing integral representations of the Bessel function (Exercise 11.1.18) and the hypergeometric function (Exercise 13.4.10).

Verification of $\pi\alpha/\sin\pi\alpha$ Relation

If we take $m = a$, $n = -a$, $-1 < a < 1$, then

$$\int_0^\infty \frac{u^a}{(1+u)^2} du = a!(-a)!. \tag{8.61}$$

By contour integration this integral may be shown to be equal to $\pi a/\sin\pi a$ (Exercise 7.1.18), thus providing another method of obtaining Eq. (8.32).

Derivation of Legendre Duplication Formula

The form of Eq. (8.58a) suggests that the beta function may be useful in deriving the doubling formula used in the preceding section. From Eq. (8.59a) with $m = n = z$ and $\Re(z) > -1$,

$$\frac{z!z!}{(2z+1)!} = \int_0^1 t^z (1-t)^z dt. \tag{8.62}$$

By substituting $t = (1+s)/2$, we have

$$\frac{z!z!}{(2z+1)!} = 2^{-2z-1}\int_{-1}^1 \left(1-s^2\right)^z ds = 2^{-2z}\int_0^1 \left(1-s^2\right)^z ds. \tag{8.63}$$

The last equality holds because the integrand is even. Evaluating this integral as a beta function (Eq. (8.59b)), we obtain

$$\frac{z!z!}{(2z+1)!} = 2^{-2z-1}\frac{z!(-\frac{1}{2})!}{(z+\frac{1}{2})!}. \tag{8.64}$$

Rearranging terms and recalling that $(-\frac{1}{2})! = \pi^{1/2}$, we reduce this equation to one form of the Legendre duplication formula,

$$z!\left(z+\tfrac{1}{2}\right)! = 2^{-2z-1}\pi^{1/2}(2z+1)!. \tag{8.65a}$$

Dividing by $(z+\frac{1}{2})$, we obtain an alternate form of the duplication formula:

$$z!\left(z-\tfrac{1}{2}\right)! = 2^{-2z}\pi^{1/2}(2z)!. \tag{8.65b}$$

Although the integrals used in this derivation are defined only for $\Re(z) > -1$, the results (Eqs. (8.65a) and (8.65b) hold for all regular points z by analytic continuation.[8]

Using the double factorial notation (Section 8.1), we may rewrite Eq. (8.65a) (with $z = n$, an integer) as

$$\left(n + \tfrac{1}{2}\right)! = \pi^{1/2}(2n+1)!!/2^{n+1}. \tag{8.65c}$$

This is often convenient for eliminating factorials of fractions.

Incomplete Beta Function

Just as there is an incomplete gamma function (Section 8.5), there is also an incomplete beta function,

$$B_x(p,q) = \int_0^x t^{p-1}(1-t)^{q-1}\,dt, \qquad 0 \le x \le 1,\ p > 0,\ q > 0 \qquad (\text{if } x = 1). \tag{8.66}$$

Clearly, $B_{x=1}(p,q)$ becomes the regular (complete) beta function, Eq. (8.59a). A power-series expansion of $B_x(p,q)$ is the subject of Exercises 5.2.18 and 5.7.8. The relation to hypergeometric functions appears in Section 13.4.

The incomplete beta function makes an appearance in probability theory in calculating the probability of at most k successes in n independent trials.[9]

Exercises

8.4.1 Derive the doubling formula for the factorial function by integrating $(\sin 2\theta)^{2n+1} = (2\sin\theta\cos\theta)^{2n+1}$ (and using the beta function).

8.4.2 Verify the following beta function identities:

(a) $B(a,b) = B(a+1,b) + B(a,b+1)$,

(b) $B(a,b) = \dfrac{a+b}{b}B(a,b+1)$,

(c) $B(a,b) = \dfrac{b-1}{a}B(a+1,b-1)$,

(d) $B(a,b)B(a+b,c) = B(b,c)B(a,b+c)$.

8.4.3 (a) Show that

$$\int_{-1}^{1}\left(1-x^2\right)^{1/2}x^{2n}\,dx = \begin{cases} \pi/2, & n = 0 \\ \pi\dfrac{(2n-1)!!}{(2n+2)!!}, & n = 1,2,3,\ldots. \end{cases}$$

[8] If $2z$ is a negative integer, we get the valid but unilluminating result $\infty = \infty$.

[9] W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd ed. New York: Wiley (1968), Section VI.10.

(b) Show that

$$\int_{-1}^{1}\left(1-x^2\right)^{-1/2}x^{2n}\,dx=\begin{cases}\pi, & n=0\\ \pi\dfrac{(2n-1)!!}{(2n)!!}, & n=1,2,3,\ldots.\end{cases}$$

8.4.4 Show that

$$\int_{-1}^{1}\left(1-x^2\right)^{n}\,dx=\begin{cases}2^{2n+1}\dfrac{n!n!}{(2n+1)!}, & n>-1\\ 2\dfrac{(2n)!!}{(2n+1)!!}, & n=0,1,2,\ldots.\end{cases}$$

8.4.5 Evaluate $\int_{-1}^{1}(1+x)^a(1-x)^b\,dx$ in terms of the beta function.

ANS. $2^{a+b+1}B(a+1,b+1)$.

8.4.6 Show, by means of the beta function, that

$$\int_t^z\frac{dx}{(z-x)^{1-\alpha}(x-t)^{\alpha}}=\frac{\pi}{\sin\pi\alpha},\qquad 0<\alpha<1.$$

8.4.7 Show that the Dirichlet integral

$$\iint x^p y^q\,dx\,dy=\frac{p!q!}{(p+q+2)!}=\frac{B(p+1,q+1)}{p+q+2},$$

where the range of integration is the triangle bounded by the positive x- and y-axes and the line $x+y=1$.

8.4.8 Show that

$$\int_0^{\infty}\int_0^{\infty}e^{-(x^2+y^2+2xy\cos\theta)}\,dx\,dy=\frac{\theta}{2\sin\theta}.$$

What are the limits on θ?
Hint. Consider oblique xy-coordinates.

ANS. $-\pi<\theta<\pi$.

8.4.9 Evaluate (using the beta function)

(a)

$$\int_0^{\pi/2}\cos^{1/2}\theta\,d\theta=\frac{(2\pi)^{3/2}}{16[(\frac{1}{4})!]^2},$$

(b)

$$\int_0^{\pi/2}\cos^n\theta\,d\theta=\int_0^{\pi/2}\sin^n\theta\,d\theta=\frac{\sqrt{\pi}[(n-1)/2]!}{2(n/2)!}$$
$$=\begin{cases}\dfrac{(n-1)!!}{n!!} & \text{for } n \text{ odd},\\ \dfrac{\pi}{2}\cdot\dfrac{(n-1)!!}{n!!} & \text{for } n \text{ even}.\end{cases}$$

8.4.10 Evaluate $\int_0^1 (1-x^4)^{-1/2}\,dx$ as a beta function.

ANS. $\dfrac{[(\frac{1}{4})!]^2 \cdot 4}{(2\pi)^{1/2}} = 1.311028777.$

8.4.11 Given

$$J_\nu(z) = \frac{2}{\pi^{1/2}(\nu-\frac{1}{2})!}\left(\frac{z}{2}\right)^\nu \int_0^{\pi/2} \sin^{2\nu}\theta\cos(z\cos\theta)\,d\theta, \qquad \Re(\nu) > -\tfrac{1}{2},$$

show, with the aid of beta functions, that this reduces to the Bessel series

$$J_\nu(z) = \sum_{s=0}^{\infty}(-1)^s \frac{1}{s!(s+\nu)!}\left(\frac{z}{2}\right)^{2s+\nu},$$

identifying the initial J_ν as an integral representation of the Bessel function, J_ν (Section 11.1).

8.4.12 Given the associated Legendre function

$$P_m^m(x) = (2m-1)!!\left(1-x^2\right)^{m/2},$$

Section 12.5, show that

(a) $$\int_{-1}^{1}\left[P_m^m(x)\right]^2 dx = \frac{2}{2m+1}(2m)!, \qquad m = 0, 1, 2, \ldots,$$

(b) $$\int_{-1}^{1}\left[P_m^m(x)\right]^2 \frac{dx}{1-x^2} = 2\cdot(2m-1)!, \qquad m = 1, 2, 3, \ldots.$$

8.4.13 Show that

(a) $$\int_0^1 x^{2s+1}\left(1-x^2\right)^{-1/2} dx = \frac{(2s)!!}{(2s+1)!!},$$

(b) $$\int_0^1 x^{2p}\left(1-x^2\right)^q dx = \frac{1}{2}\frac{(p-\frac{1}{2})!q!}{(p+q+\frac{1}{2})!}.$$

8.4.14 A particle of mass m moving in a symmetric potential that is well described by $V(x) = A|x|^n$ has a total energy $\frac{1}{2}m(dx/dt)^2 + V(x) = E$. Solving for dx/dt and integrating we find that the period of motion is

$$\tau = 2\sqrt{2m}\int_0^{x_{\max}} \frac{dx}{(E-Ax^n)^{1/2}},$$

where $x_{\max}$ is a classical turning point given by $Ax_{\max}^n = E$. Show that

$$\tau = \frac{2}{n}\sqrt{\frac{2\pi m}{E}}\left(\frac{E}{A}\right)^{1/n}\frac{\Gamma(1/n)}{\Gamma(1/n+\frac{1}{2})}.$$

8.4.15 Referring to Exercise 8.4.14,

(a) Determine the limit as $n \to \infty$ of

$$\frac{2}{n}\sqrt{\frac{2\pi m}{E}}\left(\frac{E}{A}\right)^{1/n}\frac{\Gamma(1/n)}{\Gamma(1/n+\frac{1}{2})}.$$

(b) Find $\lim_{n\to\infty} \tau$ from the behavior of the integrand $(E - Ax^n)^{-1/2}$.

(c) Investigate the behavior of the physical system (potential well) as $n \to \infty$. Obtain the period from inspection of this limiting physical system.

8.4.16 Show that

$$\int_0^\infty \frac{\sinh^\alpha x}{\cosh^\beta x}dx = \frac{1}{2}B\left(\frac{\alpha+1}{2}, \frac{\beta-\alpha}{2}\right), \qquad -1 < \alpha < \beta.$$

Hint. Let $\sinh^2 x = u$.

8.4.17 The beta distribution of probability theory has a probability density

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1},$$

with x restricted to the interval (0, 1). Show that

(a) $\langle x\rangle(\text{mean}) = \dfrac{\alpha}{\alpha+\beta}$.

(b) $\sigma^2(\text{variance}) \equiv \langle x^2\rangle - \langle x\rangle^2 = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

8.4.18 From

$$\lim_{n\to\infty}\frac{\int_0^{\pi/2}\sin^{2n}\theta\, d\theta}{\int_0^{\pi/2}\sin^{2n+1}\theta\, d\theta} = 1$$

derive the Wallis formula for π:

$$\frac{\pi}{2} = \frac{2\cdot 2}{1\cdot 3}\cdot\frac{4\cdot 4}{3\cdot 5}\cdot\frac{6\cdot 6}{5\cdot 7}\cdots.$$

8.4.19 Tabulate the beta function $B(p, q)$ for p and $q = 1.0(0.1)2.0$ independently.

Check value. $B(1.3, 1.7) = 0.40774$.

8.4.20 (a) Write a subroutine that will calculate the incomplete beta function $B_x(p, q)$. For $0.5 < x \le 1$ you will find it convenient to use the relation

$$B_x(p, q) = B(p, q) - B_{1-x}(q, p).$$

(b) Tabulate $B_x(\frac{3}{2}, \frac{3}{2})$. Spot check your results by using the Gauss–Legendre quadrature.

8.5 THE INCOMPLETE GAMMA FUNCTIONS AND RELATED FUNCTIONS

Generalizing the Euler definition of the gamma function (Eq. (8.5)), we define the incomplete gamma functions by the variable limit integrals

$$\gamma(a,x)=\int_0^x e^{-t}t^{a-1}\,dt, \qquad \Re(a)>0$$

and

$$\Gamma(a,x)=\int_x^\infty e^{-t}t^{a-1}\,dt. \tag{8.67}$$

Clearly, the two functions are related, for

$$\gamma(a,x)+\Gamma(a,x)=\Gamma(a). \tag{8.68}$$

The choice of employing $\gamma(a,x)$ or $\Gamma(a,x)$ is purely a matter of convenience. If the parameter a is a positive integer, Eq. (8.67) may be integrated completely to yield

$$\begin{aligned}\gamma(n,x)&=(n-1)!\left(1-e^{-x}\sum_{s=0}^{n-1}\frac{x^s}{s!}\right)\\ \Gamma(n,x)&=(n-1)!e^{-x}\sum_{s=0}^{n-1}\frac{x^s}{s!}, \quad n=1,2,\ldots.\end{aligned} \tag{8.69}$$

For nonintegral a, a power-series expansion of $\gamma(a,x)$ for small x and an asymptotic expansion of $\Gamma(a,x)$ (denoted as $I(x,p)$) are developed in Exercise 5.7.7 and Section 5.10:

$$\begin{aligned}\gamma(a,x)&=x^a\sum_{n=0}^\infty(-1)^n\frac{x^n}{n!(a+n)}, \qquad |x|\sim 1 \text{ (small } x),\\ \Gamma(a,x)&=x^{a-1}e^{-x}\sum_{n=0}^\infty\frac{(a-1)!}{(a-1-n)!}\cdot\frac{1}{x^n}\\ &=x^{a-1}e^{-x}\sum_{n=0}^\infty(-1)^n\frac{(n-a)!}{(-a)!}\cdot\frac{1}{x^n}, \qquad x\gg 1 \text{ (large } x).\end{aligned} \tag{8.70}$$

These incomplete gamma functions may also be expressed quite elegantly in terms of confluent hypergeometric functions (compare Section 13.5).

Exponential Integral

Although the incomplete gamma function $\Gamma(a,x)$ in its general form (Eq. (8.67)) is only infrequently encountered in physical problems, a special case is quite common and very

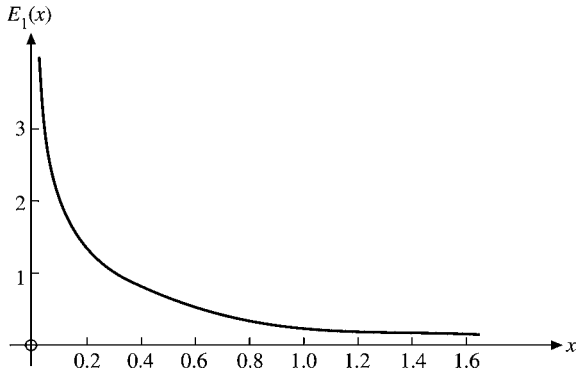

FIGURE 8.7 The exponential integral, $E_1(x) = -\mathrm{Ei}(-x)$.

useful. We define the exponential integral by[10]

$$-\mathrm{Ei}(-x) \equiv \int_x^\infty \frac{e^{-t}}{t} dt = E_1(x). \tag{8.71}$$

(See Fig. 8.7.) Caution is needed here, for the integral in Eq. (8.71) diverges logarithmically as $x \to 0$. To obtain a series expansion for small x, we start from

$$E_1(x) = \Gamma(0, x) = \lim_{a\to 0}\big[\Gamma(a) - \gamma(a, x)\big]. \tag{8.72}$$

We may split the divergent term in the series expansion for $\gamma(a, x)$,

$$E_1(x) = \lim_{a\to 0}\left[\frac{a\Gamma(a) - x^a}{a}\right] - \sum_{n=1}^{\infty} \frac{(-1)^n x^n}{n \cdot n!}. \tag{8.73}$$

Using l'Hôpital's rule (Exercise 5.6.8) and

$$\frac{d}{da}\big\{a\Gamma(a)\big\} = \frac{d}{da}a! = \frac{d}{da}e^{\ln(a!)} = a!\psi(a+1), \tag{8.74}$$

and then Eq. (8.40),[11] we obtain the rapidly converging series

$$E_1(x) = -\gamma - \ln x - \sum_{n=1}^{\infty} \frac{(-1)^n x^n}{n \cdot n!}. \tag{8.75}$$

An asymptotic expansion $E_1(x) \approx e^{-x}[\frac{1}{x} - \frac{1!}{x^2} + \cdots]$ for $x \to \infty$ is developed in Section 5.10.

[10] The appearance of the two minus signs in $-\mathrm{Ei}(-x)$ is a historical monstrosity. AMS-55, Chapter 5, denotes this integral as $E_1(x)$. See Additional Readings for the reference.

[11] $dx^a/da = x^a \ln x$.

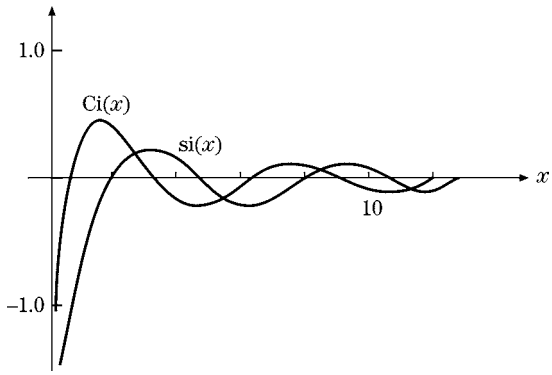

FIGURE 8.8 Sine and cosine integrals.

Further special forms related to the exponential integral are the sine integral, cosine integral (Fig. 8.8), and logarithmic integral, defined by[12]

$$\begin{aligned} \mathrm{si}(x) &= -\int_x^\infty \frac{\sin t}{t}\,dt \\ \mathrm{Ci}(x) &= -\int_x^\infty \frac{\cos t}{t}\,dt \\ \mathrm{li}(x) &= \int_0^x \frac{du}{\ln u} = \mathrm{Ei}(\ln x) \end{aligned} \tag{8.76}$$

for their principal branch, with the branch cut conventionally chosen to be along the negative real axis from the branch point at zero. By transforming from real to imaginary argument, we can show that

$$\mathrm{si}(x) = \frac{1}{2i}\big[\mathrm{Ei}(ix) - \mathrm{Ei}(-ix)\big] = \frac{1}{2i}\big[E_1(ix) - E_1(-ix)\big], \tag{8.77}$$

whereas

$$\mathrm{Ci}(x) = \frac{1}{2}\big[\mathrm{Ei}(ix) + \mathrm{Ei}(-ix)\big] = -\frac{1}{2}\big[E_1(ix) + E_1(-ix)\big], \qquad |\arg x| < \frac{\pi}{2}. \tag{8.78}$$

Adding these two relations, we obtain

$$\mathrm{Ei}(ix) = \mathrm{Ci}(x) + i\,\mathrm{si}(x), \tag{8.79}$$

to show that the relation among these integrals is exactly analogous to that among e^{ix}, $\cos x$, and $\sin x$. Reference to Eqs. (8.71) and (8.78) shows that $\mathrm{Ci}(x)$ agrees with the definitions of AMS-55 (see Additional Readings for the reference). In terms of E_1,

$$E_1(ix) = -\mathrm{Ci}(x) + i\,\mathrm{si}(x).$$

Asymptotic expansions of $\mathrm{Ci}(x)$ and $\mathrm{si}(x)$ are developed in Section 5.10. Power-series expansions about the origin for $\mathrm{Ci}(x)$, $\mathrm{si}(x)$, and $\mathrm{li}(x)$ may be obtained from those for

[12] Another sine integral is given by $\mathrm{Si}(x) = \mathrm{si}(x) + \pi/2$.

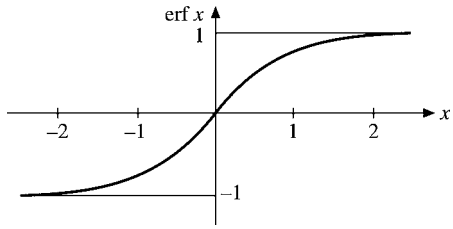

FIGURE 8.9 Error function, $\operatorname{erf} x$.

the exponential integral, $E_1(x)$, or by direct integration, Exercise 8.5.10. The exponential, sine, and cosine integrals are tabulated in AMS-55, Chapter 5, (see Additional Readings for the reference) and can also be accessed by symbolic software such as Mathematica, Maple, Mathcad, and Reduce.

Error Integrals

The error integrals

$$\boxed{\operatorname{erf} z = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt, \qquad \operatorname{erfc} z = 1 - \operatorname{erf} z = \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-t^2}\, dt} \tag{8.80a}$$

(normalized so that $\operatorname{erf} \infty = 1$) are introduced in Exercise 5.10.4 (Fig. 8.9). Asymptotic forms are developed there. From the general form of the integrands and Eq. (8.6) we expect that $\operatorname{erf} z$ and $\operatorname{erfc} z$ may be written as incomplete gamma functions with $a = \frac{1}{2}$. The relations are

$$\operatorname{erf} z = \pi^{-1/2} \gamma\left(\tfrac{1}{2}, z^2\right), \qquad \operatorname{erfc} z = \pi^{-1/2} \Gamma\left(\tfrac{1}{2}, z^2\right). \tag{8.80b}$$

The power-series expansion of $\operatorname{erf} z$ follows directly from Eq. (8.70).

Exercises

8.5.1 Show that

$$\gamma(a, x) = e^{-x} \sum_{n=0}^{\infty} \frac{(a-1)!}{(a+n)!} x^{a+n}$$

(a) by repeatedly integrating by parts.
(b) Demonstrate this relation by transforming it into Eq. (8.70).

8.5.2 Show that

(a) $\dfrac{d^m}{dx^m}\left[x^{-a}\gamma(a, x)\right] = (-1)^m x^{-a-m} \gamma(a+m, x),$

(b) $$\frac{d^m}{dx^m}\left[e^x\gamma(a,x)\right] = e^x\frac{\Gamma(a)}{\Gamma(a-m)}\gamma(a-m,x).$$

8.5.3 Show that $\gamma(a,x)$ and $\Gamma(a,x)$ satisfy the recurrence relations

(a) $\gamma(a+1,x) = a\gamma(a,x) - x^a e^{-x}$,
(b) $\Gamma(a+1,x) = a\Gamma(a,x) + x^a e^{-x}$.

8.5.4 The potential produced by a $1S$ hydrogen electron (Exercise 12.8.6) is given by

$$V(r) = \frac{q}{4\pi\varepsilon_0 a_0}\left\{\frac{1}{2r}\gamma(3,2r) + \Gamma(2,2r)\right\}.$$

(a) For $r \ll 1$, show that

$$V(r) = \frac{q}{4\pi\varepsilon_0 a_0}\left\{1 - \frac{2}{3}r^2 + \cdots\right\}.$$

(b) For $r \gg 1$, show that

$$V(r) = \frac{q}{4\pi\varepsilon_0 a_0}\cdot\frac{1}{r}.$$

Here r is expressed in units of a_0, the Bohr radius.
Note. For computation at intermediate values of r, Eqs. (8.69) are convenient.

8.5.5 The potential of a $2P$ hydrogen electron is found to be (Exercise 12.8.7)

$$V(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\cdot\frac{q}{24a_0}\left\{\frac{1}{r}\gamma(5,r) + \Gamma(4,r)\right\}$$
$$-\frac{1}{4\pi\varepsilon_0}\cdot\frac{q}{120a_0}\left\{\frac{1}{r^3}\gamma(7,r) + r^2\Gamma(2,r)\right\}P_2(\cos\theta).$$

Here r is expressed in units of a_0, the Bohr radius. $P_2(\cos\theta)$ is a Legendre polynomial (Section 12.1).

(a) For $r \ll 1$, show that

$$V(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\cdot\frac{q}{a_0}\left\{\frac{1}{4} - \frac{1}{120}r^2 P_2(\cos\theta) + \cdots\right\}.$$

(b) For $r \gg 1$, show that

$$V(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\cdot\frac{q}{a_0 r}\left\{1 - \frac{6}{r^2}P_2(\cos\theta) + \cdots\right\}.$$

8.5.6 Prove that the exponential integral has the expansion

$$\int_x^\infty \frac{e^{-t}}{t}dt = -\gamma - \ln x - \sum_{n=1}^{\infty}\frac{(-1)^n x^n}{n\cdot n!},$$

where γ is the Euler–Mascheroni constant.

8.5.7 Show that $E_1(z)$ may be written as

$$E_1(z) = e^{-z}\int_0^\infty \frac{e^{-zt}}{1+t}dt.$$

Show also that we must impose the condition $|\arg z| \le \pi/2$.

8.5.8 Related to the exponential integral (Eq. (8.71)) by a simple change of variable is the function

$$E_n(x) = \int_1^\infty \frac{e^{-xt}}{t^n}dt.$$

Show that $E_n(x)$ satisfies the recurrence relation

$$E_{n+1}(x) = \frac{1}{n}e^{-x} - \frac{x}{n}E_n(x), \qquad n = 1, 2, 3, \ldots.$$

8.5.9 With $E_n(x)$ as defined in Exercise 8.5.8, show that $E_n(0) = 1/(n-1)$, $n > 1$.

8.5.10 Develop the following power-series expansions:

(a) $\mathrm{si}(x) = -\dfrac{\pi}{2} + \displaystyle\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)(2n+1)!}$,

(b) $\mathrm{Ci}(x) = \gamma + \ln x + \displaystyle\sum_{n=1}^{\infty} \frac{(-1)^n x^{2n}}{2n(2n)!}$.

8.5.11 An analysis of a center-fed linear antenna leads to the expression

$$\int_0^x \frac{1-\cos t}{t}dt.$$

Show that this is equal to $\gamma + \ln x - \mathrm{Ci}(x)$.

8.5.12 Using the relation

$$\Gamma(a) = \gamma(a, x) + \Gamma(a, x),$$

show that if $\gamma(a, x)$ satisfies the relations of Exercise 8.5.2, then $\Gamma(a, x)$ must satisfy the same relations.

8.5.13 (a) Write a subroutine that will calculate the incomplete gamma functions $\gamma(n, x)$ and $\Gamma(n, x)$ for n a positive integer. Spot check $\Gamma(n, x)$ by Gauss–Laguerre quadratures.
(b) Tabulate $\gamma(n, x)$ and $\Gamma(n, x)$ for $x = 0.0(0.1)1.0$ and $n = 1, 2, 3$.

8.5.14 Calculate the potential produced by a $1S$ hydrogen electron (Exercise 8.5.4) (Fig. 8.10). Tabulate $V(r)/(q/4\pi\varepsilon_0 a_0)$ for $x = 0.0(0.1)4.0$. Check your calculations for $r \ll 1$ and for $r \gg 1$ by calculating the limiting forms given in Exercise 8.5.4.

8.5.15 Using Eqs. (5.182) and (8.75), calculate the exponential integral $E_1(x)$ for

(a) $x = 0.2(0.2)1.0$, (b) $x = 6.0(2.0)10.0$.

Program your own calculation but check each value, using a library subroutine if available. Also check your calculations at each point by a Gauss–Laguerre quadrature.

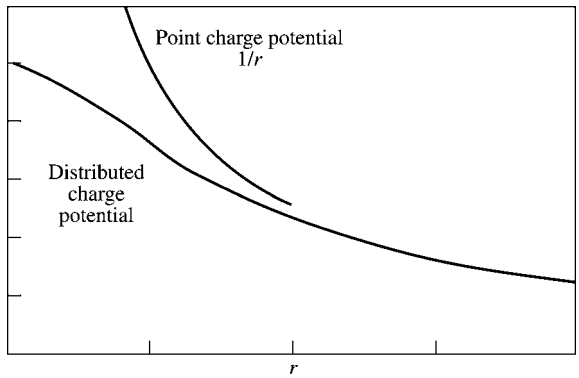

FIGURE 8.10 Distributed charge potential produced by a $1S$ hydrogen electron, Exercise 8.5.14.

You'll find that the power-series converges rapidly and yields high precision for small x. The asymptotic series, even for $x = 10$, yields relatively poor accuracy.

Check values. $E_1(1.0) = 0.219384$
$E_1(10.0) = 4.15697 \times 10^{-6}$.

8.5.16 The two expressions for $E_1(x)$, (1) Eq. (5.182), an asymptotic series and (2) Eq. (8.75), a convergent power series, provide a means of calculating the Euler–Mascheroni constant γ to high accuracy. Using double precision, calculate γ from Eq. (8.75), with $E_1(x)$ evaluated by Eq. (5.182).
Hint. As a convenient choice take x in the range 10 to 20. (Your choice of x will set a limit on the accuracy of your result.) To minimize errors in the alternating series of Eq. (8.75), accumulate the positive and negative terms separately.

ANS. For $x = 10$ and "double precision," $\gamma = 0.57721566$.

Additional Readings

Abramowitz, M., and I. A. Stegun, eds., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (AMS-55). Washington, DC: National Bureau of Standards (1972), reprinted, Dover (1974). Contains a wealth of information about gamma functions, incomplete gamma functions, exponential integrals, error functions, and related functions — Chapters 4 to 6.

Artin, E., *The Gamma Function* (translated by M. Butler). New York: Holt, Rinehart and Winston (1964). Demonstrates that if a function $f(x)$ is smooth (log convex) and equal to $(n-1)!$ when $x = n =$ integer, it is the gamma function.

Davis, H. T., *Tables of the Higher Mathematical Functions*. Bloomington, IN: Principia Press (1933). Volume I contains extensive information on the gamma function and the polygamma functions.

Gradshteyn, I. S., and I. M. Ryzhik, *Table of Integrals, Series, and Products*. New York: Academic Press (1980).

Luke, Y. L., *The Special Functions and Their Approximations*, Vol. 1. New York: Academic Press (1969).

Luke, Y. L., *Mathematical Functions and Their Approximations*. New York: Academic Press (1975). This is an updated supplement to *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (AMS-55). Chapter 1 deals with the gamma function. Chapter 4 treats the incomplete gamma function and a host of related functions.

This page intentionally left blank

CHAPTER 9

DIFFERENTIAL EQUATIONS

9.1 PARTIAL DIFFERENTIAL EQUATIONS

Introduction

In physics the knowledge of the force in an equation of motion usually leads to a differential equation. Thus, almost all the elementary and numerous advanced parts of theoretical physics are formulated in terms of differential equations. Sometimes these are ordinary differential equations in one variable (abbreviated **ODEs**). More often the equations are partial differential equations (**PDEs**) in two or more variables.

Let us recall from calculus that the operation of taking an ordinary or partial derivative is a **linear operation** ($\mathcal{L}$),[1]

$$\frac{d(a\varphi(x)+b\psi(x))}{dx}=a\frac{d\varphi}{dx}+b\frac{d\psi}{dx},$$

for ODEs involving derivatives in one variable x only and no quadratic, $(d\psi/dx)^2$, or higher powers. Similarly, for partial derivations,

$$\frac{\partial(a\varphi(x,y)+b\psi(x,y))}{\partial x}=a\frac{\partial\varphi(x,y)}{\partial x}+b\frac{\partial\psi(x,y)}{\partial x}.$$

In general

$$\mathcal{L}(a\varphi+b\psi)=a\mathcal{L}(\varphi)+b\mathcal{L}(\psi).$$

Thus, ODEs and PDEs appear as linear operator equations,

$$\mathcal{L}\psi=F, \tag{9.1}$$

[1] We are especially interested in linear operators because in quantum mechanics physical quantities are represented by linear operators operating in a complex, infinite-dimensional Hilbert space.

where F is a known (source) function of one (for ODEs) or more variables (for PDEs), $\mathcal{L}$ is a linear combination of derivatives, and ψ is the unknown function or solution. Any linear combination of solutions is again a solution if $F = 0$; this is the **superposition principle** for homogeneous PDEs.

Since the dynamics of many physical systems involve just two derivatives, for example, acceleration in classical mechanics and the kinetic energy operator, $\sim \nabla^2$, in quantum mechanics, differential equations of second order occur most frequently in physics. (Maxwell's and Dirac's equations are first order but involve two unknown functions. Eliminating one unknown yields a second-order differential equation for the other (compare Section 1.9).)

Examples of PDEs

Among the most frequently encountered PDEs are the following:

1. Laplace's equation, $\nabla^2\psi = 0$.
   This very common and very important equation occurs in studies of
   a. electromagnetic phenomena, including electrostatics, dielectrics, steady currents, and magnetostatics,
   b. hydrodynamics (irrotational flow of perfect fluid and surface waves),
   c. heat flow,
   d. gravitation.
2. Poisson's equation, $\nabla^2\psi = -\rho/\varepsilon_0$.
   In contrast to the homogeneous Laplace equation, Poisson's equation is nonhomogeneous with a source term $-\rho/\varepsilon_0$.
3. The wave (Helmholtz) and time-independent diffusion equations, $\nabla^2\psi \pm k^2\psi = 0$.
   These equations appear in such diverse phenomena as
   a. elastic waves in solids, including vibrating strings, bars, membranes,
   b. sound, or acoustics,
   c. electromagnetic waves,
   d. nuclear reactors.
4. The time-dependent diffusion equation
   $$\nabla^2\psi = \frac{1}{a^2}\frac{\partial\psi}{\partial t}$$
   and the corresponding four-dimensional forms involving the d'Alembertian, a four-dimensional analog of the Laplacian in Minkowski space,
   $$\partial^\mu\partial_\mu = \partial^2 = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2.$$
5. The time-dependent wave equation, $\partial^2\psi = 0$.
6. The scalar potential equation, $\partial^2\psi = \rho/\varepsilon_0$.
   Like Poisson's equation, this equation is nonhomogeneous with a source term ρ/ε_0.

7. The Klein–Gordon equation, $\partial^2\psi = -\mu^2\psi$, and the corresponding vector equations, in which the scalar function ψ is replaced by a vector function. Other, more complicated forms are common.
8. The Schrödinger wave equation,

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V\psi = i\hbar\frac{\partial\psi}{\partial t}$$

and

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V\psi = E\psi$$

for the time-independent case.
9. The equations for elastic waves and viscous fluids and the telegraphy equation.
10. Maxwell's coupled partial differential equations for electric and magnetic fields and those of Dirac for relativistic electron wave functions. For Maxwell's equations see the Introduction and also Section 1.9.

Some general techniques for solving second-order PDEs are discussed in this section.

1. Separation of variables, where the PDE is split into ODEs that are related by common constants that appear as eigenvalues of linear operators, $\mathcal{L}\psi = l\psi$, usually in one variable. This method is closely related to **symmetries of the PDE and a group of transformations** (see Section 4.2). The Helmholtz equation, listed example 3, has this form, where the eigenvalue k^2 may arise by separation of the time t from the spatial variables. Likewise, in example 8 the energy E is the eigenvalue that arises in the separation of t from $\mathbf{r}$ in the Schrödinger equation. This is pursued in Chapter 10 in greater detail. Section 9.2 serves as introduction. ODEs may be attacked by Frobenius' power-series method in Section 9.5. It does not always work but is often the simplest method when it does.
2. Conversion of a PDE into an integral equation using **Green's functions** applies to **inhomogeneous** PDEs, such as examples 2 and 6 given above. An introduction to the Green's function technique is given in Section 9.7.
3. Other analytical methods, such as the use of integral transforms, are developed and applied in Chapter 15.

Occasionally, we encounter equations of higher order. In both the theory of the slow motion of a viscous fluid and the theory of an elastic body we find the equation

$$\left(\nabla^2\right)^2\psi = 0.$$

Fortunately, these higher-order differential equations are relatively rare and are not discussed here.

Although not so frequently encountered and perhaps not so important as second-order ODEs, first-order ODEs do appear in theoretical physics and are sometimes intermediate steps for second-order ODEs. The solutions of some more important types of first-order ODEs are developed in Section 9.2. First-order PDEs can always be reduced to ODEs. This is a straightforward but lengthy process and involves a search for characteristics that are briefly introduced in what follows; for more details we refer to the literature.

Classes of PDEs and Characteristics

Second-order PDEs form three classes: (i) Elliptic PDEs involve ∇^2 or $c^{-2}\partial^2/\partial t^2 + \nabla^2$ (ii) parabolic PDEs, $a\partial/\partial t + \nabla^2$; (iii) hyperbolic PDEs, $c^{-2}\partial^2/\partial t^2 - \nabla^2$. These canonical operators come about by a change of variables $\xi = \xi(x, y)$, $\eta = \eta(x, y)$ in a linear operator (for two variables just for simplicity)

$$\mathcal{L} = a\frac{\partial^2}{\partial x^2} + 2b\frac{\partial^2}{\partial x \partial y} + c\frac{\partial^2}{\partial y^2} + d\frac{\partial}{\partial x} + e\frac{\partial}{\partial y} + f, \tag{9.2}$$

which can be reduced to the canonical forms (i), (ii), (iii) according to whether the discriminant $D = ac - b^2 > 0, = 0$, or < 0. If $\xi(x, y)$ is determined from the first-order, but nonlinear, PDE

$$a\left(\frac{\partial \xi}{\partial x}\right)^2 + 2b\left(\frac{\partial \xi}{\partial x}\right)\left(\frac{\partial \xi}{\partial y}\right) + c\left(\frac{\partial \xi}{\partial y}\right)^2 = 0, \tag{9.3}$$

then the coefficient of $\partial^2/\partial\xi^2$ in $\mathcal{L}$ (that is, Eq. (9.3)) is zero. If η is an independent solution of the same Eq. (9.3), then the coefficient of $\partial^2/\partial\eta^2$ is also zero. The remaining operator, $\partial^2/\partial\xi\partial\eta$, in $\mathcal{L}$ is characteristic of the hyperbolic case (iii) with $D < 0$ ($a = 0 = c$ leads to $D = -b^2 < 0$), where the quadratic form $a\lambda^2 + 2b\lambda + c$ factorizes and, therefore, Eq. (9.3) has two independent solutions $\xi(x, y)$, $\eta(x, y)$. In the elliptic case (i) with $D > 0$, the two solutions ξ, η are complex conjugate, which, when substituted into Eq. (9.2), remove the mixed second-order derivative instead of the other second-order terms, yielding the canonical form (i). In the parabolic case (ii) with $D = 0$, only $\partial^2/\partial\xi^2$ remains in $\mathcal{L}$, while the coefficients of the other two second-order derivatives vanish.

If the coefficients a, b, c in $\mathcal{L}$ are functions of the coordinates, then this classification is only local; that is, its type may change as the coordinates vary.

Let us illustrate the physics underlying the **hyperbolic case** by looking at the wave equation, Eq. (9.2) (in $1 + 1$ dimensions for simplicity)

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}\right)\psi = 0.$$

Since Eq. (9.3) now becomes

$$\left(\frac{\partial \xi}{\partial t}\right)^2 - c^2\left(\frac{\partial \xi}{\partial x}\right)^2 = \left(\frac{\partial \xi}{\partial t} - c\frac{\partial \xi}{\partial x}\right)\left(\frac{\partial \xi}{\partial t} + c\frac{\partial \xi}{\partial x}\right) = 0$$

and factorizes, we determine the solution of $\partial\xi/\partial t - c\partial\xi/\partial x = 0$. This is an arbitrary function $\xi = F(x + ct)$, and $\xi = G(x - ct)$ solves $\partial\xi/\partial t + c\partial\xi/\partial x = 0$, which is readily verified. By linear superposition a general solution of the wave equation is $\psi = F(x+ct)+ G(x - ct)$. For periodic functions F, G we recognize the lines $x + ct$ and $x - ct$ as the phases of plane waves or wave fronts, where not all second-order derivatives of ψ in the wave equation are well defined. Normal to the wave fronts are the rays of geometric optics. Thus, the lines that are solutions of Eq. (9.3) and are called **characteristics** or sometimes **bicharacteristics** (for second-order PDEs) in the mathematical literature correspond to the wave fronts of the geometric optics solution of the wave equation.

For the **elliptic case** let us consider Laplace's equation,

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}=0,$$

for a potential ψ of two variables. Here the characteristics equation,

$$\left(\frac{\partial\xi}{\partial x}\right)^2+\left(\frac{\partial\xi}{\partial y}\right)^2=\left(\frac{\partial\xi}{\partial x}+i\frac{\partial\xi}{\partial y}\right)\left(\frac{\partial\xi}{\partial x}-i\frac{\partial\xi}{\partial y}\right)=0,$$

has complex conjugate solutions: $\xi=F(x+iy)$ for $\partial\xi/\partial x+i(\partial\xi/\partial y)=0$ and $\xi=G(x-iy)$ for $\partial\xi/\partial x-i(\partial\xi/\partial y)=0$. A general solution of Laplace's equation is therefore $\psi=F(x+iy)+G(x-iy)$, as well as the real and imaginary parts of ψ, which are called **harmonic** functions, while polynomial solutions are called **harmonic polynomials**.

In quantum mechanics the Wentzel–Kramers–Brillouin (WKB) form $\psi=\exp(-iS/\hbar)$ for the solution of the Schrödinger equation, a complex parabolic PDE,

$$\left(-\frac{\hbar^2}{2m}\nabla^2+V\right)\psi=i\hbar\frac{\partial\psi}{\partial t},$$

leads to the Hamilton–Jacobi equation of classical mechanics,

$$\frac{1}{2m}(\nabla S)^2+V=\frac{\partial S}{\partial t}, \tag{9.4}$$

in the limit $\hbar\to 0$. The classical action S obeys the Hamilton–Jacobi equation, which is the analog of Eq. (9.3) of the Schrödinger equation. Substituting $\nabla\psi=-i\psi\nabla S/\hbar$, $\partial\psi/\partial t=-i\psi(\partial S/\partial t)/\hbar$ into the Schrödinger equation, dropping the overall nonvanishing factor ψ, and taking the limit of the resulting equation as $\hbar\to 0$, we indeed obtain Eq. (9.4).

Finding solutions of PDEs by solving for the characteristics is one of several general techniques. For more examples we refer to H. Bateman, *Partial Differential Equations of Mathematical Physics*, New York: Dover (1944); K. E. Gustafson, *Partial Differential Equations and Hilbert Space Methods*, 2nd ed., New York: Wiley (1987), reprinted Dover (1998).

In order to derive and appreciate more the mathematical method behind these solutions of hyperbolic, parabolic, and elliptic PDEs let us reconsider the PDE (9.2) with constant coefficients and, at first, $d=e=f=0$ for simplicity. In accordance with the form of the wave front solutions, we seek a solution $\psi=F(\xi)$ of Eq. (9.2) with a function $\xi=\xi(t,x)$ using the variables t, x instead of x, y. Then the partial derivatives become

$$\frac{\partial\psi}{\partial x}=\frac{\partial\xi}{\partial x}\frac{dF}{d\xi},\qquad \frac{\partial\psi}{\partial t}=\frac{\partial\xi}{\partial t}\frac{dF}{d\xi},\qquad \frac{\partial^2\psi}{\partial x^2}=\frac{\partial^2\xi}{\partial x^2}\frac{dF}{d\xi}+\left(\frac{\partial\xi}{\partial x}\right)^2\frac{d^2F}{d\xi^2},$$

and

$$\frac{\partial^2\psi}{\partial x\partial t}=\frac{\partial^2\xi}{\partial x\partial t}\frac{dF}{d\xi}+\frac{\partial\xi}{\partial x}\frac{\partial\xi}{\partial t}\frac{d^2F}{d\xi^2},\qquad \frac{\partial^2\psi}{\partial t^2}=\frac{\partial^2\xi}{\partial t^2}\frac{dF}{d\xi}+\left(\frac{\partial\xi}{\partial t}\right)^2\frac{d^2F}{d\xi^2},$$

using the chain rule of differentiation. When ξ depends on x and t **linearly**, these partial derivatives of ψ yield a single term only and solve our PDE (9.2) as a consequence. From the linear $\xi=\alpha x+\beta t$ we obtain

$$\frac{\partial^2\psi}{\partial x^2}=\alpha^2\frac{d^2F}{d\xi^2},\qquad \frac{\partial^2\psi}{\partial x\partial t}=\alpha\beta\frac{d^2F}{d\xi^2},\qquad \frac{\partial^2\psi}{\partial t^2}=\beta^2\frac{d^2F}{d\xi^2},$$

and our PDE (9.2) becomes equivalent to the analog of Eq. (9.3),

$$\left(\alpha^2 a + 2\alpha\beta b + \beta^2 c\right)\frac{d^2F}{d\xi^2} = 0. \tag{9.5}$$

A solution of $\frac{d^2F}{d\xi^2} = 0$ only leads to the trivial $\psi = k_1 x + k_2 t + k_3$ with constant k_i that is linear in the coordinates and for which all second derivatives vanish. From $\alpha^2 a + 2\alpha\beta b + \beta^2 c = 0$, on the other hand, we get the ratios

$$\frac{\beta}{\alpha} = \frac{1}{c}\left[-b \pm \left(b^2 - ac\right)^{1/2}\right] \equiv r_{1,2} \tag{9.6}$$

as solutions of Eq. (9.5) with $\frac{d^2F}{d\xi^2} \neq 0$ in general. The lines $\xi_1 = x + r_1 t$ and $\xi_2 = x + r_2 t$ will solve the PDE (9.2), with $\psi(x,t) = F(\xi_1) + G(\xi_2)$ corresponding to the generalization of our previous hyperbolic and elliptic PDE examples.

For the parabolic case, where $b^2 = ac$, there is only one ratio from Eq. (9.6), $\beta/\alpha = r = -b/c$, and one solution, $\psi(x,t) = F(x - bt/c)$. In order to find the second general solution of our PDE (9.2) we make the Ansatz (trial solution) $\psi(x,t) = \psi_0(x,t) \cdot G(x - bt/c)$. Substituting this into Eq. (9.2) we find

$$a\frac{\partial^2\psi_0}{\partial x^2} + 2b\frac{\partial^2\psi_0}{\partial x\partial t} + c\frac{\partial^2\psi_0}{\partial t^2} = 0$$

for ψ_0 since, upon replacing $F \to G$, G solves Eq. (9.5) with $d^2G/d\xi^2 \neq 0$ in general. The solution ψ_0 can be any solution of our PDE (9.2), including the trivial ones such as $\psi_0 = x$ and $\psi_0 = t$. Thus we obtain the **general parabolic solution**,

$$\psi(x,t) = F\left(x - \frac{b}{c}t\right) + \psi_0(x,t)G\left(x - \frac{b}{c}t\right),$$

with $\psi_0 = x$ or $\psi_0 = t$, etc.

With the same Ansatz one finds solutions of our PDE (9.2) with a source term, for example, $f \neq 0$, but still $d = e = 0$ and constant a, b, c.

Next we determine the characteristics, that is, curves where the second order derivatives of the solution ψ are not well defined. These are the wave fronts along which the solutions of our hyperbolic PDE (9.2) propagate. We solve our PDE with a source term $f \neq 0$ and Cauchy boundary conditions (see Table 9.1) that are appropriate for hyperbolic PDEs, where ψ and its normal derivative $\partial\psi/\partial n$ are specified on an open curve

$$\mathcal{C} : x = x(s), \qquad t = t(s),$$

with the parameter s the length on $\mathcal{C}$. Then $d\mathbf{r} = (dx, dt)$ is tangent and $\hat{\mathbf{n}}\,ds = (dt, -dx)$ is normal to the curve $\mathcal{C}$, and the first-order tangential and normal derivatives are given by the chain rule

$$\frac{d\psi}{ds} = \nabla\psi \cdot \frac{d\mathbf{r}}{ds} = \frac{\partial\psi}{\partial x}\frac{dx}{ds} + \frac{\partial\psi}{\partial t}\frac{dt}{ds},$$
$$\frac{d\psi}{dn} = \nabla\psi \cdot \hat{\mathbf{n}} = \frac{\partial\psi}{\partial x}\frac{dt}{ds} - \frac{\partial\psi}{\partial t}\frac{dx}{ds}.$$

From these two linear equations, $\partial\psi/\partial t$ and $\partial\psi/\partial x$ can be determined on $\mathcal{C}$, provided

$$\begin{vmatrix} \frac{dx}{ds} & \frac{dt}{ds} \\ \frac{dt}{ds} & -\frac{dx}{ds} \end{vmatrix} = -\left(\frac{dx}{ds}\right)^2 - \left(\frac{dt}{ds}\right)^2 \neq 0.$$

For the second derivatives we use the chain rule again:

$$\frac{d}{ds}\frac{\partial\psi}{\partial x} = \frac{dx}{ds}\frac{\partial^2\psi}{\partial x^2} + \frac{dt}{ds}\frac{\partial^2\psi}{\partial x\partial t}, \tag{9.7a}$$

$$\frac{d}{ds}\frac{\partial\psi}{\partial t} = \frac{dx}{ds}\frac{\partial^2\psi}{\partial x\partial t} + \frac{dt}{ds}\frac{\partial^2\psi}{\partial t^2}. \tag{9.7b}$$

From our PDE (9.2), and Eqs. (9.7a,b), which are linear in the second-order derivatives, they cannot be calculated when the determinant vanishes, that is,

$$\begin{vmatrix} a & 2b & c \\ \frac{dx}{ds} & \frac{dt}{ds} & 0 \\ 0 & \frac{dx}{ds} & \frac{dt}{ds} \end{vmatrix} = a\left(\frac{dt}{ds}\right)^2 - 2b\frac{dx}{ds}\frac{dt}{ds} + c\left(\frac{dx}{ds}\right)^2 = 0. \tag{9.8}$$

From Eq. (9.8), which defines the characteristics, we find that the tangent ratio dx/dt obeys

$$c\left(\frac{dx}{dt}\right)^2 - 2b\frac{dx}{dt} + a = 0,$$

so

$$\frac{dx}{dt} = \frac{1}{c}\left[b \pm \left(b^2 - ac\right)^{1/2}\right]. \tag{9.9}$$

For the earlier hyperbolic wave (*and elliptic potential*) *equation examples,* $b = 0$ *and* a, c *are constants, so the solutions* $\xi_i = x + tr_i$ *from Eq.* (9.6) *coincide with the characteristics of Eq.* (9.9).

Nonlinear PDEs

Nonlinear ODEs and PDEs are a rapidly growing and important field. We encountered earlier the simplest linear wave equation,

$$\frac{\partial\psi}{\partial t} + c\frac{\partial\psi}{\partial x} = 0,$$

as the first-order PDE of the wave fronts of the wave equation. The simplest nonlinear wave equation,

$$\frac{\partial\psi}{\partial t} + c(\psi)\frac{\partial\psi}{\partial x} = 0, \tag{9.10}$$

results if the local speed of propagation, c, is not constant but depends on the wave ψ. When a nonlinear equation has a solution of the form $\psi(x,t) = \mathrm{A}\cos(kx - \omega t)$, where

$\omega(k)$ varies with k so that $\omega''(k) \neq 0$, then it is called **dispersive**. Perhaps the best-known nonlinear dispersive equation is the **Korteweg–deVries** equation,

$$\frac{\partial \psi}{\partial t} + \psi \frac{\partial \psi}{\partial x} + \frac{\partial^3 \psi}{\partial x^3} = 0, \tag{9.11}$$

which models the lossless propagation of shallow water waves and other phenomena. It is widely known for its **soliton** solutions. A soliton is a traveling wave with the property of persisting through an interaction with another soliton: After they pass through each other, they emerge in the same shape and with the same velocity and acquire no more than a phase shift. Let $\psi(\xi = x - ct)$ be such a traveling wave. When substituted into Eq. (9.11) this yields the nonlinear ODE

$$(\psi - c)\frac{d\psi}{d\xi} + \frac{d^3\psi}{d\xi^3} = 0, \tag{9.12}$$

which can be integrated to yield

$$\frac{d^2\psi}{d\xi^2} = c\psi - \frac{\psi^2}{2}. \tag{9.13}$$

There is no additive integration constant in Eq. (9.13) to ensure that $d^2\psi/d\xi^2 \to 0$ with $\psi \to 0$ for large ξ, so ψ is localized at the characteristic $\xi = 0$, or $x = ct$. Multiplying Eq. (9.13) by $d\psi/d\xi$ and integrating again yields

$$\left(\frac{d\psi}{d\xi}\right)^2 = c\psi^2 - \frac{\psi^3}{3}, \tag{9.14}$$

where $d\psi/d\xi \to 0$ for large ξ. Taking the root of Eq. (9.14) and integrating once more yields the soliton solution

$$\psi(x - ct) = \frac{3c}{\cosh^2\left(\sqrt{c}\frac{x-ct}{2}\right)}. \tag{9.15}$$

Some nonlinear topics, for example, the logistic equation and the onset of chaos, are reviewed in Chapter 18. For more details and literature, see J. Guckenheimer, P. Holmes, and F. John, *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*, rev. ed., New York: Springer-Verlag (1990).

Boundary Conditions

Usually, when we know a physical system at some time and the law governing the physical process, then we are able to predict the subsequent development. Such initial values are the most common boundary conditions associated with ODEs and PDE. Finding solutions that match given points, curves, or surfaces corresponds to boundary value problems. Solutions usually are required to satisfy certain imposed (for example, asymptotic) boundary conditions. These boundary conditions may take three forms:

1. **Cauchy boundary conditions**. The value of a function and normal derivative specified on the boundary. In electrostatics this would mean φ, the potential, and E_n, the normal component of the electric field.

Table 9.1

Boundary conditions	Type of partial differential equation		
	Elliptic	Hyperbolic	Parabolic
	Laplace, Poisson in (x, y)	Wave equation in (x, t)	Diffusion equation in (x, t)
Cauchy			
Open surface	Unphysical results (instability)	**Unique, stable solution**	Too restrictive
Closed surface	Too restrictive	Too restrictive	Too restrictive
Dirichlet			
Open surface	Insufficient	Insufficient	**Unique, stable solution** in one direction
Closed surface	**Unique, stable solution**	Solution not unique	Too restrictive
Neumann			
Open surface	Insufficient	Insufficient	**Unique, stable solution** in one direction
Closed surface	**Unique, stable solution**	Solution not unique	Too restrictive

2. **Dirichlet boundary conditions**. The value of a function specified on the boundary.
3. **Neumann boundary conditions**. The normal derivative (normal gradient) of a function specified on the boundary. In the electrostatic case this would be E_n and therefore σ, the surface charge density.

A summary of the relation of these three types of boundary conditions to the three types of two-dimensional partial differential equations is given in Table 9.1. For extended discussions of these partial differential equations the reader may consult Morse and Feshbach, Chapter 6 (see Additional Readings).

Parts of Table 9.1 are simply a matter of maintaining internal consistency or of common sense. For instance, for Poisson's equation with a closed surface, Dirichlet conditions lead to a unique, stable solution. Neumann conditions, independent of the Dirichlet conditions, likewise lead to a unique stable solution independent of the Dirichlet solution. Therefore Cauchy boundary conditions (meaning Dirichlet plus Neumann) could lead to an inconsistency.

The term **boundary conditions** includes as a special case the concept of **initial conditions**. For instance, specifying the initial position x_0 and the initial velocity v_0 in some dynamical problem would correspond to the Cauchy boundary conditions. The only difference in the present usage of boundary conditions in these one-dimensional problems is that we are going to apply the conditions on **both** ends of the allowed range of the variable.

9.2 First-Order Differential Equations

Physics involves some first-order differential equations. For completeness (and review) it seems desirable to touch on them briefly. We consider here differential equations of the

general form

$$\frac{dy}{dx} = f(x, y) = -\frac{P(x, y)}{Q(x, y)}. \tag{9.16}$$

Equation (9.16) is clearly a first-order, ordinary differential equation. It is **first order** because it contains the first and no higher derivatives. It is **ordinary** because the only derivative, dy/dx, is an ordinary, or total, derivative. Equation (9.16) may or may not be **linear**, although we shall treat the linear case explicitly later, Eq. (9.25).

Separable Variables

Frequently Eq. (9.16) will have the special form

$$\frac{dy}{dx} = f(x, y) = -\frac{P(x)}{Q(y)}. \tag{9.17}$$

Then it may be rewritten as

$$P(x)\,dx + Q(y)\,dy = 0.$$

Integrating from (x_0, y_0) to (x, y) yields

$$\int_{x_0}^{x} P(x)\,dx + \int_{y_0}^{y} Q(y)\,dy = 0.$$

Since the lower limits, x_0 and y_0, contribute constants, we may ignore the lower limits of integration and simply add a constant of integration. Note that this separation of variables technique does **not** require that the differential equation be linear.

Example 9.2.1 PARACHUTIST

We want to find the velocity of the falling parachutist as a function of time and are particularly interested in the constant limiting velocity, v_0, that comes about by air drag, taken, to be quadratic, $-bv^2$, and opposing the force of the gravitational attraction, mg, of the Earth. We choose a coordinate system in which the positive direction is downward so that the gravitational force is positive. For simplicity we assume that the parachute opens immediately, that is, at time $t = 0$, where $v(t = 0) = 0$, our initial condition. Newton's law applied to the falling parachutist gives

$$m\dot{v} = mg - bv^2,$$

where m includes the mass of the parachute.

The terminal velocity, v_0, can be found from the equation of motion as $t \to \infty$; when there is no acceleration, $\dot{v} = 0$, so

$$bv_0^2 = mg, \qquad \text{or} \qquad v_0 = \sqrt{\frac{mg}{b}}.$$

The variables t and v separate

$$\frac{dv}{g - \frac{b}{m}v^2} = dt,$$

which we integrate by decomposing the denominator into partial fractions. The roots of the denominator are at $v = \pm v_0$. Hence

$$\left(g - \frac{b}{m}v^2\right)^{-1} = \frac{m}{2v_0 b}\left(\frac{1}{v + v_0} - \frac{1}{v - v_0}\right).$$

Integrating both terms yields

$$\int^{v} \frac{dV}{g - \frac{b}{m}V^2} = \frac{1}{2}\sqrt{\frac{m}{gb}} \ln \frac{v_0 + v}{v_0 - v} = t.$$

Solving for the velocity yields

$$v = \frac{e^{2t/T} - 1}{e^{2t/T} + 1} v_0 = v_0 \frac{\sinh \frac{t}{T}}{\cosh \frac{t}{T}} = v_0 \tanh \frac{t}{T},$$

where $T = \sqrt{\frac{m}{gb}}$ is the time constant governing the asymptotic approach of the velocity to the limiting velocity, v_0.

Putting in numerical values, $g = 9.8$ m/s^2 and taking $b = 700$ kg/m, $m = 70$ kg, gives $v_0 = \sqrt{9.8/10} \sim 1$ m/s ~ 3.6 km/h ~ 2.23 mi/h, the walking speed of a pedestrian at landing, and $T = \sqrt{\frac{m}{bg}} = 1/\sqrt{10 \cdot 9.8} \sim 0.1$ s. Thus, the constant speed v_0 is reached within a second. Finally, because **it is always important to check the solution**, we verify that our solution satisfies

$$\dot{v} = \frac{\cosh t/T}{\cosh t/T}\frac{v_0}{T} - \frac{\sinh^2 t/T}{\cosh^2 t/T}\frac{v_0}{T} = \frac{v_0}{T} - \frac{v^2}{T v_0} = g - \frac{b}{m}v^2,$$

that is, Newton's equation of motion. The more realistic case, where the parachutist is in free fall with an initial speed $v_i = v(0) > 0$ before the parachute opens, is addressed in Exercise 9.2.18. ■

Exact Differential Equations

We rewrite Eq. (9.16) as

$$P(x, y)\, dx + Q(x, y)\, dy = 0. \tag{9.18}$$

This equation is said to be **exact** if we can match the left-hand side of it to a differential $d\varphi$,

$$d\varphi = \frac{\partial \varphi}{\partial x} dx + \frac{\partial \varphi}{\partial y} dy. \tag{9.19}$$

Since Eq. (9.18) has a zero on the right, we look for an unknown function $\varphi(x, y) =$ constant and $d\varphi = 0$.

We have (if such a function $\varphi(x, y)$ exists)

$$P(x, y)\,dx + Q(x, y)\,dy = \frac{\partial \varphi}{\partial x}dx + \frac{\partial \varphi}{\partial y}dy \tag{9.20a}$$

and

$$\frac{\partial \varphi}{\partial x} = P(x, y), \qquad \frac{\partial \varphi}{\partial y} = Q(x, y). \tag{9.20b}$$

The necessary and sufficient condition for our equation to be exact is that the second, mixed partial derivatives of $\varphi(x, y)$ (assumed continuous) are independent of the order of differentiation:

$$\frac{\partial^2 \varphi}{\partial y \partial x} = \frac{\partial P(x, y)}{\partial y} = \frac{\partial Q(x, y)}{\partial x} = \frac{\partial^2 \varphi}{\partial x \partial y}. \tag{9.21}$$

Note the resemblance to Eqs. (1.133a) of Section 1.13, "Potential Theory." If Eq. (9.18) corresponds to a curl (equal to zero), then a potential, $\varphi(x, y)$, must exist.

If $\varphi(x, y)$ exists, then from Eqs. (9.18) and (9.20a) our solution is

$$\varphi(x, y) = C.$$

We may construct $\varphi(x, y)$ from its partial derivatives just as we constructed a magnetic vector potential in Section 1.13 from its curl. See Exercises 9.2.7 and 9.2.8.

It may well turn out that Eq. (9.18) is not exact and that Eq. (9.21) is not satisfied. However, there always exists at least one and perhaps an infinity of integrating factors $\alpha(x, y)$ such that

$$\alpha(x, y)P(x, y)\,dx + \alpha(x, y)Q(x, y)\,dy = 0$$

is exact. Unfortunately, an integrating factor is not always obvious or easy to find. Unlike the case of the linear first-order differential equation to be considered next, there is no systematic way to develop an integrating factor for Eq. (9.18).

A differential equation in which the variables have been separated is automatically exact. An exact differential equation is not necessarily separable.

The wave front method of Section 9.1 also works for a first-order PDE:

$$a(x, y)\frac{\partial \psi}{\partial x} + b(x, y)\frac{\partial \psi}{\partial y} = 0. \tag{9.22a}$$

We look for a solution of the form $\psi = F(\xi)$, where $\xi(x, y) =$ constant for varying x and y defines the wave front. Hence

$$d\xi = \frac{\partial \xi}{\partial x}dx + \frac{\partial \xi}{\partial y}dy = 0, \tag{9.22b}$$

while the PDE yields

$$\left(a\frac{\partial \xi}{\partial x} + b\frac{\partial \xi}{\partial y}\right)\frac{dF}{d\xi} = 0 \tag{9.23a}$$

with $dF/d\xi \neq 0$ in general. Comparing Eqs. (9.22b) and (9.23a) yields

$$\frac{dx}{a} = \frac{dy}{b}, \tag{9.23b}$$

which reduces the PDE to a first-order ODE for the tangent dy/dx of the wave front function $\xi(x, y)$.

When there is an additional source term in the PDE,

$$a\frac{\partial\psi}{\partial x} + b\frac{\partial\psi}{\partial y} + c\psi = 0, \tag{9.23c}$$

then we use the Ansatz $\psi = \psi_0(x, y)F(\xi)$, which converts our PDE to

$$F\left(a\frac{\partial\psi_0}{\partial x} + b\frac{\partial\psi_0}{\partial y} + c\psi_0\right) + \psi_0\frac{dF}{d\xi}\left(a\frac{\partial\xi}{\partial x} + b\frac{\partial\xi}{\partial y}\right) = 0. \tag{9.24}$$

If we can guess a solution ψ_0 of Eq. (9.23c), then Eq. (9.24) reduces to our previous equation, Eq. (9.23a), from which the ODE of Eq. (9.23b) follows.

Linear First-Order ODEs

If $f(x, y)$ in Eq. (9.16) has the form $-p(x)y + q(x)$, then Eq. (9.16) becomes

$$\frac{dy}{dx} + p(x)y = q(x). \tag{9.25}$$

Equation (9.25) is the most general **linear** first-order ODE. If $q(x) = 0$, Eq. (9.25) is homogeneous (in y). A nonzero $q(x)$ may represent a **source** or a **driving term**. Equation (9.25) is **linear**; each term is linear in y or dy/dx. There are no higher powers, that is, y^2, and no products, $y(dy/dx)$. Note that the linearity refers to the y and dy/dx; $p(x)$ and $q(x)$ need not be linear in x. Equation (9.25), the most important of these first-order ODEs for physics, may be solved exactly.

Let us look for an **integrating factor** $\alpha(x)$ so that

$$\alpha(x)\frac{dy}{dx} + \alpha(x)p(x)y = \alpha(x)q(x) \tag{9.26}$$

may be rewritten as

$$\frac{d}{dx}\big[\alpha(x)y\big] = \alpha(x)q(x). \tag{9.27}$$

The purpose of this is to make the left-hand side of Eq. (9.25) a derivative so that it can be integrated—by inspection. It also, incidentally, makes Eq. (9.25) exact. Expanding Eq. (9.27), we obtain

$$\alpha(x)\frac{dy}{dx} + \frac{d\alpha}{dx}y = \alpha(x)q(x).$$

Comparison with Eq. (9.26) shows that we must require

$$\frac{d\alpha}{dx} = \alpha(x)p(x). \tag{9.28}$$

Here is a differential equation for $\alpha(x)$, with the variables α and x **separable**. We separate variables, integrate, and obtain

$$\alpha(x) = \exp\left[\int^x p(x)\,dx\right] \tag{9.29}$$

as our integrating factor.

With $\alpha(x)$ known we proceed to integrate Eq. (9.27). This, of course, was the point of introducing α in the first place. We have

$$\int^x \frac{d}{dx}\big[\alpha(x)y(x)\big]\,dx = \int^x \alpha(x)q(x)\,dx.$$

Now integrating by inspection, we have

$$\alpha(x)y(x) = \int^x \alpha(x)q(x)\,dx + C.$$

The constants from a constant lower limit of integration are lumped into the constant C. Dividing by $\alpha(x)$, we obtain

$$y(x) = \big[\alpha(x)\big]^{-1}\left\{\int^x \alpha(x)q(x)\,dx + C\right\}.$$

Finally, substituting in Eq. (9.29) for α yields

$$y(x) = \exp\left[-\int^x p(t)\,dt\right]\left\{\int^x \exp\left[\int^s p(t)\,dt\right]q(s)\,ds + C\right\}. \tag{9.30}$$

Here the (dummy) variables of integration have been rewritten to make them unambiguous. Equation (9.30) is the complete general solution of the linear, first-order differential equation, Eq. (9.25). The portion

$$y_1(x) = C\exp\left[-\int^x p(t)\,dt\right] \tag{9.31}$$

corresponds to the case $q(x) = 0$ and is a general solution of the homogeneous differential equation. The other term in Eq. (9.30),

$$y_2(x) = \exp\left[-\int^x p(t)\,dt\right]\int^x \exp\left[\int^s p(t)\,dt\right]q(s)\,ds, \tag{9.32}$$

is a **particular solution corresponding to the specific source term** $q(x)$.

Note that if our linear first-order differential equation is homogeneous ($q = 0$), then it is separable. Otherwise, apart from special cases such as $p =$ constant, $q =$ constant, and $q(x) = ap(x)$, Eq. (9.25) is not separable.

Let us summarize this solution of the inhomogeneous ODE in terms of a **method called variation of the constant** as follows. In the first step, we solve the homogeneous ODE by separation of variables as before, giving

$$\frac{y'}{y} = -p, \qquad \ln y = -\int^x p(X)\,dX + \ln C, \qquad y(x) = Ce^{-\int^x p(X)\,dX}.$$

In the second step, we let the integration constant become x-dependent, that is, $C \to C(x)$. This is the "variation of the constant" used to solve the inhomogeneous ODE. Differentiating $y(x)$ we obtain

$$y' = -pCe^{-\int p(x)\,dx} + C'(x)e^{-\int p(x)\,dx} = -py(x) + C'(x)e^{-\int p(x)\,dx}.$$

Comparing with the inhomogeneous ODE we find the ODE for C:

$$C'e^{-\int p(x)\,dx} = q, \qquad \text{or} \qquad C(x) = \int^x e^{\int^X p(Y)\,dY} q(X)\,dX.$$

Substituting this C into $y = C(x)e^{-\int^x p(X)\,dX}$ reproduces Eq. (9.32).

Example 9.2.2 RL CIRCUIT

For a resistance-inductance circuit Kirchhoff's law leads to

$$\boxed{L\frac{dI(t)}{dt} + RI(t) = V(t)}$$

for the current $I(t)$, where L is the inductance and R is the resistance, both constant. $V(t)$ is the time-dependent input voltage.

From Eq. (9.29) our integrating factor $\alpha(t)$ is

$$\alpha(t) = \exp\int^t \frac{R}{L}dt = e^{Rt/L}.$$

Then by Eq. (9.30),

$$I(t) = e^{-Rt/L}\left[\int^t e^{Rt/L}\frac{V(t)}{L}\,dt + C\right],$$

with the constant C to be determined by an initial condition (a boundary condition).

For the special case $V(t) = V_0$, a constant,

$$I(t) = e^{-Rt/L}\left[\frac{V_0}{L}\cdot\frac{L}{R}e^{Rt/L} + C\right] = \frac{V_0}{R} + Ce^{-Rt/L}.$$

If the initial condition is $I(0) = 0$, then $C = -V_0/R$ and

$$\boxed{I(t) = \frac{V_0}{R}\left[1 - e^{-Rt/L}\right].}$$

■

Now we prove the *theorem* that **the solution of the inhomogeneous ODE is unique up to an arbitrary multiple of the solution of the homogeneous ODE**.

To show this, suppose $y_1,\ y_2$ both solve the inhomogeneous ODE, Eq. (9.25); then

$$y_1' - y_2' + p(x)(y_1 - y_2) = 0$$

follows by subtracting the ODEs and says that $y_1 - y_2$ is a solution of the homogeneous ODE. The solution of the homogeneous ODE can always be multiplied by an arbitrary constant.

We also prove the *theorem* that **a first-order homogeneous ODE has only one linearly independent solution**. This is meant in the following sense. If two solutions are **linearly dependent**, by definition they satisfy $ay_1(x) + by_2(x) = 0$ with nonzero constants $a,\ b$ for all values of x. If the only solution of this linear relation is $a = 0 = b$, then our solutions y_1 and y_2 are said to be **linearly independent**.

To prove this theorem, suppose $y_1,\ y_2$ both solve the homogeneous ODE. Then

$$\frac{y_1'}{y_1} = -p(x) = \frac{y_2'}{y_2} \qquad \text{implies} \qquad W(x) \equiv y_1' y_2 - y_1 y_2' \equiv 0. \tag{9.33}$$

The functional determinant W is called the **Wronskian of the pair** $y_1,\ y_2$. We now show that $W \equiv 0$ is the condition for them to be linearly dependent. Assuming linear dependence, that is,

$$a y_1(x) + b y_2(x) = 0$$

with nonzero constants $a,\ b$ for all values of x, we differentiate this linear relation to get another linear relation,

$$a y_1'(x) + b y_2'(x) = 0.$$

The condition for these two homogeneous linear equations in the unknowns a, b to have a nontrivial solution is that their determinant be zero, which is $W = 0$.

Conversely, from $W = 0$, there follows linear dependence, because we can find a nontrivial solution of the relation

$$\frac{y_1'}{y_1} = \frac{y_2'}{y_2}$$

by integration, which gives

$$\ln y_1 = \ln y_2 + \ln C, \qquad \text{or} \qquad y_1 = C y_2.$$

Linear dependence and the Wronskian are generalized to three or more functions in Section 9.6.

Exercises

9.2.1 From Kirchhoff's law the current I in an RC (resistance–capacitance) circuit (Fig. 9.1) obeys the equation

$$R\frac{dI}{dt} + \frac{1}{C} I = 0.$$

(a) Find $I(t)$.
(b) For a capacitance of 10,000 μF charged to 100 V and discharging through a resistance of 1 MΩ, find the current I for $t = 0$ and for $t = 100$ seconds.

Note. The initial voltage is $I_0 R$ or Q/C, where $Q = \int_0^\infty I(t)\, dt$.

9.2.2 The Laplace transform of Bessel's equation ($n = 0$) leads to

$$\left(s^2 + 1\right) f'(s) + s f(s) = 0.$$

Solve for $f(s)$.

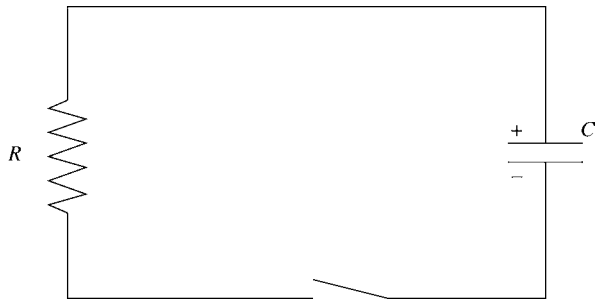

FIGURE 9.1 RC circuit.

9.2.3 The decay of a population by catastrophic two-body collisions is described by

$$\frac{dN}{dt} = -kN^2.$$

This is a first-order, **nonlinear** differential equation. Derive the solution

$$N(t) = N_0\left(1 + \frac{t}{\tau_0}\right)^{-1},$$

where $\tau_0 = (kN_0)^{-1}$. This implies an infinite population at $t = -\tau_0$.

9.2.4 The rate of a particular chemical reaction $A + B \to C$ is proportional to the concentrations of the reactants A and B:

$$\frac{dC(t)}{dt} = \alpha\big[A(0) - C(t)\big]\big[B(0) - C(t)\big].$$

(a) Find $C(t)$ for $A(0) \neq B(0)$.
(b) Find $C(t)$ for $A(0) = B(0)$.

The initial condition is that $C(0) = 0$.

9.2.5 A boat, coasting through the water, experiences a resisting force proportional to v^n, v being the boat's instantaneous velocity. Newton's second law leads to

$$m\frac{dv}{dt} = -kv^n.$$

With $v(t = 0) = v_0$, $x(t = 0) = 0$, integrate to find v as a function of time and v as a function of distance.

9.2.6 In the first-order differential equation $dy/dx = f(x, y)$ the function $f(x, y)$ is a function of the ratio y/x:

$$\frac{dy}{dx} = g(y/x).$$

Show that the substitution of $u = y/x$ leads to a separable equation in u and x.

9.2.7 The differential equation

$$P(x, y)\,dx + Q(x, y)\,dy = 0$$

is **exact**. Construct a solution

$$\varphi(x, y) = \int_{x_0}^{x} P(x, y)\,dx + \int_{y_0}^{y} Q(x_0, y)\,dy = \text{constant}.$$

9.2.8 The differential equation

$$P(x, y)\,dx + Q(x, y)\,dy = 0$$

is **exact**. If

$$\varphi(x, y) = \int_{x_0}^{x} P(x, y)\,dx + \int_{y_0}^{y} Q(x_0, y)\,dy,$$

show that

$$\frac{\partial \varphi}{\partial x} = P(x, y), \qquad \frac{\partial \varphi}{\partial y} = Q(x, y).$$

Hence $\varphi(x, y) =$ constant is a solution of the original differential equation.

9.2.9 Prove that Eq. (9.26) is exact in the sense of Eq. (9.21), provided that $\alpha(x)$ satisfies Eq. (9.28).

9.2.10 A certain differential equation has the form

$$f(x)\,dx + g(x)h(y)\,dy = 0,$$

with none of the functions $f(x)$, $g(x)$, $h(y)$ identically zero. Show that a necessary and sufficient condition for this equation to be exact is that $g(x) =$ constant.

9.2.11 Show that

$$y(x) = \exp\left[-\int^{x} p(t)\,dt\right]\left\{\int^{x} \exp\left[\int^{s} p(t)\,dt\right] q(s)\,ds + C\right\}$$

is a solution of

$$\frac{dy}{dx} + p(x)y(x) = q(x)$$

by differentiating the expression for $y(x)$ and substituting into the differential equation.

9.2.12 The motion of a body falling in a resisting medium may be described by

$$m\frac{dv}{dt} = mg - bv$$

when the retarding force is proportional to the velocity, v. Find the velocity. Evaluate the constant of integration by demanding that $v(0) = 0$.

9.2.13 Radioactive nuclei decay according to the law

$$\frac{dN}{dt} = -\lambda N,$$

N being the concentration of a given nuclide and λ, the particular decay constant. In a radioactive series of n different nuclides, starting with N_1,

$$\frac{dN_1}{dt} = -\lambda_1 N_1,$$
$$\frac{dN_2}{dt} = \lambda_1 N_1 - \lambda_2 N_2, \qquad \text{and so on.}$$

Find $N_2(t)$ for the conditions $N_1(0) = N_0$ and $N_2(0) = 0$.

9.2.14 The rate of evaporation from a particular spherical drop of liquid (constant density) is proportional to its surface area. Assuming this to be the sole mechanism of mass loss, find the radius of the drop as a function of time.

9.2.15 In the linear homogeneous differential equation

$$\frac{dv}{dt} = -av$$

the variables are separable. When the variables are separated, the equation is exact. Solve this differential equation subject to $v(0) = v_0$ by the following three methods:

(a) Separating variables and integrating.
(b) Treating the separated variable equation as exact.
(c) Using the result for a linear homogeneous differential equation.

ANS. $v(t) = v_0 e^{-at}$.

9.2.16 Bernoulli's equation,

$$\frac{dy}{dx} + f(x)y = g(x)y^n,$$

is nonlinear for $n \neq 0$ or 1. Show that the substitution $u = y^{1-n}$ reduces Bernoulli's equation to a linear equation. (See Section 18.4.)

ANS. $\dfrac{du}{dx} + (1-n)f(x)u = (1-n)g(x)$.

9.2.17 Solve the linear, first-order equation, Eq. (9.25), by assuming $y(x) = u(x)v(x)$, where $v(x)$ is a solution of the corresponding homogeneous equation [$q(x) = 0$]. This is the method of **variation of parameters** due to Lagrange. We apply it to second-order equations in Exercise 9.6.25.

9.2.18 (a) Solve Example 9.2.1 for an initial velocity $v_i = 60$ mi/h, when the parachute opens. Find $v(t)$. (b) For a skydiver in free fall use the friction coefficient $b = 0.25$ kg/m and mass $m = 70$ kg. What is the limiting velocity in this case?

9.3 SEPARATION OF VARIABLES

The equations of mathematical physics listed in Section 9.1 are all partial differential equations. Our first technique for their solution splits the partial differential equation of n variables into n ordinary differential equations. Each separation introduces an arbitrary constant of separation. If we have n variables, we have to introduce $n-1$ constants, determined by the conditions imposed in the problem being solved.

Cartesian Coordinates

In Cartesian coordinates the Helmholtz equation becomes

$$\boxed{\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}+k^2\psi=0,} \tag{9.34}$$

using Eq. (2.27) for the Laplacian. For the present let k^2 be a constant. Perhaps the simplest way of treating a partial differential equation such as Eq. (9.34) is to split it into a set of ordinary differential equations. This may be done as follows. Let

$$\boxed{\psi(x,y,z)=X(x)Y(y)Z(z)} \tag{9.35}$$

and substitute back into Eq. (9.34). How do we know Eq. (9.35) is valid? When the differential operators in various variables are additive in the PDE, that is, when there are no products of differential operators in different variables, the separation method usually works. We are proceeding in the spirit of let's try and see if it works. If our attempt succeeds, then Eq. (9.35) will be justified. If it does not succeed, we shall find out soon enough and then we shall try another attack, such as Green's functions, integral transforms, or brute-force numerical analysis. With ψ assumed given by Eq. (9.35), Eq. (9.34) becomes

$$YZ\frac{d^2X}{dx^2}+XZ\frac{d^2Y}{dy^2}+XY\frac{d^2Z}{dz^2}+k^2XYZ=0. \tag{9.36}$$

Dividing by $\psi=XYZ$ and rearranging terms, we obtain

$$\frac{1}{X}\frac{d^2X}{dx^2}=-k^2-\frac{1}{Y}\frac{d^2Y}{dy^2}-\frac{1}{Z}\frac{d^2Z}{dz^2}. \tag{9.37}$$

Equation (9.37) exhibits one separation of variables. The left-hand side is a function of x alone, whereas the right-hand side depends only on y and z and not on x. But x, y, and z are all independent coordinates. The equality of both sides depending on different variables means that the behavior of x as an independent variable is not determined by y and z. Therefore, each side must be equal to a constant, a constant of separation. We choose[2]

$$\frac{1}{X}\frac{d^2X}{dx^2}=-l^2, \tag{9.38}$$

$$-k^2-\frac{1}{Y}\frac{d^2Y}{dy^2}-\frac{1}{Z}\frac{d^2Z}{dz^2}=-l^2. \tag{9.39}$$

[2]The choice of sign, completely arbitrary here, will be fixed in specific problems by the need to satisfy specific boundary conditions.

Now, turning our attention to Eq. (9.39), we obtain

$$\frac{1}{Y}\frac{d^2Y}{dy^2} = -k^2 + l^2 - \frac{1}{Z}\frac{d^2Z}{dz^2}, \tag{9.40}$$

and a second separation has been achieved. Here we have a function of y equated to a function of z, as before. We resolve it, as before, by equating each side to another constant of separation,[2] $-m^2$,

$$\frac{1}{Y}\frac{d^2Y}{dy^2} = -m^2, \tag{9.41}$$

$$\frac{1}{Z}\frac{d^2Z}{dz^2} = -k^2 + l^2 + m^2 = -n^2, \tag{9.42}$$

introducing a constant n^2 by $k^2 = l^2 + m^2 + n^2$ to produce a symmetric set of equations. Now we have three ODEs ((9.38), (9.41), and (9.42)) to replace Eq. (9.34). Our assumption (Eq. (9.35)) has succeeded and is thereby justified.

Our solution should be labeled according to the choice of our constants l, m, and n; that is,

$$\psi_{lm}(x, y, z) = X_l(x)Y_m(y)Z_n(z). \tag{9.43}$$

Subject to the conditions of the problem being solved and to the condition $k^2 = l^2 + m^2 + n^2$, we may choose l, m, and n as we like, and Eq. (9.43) will still be a solution of Eq. (9.34), provided $X_l(x)$ is a solution of Eq. (9.38), and so on. We may develop **the most general solution** of Eq. (9.34) by taking a **linear combination of solutions** ψ_{lm},

$$\Psi = \sum_{l,m} a_{lm}\psi_{lm}. \tag{9.44}$$

The constant coefficients a_{lm} are finally chosen to permit Ψ to satisfy the boundary conditions of the problem, which, as a rule, lead to a discrete set of values l, m.

Circular Cylindrical Coordinates

With our unknown function ψ dependent on ρ, φ, and z, the Helmholtz equation becomes (see Section 2.4 for ∇^2)

$$\nabla^2\psi(\rho, \varphi, z) + k^2\psi(\rho, \varphi, z) = 0, \tag{9.45}$$

or

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial\psi}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2\psi}{\partial\varphi^2} + \frac{\partial^2\psi}{\partial z^2} + k^2\psi = 0. \tag{9.46}$$

As before, we assume a factored form for ψ,

$$\boxed{\psi(\rho, \varphi, z) = P(\rho)\Phi(\varphi)Z(z).} \tag{9.47}$$

Substituting into Eq. (9.46), we have

$$\frac{\Phi Z}{\rho}\frac{d}{d\rho}\left(\rho\frac{dP}{d\rho}\right) + \frac{PZ}{\rho^2}\frac{d^2\Phi}{d\varphi^2} + P\Phi\frac{d^2Z}{dz^2} + k^2P\Phi Z = 0. \tag{9.48}$$

All the partial derivatives have become ordinary derivatives. Dividing by $P\Phi Z$ and moving the z derivative to the right-hand side yields

$$\frac{1}{\rho P}\frac{d}{d\rho}\left(\rho\frac{dP}{d\rho}\right)+\frac{1}{\rho^2\Phi}\frac{d^2\Phi}{d\varphi^2}+k^2=-\frac{1}{Z}\frac{d^2Z}{dz^2}. \tag{9.49}$$

Again, a function of z on the right appears to depend on a function of ρ and φ on the left. We resolve this by setting each side of Eq. (9.49) equal to the same constant. Let us choose[3] $-l^2$. Then

$$\frac{d^2Z}{dz^2}=l^2Z \tag{9.50}$$

and

$$\frac{1}{\rho P}\frac{d}{d\rho}\left(\rho\frac{dP}{d\rho}\right)+\frac{1}{\rho^2\Phi}\frac{d^2\Phi}{d\varphi^2}+k^2=-l^2. \tag{9.51}$$

Setting $k^2+l^2=n^2$, multiplying by ρ^2, and rearranging terms, we obtain

$$\frac{\rho}{P}\frac{d}{d\rho}\left(\rho\frac{dP}{d\rho}\right)+n^2\rho^2=-\frac{1}{\Phi}\frac{d^2\Phi}{d\varphi^2}. \tag{9.52}$$

We may set the right-hand side to m^2 and

$$\frac{d^2\Phi}{d\varphi^2}=-m^2\Phi. \tag{9.53}$$

Finally, for the ρ dependence we have

$$\rho\frac{d}{d\rho}\left(\rho\frac{dP}{d\rho}\right)+\left(n^2\rho^2-m^2\right)P=0. \tag{9.54}$$

This is Bessel's differential equation. The solutions and their properties are presented in Chapter 11. The separation of variables of Laplace's equation in parabolic coordinates also gives rise to Bessel's equation. It may be noted that the Bessel equation is notorious for the variety of disguises it may assume. For an extensive tabulation of possible forms the reader is referred to *Tables of Functions* by Jahnke and Emde.[4]

The original Helmholtz equation, a three-dimensional PDE, has been replaced by three ODEs, Eqs. (9.50), (9.53), and (9.54). A solution of the Helmholtz equation is

$$\psi(\rho,\varphi,z)=P(\rho)\Phi(\varphi)Z(z). \tag{9.55}$$

Identifying the specific P, Φ, Z solutions by subscripts, we see that the most general solution of the Helmholtz equation is a linear combination of the product solutions:

$$\Psi(\rho,\varphi,z)=\sum_{m,n}a_{mn}P_{mn}(\rho)\Phi_m(\varphi)Z_n(z). \tag{9.56}$$

[3]The choice of sign of the separation constant is arbitrary. However, a minus sign is chosen for the axial coordinate z in expectation of a possible exponential dependence on z (from Eq. (9.50)). A positive sign is chosen for the azimuthal coordinate φ in expectation of a periodic dependence on φ (from Eq. (9.53)).

[4]E. Jahnke and F. Emde, *Tables of functions*, 4th rev. ed., New York: Dover (1945), p. 146; also, E. Jahnke, F. Emde, and F. Lösch, *Tables of Higher Functions*, 6th ed., New York: McGraw-Hill (1960).

Spherical Polar Coordinates

Let us try to separate the Helmholtz equation, again with k^2 constant, in spherical polar coordinates. Using Eq. (2.48), we obtain

$$\frac{1}{r^2\sin\theta}\left[\sin\theta\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)+\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right)+\frac{1}{\sin\theta}\frac{\partial^2\psi}{\partial\varphi^2}\right]=-k^2\psi. \tag{9.57}$$

Now, in analogy with Eq. (9.35) we try

$$\boxed{\psi(r,\theta,\varphi)=R(r)\Theta(\theta)\Phi(\varphi).} \tag{9.58}$$

By substituting back into Eq. (9.57) and dividing by $R\Theta\Phi$, we have

$$\frac{1}{Rr^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)+\frac{1}{\Theta r^2\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)+\frac{1}{\Phi r^2\sin^2\theta}\frac{d^2\Phi}{d\varphi^2}=-k^2. \tag{9.59}$$

Note that all derivatives are now ordinary derivatives rather than partials. By multiplying by $r^2\sin^2\theta$, we can isolate $(1/\Phi)(d^2\Phi/d\varphi^2)$ to obtain[5]

$$\frac{1}{\Phi}\frac{d^2\Phi}{d\varphi^2}=r^2\sin^2\theta\left[-k^2-\frac{1}{r^2R}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)-\frac{1}{r^2\sin\theta\Theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)\right]. \tag{9.60}$$

Equation (9.60) relates a function of φ alone to a function of r and θ alone. Since r,θ, and φ are independent variables, we equate each side of Eq. (9.60) to a constant. In almost all physical problems φ will appear as an azimuth angle. This suggests a periodic solution rather than an exponential. With this in mind, let us use $-m^2$ as the separation constant, which, then, must be an integer squared. Then

$$\frac{1}{\Phi}\frac{d^2\Phi(\varphi)}{d\varphi^2}=-m^2 \tag{9.61}$$

and

$$\frac{1}{r^2R}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)+\frac{1}{r^2\sin\theta\Theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)-\frac{m^2}{r^2\sin^2\theta}=-k^2. \tag{9.62}$$

Multiplying Eq. (9.62) by r^2 and rearranging terms, we obtain

$$\frac{1}{R}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)+r^2k^2=-\frac{1}{\sin\theta\Theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)+\frac{m^2}{\sin^2\theta}. \tag{9.63}$$

Again, the variables are separated. We equate each side to a constant, Q, and finally obtain

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)-\frac{m^2}{\sin^2\theta}\Theta+Q\Theta=0, \tag{9.64}$$

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)+k^2R-\frac{QR}{r^2}=0. \tag{9.65}$$

[5]The order in which the variables are separated here is not unique. Many quantum mechanics texts show the r dependence split off first.

Once more we have replaced a partial differential equation of three variables by three ODEs. The solutions of these ODEs are discussed in Chapters 11 and 12. In Chapter 12, for example, Eq. (9.64) is identified as the associated Legendre equation, in which the constant Q becomes $l(l+1)$; l is a non-negative integer because θ is an angular variable. If k^2 is a (positive) constant, Eq. (9.65) becomes the spherical Bessel equation of Section 11.7.

Again, our most general solution may be written

$$\psi_{Qm}(r,\theta,\varphi) = \sum_{Q,m} a_{Qm} R_Q(r)\Theta_{Qm}(\theta)\Phi_m(\varphi). \tag{9.66}$$

The restriction that k^2 be a constant is unnecessarily severe. The separation process will still be possible for k^2 as general as

$$k^2 = f(r) + \frac{1}{r^2}g(\theta) + \frac{1}{r^2\sin^2\theta}h(\varphi) + k'^2. \tag{9.67}$$

In the hydrogen atom problem, one of the most important examples of the Schrödinger wave equation with a closed form solution is $k^2 = f(r)$, with k^2 independent of θ, φ. Equation (9.65) for the hydrogen atom becomes the associated Laguerre equation.

The great importance of this separation of variables in spherical polar coordinates stems from the fact that the case $k^2 = k^2(r)$ covers a tremendous amount of physics: a great deal of the theories of gravitation, electrostatics, and atomic, nuclear, and particle physics. And with $k^2 = k^2(r)$, the angular dependence is isolated in Eqs. (9.61) and (9.64), **which can be solved exactly**.

Finally, as an illustration of how the constant m in Eq. (9.61) is restricted, we note that φ in cylindrical and spherical polar coordinates is an azimuth angle. If this is a classical problem, we shall certainly require that the azimuthal solution $\Phi(\varphi)$ be single-valued; that is,

$$\Phi(\varphi + 2\pi) = \Phi(\varphi). \tag{9.68}$$

This is equivalent to requiring the azimuthal solution to have a period of 2π.[6] Therefore m must be an integer. Which integer it is depends on the details of the problem. If the integer $|m| > 1$, then Φ will have the period $2\pi/m$. Whenever a coordinate corresponds to an axis of translation or to an azimuth angle, the separated equation always has the form

$$\frac{d^2\Phi(\varphi)}{d\varphi^2} = -m^2\Phi(\varphi)$$

for φ, the azimuth angle, and

$$\frac{d^2Z(z)}{dz^2} = \pm a^2 Z(z) \tag{9.69}$$

for z, an axis of translation of the cylindrical coordinate system. The solutions, of course, are $\sin az$ and $\cos az$ for $-a^2$ and the corresponding hyperbolic function (or exponentials) $\sinh az$ and $\cosh az$ for $+a^2$.

[6]This also applies in most quantum mechanical problems, but the argument is much more involved. If m is not an integer, rotation group relations and ladder operator relations (Section 4.3) are disrupted. Compare E. Merzbacher, Single valuedness of wave functions. *Am. J. Phys.* **30**: 237 (1962).

Table 9.2 Solutions in Spherical Polar Coordinates[a]

$$\psi = \sum_{l,m} a_{lm}\psi_{lm}$$

		ψ_{lm}
1.	$\nabla^2\psi = 0$	$\left\{\begin{matrix} r^l \\ r^{-l-1} \end{matrix}\right\} \left\{\begin{matrix} P_l^m(\cos\theta) \\ Q_l^m(\cos\theta) \end{matrix}\right\} \left\{\begin{matrix} \cos m\varphi \\ \sin m\varphi \end{matrix}\right\}^b$
2.	$\nabla^2\psi + k^2\psi = 0$	$\left\{\begin{matrix} j_l(kr) \\ n_l(kr) \end{matrix}\right\} \left\{\begin{matrix} P_l^m(\cos\theta) \\ Q_l^m(\cos\theta) \end{matrix}\right\} \left\{\begin{matrix} \cos m\varphi \\ \sin m\varphi \end{matrix}\right\}^b$
3.	$\nabla^2\psi - k^2\psi = 0$	$\left\{\begin{matrix} i_l(kr) \\ k_l(kr) \end{matrix}\right\} \left\{\begin{matrix} P_l^m(\cos\theta) \\ Q_l^m(\cos\theta) \end{matrix}\right\} \left\{\begin{matrix} \cos m\varphi \\ \sin m\varphi \end{matrix}\right\}^b$

[a] References for some of the functions are $P_l^m(\cos\theta)$, $m = 0$, Section 12.1; $m \neq 0$, Section 12.5; $Q_l^m(\cos\theta)$, Section 12.10; $j_l(kr)$, $n_l(kr)$, $i_l(kr)$, and $k_l(kr)$, Section 11.7.

[b] $\cos m\varphi$ and $\sin m\varphi$ may be replaced by $e^{\pm im\varphi}$.

Other occasionally encountered ODEs include the Laguerre and associated Laguerre equations from the supremely important hydrogen atom problem in quantum mechanics:

$$x\frac{d^2y}{dx^2} + (1-x)\frac{dy}{dx} + \alpha y = 0, \tag{9.70}$$

$$x\frac{d^2y}{dx^2} + (1+k-x)\frac{dy}{dx} + \alpha y = 0. \tag{9.71}$$

From the quantum mechanical theory of the linear oscillator we have Hermite's equation,

$$\frac{d^2y}{dx^2} - 2x\frac{dy}{dx} + 2\alpha y = 0. \tag{9.72}$$

Finally, from time to time we find the Chebyshev differential equation,

$$\left(1-x^2\right)\frac{d^2y}{dx^2} - x\frac{dy}{dx} + n^2 y = 0. \tag{9.73}$$

For convenient reference, the forms of the solutions of Laplace's equation, Helmholtz's equation, and the diffusion equation for spherical polar coordinates are collected in Table 9.2. The solutions of Laplace's equation in circular cylindrical coordinates are presented in Table 9.3.

General properties following from the form of the differential equations are discussed in Chapter 10. The individual solutions are developed and applied in Chapters 11–13.

The practicing physicist may and probably will meet other second-order ODEs, some of which may possibly be transformed into the examples studied here. Some of these ODEs may be solved by the techniques of Sections 9.5 and 9.6. Others may require a computer for a numerical solution.

We refer to the second edition of this text for other important coordinate systems.

- To put the separation method of solving PDEs in perspective, let us review it as a consequence of a symmetry of the PDE. Take the stationary Schrödinger equation $H\psi = E\psi$ as an example, with a potential $V(r)$ depending only on the radial distance r. Then this

Table 9.3 Solutions in Circular Cylindrical Coordinates[a]

		$\psi = \sum_{m,\alpha} a_{m\alpha}\psi_{m\alpha}$
a.	$\nabla^2\psi + \alpha^2\psi = 0$	$\psi_{m\alpha} = \left\{\begin{matrix} J_m(\alpha\rho) \\ N_m(\alpha\rho) \end{matrix}\right\} \left\{\begin{matrix} \cos m\varphi \\ \sin m\varphi \end{matrix}\right\} \left\{\begin{matrix} e^{-\alpha z} \\ e^{\alpha z} \end{matrix}\right\}$
b.	$\nabla^2\psi - \alpha^2\psi = 0$	$\psi_{m\alpha} = \left\{\begin{matrix} I_m(\alpha\rho) \\ K_m(\alpha\rho) \end{matrix}\right\} \left\{\begin{matrix} \cos m\varphi \\ \sin m\varphi \end{matrix}\right\} \left\{\begin{matrix} \cos \alpha z \\ \sin \alpha z \end{matrix}\right\}$
c.	$\nabla^2\psi = 0$	$\psi_m = \left\{\begin{matrix} \rho^m \\ \rho^{-m} \end{matrix}\right\} \left\{\begin{matrix} \cos m\varphi \\ \sin m\varphi \end{matrix}\right\}$

[a] References for the radial functions are $J_m(\alpha\rho)$, Section 11.1; $N_m(\alpha\rho)$, Section 11.3; $I_m(\alpha\rho)$ and $K_m(\alpha\rho)$, Section 11.5.

PDE is invariant under rotations that comprise the group SO(3). Its diagonal generator is the orbital angular momentum operator $L_z = -i\frac{\partial}{\partial\varphi}$, and its quadratic (Casimir) invariant is $\mathbf{L}^2$. Since both commute with H (see Section 4.3), we end up with three separate eigenvalue equations:

$$H\psi = E\psi, \qquad \mathbf{L}^2\psi = l(l+1)\psi, \qquad L_z\psi = m\psi.$$

Upon replacing L_z^2 in $\mathbf{L^2}$ by its eigenvalue m^2, the $\mathbf{L}^2$ PDE becomes Legendre's ODE, and similarly $H\psi = E\psi$ becomes the radial ODE of the separation method in spherical polar coordinates.

- For cylindrical coordinates the PDE is invariant under rotations about the z-axis only, which form a subgroup of SO(3). This invariance yields the generator $L_z = -i\partial/\partial\varphi$ and separate azimuthal ODE $L_z\psi = m\psi$, as before. If the potential V is invariant under translations along the z-axis, then the generator $-i\partial/\partial z$ gives the separate ODE in the z variable.
- In general (see Section 4.3), there are n mutually commuting generators H_i with eigenvalues m_i of the (classical) Lie group G of rank n and the corresponding Casimir invariants C_i with eigenvalues c_i (Chapter 4), which yield the separate ODEs

$$H_i\psi = m_i\psi, \qquad C_i\psi = c_i\psi$$

in addition to the (by now) radial ODE $H\psi = E\psi$.

Exercises

9.3.1 By letting the operator $\nabla^2 + k^2$ act on the general form $a_1\psi_1(x, y, z) + a_2\psi_2(x, y, z)$, show that it is linear, that is, that $(\nabla^2 + k^2)(a_1\psi_1 + a_2\psi_2) = a_1(\nabla^2 + k^2)\psi_1 + a_2(\nabla^2 + k^2)\psi_2$.

9.3.2 Show that the Helmholtz equation,

$$\nabla^2\psi + k^2\psi = 0,$$

is still separable in circular cylindrical coordinates if k^2 is generalized to $k^2 + f(\rho) + (1/\rho^2)g(\varphi) + h(z)$.

9.3.3 Separate variables in the Helmholtz equation in spherical polar coordinates, splitting off the radial dependence **first**. Show that your separated equations have the same form as Eqs. (9.61), (9.64), and (9.65).

9.3.4 Verify that

$$\nabla^2\psi(r,\theta,\varphi) + \left[k^2 + f(r) + \frac{1}{r^2}g(\theta) + \frac{1}{r^2\sin^2\theta}h(\varphi)\right]\psi(r,\theta,\varphi) = 0$$

is separable (in spherical polar coordinates). The functions f, g, and h are functions only of the variables indicated; k^2 is a constant.

9.3.5 An atomic (quantum mechanical) particle is confined inside a rectangular box of sides a, b, and c. The particle is described by a wave function ψ that satisfies the Schrödinger wave equation

$$-\frac{\hbar^2}{2m}\nabla^2\psi = E\psi.$$

The wave function is required to vanish at each surface of the box (but not to be identically zero). This condition imposes constraints on the separation constants and therefore on the energy E. What is the smallest value of E for which such a solution can be obtained?

$$\text{ANS. } E = \frac{\pi^2\hbar^2}{2m}\left(\frac{1}{a^2} + \frac{1}{b^2} + \frac{1}{c^2}\right).$$

9.3.6 For a homogeneous spherical solid with constant thermal diffusivity, K, and no heat sources, the equation of heat conduction becomes

$$\frac{\partial T(r,t)}{\partial t} = K\nabla^2 T(r,t).$$

Assume a solution of the form

$$T = R(r)T(t)$$

and separate variables. Show that the radial equation may take on the standard form

$$r^2\frac{d^2R}{dr^2} + 2r\frac{dR}{dr} + \left[\alpha^2r^2 - n(n+1)\right]R = 0; \qquad n = \text{integer}.$$

The solutions of this equation are called **spherical Bessel functions**.

9.3.7 Separate variables in the thermal diffusion equation of Exercise 9.3.6 in circular cylindrical coordinates. Assume that you can neglect end effects and take $T = T(\rho, t)$.

9.3.8 The quantum mechanical angular momentum operator is given by $\mathbf{L} = -i(\mathbf{r}\times\nabla)$. Show that

$$\mathbf{L}\cdot\mathbf{L}\psi = l(l+1)\psi$$

leads to the associated Legendre equation.

Hint. Exercises 1.9.9 and 2.5.16 may be helpful.

9.3.9 The one-dimensional Schrödinger wave equation for a particle in a potential field $V = \frac{1}{2}kx^2$ is

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + \frac{1}{2}kx^2\psi = E\psi(x).$$

(a) Using $\xi = ax$ and a constant λ, we have

$$a = \left(\frac{mk}{\hbar^2}\right)^{1/4}, \qquad \lambda = \frac{2E}{\hbar}\left(\frac{m}{k}\right)^{1/2};$$

show that

$$\frac{d^2\psi(\xi)}{d\xi^2} + \left(\lambda - \xi^2\right)\psi(\xi) = 0.$$

(b) Substituting

$$\psi(\xi) = y(\xi)e^{-\xi^2/2},$$

show that $y(\xi)$ satisfies the Hermite differential equation.

9.3.10 Verify that the following are solutions of Laplace's equation:

(a) $\psi_1 = 1/r,\ r \neq 0,$ (b) $\psi_2 = \frac{1}{2r}\ln\frac{r+z}{r-z}.$

Note. The z derivatives of $1/r$ generate the Legendre polynomials, $P_n(\cos\theta)$, Exercise 12.1.7. The z derivatives of $(1/2r)\ln[(r+z)/(r-z)]$ generate the Legendre functions, $Q_n(\cos\theta)$.

9.3.11 If Ψ is a solution of Laplace's equation, $\nabla^2\Psi = 0$, show that $\partial\Psi/\partial z$ is also a solution.

9.4 SINGULAR POINTS

In this section the concept of a singular point, or singularity (as applied to a differential equation), is introduced. The interest in this concept stems from its usefulness in (1) classifying ODEs and (2) investigating the feasibility of a series solution. This feasibility is the topic of Fuchs' theorem, Sections 9.5 and 9.6.

All the ODEs listed in Section 9.3 may be solved for d^2y/dx^2. Using the notation $d^2y/dx^2 = y''$, we have[7]

$$y'' = f(x, y, y'). \tag{9.74}$$

If we write our second-order homogeneous differential equation (in y) as

$$y'' + P(x)y' + Q(x)y = 0, \tag{9.75}$$

we are ready to define ordinary and singular points. If the functions $P(x)$ and $Q(x)$ remain finite at $x = x_0$, point $x = x_0$ is an ordinary point. However, if either $P(x)$ or $Q(x)$ (or

[7]This prime notation, $y' = dy/dx$, was introduced by Lagrange in the late 18th century as an abbreviation for Leibniz's more explicit but more cumbersome dy/dx.

both) diverges as $x \to x_0$, point x_0 is a singular point. Using Eq. (9.75), we may distinguish between two kinds of singular points.

1. If either $P(x)$ or $Q(x)$ diverges as $x \to x_0$ but $(x - x_0)P(x)$ and $(x - x_0)^2 Q(x)$ remain finite as $x \to x_0$, then $x = x_0$ is called a **regular**, or nonessential, singular point.
2. If $P(x)$ diverges faster than $1/(x - x_0)$ so that $(x - x_0)P(x)$ goes to infinity as $x \to x_0$, or $Q(x)$ diverges faster than $1/(x - x_0)^2$ so that $(x - x_0)^2 Q(x)$ goes to infinity as $x \to x_0$, then point $x = x_0$ is labeled an **irregular**, or **essential**, **singularity**.

These definitions hold for all finite values of x_0. The analysis of point $x \to \infty$ is similar to the treatment of functions of a complex variable (Section 6.6). We set $x = 1/z$, substitute into the differential equation, and then let $z \to 0$. By changing variables in the derivatives, we have

$$\frac{dy(x)}{dx} = \frac{dy(z^{-1})}{dz}\frac{dz}{dx} = -\frac{1}{x^2}\frac{dy(z^{-1})}{dz} = -z^2\frac{dy(z^{-1})}{dz}, \tag{9.76}$$

$$\begin{aligned}\frac{d^2y(x)}{dx^2} = \frac{d}{dz}\left[\frac{dy(x)}{dx}\right]\frac{dz}{dx} &= \left(-z^2\right)\left[-2z\frac{dy(z^{-1})}{dz} - z^2\frac{d^2y(z^{-1})}{dz^2}\right] \\ &= 2z^3\frac{dy(z^{-1})}{dz} + z^4\frac{d^2y(z^{-1})}{dz^2}.\end{aligned} \tag{9.77}$$

Using these results, we transform Eq. (9.75) into

$$z^4\frac{d^2y}{dz^2} + \left[2z^3 - z^2P\left(z^{-1}\right)\right]\frac{dy}{dz} + Q\left(z^{-1}\right)y = 0. \tag{9.78}$$

The behavior at $x = \infty(z = 0)$ then depends on the behavior of the new coefficients,

$$\frac{2z - P(z^{-1})}{z^2} \qquad \text{and} \qquad \frac{Q(z^{-1})}{z^4},$$

as $z \to 0$. If these two expressions remain finite, point $x = \infty$ is an ordinary point. If they diverge no more rapidly than $1/z$ and $1/z^2$, respectively, point $x = \infty$ is a regular singular point; otherwise it is an irregular singular point (an essential singularity).

Example 9.4.1

Bessel's equation is

$$x^2y'' + xy' + \left(x^2 - n^2\right)y = 0. \tag{9.79}$$

Comparing it with Eq. (9.75) we have

$$P(x) = \frac{1}{x}, \qquad Q(x) = 1 - \frac{n^2}{x^2},$$

which shows that point $x = 0$ is a regular singularity. By inspection we see that there are no other singular points in the finite range. As $x \to \infty(z \to 0)$, from Eq. (9.78) we have

Table 9.4

	Equation	Regular singularity $x =$	Irregular singularity $x =$
1.	Hypergeometric $x(x-1)y'' + [(1+a+b)x - c]y' + aby = 0.$	$0, 1, \infty$	–
2.	Legendre[a] $(1-x^2)y'' - 2xy' + l(l+1)y = 0.$	$-1, 1, \infty$	–
3.	Chebyshev $(1-x^2)y'' - xy' + n^2 y = 0.$	$-1, 1, \infty$	–
4.	Confluent hypergeometric $xy'' + (c-x)y' - ay = 0.$	0	∞
5.	Bessel $x^2y'' + xy' + (x^2 - n^2)y = 0.$	0	∞
6.	Laguerre[a] $xy'' + (1-x)y' + ay = 0.$	0	∞
7.	Simple harmonic oscillator $y'' + \omega^2 y = 0.$	–	∞
8.	Hermite $y'' - 2xy' + 2\alpha y = 0.$	–	∞

[a] The associated equations have the same singular points.

the coefficients

$$\frac{2z - z}{z^2} \qquad \text{and} \qquad \frac{1 - n^2 z^2}{z^4}.$$

Since the latter expression diverges as z^4, point $x = \infty$ is an irregular, or essential, singularity. ■

The ordinary differential equations of Section 9.3, plus two others, the hypergeometric and the confluent hypergeometric, have singular points, as shown in Table 9.4.

It will be seen that the first three equations in Table 9.4, hypergeometric, Legendre, and Chebyshev, all have three regular singular points. The hypergeometric equation, with regular singularities at 0, 1, and ∞ is taken as the standard, the canonical form. The solutions of the other two may then be expressed in terms of its solutions, the hypergeometric functions. This is done in Chapter 13.

In a similar manner, the confluent hypergeometric equation is taken as the canonical form of a linear second-order differential equation with one regular and one irregular singular point.

Exercises

9.4.1 Show that Legendre's equation has regular singularities at $x = -1$, 1, and ∞.

9.4.2 Show that Laguerre's equation, like the Bessel equation, has a regular singularity at $x = 0$ and an irregular singularity at $x = \infty$.

9.4.3 Show that the substitution

$$x \to \frac{1-x}{2}, \qquad a = -l, \qquad b = l+1, \qquad c = 1$$

converts the hypergeometric equation into Legendre's equation.

9.5 SERIES SOLUTIONS — FROBENIUS' METHOD

In this section we develop a method of obtaining one solution of the linear, second-order, homogeneous ODE. The method, a series expansion, will always work, provided the point of expansion is no worse than a regular singular point. In physics this very gentle condition is almost always satisfied.

A **linear, second-order, homogeneous** ODE may be put in the form

$$\boxed{\frac{d^2y}{dx^2} + P(x)\frac{dy}{dx} + Q(x)y = 0.} \tag{9.80}$$

The equation is **homogeneous** because each term contains $y(x)$ or a derivative; **linear** because each y, dy/dx, or d^2y/dx^2 appears as the first power — and no products. In this section we develop (at least) one solution of Eq. (9.80). In Section 9.6 we develop the **second, independent solution and prove that no third, independent solution exists**. Therefore the **most general solution** of Eq. (9.80) may be written as

$$\boxed{y(x) = c_1 y_1(x) + c_2 y_2(x).} \tag{9.81}$$

Our physical problem may lead to a **nonhomogeneous**, linear, second-order ODE,

$$\boxed{\frac{d^2y}{dx^2} + P(x)\frac{dy}{dx} + Q(x)y = F(x).} \tag{9.82}$$

The function on the right, $F(x)$, represents a source (such as electrostatic charge) or a driving force (as in a driven oscillator). Specific solutions of this nonhomogeneous equation are touched on in Exercise 9.6.25. They are explored in some detail, using Green's function techniques, in Sections 9.7 and 10.5, and with a Laplace transform technique in Section 15.11. Calling this solution y_p, we may add to it any solution of the corresponding homogeneous equation (Eq. (9.80)). Hence the **most general solution** of Eq. (9.82) is

$$\boxed{y(x) = c_1 y_1(x) + c_2 y_2(x) + y_p(x).} \tag{9.83}$$

The constants c_1 and c_2 will eventually be fixed by boundary conditions.

For the present, we assume that $F(x) = 0$ and that our differential equation is homogeneous. We shall attempt to develop a solution of our linear, second-order, homogeneous differential equation, Eq. (9.80), by substituting in a power series with undetermined coefficients. Also available as a parameter is the power of the lowest nonvanishing term of the series. To illustrate, we apply the method to two important differential equations, first the

linear (classical) oscillator equation

$$\frac{d^2y}{dx^2} + \omega^2 y = 0, \tag{9.84}$$

with known solutions $y = \sin\omega x, \cos\omega x$.

We try

$$\begin{aligned} y(x) &= x^k\left(a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots\right) \\ &= \sum_{\lambda=0}^{\infty} a_\lambda x^{k+\lambda}, \qquad a_0 \neq 0, \end{aligned} \tag{9.85}$$

with the exponent k and all the coefficients a_λ still undetermined. Note that k need not be an integer. By differentiating twice, we obtain

$$\begin{aligned} \frac{dy}{dx} &= \sum_{\lambda=0}^{\infty} a_\lambda (k+\lambda) x^{k+\lambda-1}, \\ \frac{d^2y}{dx^2} &= \sum_{\lambda=0}^{\infty} a_\lambda (k+\lambda)(k+\lambda-1) x^{k+\lambda-2}. \end{aligned}$$

By substituting into Eq. (9.84), we have

$$\sum_{\lambda=0}^{\infty} a_\lambda (k+\lambda)(k+\lambda-1) x^{k+\lambda-2} + \omega^2 \sum_{\lambda=0}^{\infty} a_\lambda x^{k+\lambda} = 0. \tag{9.86}$$

From our analysis of the uniqueness of power series (Chapter 5), the coefficients of each power of x on the left-hand side of Eq. (9.86) must vanish individually.

The lowest power of x appearing in Eq. (9.86) is x^{k-2}, for $\lambda = 0$ in the first summation. The requirement that the coefficient vanish[8] yields

$$a_0 k(k-1) = 0.$$

We had chosen a_0 as the coefficient of the lowest nonvanishing terms of the series (Eq. (9.85)), hence, by definition, $a_0 \neq 0$. Therefore we have

$$k(k-1) = 0. \tag{9.87}$$

This equation, coming from the coefficient of the lowest power of x, we call the **indicial equation**. The indicial equation and its roots are of critical importance to our analysis. If $k = 1$, the coefficient $a_1(k+1)k$ of x^{k-1} must vanish so that $a_1 = 0$. Clearly, in this example we must require either that $k = 0$ or $k = 1$.

Before considering these two possibilities for k, we return to Eq. (9.86) and demand that the remaining net coefficients, say, the coefficient of $x^{k+j}(j \geq 0)$, vanish. We set $\lambda = j+2$ in the first summation and $\lambda = j$ in the second. (They are independent summations and λ is a dummy index.) This results in

$$a_{j+2}(k+j+2)(k+j+1) + \omega^2 a_j = 0$$

[8] See the uniqueness of power series, Section 5.7.

or

$$a_{j+2} = -a_j \frac{\omega^2}{(k+j+2)(k+j+1)}. \tag{9.88}$$

This is a two-term **recurrence relation**.[9] Given a_j, we may compute a_{j+2} and then a_{j+4}, a_{j+6}, and so on up as far as desired. Note that for this example, if we start with a_0, Eq. (9.88) leads to the even coefficients a_2, a_4, and so on, and ignores a_1, a_3, a_5, and so on. Since a_1 is arbitrary if $k = 0$ and necessarily zero if $k = 1$, let us set it equal to zero (compare Exercises 9.5.3 and 9.5.4) and then by Eq. (9.88)

$$a_3 = a_5 = a_7 = \cdots = 0,$$

and all the odd-numbered coefficients vanish. The odd powers of x will actually reappear when the **second** root of the indicial equation is used.

Returning to Eq. (9.87) our indicial equation, we first try the solution $k = 0$. The recurrence relation (Eq. (9.88)) becomes

$$a_{j+2} = -a_j \frac{\omega^2}{(j+2)(j+1)}, \tag{9.89}$$

which leads to

$$\begin{aligned} a_2 &= -a_0 \frac{\omega^2}{1\cdot 2} = -\frac{\omega^2}{2!} a_0, \\ a_4 &= -a_2 \frac{\omega^2}{3\cdot 4} = +\frac{\omega^4}{4!} a_0, \\ a_6 &= -a_4 \frac{\omega^2}{5\cdot 6} = -\frac{\omega^6}{6!} a_0, \qquad \text{and so on.} \end{aligned}$$

By inspection (and mathematical induction),

$$a_{2n} = (-1)^n \frac{\omega^{2n}}{(2n)!} a_0, \tag{9.90}$$

and our solution is

$$y(x)_{k=0} = a_0 \left[1 - \frac{(\omega x)^2}{2!} + \frac{(\omega x)^4}{4!} - \frac{(\omega x)^6}{6!} + \cdots \right] = a_0 \cos \omega x. \tag{9.91}$$

If we choose the indicial equation root $k = 1$ (Eq. (9.88)), the recurrence relation becomes

$$a_{j+2} = -a_j \frac{\omega^2}{(j+3)(j+2)}. \tag{9.92}$$

[9] The recurrence relation may involve three terms, that is, a_{j+2}, depending on a_j and a_{j-2}. Equation (13.2) for the Hermite functions provides an example of this behavior.

Substituting in $j = 0, 2, 4$, successively, we obtain

$$a_2 = -a_0\frac{\omega^2}{2\cdot 3} = -\frac{\omega^2}{3!}a_0,$$

$$a_4 = -a_2\frac{\omega^2}{4\cdot 5} = +\frac{\omega^4}{5!}a_0,$$

$$a_6 = -a_4\frac{\omega^2}{6\cdot 7} = -\frac{\omega^6}{7!}a_0, \qquad \text{and so on.}$$

Again, by inspection and mathematical induction,

$$a_{2n} = (-1)^n\frac{\omega^{2n}}{(2n+1)!}a_0. \tag{9.93}$$

For this choice, $k = 1$, we obtain

$$\begin{aligned} y(x)_{k=1} &= a_0 x\left[1 - \frac{(\omega x)^2}{3!} + \frac{(\omega x)^4}{5!} - \frac{(\omega x)^6}{7!} + \cdots\right] \\ &= \frac{a_0}{\omega}\left[(\omega x) - \frac{(\omega x)^3}{3!} + \frac{(\omega x)^5}{5!} - \frac{(\omega x)^7}{7!} + \cdots\right] \\ &= \frac{a_0}{\omega}\sin\omega x. \end{aligned} \tag{9.94}$$

To summarize this approach, we may write Eq. (9.86) *schematically as shown in Fig.* 9.2. *From the uniqueness of power series* (*Section* 5.7), *the total coefficient of each power of* x *must vanish—all by itself. The requirement that the first coefficient* (1) *vanish leads to the indicial equation, Eq.* (9.87). *The second coefficient is handled by setting* $a_1 = 0$. *The vanishing of the coefficient of* x^k (*and higher powers, taken one at a time*) *leads to the recurrence relation, Eq.* (9.88).

This series substitution, known as Frobenius' method, has given us two series solutions of the linear oscillator equation. However, there are two points about such series solutions that must be strongly emphasized:

1. The series solution should always be substituted back into the differential equation, to see if it works, as a precaution against algebraic and logical errors. If it works, it is a solution.
2. The acceptability of a series solution depends on its convergence (including asymptotic convergence). It is quite possible for Frobenius' method to give a series solution that satisfies the original differential equation when substituted in the equation but that does

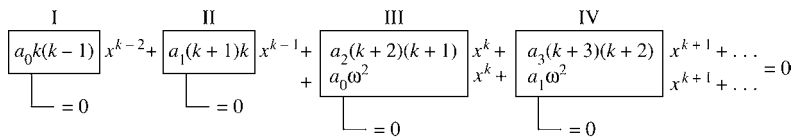

FIGURE 9.2 Recurrence relation from power series expansion.

not converge over the region of interest. Legendre's differential equation illustrates this situation.

Expansion About x_0

Equation (9.85) is an expansion about the origin, $x_0 = 0$. It is perfectly possible to replace Eq. (9.85) with

$$y(x) = \sum_{\lambda=0}^{\infty} a_\lambda (x - x_0)^{k+\lambda}, \qquad a_0 \neq 0. \tag{9.95}$$

Indeed, for the Legendre, Chebyshev, and hypergeometric equations the choice $x_0 = 1$ has some advantages. The point x_0 should not be chosen at an essential singularity — or our Frobenius method will probably fail. The resultant series (x_0 an ordinary point or regular singular point) will be valid where it converges. You can expect a divergence of some sort when $|x - x_0| = |z_s - x_0|$, where z_s is the closest singularity to x_0 (in the complex plane).

Symmetry of Solutions

Let us note that we obtained one solution of even symmetry, $y_1(x) = y_1(-x)$, and one of odd symmetry, $y_2(x) = -y_2(-x)$. This is not just an accident but a direct consequence of the form of the ODE. Writing a general ODE as

$$\mathcal{L}(x)y(x) = 0, \tag{9.96}$$

in which $\mathcal{L}(x)$ is the differential operator, we see that for the linear oscillator equation (Eq. (9.84)), $\mathcal{L}(x)$ is even under parity; that is,

$$\mathcal{L}(x) = \mathcal{L}(-x). \tag{9.97}$$

Whenever the differential operator has a specific parity or symmetry, either even or odd, we may interchange $+x$ and $-x$, and Eq. (9.96) becomes

$$\pm\mathcal{L}(x)y(-x) = 0, \tag{9.98}$$

$+$ if $\mathcal{L}(x)$ is even, $-$ if $\mathcal{L}(x)$ is odd. Clearly, if $y(x)$ is a solution of the differential equation, $y(-x)$ is also a solution. Then any solution may be resolved into even and odd parts,

$$y(x) = \tfrac{1}{2}\big[y(x) + y(-x)\big] + \tfrac{1}{2}\big[y(x) - y(-x)\big], \tag{9.99}$$

the first bracket on the right giving an even solution, the second an odd solution.

If we refer back to Section 9.4, we can see that Legendre, Chebyshev, Bessel, simple harmonic oscillator, and Hermite equations (or differential operators) all exhibit this even parity; that is, their $P(x)$ in Eq. (9.80) is odd and $Q(x)$ even. Solutions of all of them may be presented as series of even powers of x and separate series of odd powers of x. The Laguerre differential operator has neither even nor odd symmetry; hence its solutions cannot be expected to exhibit even or odd parity. Our emphasis on parity stems primarily from the importance of parity in quantum mechanics. We find that wave functions usually are either even or odd, meaning that they have a definite parity. Most interactions (beta decay is the big exception) are also even or odd, and the result is that parity is conserved.

Limitations of Series Approach — Bessel's Equation

This attack on the linear oscillator equation was perhaps a bit too easy. By substituting the power series (Eq. (9.85)) into the differential equation (Eq. (9.84)), we obtained two independent solutions with no trouble at all.

To get some idea of what can happen we try to solve Bessel's equation,

$$x^2y'' + xy' + \left(x^2 - n^2\right)y = 0, \tag{9.100}$$

using y' for dy/dx and y'' for d^2y/dx^2. Again, assuming a solution of the form

$$y(x) = \sum_{\lambda=0}^{\infty} a_\lambda x^{k+\lambda},$$

we differentiate and substitute into Eq. (9.100). The result is

$$\sum_{\lambda=0}^{\infty} a_\lambda (k+\lambda)(k+\lambda-1)x^{k+\lambda} + \sum_{\lambda=0}^{\infty} a_\lambda (k+\lambda)x^{k+\lambda} + \sum_{\lambda=0}^{\infty} a_\lambda x^{k+\lambda+2} - \sum_{\lambda=0}^{\infty} a_\lambda n^2 x^{k+\lambda} = 0. \tag{9.101}$$

By setting $\lambda = 0$, we get the coefficient of x^k, the lowest power of x appearing on the left-hand side,

$$a_0\left[k(k-1) + k - n^2\right] = 0, \tag{9.102}$$

and again $a_0 \neq 0$ by definition. Equation (9.102) therefore yields the **indicial** equation

$$k^2 - n^2 = 0 \tag{9.103}$$

with solutions $k = \pm n$.

It is of some interest to examine the coefficient of x^{k+1} also. Here we obtain

$$a_1\left[(k+1)k + k + 1 - n^2\right] = 0,$$

or

$$a_1(k+1-n)(k+1+n) = 0. \tag{9.104}$$

For $k = \pm n$, neither $k+1-n$ nor $k+1+n$ vanishes and we **must** require $a_1 = 0$.[10]

Proceeding to the coefficient of x^{k+j} for $k = n$, we set $\lambda = j$ in the first, second, and fourth terms of Eq. (9.101) and $\lambda = j-2$ in the third term. By requiring the resultant coefficient of x^{k+1} to vanish, we obtain

$$a_j\left[(n+j)(n+j-1) + (n+j) - n^2\right] + a_{j-2} = 0.$$

When j is replaced by $j+2$, this can be rewritten for $j \geq 0$ as

$$a_{j+2} = -a_j \frac{1}{(j+2)(2n+j+2)}, \tag{9.105}$$

[10] $k = \pm n = -\frac{1}{2}$ are exceptions.

which is the desired recurrence relation. Repeated application of this recurrence relation leads to

$$a_2 = -a_0\frac{1}{2(2n+2)} = -\frac{a_0 n!}{2^2 1!(n+1)!},$$
$$a_4 = -a_2\frac{1}{4(2n+4)} = \frac{a_0 n!}{2^4 2!(n+2)!},$$
$$a_6 = -a_4\frac{1}{6(2n+6)} = -\frac{a_0 n!}{2^6 3!(n+3)!}, \qquad \text{and so on,}$$

and in general,

$$a_{2p} = (-1)^p \frac{a_0 n!}{2^{2p} p!(n+p)!}. \tag{9.106}$$

Inserting these coefficients in our assumed series solution, we have

$$y(x) = a_0 x^n\left[1 - \frac{n!x^2}{2^2 1!(n+1)!} + \frac{n!x^4}{2^4 2!(n+2)!} - \cdots\right]. \tag{9.107}$$

In summation form

$$y(x) = a_0\sum_{j=0}^{\infty}(-1)^j\frac{n!x^{n+2j}}{2^{2j}\,j!(n+j)!}$$
$$= a_0 2^n n!\sum_{j=0}^{\infty}(-1)^j\frac{1}{j!(n+j)!}\left(\frac{x}{2}\right)^{n+2j}. \tag{9.108}$$

In Chapter 11 the final summation is identified as the Bessel function $J_n(x)$. Notice that this solution, $J_n(x)$, has either even or odd symmetry,[11] as might be expected from the form of Bessel's equation.

When $k = -n$ and n is not an integer, we may generate a second distinct series, to be labeled $J_{-n}(x)$. However, when $-n$ is a negative integer, trouble develops. The recurrence relation for the coefficients a_j is still given by Eq. (9.105), but with $2n$ replaced by $-2n$. Then, when $j + 2 = 2n$ or $j = 2(n-1)$, the coefficient a_{j+2} blows up and we have no series solution. This catastrophe can be remedied in Eq. (9.108), as it is done in Chapter 11, with the result that

$$J_{-n}(x) = (-1)^n J_n(x), \qquad n \text{ an integer.} \tag{9.109}$$

The second solution simply reproduces the first. We have failed to construct a second independent solution for Bessel's equation by this series technique when n is an integer.

By substituting in an infinite series, we have obtained two solutions for the linear oscillator equation and one for Bessel's equation (two if n is not an integer). To the questions "Can we always do this? Will this method always work?" the answer is no, we cannot always do this. This method of series solution will not always work.

[11] $J_n(x)$ is an even function if n is an even integer, an odd function if n is an odd integer. For nonintegral n the x^n has no such simple symmetry.

Regular and Irregular Singularities

The success of the series substitution method depends on the roots of the indicial equation and the degree of singularity of the coefficients in the differential equation. To understand better the effect of the equation coefficients on this naive series substitution approach, consider four simple equations:

$$y'' - \frac{6}{x^2}y = 0, \tag{9.110a}$$

$$y'' - \frac{6}{x^3}y = 0, \tag{9.110b}$$

$$y'' + \frac{1}{x}y' - \frac{a^2}{x^2}y = 0, \tag{9.110c}$$

$$y'' + \frac{1}{x^2}y' - \frac{a^2}{x^2}y = 0. \tag{9.110d}$$

The reader may show easily that for Eq. (9.110a) the indicial equation is

$$k^2 - k - 6 = 0,$$

giving $k = 3, -2$. Since the equation is homogeneous in x (counting d^2/dx^2 as x^{-2}), there is no recurrence relation. However,we are left with two perfectly good solutions, x^3 and x^{-2}.

Equation (9.110b) differs from Eq. (9.110a) by only one power of x, but this sends the indicial equation to

$$-6a_0 = 0,$$

with no solution at all, for we have agreed that $a_0 \neq 0$. Our series substitution worked for Eq. (9.110a), which had only a regular singularity, but broke down at Eq. (9.110b), which has an irregular singular point at the origin.

Continuing with Eq. (9.110c), we have added a term y'/x. The indicial equation is

$$k^2 - a^2 = 0,$$

but again, there is no recurrence relation. The solutions are $y = x^a, x^{-a}$, both perfectly acceptable one-term series.

When we change the power of x in the coefficient of y' from -1 to -2, Eq. (9.110d), there is a drastic change in the solution. The indicial equation (with only the y' term contributing) becomes

$$k = 0.$$

There is a recurrence relation,

$$a_{j+1} = +a_j \frac{a^2 - j(j-1)}{j+1}.$$

Unless the parameter a is selected to make the series terminate, we have

$$\lim_{j\to\infty}\left|\frac{a_{j+1}}{a_j}\right| = \lim_{j\to\infty}\frac{j(j+1)}{j+1}$$

$$= \lim_{j\to\infty}\frac{j^2}{j} = \infty.$$

Hence our series solution diverges for all $x \neq 0$. Again, our method worked for Eq. (9.110c) with a regular singularity but failed when we had the irregular singularity of Eq. (9.110d).

Fuchs' Theorem

The answer to the basic question when the method of series substitution can be expected to work is given by Fuchs' theorem, which asserts that we can always obtain at least one power-series solution, provided we are expanding about a point that is an ordinary point or at worst a regular singular point.

If we attempt an expansion about an irregular or essential singularity, our method may fail, as it did for Eqs. (9.110b) and (9.110d). Fortunately, the more important equations of mathematical physics, listed in Section 9.4, have no irregular singularities in the finite plane. Further discussion of Fuchs' theorem appears in Section 9.6.

From Table 9.4, Section 9.4, infinity is seen to be a singular point for all equations considered. As a further illustration of Fuchs' theorem, Legendre's equation (with infinity as a regular singularity) has a convergent-series solution in negative powers of the argument (Section 12.10). In contrast, Bessel's equation (with an irregular singularity at infinity) yields asymptotic series (Sections 5.10 and 11.6). These asymptotic solutions are extremely useful.

Summary

If we are expanding about an ordinary point or at worst about a regular singularity, the series substitution approach will yield at least one solution (Fuchs' theorem).

Whether we get one or two distinct solutions depends on the roots of the indicial equation.

1. If the two roots of the indicial equation are equal, we can obtain only one solution by this series substitution method.
2. If the two roots differ by a nonintegral number, two independent solutions may be obtained.
3. If the two roots differ by an integer, the larger of the two will yield a solution.

The smaller may or may not give a solution, depending on the behavior of the coefficients. In the linear oscillator equation we obtain two solutions; for Bessel's equation, we get only one solution.

The usefulness of the series solution in terms of what the solution is (that is, numbers) depends on the rapidity of convergence of the series and the availability of the coefficients. Many ODEs will not yield nice, simple recurrence relations for the coefficients. In general, the available series will probably be useful for $|x|$ (or $|x - x_0|$) very small. Computers can be used to determine additional series coefficients using a symbolic language, such as Mathematica,[12] Maple,[13] or Reduce.[14] Often, however, for numerical work a direct numerical integration will be preferred.

Exercises

9.5.1 Uniqueness theorem. The function $y(x)$ satisfies a second-order, linear, homogeneous differential equation. At $x = x_0$, $y(x) = y_0$ and $dy/dx = y_0'$. Show that $y(x)$ is unique, in that no other solution of this differential equation passes through the points (x_0, y_0) with a slope of y_0'.
Hint. Assume a second solution satisfying these conditions and compare the Taylor series expansions.

9.5.2 A series solution of Eq. (9.80) is attempted, expanding about the point $x = x_0$. If x_0 is an ordinary point, show that the indicial equation has roots $k = 0, 1$.

9.5.3 In the development of a series solution of the simple harmonic oscillator (SHO) equation, the second series coefficient a_1 was neglected except to set it equal to zero. From the coefficient of the next-to-the-lowest power of x, x^{k-1}, develop a second indicial-type equation.

(a) (SHO equation with $k = 0$). Show that a_1, may be assigned any finite value (including zero).
(b) (SHO equation with $k = 1$). Show that a_1 must be set equal to zero.

9.5.4 Analyze the series solutions of the following differential equations to see when a_1 **may** be set equal to zero without irrevocably losing anything and when a_1 **must** be set equal to zero.
(a) Legendre, (b) Chebyshev, (c) Bessel, (d) Hermite.

ANS. (a) Legendre, (b) Chebyshev, and (d) Hermite: For $k = 0$, a_1 **may** be set equal to zero; for $k = 1$, a_1 **must** be set equal to zero.
(c) Bessel: a_1 **must** be set equal to zero (except for $k = \pm n = -\frac{1}{2}$).

9.5.5 Solve the Legendre equation

$$\left(1 - x^2\right)y'' - 2xy' + n(n+1)y = 0$$

by direct series substitution.

[12] S. Wolfram, *Mathematica, A System for Doing Mathematics by Computer*, New York: Addison Wesley (1991).
[13] A. Heck, *Introduction to Maple*, New York: Springer (1993).
[14] G. Rayna, *Reduce Software for Algebraic Computation*, New York: Springer (1987).

(a) Verify that the indicial equation is

$$k(k-1)=0.$$

(b) Using $k=0$, obtain a series of even powers of x ($a_1=0$).

$$y_{\text{even}}=a_0\left[1-\frac{n(n+1)}{2!}x^2+\frac{n(n-2)(n+1)(n+3)}{4!}x^4+\cdots\right],$$

where

$$a_{j+2}=\frac{j(j+1)-n(n+1)}{(j+1)(j+2)}a_j.$$

(c) Using $k=1$, develop a series of odd powers of x ($a_1=1$).

$$y_{\text{odd}}=a_1\left[x-\frac{(n-1)(n+2)}{3!}x^3+\frac{(n-1)(n-3)(n+2)(n+4)}{5!}x^5+\cdots\right],$$

where

$$a_{j+2}=\frac{(j+1)(j+2)-n(n+1)}{(j+2)(j+3)}a_j.$$

(d) Show that both solutions, y_{even} and y_{odd}, diverge for $x=\pm1$ **if the series continue to infinity**.

(e) Finally, show that by an appropriate choice of n, one series at a time may be converted into a polynomial, thereby avoiding the divergence catastrophe. In quantum mechanics this restriction of n to integral values corresponds to **quantization of angular momentum**.

9.5.6 Develop series solutions for Hermite's differential equation

(a) $y''-2xy'+2\alpha y=0.$

ANS. $k(k-1)=0$, indicial equation.

For $k=0$,

$$a_{j+2}=2a_j\frac{j-\alpha}{(j+1)(j+2)}\qquad (j\text{ even}),$$

$$y_{\text{even}}=a_0\left[1+\frac{2(-\alpha)x^2}{2!}+\frac{2^2(-\alpha)(2-\alpha)x^4}{4!}+\cdots\right].$$

For $k=1$,

$$a_{j+2}=2a_j\frac{j+1-\alpha}{(j+2)(j+3)}\qquad (j\text{ even}),$$

$$y_{\text{odd}}=a_1\left[x+\frac{2(1-\alpha)x^3}{3!}+\frac{2^2(1-\alpha)(3-\alpha)x^5}{5!}+\cdots\right].$$

(b) Show that both series solutions are convergent for all x, the ratio of successive coefficients behaving, for large index, like the corresponding ratio in the expansion of $\exp(x^2)$.

(c) Show that by appropriate choice of α the series solutions may be cut off and converted to finite polynomials. (These polynomials, properly normalized, become the Hermite polynomials in Section 13.1.)

9.5.7 Laguerre's ODE is

$$xL_n''(x) + (1-x)L_n'(x) + nL_n(x) = 0.$$

Develop a series solution selecting the parameter n to make your series a polynomial.

9.5.8 Solve the Chebyshev equation

$$\left(1-x^2\right)T_n'' - xT_n' + n^2T_n = 0,$$

by series substitution. What restrictions are imposed on n if you demand that the series solution converge for $x = \pm 1$?

ANS. The infinite series does converge for $x = \pm 1$ and no restriction on n exists (compare Exercise 5.2.16).

9.5.9 Solve

$$\left(1-x^2\right)U_n''(x) - 3xU_n'(x) + n(n+2)U_n(x) = 0,$$

choosing the root of the indicial equation to obtain a series of **odd** powers of x. Since the series will diverge for $x = 1$, choose n to convert it into a polynomial.

$$k(k-1) = 0.$$

For $k = 1$,

$$a_{j+2} = \frac{(j+1)(j+3) - n(n+2)}{(j+2)(j+3)} a_j.$$

9.5.10 Obtain a series solution of the hypergeometric equation

$$x(x-1)y'' + \left[(1+a+b)x - c\right]y' + aby = 0.$$

Test your solution for convergence.

9.5.11 Obtain two series solutions of the confluent hypergeometric equation

$$xy'' + (c-x)y' - ay = 0.$$

Test your solutions for convergence.

9.5.12 A quantum mechanical analysis of the Stark effect (parabolic coordinates) leads to the differential equation

$$\frac{d}{d\xi}\left(\xi\frac{du}{d\xi}\right) + \left(\frac{1}{2}E\xi + \alpha - \frac{m^2}{4\xi} - \frac{1}{4}F\xi^2\right)u = 0.$$

Here α is a separation constant, E is the total energy, and F is a constant, where Fz is the potential energy added to the system by the introduction of an electric field.

Using the larger root of the indicial equation, develop a power-series solution about $\xi = 0$. Evaluate the first three coefficients in terms of a_o.

$$\text{Indicial equation} \qquad k^2 - \frac{m^2}{4} = 0,$$

$$u(\xi) = a_0 \xi^{m/2} \left\{ 1 - \frac{\alpha}{m+1}\xi + \left[\frac{\alpha^2}{2(m+1)(m+2)} - \frac{E}{4(m+2)} \right] \xi^2 + \cdots \right\}.$$

Note that the perturbation F does not appear until a_3 is included.

9.5.13 For the special case of no azimuthal dependence, the quantum mechanical analysis of the hydrogen molecular ion leads to the equation

$$\frac{d}{d\eta}\left[\left(1 - \eta^2\right) \frac{du}{d\eta} \right] + \alpha u + \beta \eta^2 u = 0.$$

Develop a power-series solution for $u(\eta)$. Evaluate the first three nonvanishing coefficients in terms of a_0.

$$\text{Indicial equation} \qquad k(k-1) = 0,$$

$$u_{k=1} = a_0 \eta \left\{ 1 + \frac{2-\alpha}{6}\eta^2 + \left[\frac{(2-\alpha)(12-\alpha)}{120} - \frac{\beta}{20} \right] \eta^4 + \cdots \right\}.$$

9.5.14 To a good approximation, the interaction of two nucleons may be described by a mesonic potential

$$V = \frac{Ae^{-ax}}{x},$$

attractive for A negative. Develop a series solution of the resultant Schrödinger wave equation

$$\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + (E - V)\psi = 0$$

through the first three nonvanishing coefficients.

$$\psi = a_0 \left\{ x + \tfrac{1}{2}A' x^2 + \tfrac{1}{6}\left[\tfrac{1}{2}A'^2 - E' - aA' \right] x^3 + \cdots \right\},$$

where the prime indicates multiplication by $2m/\hbar^2$.

9.5.15 Near the nucleus of a complex atom the potential energy of one electron is given by

$$V = \frac{Ze^2}{r}\left(1 + b_1 r + b_2 r^2\right),$$

where the coefficients b_1 and b_2 arise from screening effects. For the case of zero angular momentum show that the first three terms of the solution of the Schrödinger equation have the same form as those of Exercise 9.5.14. By appropriate translation of coefficients or parameters, write out the first three terms in a series expansion of the wave function.

9.5.16 If the parameter a^2 in Eq. (9.110d) is equal to 2, Eq. (9.110d) becomes

$$y'' + \frac{1}{x^2}y' - \frac{2}{x^2}y = 0.$$

From the indicial equation and the recurrence relation **derive** a solution $y = 1 + 2x + 2x^2$. Verify that this is indeed a solution by substituting back into the differential equation.

9.5.17 The modified Bessel function $I_0(x)$ satisfies the differential equation

$$x^2\frac{d^2}{dx^2}I_0(x) + x\frac{d}{dx}I_0(x) - x^2 I_0(x) = 0.$$

From Exercise 7.3.4 the leading term in an asymptotic expansion is found to be

$$I_0(x) \sim \frac{e^x}{\sqrt{2\pi x}}.$$

Assume a series of the form

$$I_0(x) \sim \frac{e^x}{\sqrt{2\pi x}}\left\{1 + b_1 x^{-1} + b_2 x^{-2} + \cdots\right\}.$$

Determine the coefficients b_1 and b_2.

ANS. $b_1 = \frac{1}{8}, \quad b_2 = \frac{9}{128}$.

9.5.18 The even power-series solution of Legendre's equation is given by Exercise 9.5.5. Take $a_0 = 1$ and n not an even integer, say $n = 0.5$. Calculate the partial sums of the series through $x^{200}, x^{400}, x^{600}, \ldots, x^{2000}$ for $x = 0.95(0.01)1.00$. Also, write out the individual term corresponding to each of these powers.
Note. This calculation does **not** constitute proof of convergence at $x = 0.99$ or divergence at $x = 1.00$, but perhaps you can see the difference in the behavior of the sequence of partial sums for these two values of x.

9.5.19 (a) The odd power-series solution of Hermite's equation is given by Exercise 9.5.6. Take $a_0 = 1$. Evaluate this series for $\alpha = 0, x = 1, 2, 3$. Cut off your calculation after the last term calculated has dropped below the maximum term by a factor of 10^6 or more. Set an upper bound to the error made in ignoring the remaining terms in the infinite series.

(b) As a check on the calculation of part (a), show that the Hermite series $y_{\text{odd}}(\alpha = 0)$ corresponds to $\int_0^x \exp(x^2)\,dx$.

(c) Calculate this integral for $x = 1, 2, 3$.

9.6 A SECOND SOLUTION

In Section 9.5 a solution of a second-order homogeneous ODE was developed by substituting in a power series. By Fuchs' theorem this is possible, provided the power series is an expansion about an ordinary point or a nonessential singularity.[15] There is no guarantee

[15] This is why the classification of singularities in Section 9.4 is of vital importance.

that this approach will yield the two independent solutions we expect from a linear second-order ODE. In fact, we shall prove that such an ODE has at most two linearly independent solutions. Indeed, the technique gave only one solution for Bessel's equation (n an integer). In this section we also develop two methods of obtaining a second independent solution: an integral method and a power series containing a logarithmic term. First, however, we consider the question of independence of a set of functions.

Linear Independence of Solutions

Given a set of functions φ_λ, the criterion for linear dependence is the existence of a relation of the form

$$\sum_\lambda k_\lambda \varphi_\lambda = 0, \qquad = 0. \tag{9.116}$$

(9.111)

in which not all the coefficients k_λ are zero. On the other hand, if the only solution of Eq. (9.111) is $k_\lambda = 0$ for all λ, the set of functions φ_λ is said to be linearly **independent**.

This determinant is called the **Wrońskian**.

It may be helpful to think of linear dependence of vectors. Consider **A**, **B**, and **C** in three-dimensional space, with $\mathbf{A}\cdot\mathbf{B}\times\mathbf{C}\neq 0$. Then no nontrivial relation of the form

$$a\mathbf{A}+b\mathbf{B}+c\mathbf{C}=0 \tag{9.112}$$

exists. **A**, **B**, and **C** are linearly independent. On the other hand, any fourth vector, **D**, may be expressed as a linear combination of **A**, **B**, and **C** (see Section 3.1). We can always write an equation of the form

$$\mathbf{D}-a\mathbf{A}-b\mathbf{B}-c\mathbf{C}=0, \tag{9.113}$$

and the four vectors are **not** linearly independent. The three noncoplanar vectors **A**, **B**, and **C** span our real three-dimensional space.

If a set of vectors or functions are mutually orthogonal, then they are automatically linearly independent. Orthogonality implies linear independence. This can easily be demonstrated by taking inner products (scalar or dot product for vectors, orthogonality integral of Section 10.2 for functions).

Let us assume that the functions φ_λ are differentiable as needed. Then, differentiating Eq. (9.111) repeatedly, we generate a set of equations

$$\sum_\lambda k_\lambda \varphi'_\lambda = 0, \tag{9.114}$$

$$\sum_\lambda k_\lambda \varphi''_\lambda = 0, \tag{9.115}$$

and so on. This gives us a set of homogeneous linear equations in which k_λ are the unknown quantities. By Section 3.1 there is a solution $k_\lambda \neq 0$ only if the determinant of the coefficients of the k_λ' vanishes. This means

$$\begin{vmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_n \\ \varphi'_1 & \varphi'_2 & \cdots & \varphi'_n \\ \cdots & \cdots & \cdots & \cdots \\ \varphi_1^{(n-1)} & \varphi_2^{(n-1)} & \cdots & \varphi_n^{(n-1)} \end{vmatrix}$$

1. If the Wronskian is not equal to zero, then Eq. (9.111) has no solution other than $k_\lambda = 0$. The set of functions φ_λ is therefore linearly independent.
2. If the Wronskian vanishes at isolated values of the argument, this does not necessarily prove linear dependence (unless the set of functions has only two functions). However, if the Wronskian is zero over the entire range of the variable, the functions φ_λ are linearly dependent over this range[16] (compare Exercise 9.5.2 for the simple case of two functions).

Example 9.6.1 LINEAR INDEPENDENCE

The solutions of the linear oscillator equation (9.84) are $\varphi_1 = \sin\omega x, \varphi_2 = \cos\omega x$. The Wronskian becomes

$$\begin{vmatrix} \sin\omega x & \cos\omega x \\ \omega\cos\omega x & -\omega\sin\omega x \end{vmatrix} = -\omega \neq 0.$$

These two solutions, φ_1 and φ_2, are therefore linearly independent. For just two functions this means that one is not a multiple of the other, which obviously true in this case. Hence e^x, e^{-x}, and $\cosh x$ are linearly dependent, and, indeed, we have a relation of the form of Eq. (9.111):

$$e^x + e^{-x} - 2\cosh x = 0 \qquad \text{with } k_\lambda \neq 0.$$

but this is **not** a **linear** relation, of the form of Eq. (9.111). ∎

Now we are ready to prove the **theorem that a second-order homogeneous ODE has two linearly independent solutions**.

Examples 9.6.2 LINEAR DEPENDENCE

Suppose y_1, y_2, y_3 are three solutions of the homogeneous ODE (9.80). Then we form the Wronskian $W_{jk} = y_j y_k' - y_j' y_k$ of any pair y_j, y_k of them and recall that The determinant vanishes for all x because the first and third rows are identical. Hence For an illustration of linear dependence, consider the solutions of the one-dimensional diffusion equation. We have $\varphi_1 = e^x$ and $\varphi_2 = e^{-x}$, and we add $\varphi_3 = \cosh x$, also a solution. The Wronskian is

$$\begin{vmatrix} e^x & e^{-x} & \cosh x \\ e^x & -e^{-x} & \sinh x \\ e^x & e^{-x} & \cosh x \end{vmatrix}$$

[16] Compare H. Lass, *Elements of Pure and Applied Mathematics*, New York: McGraw-Hill (1957), p. 187, for proof of this assertion. It is assumed that the functions have continuous derivatives and that at least one of the minors of the bottom row of Eq. (9.116) (Laplace expansion) does not vanish in $[a, b]$, the interval under consideration.

In the special case that $P(x) = 0$, that is,

$$y'' + Q(x)y = 0, \tag{9.121}$$

the Wronskian $= P(y_1'W_{23} - y_2'W_{13} + y_3'W_{12}) = -P$

$$W = y_1y_2' - y_1'y_2 = \text{constant}. \tag{9.122}$$

$W_{jk}' = y_jy_k'' - y_j''y_k$. We divide each ODE by y, getting $-Q$ on their right-hand side, so

$$\frac{y_j''}{y_j} + P\frac{y_j'}{y_j} = -Q(x) = \frac{y_k''}{y_k} + P\frac{y_k'}{y_k}.$$

Multiplying by y_jy_k, we find

$$(y_jy_k'' - y_j''y_k) + P(y_jy_k' - y_j'y_k) = 0, \qquad \text{or} \qquad W_{jk}' = -PW_{jk} \tag{9.117}$$

for any pair of solutions. Finally we evaluate the Wronskian of all three solutions, expanding it along the second row and using the ODEs for the W_{jk}:

$$W = \begin{vmatrix} y_1 & y_2 & y_3 \\ y_1' & y_2' & y_3' \\ y_1'' & y_2'' & y_3'' \end{vmatrix}$$

Since our original differential equation is homogeneous, we may multiply the solutions y_1 and y_2 by whatever constants we wish and arrange to have the Wronskian equal to unity (or -1). This case, $P(x) = 0$, appears more frequently than might be expected. Recall that the portion of $\nabla^2(\frac{\psi}{r})$ in spherical polar coordinates involving radial derivatives contains no first radial derivative. Finally, every linear second-order differential equation can be transformed into an equation of the form of Eq. (9.121) (compare Exercise 9.6.11).

For the general case, let us now assume that we have one solution of Eq. (9.118) by a series substitution (or by guessing). We now proceed to develop a second, independent solution for which $W \neq 0$. Rewriting Eq. (9.120) as

$$\frac{dW}{W} = -P\,dx,$$

we integrate over the variable x, from a to x, to obtain

$$\ln \frac{W(x)}{W(a)} = -\int_a^x P(x_1)\,dx_1,$$

or[17]

$$\boxed{W(x) = W(a)\exp\left[-\int_a^x P(x_1)\,dx_1\right].} \tag{9.123}$$

But

$$W(x) = y_1y_2' - y_1'y_2 = y_1^2\frac{d}{dx}\left(\frac{y_2}{y_1}\right). \tag{9.124}$$

By combining Eqs. (9.123) and (9.124), we have

$$\frac{d}{dx}\left(\frac{y_2}{y_1}\right) = W(a)\frac{\exp[-\int_a^x P(x_1)\,dx_1]}{y_1^2}. \tag{9.125}$$

Finally, by integrating Eq. (9.125) from $x_2 = b$ to $x_2 = x$ we get

$$y_2(x) = y_1(x)W(a)\int_b^x \frac{\exp[-\int_a^{x_2} P(x_1)dx_1]}{[y_1(x_2)]^2}dx_2. \tag{9.126}$$

Here a and b are arbitrary constants and a term $y_1(x)y_2(b)/y_1(b)$ has been dropped, for it leads to nothing new. Since $W(a)$, the Wronskian evaluated at $x = a$, is a constant and our solutions for the homogeneous differential equation always contain an unknown normalizing factor, we set $W(a) = 1$ and write

$$\boxed{y_2(x) = y_1(x)\int^x \frac{\exp[-\int^{x_2} P(x_1)\,dx_1]}{[y_1(x_2)]^2}dx_2.} \tag{9.127}$$

Note that the lower limits $x_1 = a$ and $x_2 = b$ have been omitted. If they are retained, they simply make a contribution equal to a constant times the known first solution, $y_1(x)$, and

[17] If $P(x)$ remains finite in the domain of interest, $W(x) \neq 0$ unless $W(a) = 0$. That is, the Wronskian of our two solutions is either identically zero or never zero. However, if $P(x)$ does not remain finite in our interval, then $W(x)$ can have isolated zeros in that domain and one must be careful to choose a so that $W(a) \neq 0$.

hence add nothing new. If we have the important special case of $P(x) = 0$, Eq. (9.127) reduces to

$$y_2(x) = y_1(x) \int^x \frac{dx_2}{[y_1(x_2)]^2}. \tag{9.128}$$

This means that by using either Eq. (9.127) or Eq. (9.128) we can take one known solution and by integrating can generate a second, independent solution of Eq. (9.118). This technique is used in Section 12.10 to generate a second solution of Legendre's differential equation.

Example 9.6.3 A Second Solution for the Linear Oscillator Equation

From $d^2y/dx^2 + y = 0$ with $P(x) = 0$ let one solution be $y_1 = \sin x$. By applying Eq. (9.128), we obtain

$$y_2(x) = \sin x \int^x \frac{dx_2}{\sin^2 x_2} = \sin x(-\cot x) = -\cos x,$$

which is clearly independent (not a linear multiple) of $\sin x$. ■

Series Form of the Second Solution

Further insight into the nature of the second solution of our differential equation may be obtained by the following sequence of operations.

1. Express $P(x)$ and $Q(x)$ in Eq. (9.118) as

$$P(x) = \sum_{i=-1}^{\infty} p_i x^i, \qquad Q(x) = \sum_{j=-2}^{\infty} q_j x^j. \tag{9.129}$$

   The lower limits of the summations are selected to create the strongest possible **regular** singularity (at the origin). These conditions just satisfy Fuchs' theorem and thus help us gain a better understanding of Fuchs' theorem.
2. Develop the first few terms of a power-series solution, as in Section 9.5.
3. Using this solution as y_1, obtain a second series type solution, y_2, with Eq. (9.127), integrating term by term.

Proceeding with Step 1, we have

$$y'' + \left(p_{-1}x^{-1} + p_0 + p_1 x + \cdots\right)y' + \left(q_{-2}x^{-2} + q_{-1}x^{-1} + \cdots\right)y = 0, \tag{9.130}$$

in which point $x = 0$ is at worst a regular singular point. If $p_{-1} = q_{-1} = q_{-2} = 0$, it reduces to an ordinary point. Substituting

$$y = \sum_{\lambda=0}^{\infty} a_\lambda x^{k+\lambda}$$

(Step 2), we obtain

$$\sum_{\lambda=0}^{\infty}(k+\lambda)(k+\lambda-1)a_\lambda x^{k+\lambda-2}+\sum_{i=-1}^{\infty}p_i x^i\sum_{\lambda=0}^{\infty}(k+\lambda)a_\lambda x^{k+\lambda-1}$$
$$+\sum_{j=-2}^{\infty}q_j x^j\sum_{\lambda=0}^{\infty}a_\lambda x^{k+\lambda}=0. \tag{9.131}$$

Assuming that $p_{-1}\neq 0, q_{-2}\neq 0$, our indicial equation is

$$k(k-1)+p_{-1}k+q_{-2}=0,$$

which sets the net coefficient of x^{k-2} equal to zero. This reduces to

$$k^2+(p_{-1}-1)k+q_{-2}=0. \tag{9.132}$$

We denote the two roots of this indicial equation by $k=\alpha$ and $k=\alpha-n$, where n is zero or a positive integer. (If n is not an integer, we expect two independent series solutions by the methods of Section 9.5 and we are done.) Then

$$(k-\alpha)(k-\alpha+n)=0, \tag{9.133}$$

or

$$k^2+(n-2\alpha)k+\alpha(\alpha-n)=0,$$

and equating coefficients of k in Eqs. (9.132) and (9.133), we have

$$p_{-1}-1=n-2\alpha. \tag{9.134}$$

The known series solution corresponding to the larger root $k=\alpha$ may be written as

$$y_1=x^\alpha\sum_{\lambda=0}^{\infty}a_\lambda x^\lambda.$$

Substituting this series solution into Eq. (9.127) (Step 3), we are faced with

$$y_2(x)=y_1(x)\int^x\frac{\exp(-\int_a^{x_2}\sum_{i=-1}^{\infty}p_i x_1^i\,dx_1)}{x_2^{2\alpha}(\sum_{\lambda=0}^{\infty}a_\lambda x_2^\lambda)^2}dx_2, \tag{9.135}$$

where the solutions y_1 and y_2 have been normalized so that the Wronskian $W(a)=1$. Tackling the exponential factor first, we have

$$\int_a^{x_2}\sum_{i=-1}^{\infty}p_i x_1^i\,dx_1=p_{-1}\ln x_2+\sum_{k=0}^{\infty}\frac{p_k}{k+1}x_2^{k+1}+f(a) \tag{9.136}$$

with $f(a)$ an integration constant that may depend on a. Hence,

$$\begin{aligned}
&\exp\left(-\int_a^{x_2}\sum_i p_i x_1^i\,dx_1\right)\\
&\quad=\exp\bigl[-f(a)\bigr]x_2^{-p_{-1}}\exp\left(-\sum_{k=0}^{\infty}\frac{p_k}{k+1}x_2^{k+1}\right)\\
&\quad=\exp\bigl[-f(a)\bigr]x_2^{-p_{-1}}\left[1-\sum_{k=0}^{\infty}\frac{p_k}{k+1}x_2^{k+1}+\frac{1}{2!}\left(-\sum_{k=0}^{\infty}\frac{p_k}{k+1}x_2^{k+1}\right)^2+\cdots\right].
\end{aligned}\tag{9.137}$$

This final series expansion of the exponential is certainly convergent if the original expansion of the coefficient $P(x)$ was uniformly convergent.

The denominator in Eq. (9.135) may be handled by writing

$$\left[x_2^{2\alpha}\left(\sum_{\lambda=0}^{\infty}a_\lambda x_2^\lambda\right)^2\right]^{-1}=x_2^{-2\alpha}\left(\sum_{\lambda=0}^{\infty}a_\lambda x_2^\lambda\right)^{-2}=x_2^{-2\alpha}\sum_{\lambda=0}^{\infty}b_\lambda x_2^\lambda.\tag{9.138}$$

Neglecting constant factors, which will be picked up anyway by the requirement that $W(a)=1$, we obtain

$$y_2(x)=y_1(x)\int^x x_2^{-p_{-1}-2\alpha}\left(\sum_{\lambda=0}^{\infty}c_\lambda x_2^\lambda\right)dx_2.\tag{9.139}$$

By Eq. (9.134),

$$x_2^{-p_{-1}-2\alpha}=x_2^{-n-1},\tag{9.140}$$

and we have assumed here that n is an integer. Substituting this result into Eq. (9.139), we obtain

$$y_2(x)=y_1(x)\int^x\left(c_0x_2^{-n-1}+c_1x_2^{-n}+c_2x_2^{-n+1}+\cdots+c_nx_2^{-1}+\cdots\right)dx_2.\tag{9.141}$$

The integration indicated in Eq. (9.141) leads to a coefficient of $y_1(x)$ consisting of two parts:

1. A power series starting with x^{-n}.
2. A logarithm term from the integration of x^{-1} (when $\lambda=n$). This term always appears when n is an integer, **unless** c_n fortuitously happens to vanish.[18]

[18] For parity considerations, $\ln x$ is taken to be $\ln|x|$, even.

Example 9.6.4 A Second Solution of Bessel's Equation

From Bessel's equation, Eq. (9.100) (divided by x^2 to agree with Eq. (9.118)), we have

$$P(x) = x^{-1} \qquad Q(x) = 1 \qquad \text{for the case } n = 0.$$

Hence $p_{-1} = 1$, $q_0 = 1$; all other p_i and q_j vanish. The Bessel indicial equation is

$$k^2 = 0$$

(Eq. (9.103) with $n = 0$). Hence we verify Eqs. (9.132) to (9.134) with n and $\alpha = 0$.

Our first solution is available from Eq. (9.108). Relabeling it to agree with Chapter 11 (and using $a_0 = 1$), we obtain[19]

$$y_1(x) = J_0(x) = 1 - \frac{x^2}{4} + \frac{x^4}{64} - O\left(x^6\right). \tag{9.142a}$$

Now, substituting all this into Eq. (9.127), we have the specific case corresponding to Eq. (9.135):

$$y_2(x) = J_0(x) \int^x \frac{\exp[-\int^{x_2} x_1^{-1}\, dx_1]}{[1 - x_2^2/4 + x_2^4/64 - \cdots]^2} dx_2. \tag{9.142b}$$

From the numerator of the integrand,

$$\exp\left[-\int^{x_2} \frac{dx_1}{x_1}\right] = \exp[-\ln x_2] = \frac{1}{x_2}.$$

This corresponds to the $x_2^{-p_{-1}}$ in Eq. (9.137). From the denominator of the integrand, using a binomial expansion, we obtain

$$\left[1 - \frac{x_2^2}{4} + \frac{x_2^4}{64}\right]^{-2} = 1 + \frac{x_2^2}{2} + \frac{5x_2^4}{32} + \cdots.$$

Corresponding to Eq. (9.139), we have

$$\begin{aligned} y_2(x) &= J_0(x) \int^x \frac{1}{x_2}\left[1 + \frac{x_2^2}{2} + \frac{5x_2^4}{32} + \cdots\right] dx_2 \\ &= J_0(x)\left\{\ln x + \frac{x^2}{4} + \frac{5x^4}{128} + \cdots\right\}. \end{aligned} \tag{9.142c}$$

Let us check this result. From Eqs. (11.62) and (11.64), which give the standard form of the second solution (higher-order terms are needed)

$$N_0(x) = \frac{2}{\pi}[\ln x - \ln 2 + \gamma] J_0(x) + \frac{2}{\pi}\left\{\frac{x^2}{4} - \frac{3x^4}{128} + \cdots\right\}. \tag{9.142d}$$

Two points arise: (1) Since Bessel's equation is homogeneous, we may multiply $y_2(x)$ by any constant. To match $N_0(x)$, we multiply our $y_2(x)$ by $2/\pi$. (2) To our second solution,

[19] The capital O (order of) as written here means terms proportional to x^6 and possibly higher powers of x.

$(2/\pi)y_2(x)$, we may add any constant multiple of the first solution. Again, to match $N_0(x)$ we add

$$\frac{2}{\pi}[-\ln 2 + \gamma]J_0(x),$$

where γ is the usual Euler–Mascheroni constant (Section 5.2).[20] Our new, modified second solution is

$$y_2(x) = \frac{2}{\pi}[\ln x - \ln 2 + \gamma]J_0(x) + \frac{2}{\pi}J_0(x)\left\{\frac{x^2}{4} + \frac{5x^4}{128} + \cdots\right\}. \tag{9.142e}$$

Now the comparison with $N_0(x)$ becomes a simple multiplication of $J_0(x)$ from Eq. (9.142a) and the curly bracket of Eq. (9.142c). The multiplication checks, through terms of order x^2 and x^4, which is all we carried. Our second solution from Eqs. (9.127) and (9.135) agrees with the standard second solution, the Neumann function, $N_0(x)$.

From the preceding analysis, the second solution of Eq. (9.118), $y_2(x)$, may be written as

$$y_2(x) = y_1(x)\ln x + \sum_{j=-n}^{\infty} d_j x^{j+\alpha}, \tag{9.142f}$$

the first solution times $\ln x$ and another power series, this one starting with $x^{\alpha-n}$, which means that we may look for a logarithmic term when the indicial equation of Section 9.5 gives only one series solution. With the form of the second solution specified by Eq. (9.142f), we can substitute Eq. (9.142f) into the original differential equation and determine the coefficients d_j exactly as in Section 9.5. It may be worth noting that no series expansion of $\ln x$ is needed. In the substitution, $\ln x$ will drop out; its derivatives will survive. ■

The second solution will usually diverge at the origin because of the logarithmic factor and the negative powers of x in the series. For this reason $y_2(x)$ is often referred to as the **irregular solution**. The first series solution, $y_1(x)$, which usually converges at the origin, is called the **regular solution**. The question of behavior at the origin is discussed in more detail in Chapters 11 and 12, in which we take up Bessel functions, modified Bessel functions, and Legendre functions.

Summary

The two solutions of both sections (together with the exercises) provide a **complete solution** of our linear, homogeneous, second-order ODE — assuming that the point of expansion is no worse than a regular singularity. At least one solution can always be obtained by series substitution (Section 9.5). A **second, linearly independent solution** can be constructed by the **Wronskian** double integral, Eq. (9.127). This is all there are: **No third, linearly independent solution exists** (compare Exercise 9.6.10).

The **nonhomogeneous**, linear, second-order ODE will have an **additional** solution: the **particular solution**. This particular solution may be obtained by the method of variation of parameters, Exercise 9.6.25, or by techniques such as Green's function, Section 9.7.

[20] The Neumann function N_0 is defined as it is in order to achieve convenient asymptotic properties, Sections 11.3 and 11.6.

Exercises

9.6.1 You know that the three unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ are mutually perpendicular (orthogonal). Show that $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ are linearly independent. Specifically, show that no relation of the form of Eq. (9.111) exists for $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$.

9.6.2 The criterion for the linear **independence** of three vectors $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ is that the equation

$$a\mathbf{A} + b\mathbf{B} + c\mathbf{C} = 0$$

(analogous to Eq. (9.111)) has no solution other than the trivial $a = b = c = 0$. Using components $\mathbf{A} = (A_1, A_2, A_3)$, and so on, set up the determinant criterion for the existence or nonexistence of a nontrivial solution for the coefficients a, b, and c. Show that your criterion is equivalent to the triple scalar product $\mathbf{A} \cdot \mathbf{B} \times \mathbf{C} \neq 0$.

9.6.3 Using the Wronskian determinant, show that the set of functions

$$\left\{1, \frac{x^n}{n!}\ (n = 1, 2, \ldots, N)\right\}$$

is linearly independent.

9.6.4 If the Wronskian of two functions y_1 and y_2 is identically zero, show by direct integration that

$$y_1 = cy_2,$$

that is, that y_1 and y_2 are dependent. Assume the functions have continuous derivatives and that at least one of the functions does not vanish in the interval under consideration.

9.6.5 The Wronskian of two functions is found to be zero at $x_0 - \varepsilon \leq x \leq x_0 + \varepsilon$ for arbitrarily small $\varepsilon > 0$. Show that this Wronskian vanishes for all x and that the functions are linearly dependent.

9.6.6 The three functions $\sin x$, e^x, and e^{-x} are linearly independent. No one function can be written as a linear combination of the other two. Show that the Wronskian of $\sin x$, e^x, and e^{-x} vanishes but only at isolated points.

ANS. $W = 4\sin x$,
$W = 0$ for $x = \pm n\pi$, $\qquad n = 0, 1, 2, \ldots$.

9.6.7 Consider two functions $\varphi_1 = x$ and $\varphi_2 = |x| = x\,\mathrm{sgn}\,x$ (Fig. 9.3). The function $\mathrm{sgn}\,x$ is the sign of x. Since $\varphi_1' = 1$ and $\varphi_2' = \mathrm{sgn}\,x$, $W(\varphi_1, \varphi_2) = 0$ for any interval, including $[-1, +1]$. Does the vanishing of the Wronskian over $[-1, +1]$ prove that φ_1 and φ_2 are linearly dependent? Clearly, they are not. What is wrong?

9.6.8 Explain that **linear independence** does not mean the absence of any dependence. Illustrate your argument with $\cosh x$ and e^x.

9.6.9 Legendre's differential equation

$$\left(1 - x^2\right)y'' - 2xy' + n(n+1)y = 0$$

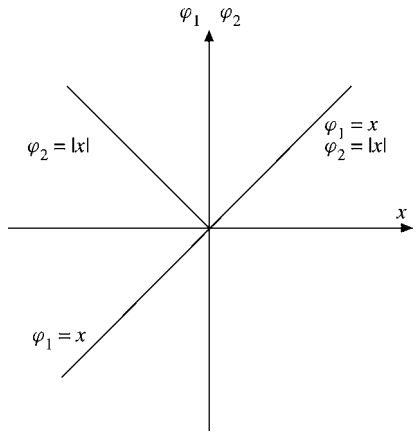

FIGURE 9.3 x and $|x|$.

has a regular solution $P_n(x)$ and an irregular solution $Q_n(x)$. Show that the Wronskian of P_n and Q_n is given by

$$P_n(x)Q_n'(x) - P_n'(x)Q_n(x) = \frac{A_n}{1-x^2},$$

with A_n **independent** of x.

9.6.10 Show, by means of the Wronskian, that a linear, second-order, homogeneous ODE of the form

$$y''(x) + P(x)y'(x) + Q(x)y(x) = 0$$

cannot have three independent solutions. (Assume a third solution and show that the Wronskian vanishes for all x.)

9.6.11 Transform our linear, second-order ODE

$$y'' + P(x)y' + Q(x)y = 0$$

by the substitution

$$y = z\exp\left[-\frac{1}{2}\int^x P(t)\,dt\right]$$

and show that the resulting differential equation for z is

$$z'' + q(x)z = 0,$$

where

$$q(x) = Q(x) - \tfrac{1}{2}P'(x) - \tfrac{1}{4}P^2(x).$$

Note. This substitution can be derived by the technique of Exercise 9.6.24.

9.6.12 Use the result of Exercise 9.6.11 to show that the replacement of $\varphi(r)$ by $r\varphi(r)$ may be expected to eliminate the first derivative from the Laplacian in spherical polar coordinates. See also Exercise 2.5.18(b).

9.6.13 By direct differentiation and substitution show that

$$y_2(x) = y_1(x) \int^x \frac{\exp[-\int^s P(t)\,dt]}{[y_1(s)]^2} ds$$

satisfies (like $y_1(x)$) the ODE

$$y_2''(x) + P(x)y_2'(x) + Q(x)y_2(x) = 0.$$

Note. The Leibniz formula for the derivative of an integral is

$$\frac{d}{d\alpha}\int_{g(\alpha)}^{h(\alpha)} f(x,\alpha)\,dx = \int_{g(\alpha)}^{h(\alpha)} \frac{\partial f(x,\alpha)}{\partial \alpha}\,dx + f\big[h(\alpha),\alpha\big]\frac{dh(\alpha)}{d\alpha} - f\big[g(\alpha),\alpha\big]\frac{dg(\alpha)}{d\alpha}.$$

9.6.14 In the equation

$$y_2(x) = y_1(x) \int^x \frac{\exp[-\int^s P(t)\,dt]}{[y_1(s)]^2} ds$$

$y_1(x)$ satisfies

$$y_1'' + P(x)y_1' + Q(x)y_1 = 0.$$

The function $y_2(x)$ is a linearly **independent** second solution of the same equation. Show that the inclusion of lower limits on the two integrals leads to nothing new, that is, that it generates only an overall constant factor and a constant multiple of the known solution $y_1(x)$.

9.6.15 Given that one solution of

$$R'' + \frac{1}{r}R' - \frac{m^2}{r^2}R = 0$$

is $R = r^m$, show that Eq. (9.127) predicts a second solution, $R = r^{-m}$.

9.6.16 Using $y_1(x) = \sum_{n=0}^{\infty}(-1)^n x^{2n+1}/(2n+1)!$ as a solution of the linear oscillator equation, follow the analysis culminating in Eq. (9.142f) and show that $c_1 = 0$ so that the second solution does not, in this case, contain a logarithmic term.

9.6.17 Show that when n is **not** an integer in Bessel's ODE, Eq. (9.100), the second solution of Bessel's equation, obtained from Eq. (9.127), does **not** contain a logarithmic term.

9.6.18 (a) One solution of Hermite's differential equation

$$y'' - 2xy' + 2\alpha y = 0$$

for $\alpha = 0$ is $y_1(x) = 1$. Find a second solution, $y_2(x)$, using Eq. (9.127). Show that your second solution is equivalent to y_{odd} (Exercise 9.5.6).

(b) Find a second solution for $\alpha = 1$, where $y_1(x) = x$, using Eq. (9.127). Show that your second solution is equivalent to y_{even} (Exercise 9.5.6).

9.6.19 One solution of Laguerre's differential equation

$$xy'' + (1-x)y' + ny = 0$$

for $n = 0$ is $y_1(x) = 1$. Using Eq. (9.127), develop a second, linearly independent solution. Exhibit the logarithmic term explicitly.

9.6.20 For Laguerre's equation with $n = 0$,

$$y_2(x) = \int^x \frac{e^s}{s} ds.$$

(a) Write $y_2(x)$ as a logarithm plus a power series.
(b) Verify that the integral form of $y_2(x)$, previously given, is a solution of Laguerre's equation ($n = 0$) by direct differentiation of the integral and substitution into the differential equation.
(c) Verify that the series form of $y_2(x)$, part (a), is a solution by differentiating the series and substituting back into Laguerre's equation.

9.6.21 One solution of the Chebyshev equation

$$\left(1 - x^2\right)y'' - xy' + n^2 y = 0$$

for $n = 0$ is $y_1 = 1$.

(a) Using Eq. (9.127), develop a second, linearly independent solution.
(b) Find a second solution by direct integration of the Chebyshev equation.

Hint. Let $v = y'$ and integrate. Compare your result with the second solution given in Section 13.3.

ANS. (a) $y_2 = \sin^{-1} x$.
(b) The second solution, $V_n(x)$, is not defined for $n = 0$.

9.6.22 One solution of the Chebyshev equation

$$\left(1 - x^2\right)y'' - xy' + n^2 y = 0$$

for $n = 1$ is $y_1(x) = x$. Set up the Wronskian double integral solution and derive a second solution, $y_2(x)$.

ANS. $y_2 = -\left(1 - x^2\right)^{1/2}$.

9.6.23 The radial Schrödinger wave equation has the form

$$\left\{-\frac{\hbar^2}{2m}\frac{d^2}{dr^2} + l(l+1)\frac{\hbar^2}{2mr^2} + V(r)\right]y(r) = Ey(r).$$

The potential energy $V(r)$ may be expanded about the origin as

$$V(r) = \frac{b_{-1}}{r} + b_0 + b_1 r + \cdots.$$

(a) Show that there is one (regular) solution starting with r^{l+1}.
(b) From Eq. (9.128) show that the irregular solution diverges at the origin as r^{-l}.

9.6.24 Show that if a second solution, y_2, is assumed to have the form $y_2(x) = y_1(x)f(x)$, substitution back into the original equation

$$y_2'' + P(x)y_2' + Q(x)y_2 = 0$$

leads to

$$f(x) = \int^x \frac{\exp[-\int^s P(t)dt]}{[y_1(s)]^2} ds,$$

in agreement with Eq. (9.127).

9.6.25 If our linear, second-order ODE is nonhomogeneous, that is, of the form of Eq. (9.82), the **most general solution is**

$$y(x) = y_1(x) + y_2(x) + y_p(x).$$

(y_1 and y_2 are independent solutions of the homogeneous equation.)
Show that

$$y_p(x) = y_2(x) \int^x \frac{y_1(s)F(s)\,ds}{W\{y_1(s), y_2(s)\}} - y_1(x) \int^x \frac{y_2(s)F(s)\,ds}{W\{y_1(s), y_2(s)\}},$$

with $W\{y_1(x), y_2(x)\}$ the Wronskian of $y_1(s)$ and $y_2(s)$.
Hint. As in Exercise 9.6.24, let $y_p(x) = y_1(x)v(x)$ and develop a first-order differential equation for $v'(x)$.

9.6.26 (a) Show that

$$y'' + \frac{1-\alpha^2}{4x^2} y = 0$$

has two solutions:

$$y_1(x) = a_0 x^{(1+\alpha)/2},$$
$$y_2(x) = a_0 x^{(1-\alpha)/2}.$$

(b) For $\alpha = 0$ the two linearly independent solutions of part (a) reduce to $y_{10} = a_0 x^{1/2}$. Using Eq. (9.128) derive a second solution,

$$y_{20}(x) = a_0 x^{1/2} \ln x.$$

Verify that y_{20} is indeed a solution.
(c) Show that the second solution from part (b) may be obtained as a limiting case from the two solutions of part (a):

$$y_{20}(x) = \lim_{\alpha \to 0} \left(\frac{y_1 - y_2}{\alpha} \right).$$

9.7 NONHOMOGENEOUS EQUATION — GREEN'S FUNCTION

The series substitution of Section 9.5 and the Wronskian double integral of Section 9.6 provide the most general solution of the **homogeneous**, linear, second-order ODE. The specific solution, y_p, linearly dependent on the source term ($F(x)$ of Eq. (9.82)) may be cranked out by the variation of parameters method, Exercise 9.6.25. In this section we turn to a different method of solution — Green's function.

For a brief introduction to Green's function method, as applied to the solution of a nonhomogeneous PDE, it is helpful to use the electrostatic analog. In the presence of charges

the electrostatic potential ψ satisfies Poisson's nonhomogeneous equation (compare Section 1.14),

$$\nabla^2\psi = -\frac{\rho}{\varepsilon_0} \qquad \text{(mks units)}, \tag{9.143}$$

and Laplace's homogeneous equation,

$$\nabla^2\psi = 0, \tag{9.144}$$

in the absence of electric charge ($\rho = 0$). If the charges are point charges q_i, we know that the solution is

$$\psi = \frac{1}{4\pi\varepsilon_0}\sum_i \frac{q_i}{r_i}, \tag{9.145}$$

a superposition of single-point charge solutions obtained from Coulomb's law for the force between two point charges q_1 and q_2,

$$\mathbf{F} = \frac{q_1 q_2 \hat{\mathbf{r}}}{4\pi\varepsilon_0 r^2}. \tag{9.146}$$

By replacement of the discrete point charges with a smeared-out distributed charge, charge density ρ, Eq. (9.145) becomes

$$\psi(r=0) = \frac{1}{4\pi\varepsilon_0}\int \frac{\rho(\mathbf{r})}{r}d\tau \tag{9.147}$$

or, for the potential at $\mathbf{r} = \mathbf{r}_1$ away from the origin and the charge at $\mathbf{r} = \mathbf{r}_2$,

$$\psi(\mathbf{r}_1) = \frac{1}{4\pi\varepsilon_0}\int \frac{\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}d\tau_2. \tag{9.148}$$

We use ψ as the potential corresponding to the given distribution of charge and therefore satisfying Poisson's equation (9.143), whereas a function G, which we label Green's function, is required to satisfy Poisson's equation with a point source at the point defined by $\mathbf{r}_2$:

$$\boxed{\nabla^2 G = -\delta(\mathbf{r}_1 - \mathbf{r}_2).} \tag{9.149}$$

Physically, then, G is the potential at $\mathbf{r}_1$ corresponding to a unit source at $\mathbf{r}_2$. By Green's theorem (Section 1.11, Eq. (11.104))

$$\int \left(\psi\nabla^2 G - G\nabla^2\psi\right)d\tau_2 = \int (\psi\nabla G - G\nabla\psi)\cdot d\boldsymbol{\sigma}. \tag{9.150}$$

Assuming that the integrand falls off faster than r^{-2} we may simplify our problem by taking the volume so large that the surface integral vanishes, leaving

$$\int \psi\nabla^2 G\, d\tau_2 = \int G\nabla^2\psi\, d\tau_2, \tag{9.151}$$

or, by substituting in Eqs. (9.143) and (9.149), we have

$$-\int \psi(\mathbf{r}_2)\delta(\mathbf{r}_1 - \mathbf{r}_2)\, d\tau_2 = -\int \frac{G(\mathbf{r}_1, \mathbf{r}_2)\rho(\mathbf{r}_2)}{\varepsilon_0}d\tau_2. \tag{9.152}$$

Integration by employing the defining property of the Dirac delta function (Eq. (1.171b)) produces

$$\boxed{\psi(\mathbf{r}_1) = \frac{1}{\varepsilon_0}\int G(\mathbf{r}_1, \mathbf{r}_2)\rho(\mathbf{r}_2)\, d\tau_2.} \tag{9.153}$$

Note that we have used Eq. (9.149) to eliminate $\nabla^2 G$ but that the function G itself is still unknown. In Section 1.14, Gauss law, we found that

$$\int \nabla^2\left(\frac{1}{r}\right) d\tau = \begin{cases} 0, \\ -4\pi, \end{cases} \tag{9.154}$$

0 if the volume did not include the origin and -4π if the origin were included. This result from Section 1.14 may be rewritten as in Eq. (1.170), or

$$\nabla^2\left(\frac{1}{4\pi r}\right) = -\delta(\mathbf{r}), \qquad \text{or} \qquad \nabla^2\left(\frac{1}{4\pi r_{12}}\right) = -\delta(\mathbf{r}_1 - \mathbf{r}_2), \tag{9.155}$$

corresponding to a shift of the electrostatic charge from the origin to the position $\mathbf{r} = \mathbf{r}_2$. Here $r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$, and the Dirac delta function $\delta(\mathbf{r}_1 - \mathbf{r}_2)$ vanishes unless $\mathbf{r}_1 = \mathbf{r}_2$. Therefore in a comparison of Eqs. (9.149) and (9.155) the function G (Green's function) is given by

$$G(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{4\pi|\mathbf{r}_1 - \mathbf{r}_2|}. \tag{9.156}$$

The solution of our differential equation (Poisson's equation) is

$$\psi(\mathbf{r}_1) = \frac{1}{4\pi\varepsilon_0}\int \frac{\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\tau_2, \tag{9.157}$$

in complete agreement with Eq. (9.148). Actually $\psi(\mathbf{r}_1)$, Eq. (9.157), is the particular solution of Poisson's equation. We may add solutions of Laplace's equation (compare Eq. (9.83)). Such solutions could describe an external field.

These results will be generalized to the second-order, linear, but nonhomogeneous differential equation

$$\boxed{\mathcal{L}y(\mathbf{r}_1) = -f(\mathbf{r}_1),} \tag{9.158}$$

where $\mathcal{L}$ is a linear differential operator. The Green's function is taken to be a solution of

$$\boxed{\mathcal{L}G(\mathbf{r}_1, \mathbf{r}_2) = -\delta(\mathbf{r}_1 - \mathbf{r}_2),} \tag{9.159}$$

analogous to Eq. (9.149). The Green's function depends on boundary conditions that may no longer be those of electrostatics in a region of infinite extent. Then the particular solution $y(\mathbf{r}_1)$ becomes

$$\boxed{y(\mathbf{r}_1) = \int G(\mathbf{r}_1, \mathbf{r}_2) f(\mathbf{r}_2)\, d\tau_2.} \tag{9.160}$$

(There may also be an integral over a bounding surface, depending on the conditions specified.)

In summary, Green's function, often written $G(\mathbf{r}_1, \mathbf{r}_2)$ *as a reminder of the name, is a solution of Eq.* (9.149) *or Eq.* (9.159) *more generally. It enters in an integral solution of our differential equation, as in Eqs.* (9.148) *and* (9.153). *For the simple, but important, electrostatic case we obtain Green's function,* $G(\mathbf{r}_1, \mathbf{r}_2)$, *by Gauss' law, comparing Eqs.* (9.149) *and* (9.155). *Finally, from the final solution* (*Eq.* (9.157)) *it is possible to develop a physical interpretation of Green's function. It occurs as a weighting function or propagator function that enhances or reduces the effect of the charge element* $\rho(\mathbf{r}_2)\, d\tau_2$ *according to its distance from the field point* $\mathbf{r}_1$. *Green's function,* $G(\mathbf{r}_1, \mathbf{r}_2)$, *gives the effect of a unit point source at* $\mathbf{r}_2$ *in producing a potential at* $\mathbf{r}_1$. *This is how it was introduced in Eq.* (9.149); *this is how it appears in Eq.* (9.157).

Symmetry of Green's Function

An important property of Green's function is the symmetry of its two variables; that is,

$$G(\mathbf{r}_1, \mathbf{r}_2) = G(\mathbf{r}_2, \mathbf{r}_1). \tag{9.161}$$

Although this is obvious in the electrostatic case just considered, it can be proved under more general conditions. In place of Eq. (9.149), let us require that $G(\mathbf{r}, \mathbf{r}_1)$ satisfy[21]

$$\nabla \cdot \big[p(\mathbf{r})\nabla G(\mathbf{r}, \mathbf{r}_1)\big] + \lambda q(\mathbf{r}) G(\mathbf{r}, \mathbf{r}_1) = -\delta(\mathbf{r} - \mathbf{r}_1), \tag{9.162}$$

corresponding to a mathematical point source at $\mathbf{r} = \mathbf{r}_1$. Here the functions $p(\mathbf{r})$ and $q(\mathbf{r})$ are well-behaved but otherwise arbitrary functions of $\mathbf{r}$. The Green's function, $G(\mathbf{r}, \mathbf{r}_2)$, satisfies the same equation, but the subscript 1 is replaced by subscript 2. The Green's functions, $G(\mathbf{r}, \mathbf{r}_1)$ and $G(\mathbf{r}, \mathbf{r}_2)$, have the same values over a given surface S of some volume of finite or infinite extent, and their normal derivatives have the same values over the surface S, or these Green's functions vanish on S (Dirichlet boundary conditions, Section 9.1).[22] Then $G(\mathbf{r}, \mathbf{r}_2)$ is a sort of potential at $\mathbf{r}$, created by a unit point source at $\mathbf{r}_2$.

We multiply the equation for $G(\mathbf{r}, \mathbf{r}_1)$ by $G(\mathbf{r}, \mathbf{r}_2)$ and the equation for $G(\mathbf{r}, \mathbf{r}_2)$ by $G(\mathbf{r}, \mathbf{r}_1)$ and then subtract the two:

$$\begin{aligned} &G(\mathbf{r}, \mathbf{r}_2)\nabla \cdot \big[p(\mathbf{r})\nabla G(\mathbf{r}, \mathbf{r}_1)\big] - G(\mathbf{r}, \mathbf{r}_1)\nabla \cdot \big[p(\mathbf{r})\nabla G(\mathbf{r}, \mathbf{r}_2)\big] \\ &\quad = -G(\mathbf{r}, \mathbf{r}_2)\delta(\mathbf{r} - \mathbf{r}_1) + G(\mathbf{r}, \mathbf{r}_1)\delta(\mathbf{r} - \mathbf{r}_2). \end{aligned} \tag{9.163}$$

The first term in Eq. (9.163),

$$G(\mathbf{r}, \mathbf{r}_2)\nabla \cdot \big[p(\mathbf{r})\nabla G(\mathbf{r}, \mathbf{r}_1)\big],$$

may be replaced by

$$\nabla \cdot \big[G(\mathbf{r}, \mathbf{r}_2)\, p(\mathbf{r})\nabla G(\mathbf{r}, \mathbf{r}_1)\big] - \nabla G(\mathbf{r}, \mathbf{r}_2) \cdot p(\mathbf{r})\nabla G(\mathbf{r}, \mathbf{r}_1).$$

[21] Equation (9.162) is a three-dimensional, inhomogeneous version of the **self-adjoint** eigenvalue equation, Eq. (10.8).

[22] Any attempt to demand that the normal derivatives vanish at the surface (Neumann's conditions, Section 9.1) leads to trouble with Gauss' law. It is like demanding that $\int \mathbf{E} \cdot d\boldsymbol{\sigma} = 0$ when you know perfectly well that there is some electric charge inside the surface.

A similar transformation is carried out on the second term. Then integrating over the volume whose surface is S and using Green's theorem, we obtain a surface integral:

$$\int_S \big[G(\mathbf{r},\mathbf{r}_2)\,p(\mathbf{r})\nabla G(\mathbf{r},\mathbf{r}_1) - G(\mathbf{r},\mathbf{r}_1)\,p(\mathbf{r})\nabla G(\mathbf{r},\mathbf{r}_2)\big]\cdot d\boldsymbol{\sigma}$$
$$= -G(\mathbf{r}_1,\mathbf{r}_2) + G(\mathbf{r}_2,\mathbf{r}_1). \tag{9.164}$$

The terms on the right-hand side appear when we use the Dirac delta functions in Eq. (9.163) and carry out the volume integration. With the boundary conditions earlier imposed on the Green's function, the surface integral vanishes and

$$G(\mathbf{r}_1,\mathbf{r}_2) = G(\mathbf{r}_2,\mathbf{r}_1), \tag{9.165}$$

which shows that Green's function is symmetric. If the eigenfunctions are complex, boundary conditions corresponding to Eqs. (10.19) to (10.20) are appropriate. Equation (9.165) becomes

$$G(\mathbf{r}_1,\mathbf{r}_2) = G^*(\mathbf{r}_2,\mathbf{r}_1). \tag{9.166}$$

Note that this symmetry property holds for Green's functions in every equation in the form of Eq. (9.162). In Chapter 10 we shall call equations in this form *self-adjoint*. The symmetry is the basis of various reciprocity theorems; the effect of a charge at $\mathbf{r}_2$ on the potential at $\mathbf{r}_1$ is the same as the effect of a charge at $\mathbf{r}_1$ on the potential at $\mathbf{r}_2$.

This use of Green's functions is a powerful technique for solving many of the more difficult problems of mathematical physics.

Form of Green's Functions

Let us assume that $\mathcal{L}$ is a self-adjoint differential operator of the general form[23]

$$\mathcal{L}_1 = \nabla_1\cdot\big[p(\mathbf{r}_1)\nabla_1\big] + q(\mathbf{r}_1). \tag{9.167}$$

Here the subscript 1 on $\mathcal{L}$ emphasizes that $\mathcal{L}$ operates on $\mathbf{r}_1$. Then, as a simple generalization of Green's theorem, Eq. (1.104), we have

$$\int (v\mathcal{L}_2 u - u\mathcal{L}_2 v)\,d\tau_2 = \int p(v\nabla_2 u - u\nabla_2 v)\cdot d\boldsymbol{\sigma}_2, \tag{9.168}$$

in which all quantities have $\mathbf{r}_2$ as their argument. (To verify Eq. (9.168), take the divergence of the integrand of the surface integral.) We let $u(\mathbf{r}_2) = y(\mathbf{r}_2)$ so that Eq. (9.158) applies and $v(\mathbf{r}_2) = G(\mathbf{r}_1,\mathbf{r}_2)$ so that Eq. (9.159) applies. (Remember, $G(\mathbf{r}_1,\mathbf{r}_2) = G(\mathbf{r}_2,\mathbf{r}_1)$.) Substituting into Green's theorem we get

$$\int \big\{-G(\mathbf{r}_1,\mathbf{r}_2)\,f(\mathbf{r}_2) + y(\mathbf{r}_2)\delta(\mathbf{r}_1-\mathbf{r}_2)\big\}\,d\tau_2$$
$$= \int p(\mathbf{r}_2)\big\{G(\mathbf{r}_1,\mathbf{r}_2)\nabla_2 y(\mathbf{r}_2) - y(\mathbf{r}_2)\nabla_2 G(\mathbf{r}_1,\mathbf{r}_2)\big\}\cdot d\boldsymbol{\sigma}_2. \tag{9.169}$$

[23] $\mathcal{L}_1$ may be in 1, 2, or 3 dimensions (with appropriate interpretation of ∇_1).

When we integrate over the Dirac delta function

$$y(\mathbf{r}_1) = \int G(\mathbf{r}_1, \mathbf{r}_2) f(\mathbf{r}_2)\, d\tau_2 + \int p(\mathbf{r}_2)\big\{G(\mathbf{r}_1, \mathbf{r}_2)\nabla_2 y(\mathbf{r}_2) - y(\mathbf{r}_2)\nabla_2 G(\mathbf{r}_1, \mathbf{r}_2)\big\}\cdot d\boldsymbol{\sigma}_2, \tag{9.170}$$

our solution to Eq. (9.158) appears as a volume integral plus a surface integral. If y and G **both** satisfy Dirichlet boundary conditions or if both satisfy Neumann boundary conditions, the surface integral vanishes and we regain Eq. (9.160). The volume integral is a weighted integral over the source term $f(\mathbf{r}_2)$ with our Green's function $G(\mathbf{r}_1, \mathbf{r}_2)$ as the weighting function.

For the special case of $p(\mathbf{r}_1) = 1$ and $q(\mathbf{r}_1) = 0$, $\mathcal{L}$ is ∇^2, the Laplacian. Let us integrate

$$\nabla_1^2 G(\mathbf{r}_1, \mathbf{r}_2) = -\delta(\mathbf{r}_1 - \mathbf{r}_2) \tag{9.171}$$

over a small volume including the point source. Then

$$\int \nabla_1 \cdot \nabla_1 G(\mathbf{r}_1, \mathbf{r}_2)\, d\tau_1 = -\int \delta(\mathbf{r}_1 - \mathbf{r}_2)\, d\tau_1 = -1. \tag{9.172}$$

The volume integral on the left may be transformed by Gauss' theorem, as in the development of Gauss' law — Section 1.14. We find that

$$\int \nabla_1 G(\mathbf{r}_1, \mathbf{r}_2) \cdot d\boldsymbol{\sigma}_1 = -1. \tag{9.173}$$

This shows, incidentally, that it may not be possible to impose a Neumann boundary condition, that the normal derivative of the Green's function, $\partial G/\partial n$, vanishes over the entire surface.

If we are in three-dimensional space, Eq. (9.173) is satisfied by taking

$$\frac{\partial}{\partial r_{12}} G(\mathbf{r}_1, \mathbf{r}_2) = -\frac{1}{4\pi} \cdot \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|^2}, \qquad r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|. \tag{9.174}$$

The integration is over the surface of a sphere centered at $\mathbf{r}_2$. The integral of Eq. (9.174) is

$$G(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{4\pi} \cdot \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}, \tag{9.175}$$

in agreement with Section 1.14.

If we are in two-dimensional space, Eq. (9.173) is satisfied by taking

$$\frac{\partial}{\partial \rho_{12}} G(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = \frac{1}{2\pi} \cdot \frac{1}{|\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2|}, \tag{9.176}$$

with $\mathbf{r}$ being replaced by $\boldsymbol{\rho}$, $\rho = (x^2 + y^2)^{1/2}$, and the integration being over the circumference of a circle centered on $\boldsymbol{\rho}_2$. Here $\rho_{12} = |\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2|$. Integrating Eq. (9.176), we obtain

$$G(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = -\frac{1}{2\pi} \ln |\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2|. \tag{9.177}$$

To $G(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)$ (and to $G(\mathbf{r}_1, \mathbf{r}_2)$) we may add any multiple of the regular solution of the homogeneous (Laplace's) equation as needed to satisfy boundary conditions.

Table 9.5 Green's Functions[a]

	Laplace ∇^2	Helmholtz $\nabla^2 + k^2$	Modified Helmholtz $\nabla^2 - k^2$
One-dimensional space	No solution for $(-\infty, \infty)$	$\frac{i}{2k}\exp(ik\lvert x_1 - x_2\rvert)$	$\frac{1}{2k}\exp(-k\lvert x_1 - x_2\rvert)$
Two-dimensional space	$-\frac{1}{2\pi}\ln\lvert\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2\rvert$	$\frac{i}{4}H_0^{(1)}(k\lvert\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2\rvert)$	$\frac{1}{2\pi}K_0(k\lvert\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2\rvert)$
Three-dimensional space	$\frac{1}{4\pi}\cdot\frac{1}{\lvert\mathbf{r}_1 - \mathbf{r}_2\rvert}$	$\frac{\exp(ik\lvert\mathbf{r}_1 - \mathbf{r}_2\rvert)}{4\pi\lvert\mathbf{r}_1 - \mathbf{r}_2\rvert}$	$\frac{\exp(-k\lvert\mathbf{r}_1 - \mathbf{r}_2\rvert)}{4\pi\lvert\mathbf{r}_1 - \mathbf{r}_2\rvert}$

[a]These are the Green's functions satisfying the boundary condition $G(\mathbf{r}_1, \mathbf{r}_2) = 0$ as $\mathbf{r}_1 \to \infty$ for the Laplace and modified Helmholtz operators. For the Helmholtz operator, $G(\mathbf{r}_1, \mathbf{r}_2)$ corresponds to an outgoing wave. $H_0^{(1)}$ is the Hankel function of Section 11.4. K_0 is he modified Bessel function of Section 11.5.

The behavior of the Laplace operator Green's function in the vicinity of the source point $\mathbf{r}_1 = \mathbf{r}_2$ shown by Eqs. (9.175) and (9.177) facilitates the identification of the Green's functions for the other cases, such as the Helmholtz and modified Helmholtz equations.

1. For $\mathbf{r}_1 \neq \mathbf{r}_2$, $G(\mathbf{r}_1, \mathbf{r}_2)$ must satisfy the **homogeneous** differential equation

$$\mathcal{L}_1 G(\mathbf{r}_1, \mathbf{r}_2) = 0, \qquad \mathbf{r}_1 \neq \mathbf{r}_2. \tag{9.178}$$

2. As $\mathbf{r}_1 \to \mathbf{r}_2$ (or $\boldsymbol{\rho}_1 \to \boldsymbol{\rho}_2$),

$$G(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) \approx -\frac{1}{2\pi}\ln\lvert\boldsymbol{\rho}_1 - \boldsymbol{\rho}_2\rvert, \qquad \text{two-dimensional space,} \tag{9.179}$$

$$G(\mathbf{r}_1, \mathbf{r}_2) \approx \frac{1}{4\pi}\cdot\frac{1}{\lvert\mathbf{r}_1 - \mathbf{r}_2\rvert}, \qquad \text{three-dimensional space.} \tag{9.180}$$

The term $\pm k^2$ in the operator does not affect the behavior of G near the singular point $\mathbf{r}_1 = \mathbf{r}_2$. For convenience, the Green's functions for the Laplace, Helmholtz, and modified Helmholtz operators are listed in Table 9.5.

Spherical Polar Coordinate Expansion[24]

As an alternate determination of the Green's function of the Laplace operator, let us assume a spherical harmonic expansion of the form

$$G(\mathbf{r}_1, \mathbf{r}_2) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} g_l(r_1, r_2) Y_l^m(\theta_1, \varphi_1) Y_l^{m*}(\theta_2, \varphi_2), \tag{9.181}$$

where the summation index l is the same for the spherical harmonics, as a consequence of the symmetry of the Green's function. We will now determine the radial functions

[24]This section is optional here and may be postponed to Chapter 12.

$g_l(r_1, r_2)$. From Exercises 1.15.11 and 12.6.6,

$$\delta(\mathbf{r}_1 - \mathbf{r}_2) = \frac{1}{r_1^2}\delta(r_1 - r_2)\delta(\cos\theta_1 - \cos\theta_2)\delta(\varphi_1 - \varphi_2)$$
$$= \frac{1}{r_1^2}\delta(r_1 - r_2)\sum_{l=0}^{\infty}\sum_{m=-l}^{l} Y_l^m(\theta_1, \varphi_1)Y_l^{m*}(\theta_2, \varphi_2). \tag{9.182}$$

Substituting Eqs. (9.181) and (9.182) into the Green's function differential equation, Eq. (9.171), and making use of the orthogonality of the spherical harmonics, we obtain a radial equation:

$$r_1\frac{d^2}{dr_1^2}\big[r_1 g_l(r_1, r_2)\big] - l(l+1)g_l(r_1, r_2) = -\delta(r_1 - r_2). \tag{9.183}$$

This is now a one-dimensional problem. The solutions[25] of the corresponding homogeneous equation are r_1^l and r_1^{-l-1}. If we demand that g_l remain finite as $r_1 \to 0$ and vanish as $r_1 \to \infty$, the technique of Section 10.5 leads to

$$g_l(r_1, r_2) = \frac{1}{2l+1}\begin{cases} \dfrac{r_1^l}{r_2^{l+1}}, & r_1 < r_2, \\ \dfrac{r_2^l}{r_1^{l+1}}, & r_1 > r_2, \end{cases} \tag{9.184}$$

or

$$g_l(r_1, r_2) = \frac{1}{2l+1}\cdot\frac{r_<^l}{r_>^{l+1}}. \tag{9.185}$$

Hence our Green's function is

$$G(\mathbf{r}_1, \mathbf{r}_2) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{1}{2l+1}\frac{r_<^l}{r_>^{l+1}}Y_l^m(\theta_1, \varphi_1)Y_l^{m*}(\theta_2, \varphi_2). \tag{9.186}$$

Since we already have $G(r_1, r_2)$ in closed form, Eq. (9.175), we may write

$$\frac{1}{4\pi}\cdot\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{1}{2l+1}\frac{r_<^l}{r_>^{l+1}}Y_l^m(\theta_1, \varphi_1)Y_l^{m*}(\theta_2, \varphi_2). \tag{9.187}$$

One immediate use for this spherical harmonic expansion of the Green's function is in the development of an electrostatic multipole expansion. The potential for an arbitrary charge distribution is

$$\psi(\mathbf{r}_1) = \frac{1}{4\pi\varepsilon_0}\int\frac{\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}d\tau_2$$

[25] Compare Table 9.2.

(which is Eq. (9.148)). Substituting Eq. (9.187), we get

$$\psi(\mathbf{r}_1) = \frac{1}{\varepsilon_0}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left\{\frac{1}{2l+1}\frac{Y_l^m(\theta_1,\varphi_1)}{r_1^{l+1}} \cdot \int \rho(\mathbf{r}_2)Y_l^{m*}(\theta_2,\varphi_2)r_2^l\,d\varphi_2\sin\theta_2\,d\theta_2\,r_2^2\,dr_2\right\}, \qquad \text{for } r_1 > r_2.$$

This is the **multipole expansion**. The relative importance of the various terms in the double sum depends on the form of the source, $\rho(\mathbf{r}_2)$.

Legendre Polynomial Addition Theorem[26]

From the generating expression for Legendre polynomials, Eq. (12.4a),

$$\frac{1}{4\pi}\cdot\frac{1}{|\mathbf{r}_1-\mathbf{r}_2|} = \frac{1}{4\pi}\sum_{l=0}^{\infty}\frac{r_<^l}{r_>^{l+1}}P_l(\cos\gamma), \tag{9.188}$$

where γ is the angle included between vectors $\mathbf{r}_1$ and $\mathbf{r}_2$, Fig. 9.4. Equating Eqs. (9.187) and (9.188), we have the Legendre polynomial addition theorem:

$$P_l(\cos\gamma) = \frac{4\pi}{2l+1}\sum_{m=-l}^{l}Y_l^m(\theta_1,\varphi_1)Y_l^{m*}(\theta_2,\varphi_2). \tag{9.189}$$

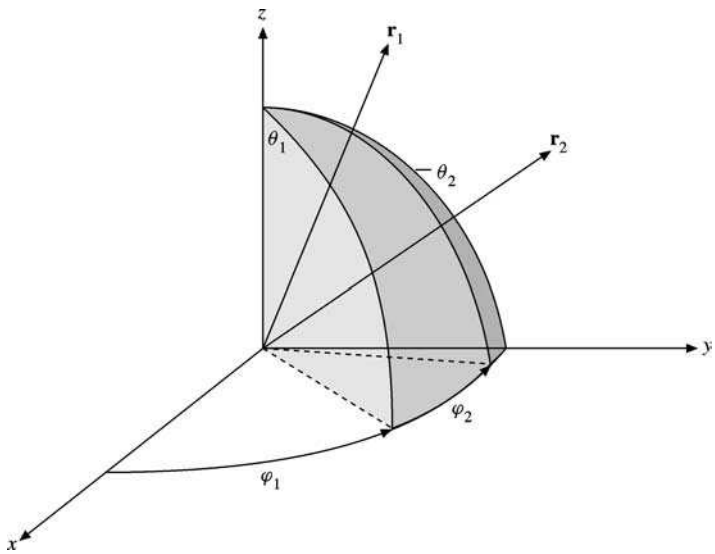

FIGURE 9.4 Spherical polar coordinates.

[26]This section is optional here and may be postponed to Chapter 12.

It is instructive to compare this derivation with the relatively cumbersome derivation of Section 12.8 leading to Eq. (12.177).

Circular Cylindrical Coordinate Expansion[27]

In analogy with the preceding spherical polar coordinate expansion, we write

$$\begin{aligned}\delta(\mathbf{r}_1-\mathbf{r}_2) &= \frac{1}{\rho_1}\delta(\rho_1-\rho_2)\delta(\varphi_1-\varphi_2)\delta(z_1-z_2)\\ &= \frac{1}{\rho_1}\delta(\rho_1-\rho_2)\frac{1}{4\pi^2}\sum_{m=-\infty}^{\infty} e^{im(\varphi_1-\varphi_2)}\int_{-\infty}^{\infty} e^{ik(z_1-z_2)}\,dk,\end{aligned} \tag{9.190}$$

using Exercise 12.6.5 and Eq. (1.193c) and the Cauchy principal value. But why a summation for the φ-dependence and an integration for the z-dependence? The requirement that the azimuthal dependence be single-valued quantizes m, hence the summation. No such restriction applies to k.

To avoid problems later with negative values of k, we rewrite Eq. (9.190) as

$$\delta(\mathbf{r}_1-\mathbf{r}_2) = \frac{1}{\rho_1}\delta(\rho_1-\rho_2)\frac{1}{2\pi}\sum_{m=-\infty}^{\infty} e^{im(\varphi_1-\varphi_2)}\frac{1}{\pi}\int_0^{\infty}\cos k(z_1-z_2)\,dk. \tag{9.191}$$

We assume a similar expansion of the Green's function,

$$G(r_1,r_2) = \frac{1}{2\pi^2}\sum_{m=-\infty}^{\infty} g_m(\rho_1,\rho_2)e^{im(\varphi_1-\varphi_2)}\int_0^{\infty}\cos k(z_1-z_2)\,dk, \tag{9.192}$$

with the ρ-dependent coefficients $g_m(\rho_1,\rho_2)$ to be determined. Substituting into Eq. (9.171), now in circular cylindrical coordinates, we find that if $g(\rho_1,\rho_2)$ satisfies

$$\frac{d}{d\rho_1}\left[\rho_1\frac{dg_m}{d\rho_1}\right]-\left[k^2\rho_1+\frac{m^2}{\rho_1}\right]g_m = -\delta(\rho_1-\rho_2), \tag{9.193}$$

then Eq. (9.171) is satisfied.

The operator in Eq. (9.193) is identified as the modified Bessel operator (in self-adjoint form). Hence the solutions of the corresponding homogeneous equation are $u_1 = I_m(k\rho), u_2 = K_m(k\rho)$. As in the spherical polar coordinate case, we demand that G be finite at $\rho_1 = 0$ and vanish as $\rho_1 \to \infty$. Then the technique of Section 10.5 yields

$$g_m(\rho_1,\rho_2) = -\frac{1}{A}I_m(k\rho_<)K_m(k\rho_>). \tag{9.194}$$

This corresponds to Eq. (9.155). The constant A comes from the Wronskian (see Eq. (9.120)):

$$I_m(k\rho)K_m'(k\rho) - I_m'(k\rho)K_m(k\rho) = \frac{A}{P(k\rho)}. \tag{9.195}$$

[27] This section is optional here and may be postponed to Chapter 11.

From Exercise 11.5.10, $A = -1$ and

$$g_m(\rho_1, \rho_2) = I_m(k\rho_<) K_m(k\rho_>). \tag{9.196}$$

Therefore our circular cylindrical coordinate Green's function is

$$\begin{aligned} G(\mathbf{r}_1, \mathbf{r}_2) &= \frac{1}{4\pi} \cdot \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \\ &= \frac{1}{2\pi^2} \sum_{m=-\infty}^{\infty} \int_0^{\infty} I_m(k\rho_<) K_m(k\rho_>) e^{im(\varphi_1 - \varphi_2)} \cos k(z_1 - z_2)\, dk. \end{aligned} \tag{9.197}$$

Exercise 9.7.14 is a special case of this result.

Example 9.7.1 QUANTUM MECHANICAL SCATTERING — NEUMANN SERIES SOLUTION

The quantum theory of scattering provides a nice illustration of integral equation techniques and an application of a Green's function. Our physical picture of scattering is as follows. A beam of particles moves along the negative z-axis toward the origin. A small fraction of the particles is scattered by the potential $V(\mathbf{r})$ and goes off as an outgoing spherical wave. Our wave function $\psi(\mathbf{r})$ must satisfy the time-independent Schrödinger equation

$$-\frac{\hbar^2}{2m} \nabla^2 \psi(\mathbf{r}) + V(\mathbf{r}) \psi(\mathbf{r}) = E \psi(\mathbf{r}), \tag{9.198a}$$

or

$$\nabla^2 \psi(\mathbf{r}) + k^2 \psi(\mathbf{r}) = -\left[-\frac{2m}{\hbar^2} V(\mathbf{r}) \psi(\mathbf{r}) \right], \qquad k^2 = \frac{2mE}{\hbar^2}. \tag{9.198b}$$

From the physical picture just presented we look for a solution having an **asymptotic** form

$$\psi(\mathbf{r}) \sim e^{i\mathbf{k_0}\cdot\mathbf{r}} + f_k(\theta, \varphi) \frac{e^{ikr}}{r}. \tag{9.199}$$

Here $e^{i\mathbf{k_0}\cdot\mathbf{r}}$ is the incident plane wave[28] with $\mathbf{k}_0$ the propagation vector carrying the subscript 0 to indicate that it is in the $\theta = 0$ (z-axis) direction. The magnitudes k_0 and k are equal (ignoring recoil), and e^{ikr}/r is the outgoing spherical wave with an angular (and energy) dependent amplitude factor $f_k(\theta, \varphi)$.[29] Vector $\mathbf{k}$ has the direction of the outgoing scattered wave. In quantum mechanics texts it is shown that the differential probability of scattering, $d\sigma/d\Omega$, the scattering cross section per unit solid angle, is given by $|f_k(\theta, \varphi|^2$.

Identifying $[-(2m/\hbar^2) V(\mathbf{r}) \psi(\mathbf{r})]$ with $f(\mathbf{r})$ of Eq. (9.158), we have

$$\psi(\mathbf{r}_1) = -\int \frac{2m}{\hbar^2} V(\mathbf{r}_2) \psi(\mathbf{r}_2) G(\mathbf{r}_1, \mathbf{r}_2)\, d^3 r_2 \tag{9.200}$$

[28] For simplicity we assume a continuous incident beam. In a more sophisticated and more realistic treatment, Eq. (9.199) would be one component of a Fourier wave packet.

[29] If $V(\mathbf{r})$ represents a central force, f_k will be a function of θ only, independent of azimuth.

by Eq. (9.170). This does not have the desired asymptotic form of Eq. (9.199), but we may add to Eq. (9.200) $e^{i\mathbf{k_0}\cdot\mathbf{r_1}}$, a solution of the homogeneous equation, and put $\psi(\mathbf{r})$ into the desired form:

$$\psi(\mathbf{r}_1) = e^{i\mathbf{k_0}\cdot\mathbf{r_1}} - \int \frac{2m}{\hbar^2} V(\mathbf{r}_2)\psi(\mathbf{r}_2)G(\mathbf{r}_1,\mathbf{r}_2)\,d^3r_2. \tag{9.201}$$

Our Green's function is the Green's function of the operator $\mathcal{L} = \nabla^2 + k^2$ (Eq. (9.198)), satisfying the boundary condition that it describe an outgoing wave. Then, from Table 9.5, $G(\mathbf{r}_1,\mathbf{r}_2) = \exp(ik|\mathbf{r}_1-\mathbf{r}_2|)/(4\pi|\mathbf{r}_1-\mathbf{r}_2|)$ and

$$\psi(\mathbf{r}_1) = e^{i\mathbf{k_0}\cdot\mathbf{r_1}} - \int \frac{2m}{\hbar^2} V(\mathbf{r}_2)\psi(\mathbf{r}_2)\frac{e^{ik|\mathbf{r}_1-\mathbf{r}_2|}}{4\pi|\mathbf{r}_1-\mathbf{r}_2|}d^3r_2. \tag{9.202}$$

This integral equation analog of the original Schrödinger wave equation is **exact**. Employing the Neumann series technique of Section 16.3 (remember, the scattering probability is very small), we have

$$\psi_0(\mathbf{r}_1) = e^{i\mathbf{k}_0\cdot\mathbf{r}_1}, \tag{9.203a}$$

which has the physical interpretation of no scattering.

Substituting $\psi_0(\mathbf{r}_2) = e^{i\mathbf{k}_0\cdot\mathbf{r}_2}$ into the integral, we obtain the first correction term,

$$\psi_1(\mathbf{r}_1) = e^{i\mathbf{k}_0\cdot\mathbf{r}_1} - \int \frac{2m}{\hbar^2} V(\mathbf{r}_2)\frac{e^{ik|\mathbf{r}_1-\mathbf{r}_2|}}{4\pi|\mathbf{r}_1-\mathbf{r}_2|}e^{i\mathbf{k}_0\cdot\mathbf{r}_2}\,d^3r_2. \tag{9.203b}$$

This is the famous **Born approximation**. It is expected to be most accurate for weak potentials and high incident energy. If a more accurate approximation is desired, the Neumann series may be continued.[30] ■

Example 9.7.2 QUANTUM MECHANICAL SCATTERING — GREEN'S FUNCTION

Again, we consider the Schrödinger wave equation (Eq. (9.198b)) for the scattering problem. This time we use Fourier transform techniques and derive the desired form of the Green's function by contour integration. Substituting the desired asymptotic form of the solution (with k replaced by k_0),

$$\psi(\mathbf{r}) \sim e^{ik_0z} + f_{k_0}(\theta,\varphi)\frac{e^{ik_0r}}{r} = e^{ik_0z} + \Phi(\mathbf{r}), \tag{9.204}$$

into the Schrödinger wave equation, Eq. (9.198b), yields

$$\left(\nabla^2 + k_0^2\right)\Phi(\mathbf{r}) = U(\mathbf{r})e^{ik_0z} + U(\mathbf{r})\Phi(\mathbf{r}). \tag{9.205a}$$

Here

$$\frac{\hbar^2}{2m}U(\mathbf{r}) = V(\mathbf{r}),$$

[30] This assumes the Neumann series is convergent. In some physical situations it is not convergent and then other techniques are needed.

the scattering (perturbing) potential. Since the probability of scattering is much less than 1, the second term on the right-hand side of Eq. (9.205a) is expected to be negligible (relative to the first term on the right-hand side) and thus we drop it. Note that we are **approximating** our differential equation with

$$\left(\nabla^2 + k_0^2\right)\Phi(\mathbf{r}) = U(\mathbf{r})e^{ik_0 z}. \tag{9.205b}$$

We now proceed to solve Eq. (9.205b), a nonhomogeneous PDE. The differential operator ∇^2 generates a continuous set of eigenfunctions

$$\nabla^2\psi_{\mathbf{k}}(\mathbf{r}) = -k^2\psi_{\mathbf{k}}(\mathbf{r}), \tag{9.206}$$

where

$$\psi_{\mathbf{k}}(\mathbf{r}) = (2\pi)^{-3/2}e^{i\mathbf{k}\cdot\mathbf{r}}.$$

These plane-wave eigenfunctions form a continuous but orthonormal set, in the sense that

$$\int \psi^*_{\mathbf{k}_1}(\mathbf{r})\psi_{\mathbf{k}_2}(\mathbf{r})\,d^3r = \delta(\mathbf{k}_1 - \mathbf{k}_2)$$

(compare Eq. (15.21d)).[31] We use these eigenfunctions to derive a Green's function.

We expand the unknown function $\Phi(\mathbf{r}_1)$ in these eigenfunctions,

$$\Phi(\mathbf{r}_1) = \int A_{\mathbf{k}_1}\psi_{\mathbf{k}_1}(\mathbf{r}_1)\,d^3k_1, \tag{9.207}$$

a Fourier integral with $A_{\mathbf{k}_1}$, the unknown coefficients. Substituting Eq. (9.207) into Eq. (9.205b) and using Eq. (9.206), we obtain

$$\int A_{\mathbf{k}}\left(k_0^2 - k^2\right)\psi_{\mathbf{k}}(\mathbf{r})\,d^3k = U(\mathbf{r})e^{ik_0 z}. \tag{9.208}$$

Using the now-familiar technique of multiplying by $\psi^*_{\mathbf{k}_2}(\mathbf{r})$ and integrating over the space coordinates, we have

$$\int A_{\mathbf{k}_1}\left(k_0^2 - k_1^2\right)d^3k_1 \int \psi^*_{\mathbf{k}_2}(\mathbf{r})\psi_{\mathbf{k}_1}(\mathbf{r})\,d^3r = A_{\mathbf{k}_2}\left(k_0^2 - k_2^2\right)$$
$$= \int \psi^*_{\mathbf{k}_2}(\mathbf{r})U(\mathbf{r})e^{ik_0 z}\,d^3r. \tag{9.209}$$

Solving for $A_{\mathbf{k}_2}$ and substituting into Eq. (9.207) we have

$$\Phi(\mathbf{r}_2) = \int\left[\left(k_0^2 - k_2^2\right)^{-1}\int \psi^*_{\mathbf{k}_2}(\mathbf{r}_1)U(\mathbf{r}_1)e^{ik_0 z_1}\,d^3r_1\right]\psi_{\mathbf{k}_2}(\mathbf{r}_2)\,d^3k_2. \tag{9.210}$$

Hence

$$\Phi(\mathbf{r}_1) = \int \psi_{\mathbf{k}_1}(\mathbf{r}_1)\left(k_0^2 - k_1^2\right)^{-1}d^3k_1 \int \psi^*_{\mathbf{k}_1}(\mathbf{r}_2)U(\mathbf{r}_2)e^{ik_0 z_2}\,d^3r_2, \tag{9.211}$$

[31] $d^3r = dx\,dy\,dz$, a (three-dimensional) volume element in $\mathbf{r}$-space.

replacing $\mathbf{k}_2$ by $\mathbf{k}_1$ and $\mathbf{r}_1$ by $\mathbf{r}_2$ to agree with Eq. (9.207). Reversing the order of integration, we have

$$\Phi(\mathbf{r}_1) = -\int G_{k_0}(\mathbf{r}_1, \mathbf{r}_2) U(\mathbf{r}_2) e^{ik_0 z_2}\, d^3 r_2, \tag{9.212}$$

where $G_{k_0}(\mathbf{r}_1, \mathbf{r}_2)$, our Green's function, is given by

$$G_{k_0}(\mathbf{r}_1, \mathbf{r}_2) = \int \frac{\psi^*_{\mathbf{k}_1}(\mathbf{r}_2)\psi_{\mathbf{k}_1}(\mathbf{r}_1)}{k_1^2 - k_0^2} d^3 k_1, \tag{9.213}$$

analogous to Eq. (10.90) of Section 10.5 for discrete eigenfunctions. Equation (9.212) should be compared with the Green's function solution of Poisson's equation (9.157).

It is perhaps worth evaluating this integral to emphasize once more the vital role played by the boundary conditions. Using the eigenfunctions from Eq. (9.206) and

$$d^3 k = k^2\, dk \sin\theta\, d\theta\, d\varphi,$$

we obtain

$$G_{k_0}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{(2\pi)^3} \int_0^\infty \int_0^\pi \int_0^{2\pi} \frac{e^{ik\rho\cos\theta}}{k^2 - k_0^2}\, d\varphi \sin\theta\, d\theta\, k^2\, dk. \tag{9.214}$$

Here $k\rho\cos\theta$ has replaced $\mathbf{k}\cdot(\mathbf{r}_1 - \mathbf{r}_2)$, with $\boldsymbol{\rho} = \mathbf{r}_1 - \mathbf{r}_2$ indicating the polar axis in k-space. Integrating over φ by inspection, we pick up a 2π. The θ-integration then leads to

$$G_{k_0}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{4\pi^2 \rho i} \int_0^\infty \frac{e^{ik\rho} - e^{-ik\rho}}{k^2 - k_0^2} k\, dk, \tag{9.215}$$

and since the integrand is an even function of k, we may set

$$G_{k_0}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{8\pi^2 \rho i} \int_{-\infty}^\infty \frac{(e^{i\kappa} - e^{-i\kappa})}{\kappa^2 - \sigma^2} \kappa\, d\kappa. \tag{9.216}$$

The latter step is taken in anticipation of the evaluation of $G_k(\mathbf{r}_1, \mathbf{r}_2)$ as a contour integral. The symbols κ and $\sigma\,(\sigma > 0)$ represent $k\rho$ and $k_0\rho$, respectively.

If the integral in Eq. (9.216) is interpreted as a Riemann integral, the integral does not exist. This implies that $\mathcal{L}^{-1}$ does not exist, and in a literal sense it does not. $\mathcal{L} = \nabla^2 + k^2$ is singular since there exist nontrivial solutions ψ for which the homogeneous equation $\mathcal{L}\psi = 0$. We avoid this problem by introducing a parameter γ, defining a different operator $\mathcal{L}_\gamma^{-1}$, and taking the limit as $\gamma \to 0$.

Splitting the integral into two parts so that each part may be written as a suitable contour integral gives us

$$G(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{8\pi^2 \rho i} \oint_{C_1} \frac{\kappa e^{i\kappa}\, d\kappa}{\kappa^2 - \sigma^2} + \frac{1}{8\pi^2 \rho i} \oint_{C_2} \frac{\kappa e^{-i\kappa}\, d\kappa}{\kappa^2 - \sigma^2}. \tag{9.217}$$

Contour C_1 is closed by a semicircle in the upper half-plane, C_2 by a semicircle in the lower half-plane. These integrals were evaluated in Chapter 7 by using appropriately chosen infinitesimal semicircles to go around the singular points $\kappa = \pm\sigma$. As an alternative procedure, let us first displace the singular points from the real axis by replacing σ by $\sigma + i\gamma$ and then, after evaluation, taking the limit as $\gamma \to 0$ (Fig. 9.5).

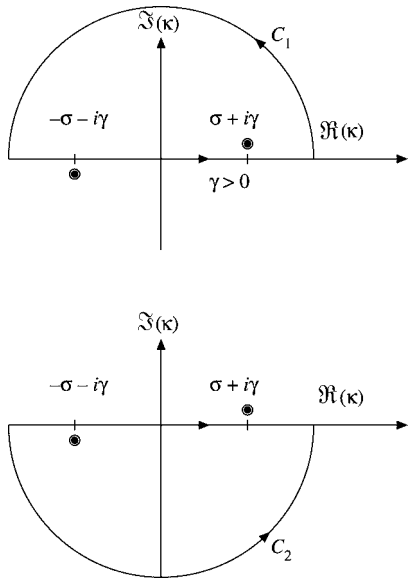

FIGURE 9.5 Possible Green's function contours of integration.

For γ positive, contour C_1 encloses the singular point $\kappa = \sigma + i\gamma$ and the first integral contributes

$$2\pi i \cdot \frac{1}{2} e^{i(\sigma + i\gamma)}.$$

From the second integral we also obtain

$$2\pi i \cdot \frac{1}{2} e^{i(\sigma + i\gamma)},$$

the enclosed singularity being $\kappa = -(\sigma + i\gamma)$. Returning to Eq. (9.217) and letting $\gamma \to 0$, we have

$$G(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{4\pi\rho} e^{i\sigma} = \frac{e^{ik_0|\mathbf{r}_1 - \mathbf{r}_2|}}{4\pi|\mathbf{r}_1 - \mathbf{r}_2|}, \tag{9.218}$$

in full agreement with Exercise 9.7.16. This result depends on starting with γ positive. Had we chosen γ negative, our Green's function would have included $e^{-i\sigma}$, which corresponds to an **incoming** wave. The choice of positive γ is dictated by the boundary conditions we wish to satisfy.

Equations (9.212) and (9.218) reproduce the scattered wave in Eq. (9.203b) and constitute an exact solution of the approximate Eq. (9.205b). Exercises 9.7.18 and 9.7.20 extend these results. ■

Exercises

9.7.1 Verify Eq. (9.168),

$$\int (v\mathcal{L}_2 u - u\mathcal{L}_2 v)\, d\tau_2 = \int p(v\nabla_2 u - u\nabla_2 v)\cdot d\boldsymbol{\sigma}_2.$$

9.7.2 Show that the terms $+k^2$ in the Helmholtz operator and $-k^2$ in the modified Helmholtz operator do not affect the behavior of $G(\mathbf{r}_1, \mathbf{r}_2)$ in the immediate vicinity of the singular point $\mathbf{r}_1 = \mathbf{r}_2$. Specifically, show that

$$\lim_{|\mathbf{r}_1 - \mathbf{r}_2|\to 0} \int k^2 G(\mathbf{r}_1, \mathbf{r}_2)\, d\tau_2 = 1.$$

9.7.3 Show that

$$\frac{\exp(ik|\mathbf{r}_1 - \mathbf{r}_2|)}{4\pi|\mathbf{r}_1 - \mathbf{r}_2|}$$

satisfies the two appropriate criteria and therefore is a Green's function for the Helmholtz equation.

9.7.4 (a) Find the Green's function for the three-dimensional Helmholtz equation, Exercise 9.7.3, when the wave is a standing wave.
(b) How is this Green's function related to the spherical Bessel functions?

9.7.5 The homogeneous Helmholtz equation

$$\nabla^2\varphi + \lambda^2\varphi = 0$$

has eigenvalues λ_i^2 and eigenfunctions φ_i. Show that the corresponding Green's function that satisfies

$$\nabla^2 G(\mathbf{r}_1, \mathbf{r}_2) + \lambda^2 G(\mathbf{r}_1, \mathbf{r}_2) = -\delta(\mathbf{r}_1 - \mathbf{r}_2)$$

may be written as

$$G(\mathbf{r}_1, \mathbf{r}_2) = \sum_{i=1}^{\infty} \frac{\varphi_i(\mathbf{r}_1)\varphi_i(\mathbf{r}_2)}{\lambda_i^2 - \lambda^2}.$$

An expansion of this form is called a **bilinear** expansion. If the Green's function is available in **closed** form, this provides a means of generating functions.

9.7.6 An electrostatic potential (mks units) is

$$\varphi(\mathbf{r}) = \frac{Z}{4\pi\varepsilon_0}\cdot\frac{e^{-ar}}{r}.$$

Reconstruct the electrical charge distribution that will produce this potential. Note that $\varphi(r)$ vanishes exponentially for large r, showing that the net charge is zero.

$$\text{ANS. } \rho(r) = Z\delta(r) - \frac{Za^2}{4\pi}\frac{e^{-ar}}{r}.$$

9.7.7 Transform the ODE

$$\frac{d^2y(r)}{dr^2} - k^2 y(r) + V_0 \frac{e^{-r}}{r} y(r) = 0$$

and the boundary conditions $y(0) = y(\infty) = 0$ into a Fredholm integral equation of the form

$$y(r) = \lambda \int_0^\infty G(r,t) \frac{e^{-t}}{t} y(t)\, dt.$$

The quantities $V_0 = \lambda$ and k^2 are constants. The ODE is derived from the Schrödinger wave equation with a mesonic potential:

$$G(r,t) = \begin{cases} \dfrac{1}{k} e^{-kt} \sinh kr, & 0 \le r < t, \\ \dfrac{1}{k} e^{-kr} \sinh kt, & t < r < \infty. \end{cases}$$

9.7.8 A charged conducting ring of radius a (Example 12.3.3) may be described by

$$\rho(\mathbf{r}) = \frac{q}{2\pi a^2} \delta(r-a)\delta(\cos\theta).$$

Using the known Green's function for this system, Eq. (9.187) find the electrostatic potential.
Hint. Exercise 12.6.3 will be helpful.

9.7.9 Changing a separation constant from k^2 to $-k^2$ and putting the discontinuity of the first derivative into the z-dependence, show that

$$\frac{1}{4\pi|\mathbf{r}_1 - \mathbf{r}_2|} = \frac{1}{4\pi} \sum_{m=-\infty}^{\infty} \int_0^\infty e^{im(\varphi_1 - \varphi_2)} J_m(k\rho_1) J_m(k\rho_2) e^{-k|z_1 - z_2|}\, dk.$$

Hint. The required $\delta(\rho_1 - \rho_2)$ may be obtained from Exercise 15.1.2.

9.7.10 Derive the expansion

$$\frac{\exp[ik|\mathbf{r}_1 - \mathbf{r}_2|]}{4\pi|\mathbf{r}_1 - \mathbf{r}_2|} = ik \sum_{l=0}^{\infty} \left\{ \begin{array}{ll} j_l(kr_1) h_l^{(1)}(kr_2), & r_1 < r_2 \\ j_l(kr_2) h_l^{(1)}(kr_1), & r_1 > r_2 \end{array} \right\} \times \sum_{m=-l}^{l} Y_l^m(\theta_1, \varphi_1) Y_l^{m*}(\theta_2, \varphi_2).$$

Hint. The left side is a known Green's function. Assume a spherical harmonic expansion and work on the remaining radial dependence. The spherical harmonic closure relation, Exercise 12.6.6, covers the angular dependence.

9.7.11 Show that the modified Helmholtz operator Green's function

$$\frac{\exp(-k|\mathbf{r}_1 - \mathbf{r}_2|)}{4\pi|\mathbf{r}_1 - \mathbf{r}_2|}$$

has the spherical polar coordinate expansion

$$\frac{\exp(-k|\mathbf{r}_1-\mathbf{r}_2|)}{4\pi|\mathbf{r}_1-\mathbf{r}_2|}=k\sum_{l=0}^{\infty}i_l(kr_<)k_l(kr_>)\sum_{m=-l}^{l}Y_l^m(\theta_1,\varphi_1)Y_l^{m*}(\theta_2,\varphi_2).$$

Note. The modified spherical Bessel functions $i_l(kr)$ and $k_l(kr)$ are defined in Exercise 11.7.15.

9.7.12 From the spherical Green's function of Exercise 9.7.10, derive the plane-wave expansion

$$e^{i\mathbf{k}\cdot\mathbf{r}}=\sum_{l=0}^{\infty}i^l(2l+1)j_l(kr)P_l(\cos\gamma),$$

where γ is the angle included between $\mathbf{k}$ and $\mathbf{r}$. This is the Rayleigh equation of Exercise 12.4.7.
Hint. Take $\mathbf{r}_2\gg\mathbf{r}_1$ so that

$$|\mathbf{r}_1-\mathbf{r}_2|\to r_2-\mathbf{r}_{20}\cdot\mathbf{r}_1=r_2-\frac{\mathbf{k}\cdot\mathbf{r}_1}{k}.$$

Let $r_2\to\infty$ and cancel a factor of e^{ikr_2}/r_2.

9.7.13 From the results of Exercises 9.7.10 and 9.7.12, show that

$$e^{ix}=\sum_{l=0}^{\infty}i^l(2l+1)j_l(x).$$

9.7.14 (a) From the circular cylindrical coordinate expansion of the Laplace Green's function (Eq. (9.197)), show that

$$\frac{1}{(\rho^2+z^2)^{1/2}}=\frac{2}{\pi}\int_0^{\infty}K_0(k\rho)\cos kz\,dk.$$

This same result is obtained directly in Exercise 15.3.11.

(b) As a special case of part (a) show that

$$\int_0^{\infty}K_0(k)\,dk=\frac{\pi}{2}.$$

9.7.15 Noting that

$$\psi_k(\mathbf{r})=\frac{1}{(2\pi)^{3/2}}e^{i\mathbf{k}\cdot\mathbf{r}}$$

is an eigenfunction of

$$\left(\nabla^2+k^2\right)\psi_k(\mathbf{r})=0$$

(Eq. (9.206)), show that the Green's function of $\mathcal{L}=\nabla^2$ may be expanded as

$$\frac{1}{4\pi|\mathbf{r}_1-\mathbf{r}_2|}=\frac{1}{(2\pi)^3}\int e^{i\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)}\frac{d^3k}{k^2}.$$

9.7.16 Using Fourier transforms, show that the Green's function satisfying the nonhomogeneous Helmholtz equation

$$\left(\nabla^2 + k_0^2\right) G(\mathbf{r}_1, \mathbf{r}_2) = -\delta(\mathbf{r}_1 - \mathbf{r}_2)$$

is

$$G(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{(2\pi)^3} \int \frac{e^{i\mathbf{k}\cdot(\mathbf{r}_1 - \mathbf{r}_2)}}{k^2 - k_0^2} d^3k,$$

in agreement with Eq. (9.213).

9.7.17 The basic equation of the scalar Kirchhoff diffraction theory is

$$\psi(\mathbf{r}_1) = \frac{1}{4\pi} \int_{S_2} \left[\frac{e^{ikr}}{r} \nabla\psi(\mathbf{r}_2) - \psi(\mathbf{r}_2) \nabla\left(\frac{e^{ikr}}{r}\right) \right] \cdot d\boldsymbol{\sigma}_2,$$

where ψ satisfies the homogeneous Helmholtz equation and $r = |\mathbf{r}_1 - \mathbf{r}_2|$. Derive this equation. Assume that $\mathbf{r}_1$ is interior to the closed surface S_2.
Hint. Use Green's theorem.

9.7.18 The Born approximation for the scattered wave is given by Eq. (9.203b) (and Eq. (9.211)). From the asymptotic form, Eq. (9.199),

$$f_k(\theta, \varphi) \frac{e^{ikr}}{r} = -\frac{2m}{\hbar^2} \int V(\mathbf{r}_2) \frac{e^{ik|\mathbf{r} - \mathbf{r}_2|}}{4\pi |\mathbf{r} - \mathbf{r}_2|} e^{i\mathbf{k}_0 \cdot \mathbf{r}_2} d^3 r_2.$$

For a scattering potential $V(\mathbf{r}_2)$ that is independent of angles and for $r \gg r_2$, show that

$$f_k(\theta, \varphi) = -\frac{2m}{\hbar^2} \int_0^\infty r_2 V(\mathbf{r}_2) \frac{\sin(|\mathbf{k}_0 - \mathbf{k}| r_2)}{|\mathbf{k}_0 - \mathbf{k}|} dr_2.$$

Here $\mathbf{k}_0$ is in the $\theta = 0$ (original z-axis) direction, whereas $\mathbf{k}$ is in the (θ, φ) direction. The magnitudes are equal: $|\mathbf{k}_0| = |\mathbf{k}|$; m is the reduced mass.
Hint. You have Exercise 9.7.12 to simplify the exponential and Exercise 15.3.20 to transform the three-dimensional Fourier exponential transform into a one-dimensional Fourier sine transform.

9.7.19 Calculate the scattering amplitude $f_k(\theta, \varphi)$ for a mesonic potential $\mathbf{V}(r) = V_0(e^{-\alpha r}/\alpha r)$.
Hint. This particular potential permits the Born integral, Exercise 9.7.18, to be evaluated as a Laplace transform.

$$\text{ANS. } f_k(\theta, \varphi) = -\frac{2m V_0}{\hbar^2 \alpha} \frac{1}{\alpha^2 + (\mathbf{k}_0 - \mathbf{k})^2}.$$

9.7.20 The mesonic potential $V(r) = V_0(e^{-\alpha r}/\alpha r)$ may be used to describe the Coulomb scattering of two charges q_1 and q_2. We let $\alpha \to 0$ and $V_0 \to 0$ but take the ratio V_0/α to be $q_1 q_2 / 4\pi\varepsilon_0$. (For Gaussian units omit the $4\pi\varepsilon_0$.) Show that the differential scattering cross section $d\sigma/d\Omega = |f_k(\theta, \varphi)|^2$ is given by

$$\frac{d\sigma}{d\Omega} = \left(\frac{q_1 q_2}{4\pi\varepsilon_0}\right)^2 \frac{1}{16E^2 \sin^4(\theta/2)}, \qquad E = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m}.$$

It happens (coincidentally) that this Born approximation is in exact agreement with both the exact quantum mechanical calculations and the classical Rutherford calculation.

9.8 HEAT FLOW, OR DIFFUSION, PDE

Here we return to a special PDE to develop fairly general methods to adapt a special solution of a PDE to boundary conditions by introducing parameters that apply to other second-order PDEs with constant coefficients as well. To some extent, they are complementary to the earlier basic separation method for finding solutions in a systematic way.

We select the full time-dependent diffusion PDE for an isotropic medium. Assuming isotropy actually is not much of a restriction because, in case we have different (constant) rates of diffusion in different directions, for example in wood, our heat flow PDE takes the form

$$\frac{\partial \psi}{\partial t} = a^2 \frac{\partial^2 \psi}{\partial x^2} + b^2 \frac{\partial^2 \psi}{\partial y^2} + c^2 \frac{\partial^2 \psi}{\partial z^2}, \tag{9.219}$$

if we put the coordinate axes along the principal directions of anisotropy. Now we simply rescale the coordinates using the substitutions $x = a\xi$, $y = b\eta$, $z = c\zeta$ to get back the original isotropic form of Eq. (9.219),

$$\frac{\partial \Phi}{\partial t} = \frac{\partial^2 \Phi}{\partial \xi^2} + \frac{\partial^2 \Phi}{\partial \eta^2} + \frac{\partial^2 \Phi}{\partial \zeta^2} \tag{9.220}$$

for the temperature distribution function $\Phi(\xi, \eta, \zeta, t) = \psi(x, y, z, t)$.

For simplicity, we first solve the time-dependent PDE for a homogeneous one-dimensional medium, a long metal rod in the x-direction, say,

$$\frac{\partial \psi}{\partial t} = a^2 \frac{\partial^2 \psi}{\partial x^2}, \tag{9.221}$$

where the constant a measures the diffusivity, or heat conductivity, of the medium. We attempt to solve this linear PDE with constant coefficients with the relevant **exponential product Ansatz** $\psi = e^{\alpha x} \cdot e^{\beta t}$, which, when substituted into Eq. (9.221), solves the PDE with the constraint $\beta = a^2\alpha^2$ for the parameters. We seek exponentially decaying solutions for large times, that is, solutions with negative β values, and therefore set $\alpha = i\omega$, $\alpha^2 = -\omega^2$ for real ω and have

$$\psi(x, t) = e^{i\omega x} e^{-\omega^2 a^2 t} = (\cos \omega x + i \sin \omega x) e^{-\omega^2 a^2 t}.$$

Forming real linear combinations we obtain the solution

$$\psi(\mathbf{x}, t) = (A \cos \omega x + B \sin \omega x) e^{-\omega^2 a^2 t},$$

for any choice of A, B, ω, which are introduced to satisfy boundary conditions. Upon summing over multiples $n\omega$ of the basic frequency for periodic boundary conditions or **integrating over the parameter** ω for general (nonperiodic boundary conditions), we find a solution,

$$\psi(x, t) = \int \left[A(\omega) \cos \omega x + B(\omega) \sin \omega x \right] e^{-a^2 \omega^2 t} \, d\omega, \tag{9.222}$$

that is general enough to be adapted to boundary conditions at $t = 0$, say. When the boundary condition gives a nonzero temperature ψ_0, as for our rod, then the summation method

applies (Fourier expansion of the boundary condition). If the space is unrestricted (as for an infinitely extended rod), the Fourier integral applies.

- This summation or integration over parameters is one of the standard methods for generalizing specific PDE solutions in order to adapt them to boundary conditions.

Example 9.8.1 A Specific Boundary Condition

Let us solve a one-dimensional case explicitly, where the temperature at time $t = 0$ is $\psi_0(x) = 1 = \text{const.}$ in the interval between $x = +1$ and $x = -1$ and zero for $x > 1$ and $x < 1$. At the ends, $x = \pm 1$, the temperature is always held at zero.

For a finite interval we choose the $\cos(l\pi x/2)$ spatial solutions of Eq. (9.221) for integer l, because they vanish at $x = \pm 1$. Thus, at $t = 0$ our solution is a Fourier series,

$$\psi(x,0) = \sum_{l=1}^{\infty} a_l \cos\frac{\pi l x}{2} = 1, \qquad -1 < x < 1$$

with coefficients (see Section 14.1.)

$$\begin{aligned} a_l &= \int_{-1}^{1} 1 \cdot \cos\frac{\pi l x}{2} = \frac{2}{l\pi}\sin\frac{\pi l x}{2}\bigg|_{x=-1}^{1} \\ &= \frac{4}{\pi l}\sin\frac{l\pi}{2} = \frac{4(-1)^m}{(2m+1)\pi}, \qquad l = 2m+1; \\ a_l &= 0, \qquad l = 2m. \end{aligned}$$

Including its time dependence, the full solution is given by the series

$$\psi(x,t) = \frac{4}{\pi}\sum_{m=0}^{\infty}\frac{(-1)^m}{2m+1}\cos\left[(2m+1)\frac{\pi x}{2}\right]e^{-t((2m+1)\pi a/2)^2}, \tag{9.223}$$

which converges absolutely for $t > 0$ but only conditionally at $t = 0$, as a result of the discontinuity at $x = \pm 1$.

Without the restriction to zero temperature at the endpoints of the given finite interval, the Fourier series is replaced by a Fourier integral. The general solution is then given by Eq. (9.222). At $t = 0$ the given temperature distribution $\psi_0 = 1$ gives the coefficients as (see Section 15.3)

$$A(\omega) = \frac{1}{\pi}\int_{-1}^{1}\cos\omega x\, dx = \frac{1}{\pi}\frac{\sin\omega x}{\omega}\bigg|_{x=-1}^{1} = \frac{2\sin\omega}{\pi\omega}, \qquad B(\omega) = 0.$$

Therefore

$$\psi(x,t) = \frac{2}{\pi}\int_0^{\infty}\frac{\sin\omega}{\omega}\cos(\omega x)e^{-a^2\omega^2 t}\, d\omega. \tag{9.224}$$

■

In **three dimensions** the corresponding exponential Ansatz $\psi = e^{i\mathbf{k}\cdot\mathbf{r}/a+\beta t}$ leads to a solution with the relation $\beta = -\mathbf{k}^2 = -k^2$ for its parameter, and the three-dimensional

form of Eq. (9.221) becomes

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}+k^2\psi=0, \tag{9.225}$$

which is the **Helmholtz** equation, which may be solved by the separation method just like the earlier Laplace equation in Cartesian, cylindrical, or spherical coordinates under appropriately generalized boundary conditions.

In Cartesian coordinates, with the product Ansatz of Eq. (9.35), the separated x- and y-ODEs from Eq. (9.221) are the same as Eqs. (9.38) and (9.41), while the z-ODE, Eq. (9.42), generalizes to

$$\frac{1}{Z}\frac{d^2Z}{dz^2}=-k^2+l^2+m^2=n^2>0, \tag{9.226}$$

where we introduce another separation constant, n^2, constrained by

$$k^2=l^2+m^2-n^2 \tag{9.227}$$

to produce a symmetric set of equations. Now, our solution of Helmholtz's Eq. (9.225) is labeled according to the choice of all three separation constants l, m, n subject to the constraint Eq. (9.227). As before the z-ODE, Eq. (9.226), yields exponentially decaying solutions $\sim e^{-nz}$. The boundary condition at $z=0$ fixes the expansion coefficients a_{lm}, as in Eq. (9.44).

In cylindrical coordinates, we now use the separation constant l^2 for the z-ODE, with an exponentially decaying solution in mind,

$$\frac{d^2Z}{dz^2}=l^2Z>0, \tag{9.228}$$

so $Z\sim e^{-lz}$, because the temperature goes to zero at large z. If we set $k^2+l^2=n^2$, Eqs. (9.53) to (9.54) stay the same, so we end up with the same Fourier–Bessel expansion, Eq. (9.56), as before.

In spherical coordinates with radial boundary conditions, the separation method leads to the same angular ODEs in Eqs. (9.61) and (9.64), and the radial ODE now becomes

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right)+k^2R-\frac{QR}{r^2}=0, \qquad Q=l(l+1), \tag{9.229}$$

that is, of Eq. (9.65), whose solutions are the spherical Bessel functions of Section 11.7. They are listed in Table 9.2.

The restriction that k^2 be a constant is unnecessarily severe. The separation process will still work with Helmholtz's PDE for k^2 as general as

$$k^2=f(r)+\frac{1}{r^2}g(\theta)+\frac{1}{r^2\sin^2\theta}h(\varphi)+k'^2. \tag{9.230}$$

In the hydrogen atom we have $k^2=f(r)$ in the Schrödinger wave equation, and this leads to a closed-form solution involving Laguerre polynomials.

Alternate Solutions

In a new approach to the heat flow PDE suggested by experiments, we now return to the one-dimensional PDE, Eq. (9.221), seeking solutions of a new functional form $\psi(x,t) = u(x/\sqrt{t})$, which is suggested by Example 15.1.1. Substituting $u(\xi)$, $\xi = x/\sqrt{t}$, into Eq. (9.221) using

$$\frac{\partial\psi}{\partial x} = \frac{u'}{\sqrt{t}}, \qquad \frac{\partial^2\psi}{\partial x^2} = \frac{u''}{t}, \qquad \frac{\partial\psi}{\partial t} = -\frac{x}{2\sqrt{t^3}}u' \tag{9.231}$$

with the notation $u'(\xi) \equiv \frac{du}{d\xi}$, the PDE is reduced to the ODE

$$2a^2u''(\xi) + \xi u'(\xi) = 0. \tag{9.232}$$

Writing this ODE as

$$\frac{u''}{u'} = -\frac{\xi}{2a^2},$$

we can integrate it once to get $\ln u' = -\frac{\xi^2}{4a^2} + \ln C_1$, with an integration constant C_1. Exponentiating and integrating again we find the solution

$$u(\xi) = C_1\int_0^\xi e^{-\frac{\xi^2}{4a^2}}\,d\xi + C_2, \tag{9.233}$$

involving two integration constants C_i. Normalizing this solution at time $t = 0$ to temperature $+1$ for $x > 0$ and -1 for $x < 0$, our boundary conditions, fixes the constants C_i, so

$$\psi = \frac{1}{a\sqrt{\pi}}\int_0^{\frac{x}{\sqrt{t}}} e^{-\frac{\xi^2}{4a^2}}\,d\xi = \frac{2}{\sqrt{\pi}}\int_0^{\frac{x}{2a\sqrt{t}}} e^{-v^2}\,dv = \Phi\left(\frac{x}{2a\sqrt{t}}\right), \tag{9.234}$$

where Φ denotes Gauss' error function (see Exercise 5.10.4). See Example 15.1.1 for a derivation using a Fourier transform. We need to generalize this specific solution to adapt it to boundary conditions.

To this end we now generate **new solutions of the PDE with constant coefficients by differentiating a special solution**, Eq. (9.234). In other words, if $\psi(x,t)$ solves the PDE in Eq. (9.221), so do $\frac{\partial\psi}{\partial t}$ and $\frac{\partial\psi}{\partial x}$, because these derivatives and the differentiations of the PDE commute; that is, the order in which they are carried out does not matter. Note carefully that this method no longer works if any coefficient of the PDE depends on t or x explicitly. However, PDEs with constant coefficients dominate in physics. Examples are Newton's equations of motion (ODEs) in classical mechanics, the wave equations of electrodynamics, and Poisson's and Laplace's equations in electrostatics and gravity. Even Einstein's nonlinear field equations of general relativity take on this special form in local geodesic coordinates.

Therefore, by differentiating Eq. (9.234) with respect to x, we find the simpler, more basic solution

$$\psi_1(x,t) = \frac{1}{a\sqrt{t\pi}}e^{-\frac{x^2}{4a^2t}}, \tag{9.235}$$

and, repeating the process, another basic solution,

$$\psi_2(x,t) = \frac{x}{2a^3\sqrt{t^3\pi}} e^{-\frac{x^2}{4a^2t}}. \tag{9.236}$$

Again, these solutions have to be generalized to adapt them to boundary conditions. And there is yet another method of generating new solutions of a PDE with constant coefficients: We can **translate** a given solution, for example, $\psi_1(x,t) \to \psi_1(x-\alpha,t)$, and then **integrate over the translation parameter** α. Therefore

$$\psi(x,t) = \frac{1}{2a\sqrt{t\pi}} \int_{-\infty}^{\infty} C(\alpha) e^{-\frac{(x-\alpha)^2}{4a^2t}}\, d\alpha \tag{9.237}$$

is again a solution, which we rewrite using the substitution

$$\xi = \frac{x-\alpha}{2a\sqrt{t}}, \qquad \alpha = x - 2a\xi\sqrt{t}, \qquad d\alpha = -2a\, d\xi\, \sqrt{t}. \tag{9.238}$$

Thus, we find that

$$\psi(x,t) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} C(x - 2a\xi\sqrt{t}\,) e^{-\xi^2}\, d\xi \tag{9.239}$$

is a solution of our PDE. In this form we recognize the significance of the weight function $C(x)$ from the translation method because, at $t = 0$, $\psi(x,0) = C(x) = \psi_0(x)$ is determined by the boundary condition, and $\int_{-\infty}^{\infty} e^{-\xi^2} d\xi = \sqrt{\pi}$. Therefore, we can also write the solution as

$$\psi(x,t) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \psi_0(x - 2a\xi\sqrt{t}\,) e^{-\xi^2}\, d\xi, \tag{9.240}$$

displaying the role of the boundary condition explicitly. From Eq. (9.240) we see that the initial temperature distribution, $\psi_0(x)$, spreads out over time and is damped by the Gaussian weight function.

Example 9.8.2 SPECIAL BOUNDARY CONDITION AGAIN

Let us express the solution of Example 9.8.1 in terms of the error function solution of Eq. (9.234). The boundary condition at $t = 0$ is $\psi_0(x) = 1$ for $-1 < x < 1$ and zero for $|x| > 1$. From Eq. (9.240) we find the limits on the integration variable ξ by setting $x - 2a\xi\sqrt{t} = \pm 1$. This yields the integration endpoints $\xi = (\pm 1 + x)/2a\sqrt{t}$. Therefore our solution becomes

$$\psi(x,t) = \frac{1}{\sqrt{\pi}} \int_{\frac{x-1}{2a\sqrt{t}}}^{\frac{x+1}{2a\sqrt{t}}} e^{-\xi^2}\, d\xi.$$

Using the error function defined in Eq. (9.234) we can also write this solution as follows

$$\psi(x,t) = \frac{1}{2}\left[\operatorname{erf}\left(\frac{x+1}{2a\sqrt{t}}\right) - \operatorname{erf}\left(\frac{x-1}{2a\sqrt{t}}\right)\right]. \tag{9.241}$$

Comparing this form of our solution with that from Example 9.8.1 we see that we can express Eq. (9.241) as the Fourier integral of Example 9.8.1, an identity that gives the Fourier integral, Eq. (9.224), in closed form of the tabulated error function. ■

Finally, we consider the heat flow case for an extended **spherically symmetric** medium centered at the origin, which prescribes polar coordinates r, θ, φ. We expect a solution of the form $\psi(\mathbf{r}, t) = u(r, t)$. Using Eq. (2.48) we find the PDE

$$\frac{\partial u}{\partial t} = a^2\left(\frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r}\right), \tag{9.242}$$

which we transform to the one-dimensional heat flow PDE by the substitution

$$u = \frac{v(r,t)}{r}, \qquad \frac{\partial u}{\partial r} = \frac{1}{r}\frac{\partial v}{\partial r} - \frac{v}{r^2}, \qquad \frac{\partial u}{\partial t} = \frac{1}{r}\frac{\partial v}{\partial t},$$
$$\frac{\partial^2 u}{\partial r^2} = \frac{1}{r}\frac{\partial^2 v}{\partial r^2} - \frac{2}{r^2}\frac{\partial v}{\partial r} + \frac{2v}{r^3}. \tag{9.243}$$

This yields the PDE

$$\frac{\partial v}{\partial t} = a^2\frac{\partial^2 v}{\partial r^2}. \tag{9.244}$$

Example 9.8.3 SPHERICALLY SYMMETRIC HEAT FLOW

Let us apply the one-dimensional heat flow PDE with the solution Eq. (9.234) to a spherically symmetric heat flow under fairly common boundary conditions, where x is released by the radial variable. Initially we have zero temperature everywhere. Then, at time $t = 0$, a finite amount of heat energy Q is released at the origin, spreading evenly in all directions. What is the resulting spatial and temporal temperature distribution?

Inspecting our special solution in Eq. (9.236) we see that, for $t \to 0$, the temperature

$$\frac{v(r,t)}{r} = \frac{C}{\sqrt{t^3}}e^{-\frac{r^2}{4a^2t}} \tag{9.245}$$

goes to zero for all $r \neq 0$, so zero initial temperature is guaranteed. As $t \to \infty$, the temperature $v/r \to 0$ for all r including the origin, which is implicit in our boundary conditions. The constant C can be determined from energy conservation, which gives the constraint

$$Q = \sigma\rho\int\frac{v}{r}d^3r = \frac{4\pi\sigma\rho C}{\sqrt{t^3}}\int_0^\infty r^2 e^{-\frac{r^2}{4a^2t}}\,dr = 8\sqrt{\pi^3}\,\sigma\rho a^3 C, \tag{9.246}$$

where ρ is the constant density of the medium and σ is its specific heat. Here we have rescaled the integration variable and integrated by parts to get

$$\int_0^\infty e^{-\frac{r^2}{4a^2t}}r^2\,dr = (2a\sqrt{t}\,)^3\int_0^\infty e^{-\xi^2}\xi^2\,d\xi,$$
$$\int_0^\infty e^{-\xi^2}\xi^2\,d\xi = -\frac{\xi}{2}e^{-\xi^2}\Big|_0^\infty + \frac{1}{2}\int_0^\infty e^{-\xi^2}\,d\xi = \frac{\sqrt{\pi}}{4}.$$

The temperature, as given by Eq. (9.245) at any moment, which is at fixed t, is a Gaussian distribution that flattens out as time increases, because its width is proportional to $\sqrt{t}$. As a function of time the temperature is proportional to $t^{-3/2}e^{-T/t}$, with $T \equiv r^2/4a^2$, which rises from zero to a maximum and then falls off to zero again for large times. To find the maximum, we set

$$\frac{d}{dt}\left(t^{-3/2}e^{-T/t}\right) = t^{-5/2}e^{-T/t}\left(\frac{T}{t} - \frac{3}{2}\right) = 0, \tag{9.247}$$

from which we find $t = 2T/3$. ■

In the case of **cylindrical symmetry** (in the plane $z = 0$ in plane polar coordinates $\rho = \sqrt{x^2 + y^2}, \varphi$) we look for a temperature $\psi = u(\rho, t)$ that then satisfies the ODE (using Eq. (2.35) in the diffusion equation)

$$\frac{\partial u}{\partial t} = a^2\left(\frac{\partial^2 u}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial u}{\partial \rho}\right), \tag{9.248}$$

which is the planar analog of Eq. (9.244). This ODE also has solutions with the functional dependence $\rho/\sqrt{t} \equiv r$. Upon substituting

$$u = v\left(\frac{\rho}{\sqrt{t}}\right), \qquad \frac{\partial u}{\partial t} = -\frac{\rho v'}{2t^{3/2}}, \qquad \frac{\partial u}{\partial \rho} = \frac{v'}{\sqrt{t}}, \qquad \frac{\partial^2 u}{\partial \rho^2} = \frac{v'}{t} \tag{9.249}$$

into Eq. (9.248) with the notation $v' \equiv \frac{dv}{dr}$, we find the ODE

$$a^2 v'' + \left(\frac{a^2}{r} + \frac{r}{2}\right)v' = 0. \tag{9.250}$$

This is a first-order ODE for v', which we can integrate when we separate the variables v and r as

$$\frac{v''}{v'} = -\left(\frac{1}{r} + \frac{r}{2a^2}\right). \tag{9.251}$$

This yields

$$v(r) = \frac{C}{r}e^{-\frac{r^2}{4a^2}} = C\frac{\sqrt{t}}{\rho}e^{-\frac{\rho^2}{4a^2 t}}. \tag{9.252}$$

This special solution for cylindrical symmetry can be similarly generalized and adapted to boundary conditions, as for the spherical case. Finally, the z-dependence can be factored in, because z separates from the plane polar radial variable ρ.

In summary, PDEs can be solved with initial conditions, just like ODEs, or with boundary conditions prescribing the value of the solution or its derivative on boundary surfaces, curves, or points. When the solution is prescribed on the boundary, the PDE is called a **Dirichlet** *problem*; *if the normal derivative of the solution is prescribed on the boundary, the PDE is called a* **Neumann** *problem.*

When the initial temperature is prescribed for the one-dimensional or three-dimensional heat equation (*with spherical or cylindrical symmetry*) *it becomes a weight function of the solution, in terms of an integral over the generic Gaussian solution. The three-dimensional*

heat equation, with spherical or cylindrical boundary conditions, is solved by separation of the variables, leading to eigenfunctions in each separated variable and eigenvalues as separation constants. For finite boundary intervals in each spatial coordinate, the sum over separation constants leads to a Fourier-series solution, while infinite boundary conditions lead to a Fourier-integral solution. The separation of variables method attempts to solve a PDE by writing the solution as a product of functions of one variable each. General conditions for the separation method to work are provided by the symmetry properties of the PDE, to which continuous group theory applies.

Additional Readings

Bateman, H., *Partial Differential Equations of Mathematical Physics*. New York: Dover (1944), 1st ed. (1932). A wealth of applications of various partial differential equations in classical physics. Excellent examples of the use of different coordinate systems — ellipsoidal, paraboloidal, toroidal coordinates, and so on.

Cohen, H., *Mathematics for Scientists and Engineers*. Englewood Cliffs, NJ: Prentice-Hall (1992).

Courant, R., and D. Hilbert, *Methods of Mathematical Physics*, Vol. 1 (English edition). New York: Interscience (1953), Wiley (1989). This is one of the classic works of mathematical physics. Originally published in German in 1924, the revised English edition is an excellent reference for a rigorous treatment of Green's functions and for a wide variety of other topics on mathematical physics.

Davis, P. J., and P. Rabinowitz, *Numerical Integration*. Waltham, MA: Blaisdell (1967). This book covers a great deal of material in a relatively easy-to-read form. Appendix 1 (*On the Practical Evaluation of Integrals* by M. Abramowitz) is excellent as an overall view.

Garcia, A. L., *Numerical Methods for Physics*. Englewood Cliffs, NJ: Prentice-Hall (1994).

Hamming, R. W., *Numerical Methods for Scientists and Engineers*, 2nd ed. New York: McGraw-Hill (1973), reprinted Dover (1987). This well-written text discusses a wide variety of numerical methods from zeros of functions to the fast Fourier transform. All topics are selected and developed with a modern computer in mind.

Hubbard, J., and B. H. West, *Differential Equations*. Berlin: Springer (1995).

Ince, E. L., *Ordinary Differential Equations*. New York: Dover (1956). The classic work in the theory of ordinary differential equations.

Lapidus, L., and J. H. Seinfeld, *Numerical Solutions of Ordinary Differential Equations*. New York: Academic Press (1971). A detailed and comprehensive discussion of numerical techniques, with emphasis on the Runge–Kutta and predictor–corrector methods. Recent work on the improvement of characteristics such as stability is clearly presented.

Margenau, H., and G. M. Murhpy, *The Mathematics of Physics and Chemistry*, 2nd ed. Princeton, NJ: Van Nostrand (1956). Chapter 5 covers curvilinear coordinates and 13 specific coordinate systems.

Miller, R. K., and A.N. Michel, *Ordinary Differential Equations*. New York: Academic Press (1982).

Morse, P. M., and H. Feshbach, *Methods of Theoretical Physics*. New York: McGraw-Hill (1953). Chapter 5 includes a description of several different coordinate systems. Note that Morse and Feshbach are not above using left-handed coordinate systems even for Cartesian coordinates. Elsewhere in this excellent (and difficult) book are many examples of the use of the various coordinate systems in solving physical problems. Chapter 7 is a particularly detailed, complete discussion of Green's functions from the point of view of mathematical physics. Note, however, that Morse and Feshbach frequently choose a source of $4\pi\delta(\mathbf{r}-\mathbf{r}')$ in place of our $\delta(\mathbf{r}-\mathbf{r}')$. Considerable attention is devoted to bounded regions.

Murphy, G. M., *Ordinary Differential Equations and Their Solutions*. Princeton, NJ: Van Nostrand (1960). A thorough, relatively readable treatment of ordinary differential equations, both linear and nonlinear.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*, 2nd ed. Cambridge, UK: Cambridge University Press (1992).

Ralston, A., and H. Wilf, eds., *Mathematical Methods for Digital Computers*. New York: Wiley (1960).

Ritger, P. D., and N. J. Rose, *Differential Equations with Applications*. New York: McGraw-Hill (1968).

Stakgold, I., *Green's Functions and Boundary Value Problems*, 2nd ed. New York: Wiley (1997).

Stoer, J., and R. Burlirsch, *Introduction to Numerical Analysis*. New York: Springer-Verlag (1992).

Stroud, A. H., *Numerical Quadrature and Solution of Ordinary Differential Equations*, Applied Mathematics Series, Vol. 10. New York: Springer-Verlag (1974). A balanced, readable, and very helpful discussion of various methods of integrating differential equations. Stroud is familiar with the work in this field and provides numerous references.

This page intentionally left blank

CHAPTER 10

STURM–LIOUVILLE THEORY — ORTHOGONAL FUNCTIONS

In the preceding chapter we developed two linearly independent solutions of the second-order linear homogeneous differential equation and proved that no third, linearly independent solution existed. In this chapter the emphasis shifts from solving the differential equation to developing and understanding general properties of the solutions. There is a close analogy between the concepts in this chapter and those of linear algebra in Chapter 3. Functions here play the role of vectors there, and linear operators that of matrices in Chapter 3. The diagonalization of a real symmetric matrix in Chapter 3 corresponds here to the solution of an ODE defined by a **self-adjoint** operator $\mathcal{L}$ in terms of its eigenfunctions, which are the "continuous" analog of the eigenvectors in Chapter 3. Examples for the corresponding analogy between Hermitian matrices and Hermitian operators are Hamiltonians in quantum mechanics and their energy eigenfunctions.

In Section 10.1 the concepts of self-adjoint operator, eigenfunction, eigenvalue, and Hermitian operator are presented. The concept of adjoint operator, given first in terms of differential equations, is then redefined in accordance with usage in quantum mechanics, where eigenfunctions take complex values. The vital properties of reality of eigenvalues and orthogonality of eigenfunctions are derived in Section 10.2. In Section 10.3 we discuss the Gram–Schmidt procedure for systematically constructuring sets of orthogonal functions. Finally, the general property of the completeness of a set of eigenfunctions is explored in Section 10.4, and Green's functions from Chapter 9 are continued in Section 10.5.

10.1 SELF-ADJOINT ODEs

In Chapter 9 we studied, classified, and solved linear, second-order ODEs corresponding to linear, second-order differential operators of the general form

$$\mathcal{L}u(x) = p_0(x)\frac{d^2}{dx^2}u(x) + p_1(x)\frac{d}{dx}u(x) + p_2(x)u(x). \tag{10.1}$$

The coefficients $p_0(x)$, $p_1(x)$, and $p_2(x)$ are real functions of x, and over the region of interest, $a \le x \le b$, the first $2 - i$ derivatives of $p_i(x)$ are continuous. Reference to Eq. (9.118) shows that $P(x) = p_1(x)/p_0(x)$ and $Q(x) = p_2(x)/p_0(x)$. Hence, $p_0(x)$ must not vanish for $a < x < b$. Now, the zeros of $p_0(x)$ are singular points (Section 9.4), and the preceding statement means that our interval $[a, b]$ must be given so that there are no singular points in the interior of the interval. There may be and often are singular points on the boundaries.

For a linear operator $\mathcal{L}$, the analog of a quadratic form for a matrix in Chapter 3 is the integral

$$\begin{aligned}\langle u|\mathcal{L}|u\rangle \equiv \langle u|\mathcal{L}u\rangle &\equiv \int_a^b u(x)\mathcal{L}u(x)\,dx \\ &= \int_a^b u\{p_0u'' + p_1u' + p_2u\}\,dx,\end{aligned} \tag{10.2}$$

where the primes on the real function $u(x)$ denote derivatives, as usual, and, for simplicity, $u(x)$ is taken to be real. If we shift the derivatives to the first factor, u, in Eq. (10.2) by integrating by parts once or twice, we are led to the equivalent expression,

$$\begin{aligned}\langle u|\mathcal{L}|u\rangle = &\big[u(x)(p_1 - p_0')u(x)\big]_{x=a}^b \\ &+ \int_a^b \left\{\frac{d^2}{dx^2}[p_0u] - \frac{d}{dx}[p_1u] + p_2u\right\}u\,dx.\end{aligned} \tag{10.3}$$

If we require that the integrals in Eqs. (10.2) and (10.3) be identical **for all (twice differentiable) functions** u, then the integrands have to be equal. The comparison then yields

$$u(p_0'' - p_1')u + 2u(p_0' - p_1)u' = 0,$$

or

$$p_0'(x) = p_1(x), \tag{10.4}$$

and, as a bonus, the terms at the boundaries $x = a$ and $x = b$ in Eq. (10.3) then also vanish.

Because of the analogy with the transposed matrix in Chapter 3, it is convenient to define the linear operator in Eq. (10.3),

$$\begin{aligned}\tilde{\mathcal{L}}u &= \frac{d^2}{dx^2}[p_0u] - \frac{d}{dx}[p_1u] + p_2u \\ &= p_0\frac{d^2u}{dx^2} + (2p_0' - p_1)\frac{du}{dx} + (p_0'' - p_1' + p_2)u,\end{aligned} \tag{10.5}$$

as the **adjoint**[1] **operator** $\bar{\mathcal{L}}$. We have defined the adjoint operator $\bar{\mathcal{L}}$ and have shown that if Eq. (10.4) is satisfied, $\langle \bar{\mathcal{L}}u|u\rangle = \langle u|\mathcal{L}u\rangle$. Following the same procedure we can show more generally that $\langle v|\mathcal{L}u\rangle = \langle \mathcal{L}v|u\rangle$. When this condition is satisfied,

$$\boxed{\bar{\mathcal{L}}u = \mathcal{L}u = \frac{d}{dx}\left[p(x)\frac{du(x)}{dx}\right] + q(x)u(x),} \tag{10.6}$$

the operator $\mathcal{L}$ is said to be **self-adjoint**. Here, for the self-adjoint case, $p_0(x)$ is replaced by $p(x)$ and $p_2(x)$ by $q(x)$ to avoid unnecessary subscripts. The form of Eq. (10.6) allows carrying out two integrations by parts in Eq. (10.3) (and Eq. (10.22) and following) without integrated terms.[2] Note that a given operator is not inherently self-adjoint; its self-adjointness depends on the properties of the function space in which it acts and on the boundary conditions.

In a survey of the ODEs introduced in Section 9.3, Legendre's equation and the linear oscillator equation are self-adjoint, but others, such as the Laguerre and Hermite equations, are not. However, the theory of linear, second-order, self-adjoint differential equations is perfectly general because we can **always** transform the non-self-adjoint operator into the required self-adjoint form. Consider Eq. (10.1) with $p_0' \neq p_1$. If we multiply $\mathcal{L}$ by[3]

$$\frac{1}{p_0(x)}\exp\left[\int^x \frac{p_1(t)}{p_0(t)}dt\right],$$

we obtain

$$\begin{aligned}\frac{1}{p_0(x)}\exp\left[\int^x \frac{p_1(t)}{p_0(t)}dt\right]\mathcal{L}u(x) = \frac{d}{dx}\left\{\exp\left[\int^x \frac{p_1(t)}{p_0(t)}dt\right]\frac{du(x)}{dx}\right\}\\ + \frac{p_2(x)}{p_0(x)}\cdot\exp\left[\int^x \frac{p_1(t)}{p_0(t)}dt\right]u,\end{aligned} \tag{10.7}$$

which is clearly self-adjoint (see Eq. (10.6)). Notice the $p_0(x)$ in the denominator. This is why we require $p_0(x) \neq 0$, $a < x < b$. In the following development we assume that $\mathcal{L}$ has been put into self-adjoint form.

[1] The **adjoint** operator bears a somewhat forced relationship to the **adjoint** matrix. A better justification for the nomenclature is found in a comparison of the **self-adjoint** operator (plus appropriate boundary conditions) with the **self-adjoint** matrix. The significant properties are developed in Section 10.2. Because of these properties, we are interested in **self-adjoint** operators.

[2] The full importance of the self-adjoint form (plus boundary conditions) will become apparent in Section 10.2. In addition, self-adjoint forms will be required for developing Green's functions in Section 10.5.

[3] If we multiply $\mathcal{L}$ by $f(x)/p_0(x)$ and then demand that

$$f'(x) = \frac{fp_1}{p_0},$$

so that the new operator will be self-adjoint, we obtain

$$f(x) = \exp\left[\int^x \frac{p_1(t)}{p_0(t)}\,dt\right].$$

Eigenfunctions, Eigenvalues

Schrödinger's wave equation

$$H\psi(x) = E\psi(x)$$

is the major example of an eigenvalue equation in physics; here the differential operator $\mathcal{L}$ is defined by the Hamiltonian H and may no longer be real, and the eigenvalue becomes the total energy E of the system. The eigenfunction $\psi(x)$ may be complex and is usually called a **wave function**. A variational formulation of this Schrödinger equation appears in Section 17.7. Based on spherical, cylindrical, or some other symmetry properties, a three- or four-dimensional PDE or eigenvalue equation such as the Schrödinger equation may separate into eigenvalue equations in a single variable each. Examples are Eqs. (9.41), (9.42), (9.50), and (9.53). However, sometimes an eigenvalue equation takes the more general self-adjoint form

$$\mathcal{L}u(x) + \lambda w(x)u(x) = 0, \tag{10.8}$$

where the constant λ is the eigenvalue[4] and $w(x)$ is a known weight or density function; $w(x) > 0$ except possibly at isolated points at which $w(x) = 0$. (In Section 10.1, $w(x) \equiv 1$.) For a given choice of the parameter λ, a function $u_\lambda(x)$, which satisfies Eq. (10.8) **and the imposed boundary conditions**, is called an **eigenfunction** corresponding to λ. The constant λ is then called an **eigenvalue** by mathematicians. There is no guarantee that an eigenfunction $u_\lambda(x)$ will exist for an arbitrary choice of the parameter λ. Indeed, the requirement that there be an eigenfunction often restricts the acceptable values of λ to a discrete set. Examples of this for the Legendre, Hermite, and Chebyshev equations appear in the exercises of Section 9.5. Here we have the mathematical approach to the process of quantization in quantum mechanics.

The inner product of two functions, $\langle v|u\rangle = \int_a^b v^*(x)w(x)u(x)\,dx$, depends on the weight function and generalizes our previous definition, where $w(x) \equiv 1$. The weight function also modifies the definition of **orthogonality** of two eigenfunctions: They are orthogonal if their inner product $\langle u_{\lambda'}|u_\lambda\rangle = 0$. The extra weight function $w(x)$ appears sometimes as an asymptotic wave function ψ_∞ that is a common factor in all solutions of a PDE such as the Schrödinger equation, for example, when the potential $V(x) \to 0$ as $x \to \infty$ in $H = T + V$. We can find ψ_∞ when we set $V = 0$ in the Schrödinger equation. Another source for $w(x)$ may be a nonzero angular momentum barrier $l(l+1)/x^2$ in a PDE or separated ODE Eq. (9.65) that has a regular singularity and dominates at $x \to 0$. In such a case the indicial equation, such as Eq. (9.87) or (9.103), shows that the wave function has x^l as an overall factor. Since the wave function enters twice in matrix elements and orthogonality relations, the weight functions in Table 10.1 come from these common factors in both radial wave functions. This is how the $\exp(-x)$ for Laguerre polynomials arises and $x^k \exp(-x)$ for associated Laguerre polynomials in Table 10.1.

[4] Note that this mathematical definition of the eigenvalue differs by a sign from the usage in physics.

Table 10.1

Equation	$p(x)$	$q(x)$	λ	$w(x)$
Legendre[a]	$1-x^2$	0	$l(l+1)$	1
Shifted Legendre[a]	$x(1-x)$	0	$l(l+1)$	1
Associated Legendre[a]	$1-x^2$	$-m^2/(1-x^2)$	$l(l+1)$	1
Chebyshev I	$(1-x^2)^{1/2}$	0	n^2	$(1-x^2)^{-1/2}$
Shifted Chebyshev I	$[x(1-x)]^{1/2}$	0	n^2	$[x(1-x)]^{-1/2}$
Chebyshev II	$(1-x^2)^{3/2}$	0	$n(n+2)$	$(1-x^2)^{1/2}$
Ultraspherical (Gegenbauer)	$(1-x^2)^{\alpha+1/2}$	0	$n(n+2\alpha)$	$(1-x^2)^{\alpha-1/2}$
Bessel[b], $0 \le x \le a$	x	$-n^2/x$	a^2	x
Laguerre, $0 \le x < \infty$	xe^{-x}	0	α	e^{-x}
Associated Laguerre[c]	$x^{k+1}e^{-x}$	0	$\alpha-k$	$x^k e^{-x}$
Hermite, $0 \le x < \infty$	e^{-x^2}	0	2α	e^{-x^2}
Simple harmonic oscillator[d]	1	0	n^2	1

[a] $l = 0, 1, \ldots, -l \le m \le l$ are integers and $-1 \le x \le 1$, $0 \le x \le 1$ for shifted Legendre.

[b] Orthogonality of Bessel functions is rather special. Compare Section 11.2. for details. A second type of orthogonality is developed in Eq. (11.174).

[c] k is a non-negative integer. For more details, see Table 10.2.

[d] This will form the basis for Chapter 14, Fourier series.

Example 10.1.1 LEGENDRE'S EQUATION

Legendre's equation is given by

$$\left(1-x^2\right)u'' - 2xu' + n(n+1)u = 0, \qquad -1 \le x \le 1. \tag{10.9}$$

From Eqs. (10.1), (10.8), and (10.9),

$$\begin{aligned} p_0(x) &= 1-x^2 = p, \qquad w(x) = 1, \\ p_1(x) &= -2x = p', \qquad \lambda = n(n+1), \\ p_2(x) &= 0 = q. \end{aligned}$$

Recall that our series solutions of Legendre's equation (Exercise 9.5.5)[5] diverged unless n was restricted to one of the integers. This represents a quantization of the eigenvalue λ. ■

When the equations of Chapter 9 are transformed into the self-adjoint form, we find the following values of the coefficients and parameters (Table 10.1). The coefficient $p(x)$ is the coefficient of the second derivative of the eigenfunction. The eigenvalue λ is the parameter that is available in a term of the form $\lambda w(x)u(x)$; any x dependence apart from the eigenfunction becomes the weighting function $w(x)$. If there is another term containing the eigenfunction (not the derivatives), the coefficient of the eigenfunction in this additional term is identified as $q(x)$. If no such term is present, $q(x)$ is zero.

[5] Compare also Exercise 5.2.15 and 12.10.

Example 10.1.2 DEUTERON

Further insight into the concepts of eigenfunction and eigenvalue may be provided by an extremely simple model of the deuteron, a bound state of a neutron and proton. From experiment, the binding energy of about 2 MeV$\ll Mc^2$, with $M = M_p = M_n$, the common neutron and proton mass whose small mass difference we neglect. Due to the short range of the nuclear force, the deuteron properties do not depend much on the detailed shape of the interaction potential. Thus, the neutron–proton nuclear interaction may be modeled by a spherically symmetric square well potential: $V = V_0 < 0$ for $0 \le r < a$, $V = 0$ for $r > a$. The Schrödinger wave equation is

$$-\frac{\hbar^2}{M}\nabla^2\psi + V\psi = E\psi, \tag{10.10}$$

where the energy eigenvalue $E < 0$ for a bound state. For the ground state the orbital angular momentum $l = 0$ because for $l \neq 0$ there is the additional positive angular momentum barrier. So, with $\psi = \psi(r)$, we may write $u(r) = r\psi(r)$, and, using Exercise 2.5.18, the wave equation becomes

$$\frac{d^2u}{dr^2} + k_1^2 u = 0, \tag{10.11}$$

with

$$k_1^2 = \frac{M}{\hbar^2}(E - V_0) > 0 \tag{10.12}$$

for the interior range, $0 \le r < a$. For $a < r < \infty$, we have

$$\frac{d^2u}{dr^2} - k_2^2 u = 0, \tag{10.13}$$

with

$$k_2^2 = -\frac{ME}{\hbar^2} > 0. \tag{10.14}$$

The boundary condition that ψ remain finite at $r = 0$ implies $u(0) = 0$ and

$$u_1(r) = \sin k_1 r, \qquad 0 \le r < a. \tag{10.15}$$

In the range outside the potential well, we have a linear combination of the two exponentials,

$$u_2(r) = A\exp k_2 r + B\exp(-k_2 r), \qquad a < r < \infty. \tag{10.16}$$

Continuity of particle and current density demand that $u_1(a) = u_2(a)$ and that $u_1'(a) = u_2'(a)$. These **joining**, or **matching, conditions** give

$$\begin{aligned} \sin k_1 a &= A\exp k_2 a + B\exp(-k_2 a), \\ k_1\cos k_1 a &= k_2 A\exp k_2 a - k_2 B\exp(-k_2 a). \end{aligned} \tag{10.17}$$

The condition that we actually have a bound proton–neutron combination is that $\int_0^\infty u^2(r)\,dr = 1$. This constraint can be met if we impose a boundary condition that $\psi(r)$

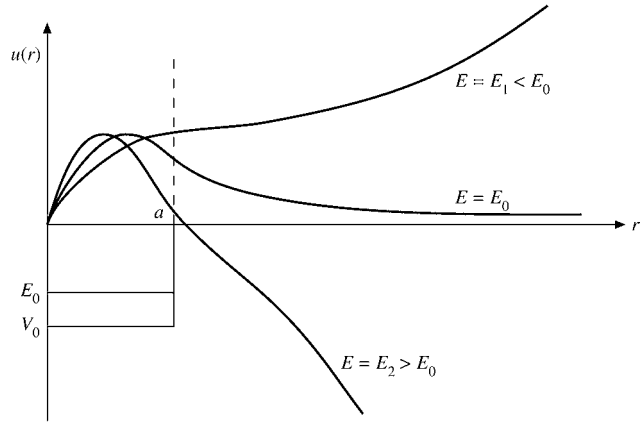

FIGURE 10.1 A deuteron eigenfunction.

remain finite as $r \to \infty$. And this, in turn, means that $A = 0$. Dividing the preceding pair of equations (to cancel B), we obtain

$$\tan k_1 a = -\frac{k_1}{k_2} = -\sqrt{\frac{E - V_0}{-E}}, \tag{10.18}$$

a transcendental equation for the energy E with only certain discrete solutions. If E is such that Eq. (10.18) can be satisfied, our solutions $u_1(r)$ and $u_2(r)$ can satisfy the boundary conditions. If Eq. (10.18) is not satisfied, **no acceptable solution exists**. The values of E for which Eq. (10.18) is satisfied are the eigenvalues; the corresponding functions u_1 and u_2 (or ψ) are the eigenfunctions. For the deuteron, problem there is one (and only one) negative value of E satisfying Eq. (10.18); that is, the deuteron has one and only one bound state.

Now, what happens if E does not satisfy Eq. (10.18), that is, if $E \neq E_0$ is not an eigenvalue? In graphical form, imagine that E and therefore k_1 are varied slightly. For $E = E_1 < E_0$, k_1 is reduced and $\sin k_1 a$ has not turned down enough to match $\exp(-k_2 a)$. The joining conditions, Eq. (10.17), require $A > 0$ and the wave function goes to $+\infty$ exponentially. For $E = E_2 > E_0$, k_1 is larger, $\sin k_1 a$ peaks sooner and has descended more rapidly at $r = a$. The joining conditions demand $A < 0$, and the wave function goes to $-\infty$ exponentially. Only for $E = E_0$, an eigenvalue, will the wave function have the required negative exponential asymptotic behavior (see Fig. 10.1). ■

Boundary Conditions

In the foregoing definition of eigenfunction, it was noted that the eigenfunction $u_\lambda(x)$ was required to satisfy certain imposed boundary conditions. The term **boundary conditions** includes as a special case the concept of **initial conditions**. For instance, specifying the initial position x_0 and the initial velocity v_0 in some dynamical problem would correspond to the Cauchy boundary conditions. The only difference in the present usage of boundary

conditions in these one-dimensional problems is that we are going to apply the conditions on **both** ends of the allowed range of the variable.

Usually the form of the differential equation or the boundary conditions on the solutions will guarantee that at the ends of our interval (that is, at the boundary, as suggested by Eq. (10.3)) the following products will vanish:

$$p(x)v^*(x)\frac{du(x)}{dx}\bigg|_{x=a} = 0 \qquad \text{and} \qquad p(x)v^*(x)\frac{du(x)}{dx}\bigg|_{x=b} = 0. \tag{10.19}$$

Here $u(x)$ and $v(x)$ are solutions of the particular ODE (Eq. (10.8)) being considered. A reason for this particular form of Eq. (10.19) is suggested shortly. If we recall the radial wave function u of the hydrogen atom with $u(0) = 0$ and $du/dr \sim e^{-kr} \to 0$ as $r \to \infty$, then both boundary conditions are satisfied. Similarly in the deuteron Example 10.1.2, $\sin k_1 r \to 0$ as $r \to 0$ and $d(e^{-k_2 r})/dr \to 0$ as $r \to \infty$, both boundary conditions are obeyed. We can, however, work with a somewhat less restrictive set of boundary conditions,

$$v^* pu'\big|_{x=a} = v^* pu'\big|_{x=b}, \tag{10.20}$$

in which $u(x)$ and $v(x)$ are solutions of the differential equation corresponding to the same or to different eigenvalues. Equation (10.20) might well be satisfied if we were dealing with a periodic physical system, such as a crystal lattice.

Equations (10.19) and (10.20) are written in terms of v^*, complex conjugate. When the solutions are real, $v = v^*$ and the asterisk may be ignored. However, in Fourier exponential expansions and in quantum mechanics the functions will be complex and the complex conjugate will be needed.

Example 10.1.3 INTEGRATION INTERVAL $[a, b]$

For $\mathcal{L} = d^2/dx^2$, a possible eigenvalue equation is

$$\frac{d^2}{dx^2}u(x) + n^2 u(x) = 0, \tag{10.21}$$

with eigenfunctions

$$u_n = \cos nx, \qquad v_m = \sin mx.$$

Equation (10.20) becomes

$$-n \sin mx \sin nx\big|_a^b = 0, \qquad \text{or} \qquad m \cos mx \cos nx\big|_a^b = 0,$$

interchanging u_n and v_m. Since $\sin mx$ and $\cos nx$ are periodic with period 2π (for n and m integral), Eq. (10.20) is clearly satisfied if $a = x_0$ and $b = x_0 + 2\pi$. If a problem prescribes a different interval, the eigenfunctions and eigenvalues will change along with the boundary conditions. The functions must always be chosen so that the boundary conditions (Eq. (10.20) etc.) are satisfied. For this case (Fourier series) the usual choices are $x_0 = 0$ leading to $(0, 2\pi)$ and $x_0 = -\pi$ leading to $(-\pi, \pi)$. Here and throughout the following several chapters **the orthogonality interval is so that the boundary conditions** (Eq. (10.20)) **will be satisfied**. The interval $[a, b]$ and the weighting factor $w(x)$ for the most commonly encountered second-order differential equations are listed in Table 10.2. ■

Table 10.2

Equation	a	b	$w(x)$
Legendre	-1	1	1
Shifted Legendre	0	1	1
Associated Legendre	-1	1	1
Chebyshev I	-1	1	$(1-x^2)^{-1/2}$
Shifted Chebyshev I	0	1	$[x(1-x)]^{-1/2}$
Chebyshev II	-1	1	$(1-x^2)^{1/2}$
Laguerre	0	∞	e^{-x}
Associated Laguerre	0	∞	$x^k e^{-x}$
Hermite	$-\infty$	∞	e^{-x^2}
Simple harmonic oscillator	0	2π	1
	$-\pi$	π	1

1. The orthogonality interval $[a, b]$ is determined by the boundary conditions of Section 10.1.
2. The weighting function is established by putting the ODE in self-adjoint form.

Hermitian Operators

We now prove an important property of the self-adjoint, second-order differential operator (Eq. (10.8)), in conjunction with solutions $u(x)$ and $v(x)$ that satisfy boundary conditions given by Eq. (10.20). This is motivated by applications in quantum mechanics.

By integrating v^* (complex conjugate) times the second-order self-adjoint differential operator $\mathcal{L}$ (operating on u) over the range $a \le x \le b$, we obtain

$$\int_a^b v^* \mathcal{L}u \, dx = \int_a^b v^* (pu')' \, dx + \int_a^b v^* q u \, dx \tag{10.22}$$

using Eq. (10.6). Integrating by parts, we have

$$\int_a^b v^* (pu')' \, dx = v^* pu' \Big|_a^b - \int_a^b v^{*\prime} pu' \, dx. \tag{10.23}$$

The integrated part vanishes on application of the boundary conditions (Eq. (10.20)). Integrating the remaining integral by parts a second time, we have

$$-\int_a^b v^{*\prime} pu' \, dx = -v^{*'} pu \Big|_a^b + \int_a^b u(pv^{*\prime})' \, dx. \tag{10.24}$$

Again, the integrated part vanishes in an application of Eq. (10.20). A combination of Eqs. (10.22) to (10.24) gives us

$$\int_a^b v^* \mathcal{L}u \, dx = \int_a^b u(\mathcal{L}v)^* \, dx. \tag{10.25}$$

This property, given by Eq. (10.25), is expressed by saying that the operator $\mathcal{L}$ is Hermitian with respect to the functions $u(x)$ and $v(x)$, which satisfy the boundary conditions specified by Eq. (10.20). Note that if this Hermitian property follows from self-adjointness in a Hilbert space, then it includes that boundary conditions are imposed on all functions of that space.

Hermitian Operators in Quantum Mechanics

The proceeding development in this section has focused on the classical second-order differential operators of mathematical physics. Generalizing our Hermitian operator theory as required in quantum mechanics, we have an extension: The operators need be neither second-order differential operators nor real. $p_x = -i\hbar(\partial/\partial x)$ will be a Hermitian operator. We simply assume (as is customary in quantum mechanics) that the wave functions satisfy appropriate boundary conditions: vanishing sufficiently strongly at infinity or having periodic behavior (as in a crystal lattice, or unit intensity for scattering problems). The operator $\mathcal{L}$ is called **Hermitian** if

$$\boxed{\int \psi_1^* \mathcal{L}\psi_2 \, d\tau = \int (\mathcal{L}\psi_1)^* \psi_2 \, d\tau.} \tag{10.26}$$

Apart from the simple extension to complex quantities, this definition is identical with Eq. (10.25).

The **adjoint** $A^\dagger$ of an operator A is defined by

$$\boxed{\int \psi_1^* A^\dagger \psi_2 \, d\tau \equiv \int (A\psi_1)^* \psi_2 \, d\tau.} \tag{10.27}$$

This generalizes our classical, second-derivative-operator–oriented definition, Eq. (10.5). Here the adjoint is defined in terms of the resultant integral, with the $A^\dagger$ as part of the integrand. Clearly, if $A = A^\dagger$ (**self-adjoint**) and satisfies the aforementioned boundary conditions, then A is Hermitian.

The **expectation value** of an operator $\mathcal{L}$ is defined as

$$\boxed{\langle \mathcal{L} \rangle = \int \psi^* \mathcal{L}\psi \, d\tau.} \tag{10.28a}$$

In the framework of quantum mechanics $\langle \mathcal{L} \rangle$ corresponds to the result of a measurement of the physical quantity represented by $\mathcal{L}$ when the physical system is in a state described by the wave function ψ. If we require $\mathcal{L}$ to be Hermitian, it is easy to show that $\langle \mathcal{L} \rangle$ is real (as would be expected from a measurement in a physical theory). Taking the complex conjugate of Eq. (10.28a), we obtain

$$\langle \mathcal{L} \rangle^* = \left[\int \psi^* \mathcal{L}\psi \, d\tau \right]^* = \int \psi \mathcal{L}^* \psi^* \, d\tau.$$

Rearranging the factors in the integrand, we have

$$\langle \mathcal{L} \rangle^* = \int (\mathcal{L}\psi)^* \psi \, d\tau.$$

Then, applying our definition of Hermitian operator, Eq. (10.26), we get

$$\langle \mathcal{L} \rangle^* = \int \psi^* \mathcal{L}\psi \, d\tau = \langle \mathcal{L} \rangle, \tag{10.28b}$$

or $\langle \mathcal{L} \rangle$ is real. It is worth noting that ψ is not necessarily an eigenfunction of $\mathcal{L}$.

Exercises

10.1.1 Show that Laguerre's ODE, Eq. (13.52), may be put into self-adjoint form by multiplying by e^{-x} and that $w(x) = e^{-x}$ is the weighting function.

10.1.2 Show that the Hermite ODE, Eq. (13.10), may be put into self-adjoint form by multiplying by e^{-x^2} and that this gives $w(x) = e^{-x^2}$ as the appropriate density function.

10.1.3 Show that the Chebyshev (type I) ODE, Eq. (13.100), may be put into self-adjoint form by multiplying by $(1 - x^2)^{-1/2}$ and that this gives $w(x) = (1 - x^2)^{-1/2}$ as the appropriate density function.

10.1.4 Show the following when the linear second-order differential equation is expressed in self-adjoint form:

(a) The Wronskian is equal to a constant divided by the initial coefficient p:

$$W(x) = \frac{C}{p(x)}.$$

(b) A second solution is given by

$$y_2(x) = Cy_1(x) \int^x \frac{dt}{p(t)[y_1(t)]^2}.$$

10.1.5 $U_n(x)$, the Chebyshev polynomial (type II), satisfies the ODE, Eq. (13.101),

$$\left(1 - x^2\right)U_n''(x) - 3xU_n'(x) + n(n+2)U_n(x) = 0.$$

(a) Locate the singular points that appear in the finite plane, and show whether they are regular or irregular.
(b) Put this equation in self-adjoint form.
(c) Identify the complete eigenvalue.
(d) Identify the weighting function.

10.1.6 For the very special case $\lambda = 0$ and $q(x) = 0$ the self-adjoint eigenvalue equation becomes

$$\frac{d}{dx}\left[p(x)\frac{du(x)}{dx}\right] = 0,$$

satisfied by

$$\frac{du}{dx} = \frac{1}{p(x)}.$$

Use this to obtain a "second" solution of the following:

(a) Legendre's equation,
(b) Laguerre's equation,
(c) Hermite's equation.

ANS. (a) $u_2(x) = \dfrac{1}{2}\ln\dfrac{1+x}{1-x}$,

(b) $u_2(x) - u_2(x_0) = \displaystyle\int_{x_0}^{x} e^t \frac{dt}{t}$,

(c) $u_2(x) = \displaystyle\int_0^x e^{t^2} dt$.

These second solutions illustrate the divergent behavior usually found in a second solution.

Note. In all three cases $u_1(x) = 1$.

10.1.7 Given that $\mathcal{L}u = 0$ and $g\mathcal{L}u$ is self-adjoint, show that for the adjoint operator $\tilde{\mathcal{L}}$, $\tilde{\mathcal{L}}(gu) = 0$.

10.1.8 For a second-order differential operator $\mathcal{L}$ that is self-adjoint show that

$$\int_a^b [y_2\mathcal{L}y_1 - y_1\mathcal{L}y_2]\,dx = p(y_1'y_2 - y_1y_2')\Big|_a^b.$$

10.1.9 Show that if a function ψ is required to satisfy Laplace's equation in a finite region of space and to satisfy Dirichlet boundary conditions over the entire closed bounding surface, then ψ is unique.

Hint. One of the forms of Green's theorem, Section 1.11, will be helpful.

10.1.10 Consider the solutions of the Legendre, Chebyshev, Hermite, and Laguerre equations to be polynomials. Show that the ranges of integration that guarantee that the Hermitian operator boundary conditions will be satisfied are

(a) Legendre $[-1, 1]$, (b) Chebyshev $[-1, 1]$,
(c) Hermite $(-\infty, \infty)$, (d) Laguerre $[0, \infty)$.

10.1.11 Within the framework of quantum mechanics (Eqs. (10.26) and following), show that the following are Hermitian operators:

(a) momentum $\mathbf{p} = -i\hbar\nabla \equiv -i\dfrac{h}{2\pi}\nabla$

(b) angular momentum $\mathbf{L} = -i\hbar\mathbf{r}\times\nabla \equiv -i\frac{h}{2\pi}\mathbf{r}\times\nabla$.

Hint. In Cartesian form $\mathbf{L}$ is a linear combination of noncommuting Hermitian operators.

10.1.12 (a) A is a non-Hermitian operator. In the sense of Eqs. (10.26) and (10.27), show that

$$A + A^\dagger \qquad \text{and} \qquad i(A - A^\dagger)$$

are Hermitian operators.

(b) Using the preceding result, show that every non-Hermitian operator may be written as a linear combination of two Hermitian operators.

10.1.13 U and V are two arbitrary operators, not necessarily Hermitian. In the sense of Eq. (10.27), show that

$$(UV)^\dagger = V^\dagger U^\dagger.$$

Note the resemblance to Hermitian adjoint matrices.
Hint. Apply the definition of adjoint operator, Eq. (10.27).

10.1.14 Prove that the product of two Hermitian operators is Hermitian (Eq. (10.26)) if and only if the two operators commute.

10.1.15 A and B are noncommuting quantum mechanical operators:

$$AB - BA = iC.$$

Show that C is Hermitian. Assume that appropriate boundary conditions are satisfied.

10.1.16 The operator $\mathcal{L}$ is Hermitian. Show that $\langle \mathcal{L}^2 \rangle \geq 0$.

10.1.17 A quantum mechanical expectation value is defined by

$$\langle A \rangle = \int \psi^*(x) A \psi(x)\, dx,$$

where A is a linear operator. Show that demanding that $\langle A \rangle$ be real means that A must be Hermitian — with respect to $\psi(x)$.

10.1.18 From the definition of adjoint, Eq. (10.27), show that $A^{\dagger\dagger} = A$ in the sense that $\int \psi_1^* A^{\dagger\dagger} \psi_2\, d\tau = \int \psi_1^* A \psi_2\, d\tau$. The adjoint of the adjoint is the original operator.
Hint. The functions ψ_1 and ψ_2 of Eq. (10.27) represent a class of functions. The subscripts 1 and 2 may be interchanged or replaced by other subscripts.

10.1.19 The Schrödinger wave equation for the deuteron (with a Woods–Saxon potential) is

$$-\frac{\hbar^2}{2M}\nabla^2\psi + \frac{V_0}{1+\exp[(r-r_0)/a]}\psi = E\psi.$$

Here $E = -2.224$ MeV, a is a "thickness parameter," 0.4×10^{-13} cm. Expressing lengths in fermis (10^{-13} cm) and energies in million electron volts (MeV), we may rewrite the wave equation as

$$\frac{d^2}{dr^2}(r\psi) + \frac{1}{41.47}\left[E - \frac{V_0}{1+\exp((r-r_0)/a)}\right](r\psi) = 0.$$

E is assumed known from experiment. The goal is to find V_0 for a specified value of r_0 (say, $r_0 = 2.1$). If we let $y(r) = r\psi(r)$, then $y(0) = 0$ and we take $y'(0) = 1$. Find V_0 such that $y(20.0) = 0$. (This should be $y(\infty)$, but $r = 20$ is far enough beyond the range of nuclear forces to approximate infinity.)

ANS. For $a = 0.4$ and $r_0 = 2.1$ fm, $V_0 = -34.159$ MeV.

10.1.20 Determine the nuclear potential well parameter V_0 of Exercise 10.1.19 as a function of r_0 for $r = 2.00(0.05)2.25$ fermis. Express your results as a power law

$$|V_0| r_0^\nu = k.$$

Determine the exponent ν and the constant k. This power-law formulation is useful for accurate interpolation.

10.1.21 In Exercise 10.1.19 it was assumed that 20 fermis was a good approximation to infinity. Check on this by calculating V_0 for $r\psi(r) = 0$ at (a) $r = 15$, (b) $r = 20$, (c) $r = 25$, and (d) $r = 30$. Sketch your results. Take $r_0 = 2.10$ and $a = 0.4$ (fermis).

10.1.22 For a quantum particle moving in a potential well, $V(x) = \frac{1}{2}m\omega^2x^2$, the Schrödinger wave equation is

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + \frac{1}{2}m\omega^2x^2\psi(x) = E\psi(x),$$

or

$$\frac{d^2\psi(z)}{dz^2} - z^2\psi(z) = -\frac{2E}{\hbar\omega}\psi(z),$$

where $z = (m\omega/\hbar)^{1/2}x$. Since this operator is even, we expect solutions of definite parity. For the initial conditions that follow, integrate out from the origin and determine the minimum constant $2E/\hbar\omega$ that will lead to $\psi(\infty) = 0$ in each case. (You may take $z = 6$ as an approximation of infinity.)

(a) For an even eigenfunction,

$$\psi(0) = 1, \qquad \psi'(0) = 0.$$

(b) For an odd eigenfunction,

$$\psi(0) = 0, \qquad \psi'(0) = 1.$$

Note. Analytical solutions appear in Section 13.1.

10.2 HERMITIAN OPERATORS

Hermitian, or self-adjoint, operators with appropriate boundary conditions have three properties that are of extreme importance in physics, both classical and quantum.

1. The eigenvalues of a Hermitian operator are real.
2. A Hermitian operator possesses an orthogonal set of eigenfunctions.
3. The eigenfunctions of a Hermitian operator form a complete set.[6]

Real Eigenvalues

We proceed to prove the first two of these three properties. Let

$$\mathcal{L}u_i + \lambda_i w u_i = 0. \tag{10.29}$$

[6]This third property is not universal. It **does** hold for our linear, second-order differential operators in Sturm–Liouville (self-adjoint) form. Completeness is defined and discussed in Section 10.4. A proof that the eigenfunctions of our linear, second-order, self-adjoint, differential equations form a complete set may be developed from the calculus of variations of Section 17.8.

Assuming the existence of a second eigenvalue and eigenfunction,

$$\mathcal{L}u_j + \lambda_j w u_j = 0. \tag{10.30}$$

Then, taking the complex conjugate, we obtain

$$\mathcal{L}^* u_j^* + \lambda_j^* w u_j^* = 0. \tag{10.31}$$

Here $w(x) \geq 0$ is a real function. But we permit λ_k, the eigenvalues, and u_k, the eigenfunctions, to be complex. Multiplying Eq. (10.29) by u_j^* and Eq. (10.31) by u_i and then subtracting, we have

$$u_j^* \mathcal{L}u_i - u_i \mathcal{L}^* u_j^* = (\lambda_j^* - \lambda_i) w u_i u_j^*. \tag{10.32}$$

We integrate over the range $a \leq x \leq b$:

$$\int_a^b \left(u_j^* \mathcal{L}u_i - u_i \mathcal{L}^* u_j^*\right) dx = (\lambda_j^* - \lambda_i) \int_a^b u_i u_j^* w \, dx. \tag{10.33}$$

Since $\mathcal{L}$ is Hermitian, the left-hand side vanishes by Eq. (10.26) and

$$(\lambda_j^* - \lambda_i) \int_a^b u_i u_j^* w \, dx = 0. \tag{10.34}$$

If $i = j$, the integral cannot vanish [$w(x) \geq 0$, apart from isolated points], except in the trivial case $u_i = 0$. Hence the coefficient $(\lambda_j^* - \lambda_i)$ must be zero,

$$\lambda_i^* = \lambda_i, \tag{10.35}$$

which says that the eigenvalue is real. Since λ_i can represent any one of the eigenvalues, this proves the first property. This is an exact analog of the nature of the eigenvalues of real symmetric (and of Hermitian) matrices (compare Section 3.5).

The analog of the spectral decomposition of a real symmetric matrix in Section 3.5 for a Hermitian operator $\mathcal{L}$ with a discrete set of eigenvalues λ_i takes the form

$$\mathcal{L} = \sum_i \lambda_i |u_i\rangle\langle u_i|, \qquad f(\mathcal{L}) = \sum_i f(\lambda_i)|u_i\rangle\langle u_i|$$

with eigenvectors $|u_i\rangle$ and any infinitely differentiable function f.

Real eigenvalues of Hermitian operators have a fundamental significance in quantum mechanics. In quantum mechanics the eigenvalues correspond to precisely measurable quantities, such as energy and angular momentum. With the theory formulated in terms of Hermitian operators, this proof of real eigenvalues guarantees that the theory will predict real numbers for these measurable physical quantities. In Section 17.8 it will be seen that the set of real eigenvalues has a lower bound (for nonrelativistic problems).

Orthogonal Eigenfunctions

If we now take $i \neq j$ and if $\lambda_i \neq \lambda_j$ in Eq. (10.34), the integral of the product of the two different eigenfunctions must vanish:

$$\int_a^b u_i u_j^* w \, dx = 0. \tag{10.36}$$

This condition, called **orthogonality**, is the continuum analog of the vanishing of a scalar product of two vectors.[7] We say that the eigenfunctions $u_i(x)$ and $u_j(x)$ are orthogonal with respect to the weighting function $w(x)$ over the interval $[a, b]$. Equation (10.36) constitutes a partial proof of the second property of our Hermitian operators. Again, the precise analogy with matrix analysis should be noted. Indeed, we can establish a one-to-one correspondence between this Sturm–Liouville theory of differential equations and the treatment of Hermitian matrices. Historically, this correspondence has been significant in establishing the mathematical equivalence of matrix mechanics developed by Heisenberg and wave mechanics developed by Schrödinger. Today, the two diverse approaches are merged into the theory of quantum mechanics, and the mathematical formulation that is more convenient for a particular problem is used for that problem. Actually the mathematical alternatives do not end here. Integral equations, Chapter 16, form a third equivalent and sometimes more convenient or more powerful approach.

This proof of orthogonality is not quite complete. There is a loophole, because we may have $u_i \neq u_j$ but still have $\lambda_i = \lambda_j$. Such a case is labeled **degenerate**. Illustrations of degeneracy are given at the end of this section. If $\lambda_i = \lambda_j$, the integral in Eq. (10.34) need not vanish. This means that linearly independent eigenfunctions corresponding to the same eigenvalue are not automatically orthogonal and that some other method must be sought to obtain an orthogonal set. Although the eigenfunctions in this degenerate case may not be orthogonal, they can always be made orthogonal. One method is developed in the next section. See also Eq. (4.21) for degeneracy due to symmetry.

We shall see in succeeding chapters that it is just as desirable to have a given set of functions orthogonal as it is to have an orthogonal coordinate system. We can work with nonorthogonal functions, but they are likely to prove as messy as an oblique coordinate system.

Example 10.2.1 FOURIER SERIES — ORTHOGONALITY

To continue Example 10.1.3, the eigenvalue equation, Eq. (10.21),

$$\frac{d^2}{dx^2} y(x) + n^2 y(x) = 0,$$

[7] From the definition of Riemann integral,

$$\int_a^b f(x)g(x)\,dx = \lim_{N\to\infty} \left(\sum_{i=1}^{N} f(x_i)g(x_i)\Delta x \right),$$

where $x_0 = a$, $x_N = b$, and $x_i - x_{i-1} = \Delta x$. If we interpret $f(x_i)$ and $g(x_i)$ as the ith components of an N-component vector, then this sum (and therefore this integral) corresponds directly to a scalar product of vectors, Eq. (1.24). The vanishing of the scalar product is the condition for **orthogonality** of the vectors — or functions.

may describe a quantum mechanical particle in a box, or perhaps a vibrating violin string, a classical harmonic oscillator with degenerate eigenfunctions — $\cos nx$, $\sin nx$ — and eigenvalues n^2, n an integer.

With n real (here taken to be integral), the orthogonality integrals become

$$\text{(a)} \quad \int_{x_0}^{x_0+2\pi} \sin mx \sin nx\, dx = C_n \delta_{nm},$$

$$\text{(b)} \quad \int_{x_0}^{x_0+2\pi} \cos mx \cos nx\, dx = D_n \delta_{nm},$$

$$\text{(c)} \quad \int_{x_0}^{x_0+2\pi} \sin mx \cos nx\, dx = 0.$$

For an interval of 2π the preceding analysis guarantees the Kronecker delta in (a) and (b) but not the zero in (c) because (c) may involve degenerate eigenfunctions. However, inspection shows that (c) always vanishes for all integral m and n.

Our Sturm–Liouville theory says nothing about the values of C_n and D_n because homogeneous ODEs have solutions whose scaling is arbitrary. Actual calculation yields

$$C_n = \begin{cases} \pi, & n \neq 0, \\ 0, & n = 0, \end{cases} \qquad D_n = \begin{cases} \pi, & n \neq 0, \\ 2\pi, & n = 0. \end{cases}$$

These orthogonality integrals form the basis of the Fourier series developed in Chapter 14. ■

Example 10.2.2 EXPANSION IN ORTHOGONAL EIGENFUNCTIONS—SQUARE WAVE

The property of completeness (see Eq. (1.190) and Section 10.4) means that certain classes of functions (for example, sectionally or piecewise continuous) may be represented by a series of orthogonal eigenfunctions. Consider the square-wave shape

$$f(x) = \begin{cases} \dfrac{h}{2}, & 0 < x < \pi, \\ -\dfrac{h}{2}, & -\pi < x < 0. \end{cases} \tag{10.37}$$

This function may be expanded in any of a variety of eigenfunctions — Legendre, Hermite, Chebyshev, and so on. The choice of eigenfunction is made on the basis of convenience or an application. To illustrate the expansion technique, let us choose the eigenfunctions of Example 10.2.1, $\cos nx$ and $\sin nx$.

The eigenfunction series is conveniently (and conventionally) written as

$$f(x) = \frac{a_0}{2} + \sum_{m=1}^{\infty} (a_m \cos mx + b_m \sin mx).$$

Upon multiplying $f(t)$ by $\cos nt$ or $\sin nt$ and integrating, only the nth term survives, by the orthogonality integrals of Example 10.2.1, thus yielding the coefficients

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos nt\, dt, \qquad b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin nt\, dt, \qquad n = 0, 1, 2 \ldots.$$

Direct substitution of $\pm h/2$ for $f(t)$ yields

$$a_n = 0,$$

which is expected here because of the antisymmetry, $f(-x) = -f(x)$, and

$$b_n = \frac{h}{n\pi}(1 - \cos n\pi) = \begin{cases} 0, & n \text{ even}, \\ \dfrac{2h}{n\pi}, & n \text{ odd}. \end{cases}$$

Hence the eigenfunction (Fourier) expansion of the square wave is

$$f(x) = \frac{2h}{\pi} \sum_{n=0}^{\infty} \frac{\sin(2n+1)x}{2n+1}. \tag{10.38}$$

Additional examples, using other eigenfunctions, appear in Chapters 11 and 12. ■

Degeneracy

The concept of degeneracy was introduced earlier. If N linearly independent eigenfunctions correspond to the same eigenvalue, the eigenvalue is said to be N-fold degenerate. A particularly simple illustration is provided by the eigenvalues and eigenfunctions of the classical harmonic oscillator equation, Example 10.2.1. For each eigenvalue n^2, there are two possible solutions: $\sin nx$ and $\cos nx$ (and any linear combination, n an integer). We say the eigenfunctions are degenerate or the eigenvalue is degenerate.

A more involved example is furnished by the physical system of an electron in an atom (nonrelativistic treatment, spin neglected). From the Schrödinger equation, Eq. (13.84) for hydrogen, the total energy of the electron is our eigenvalue. We may label it E_{nLM} by using the quantum numbers n, L, and M as subscripts. For each distinct set of quantum numbers (n, L, M) there is a distinct, linearly independent eigenfunction $\psi_{nLM}(r, \theta, \varphi)$. For hydrogen, the energy E_{nLM} is independent of L and M, reflecting the spherical (and SO(4)) symmetry of the Coulomb potential. With $0 \le L \le n-1$ and $-L \le M \le L$, the eigenvalue is n^2-fold degenerate (including the electron spin would raise this to $2n^2$). In atoms with more than one electron, the electrostatic potential is no longer a simple r^{-1} potential. The energy depends on L as well as on n, although **not** on M; E_{nLM} is still $(2L+1)$-fold degenerate. This degeneracy — due to rotational invariance of the potential — may be removed by applying an external magnetic field, breaking spherical symmetry and giving rise to the Zeeman effect. As a rule, the eigenfunctions form a Hilbert space, that is, a complete vector space of functions with a metric defined by the inner product (see Section 10.4 for more details and examples).

Often an underlying symmetry, such as rotational invariance, is causing the degeneracies. States belonging to the same energy eigenvalue then will form a multiplet or representation of the symmetry group. The powerful group-theoretical methods are treated in Chapter 4 in some detail.

Exercises

10.2.1 The functions $u_1(x)$ and $u_2(x)$ are eigenfunctions of the same Hermitian operator but for distinct eigenvalues λ_1 and λ_2. Prove that $u_1(x)$ and $u_2(x)$ are linearly independent.

10.2.2 (a) The vectors $\mathbf{e}_n$ are orthogonal to each other: $\mathbf{e}_n \cdot \mathbf{e}_m = 0$ for $n \neq m$. Show that they are linearly independent.

(b) The functions $\psi_n(x)$ are orthogonal to each other over the interval $[a, b]$ and with respect to the weighting function $w(x)$. Show that the $\psi_n(x)$ are linearly independent.

10.2.3 Given that

$$P_1(x) = x \qquad \text{and} \qquad Q_0(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right)$$

are solutions of Legendre's differential equation corresponding to different eigenvalues:

(a) Evaluate their orthogonality integral

$$\int_{-1}^{1} \frac{x}{2}\ln\left(\frac{1+x}{1-x}\right) dx.$$

(b) Explain why these two functions are not orthogonal, that is, why the proof of orthogonality does not apply.

10.2.4 $T_0(x) = 1$ and $V_1(x) = (1-x^2)^{1/2}$ are solutions of the Chebyshev differential equation corresponding to different eigenvalues. Explain, in terms of the boundary conditions, why these two functions are not orthogonal.

10.2.5 (a) Show that the first derivatives of the Legendre polynomials satisfy a self-adjoint differential equation with eigenvalue $\lambda = n(n+1) - 2$.

(b) Show that these Legendre polynomial derivatives satisfy an orthogonality relation

$$\int_{-1}^{1} P_m'(x) P_n'(x)\left(1-x^2\right) dx = 0, \qquad m \neq n.$$

Note. In Section 12.5, $(1-x^2)^{1/2} P_n'(x)$ will be labeled an associated Legendre polynomial, $P_n^1(x)$.

10.2.6 A set of functions $u_n(x)$ satisfies the Sturm–Liouville equation

$$\frac{d}{dx}\left[p(x)\frac{d}{dx}u_n(x)\right] + \lambda_n w(x) u_n(x) = 0.$$

The functions $u_m(x)$ and $u_n(x)$ satisfy boundary conditions that lead to orthogonality. The corresponding eigenvalues λ_m and λ_n are distinct. Prove that for appropriate boundary conditions, $u_m'(x)$ and $u_n'(x)$ are orthogonal with $p(x)$ as a weighting function.

10.2.7 A linear operator A has n distinct eigenvalues and n corresponding eigenfunctions: $A\psi_i = \lambda_i \psi_i$. Show that the n eigenfunctions are linearly independent. A is not necessarily Hermitian.
Hint. Assume linear dependence — that $\psi_n = \sum_{i=1}^{n-1} a_i \psi_i$. Use this relation and the operator–eigenfunction equation first in one order and then in the reverse order. Show that a contradiction results.

10.2.8 (a) Show that the Liouville substitution

$$u(x) = v(\xi)\big[p(x)w(x)\big]^{-1/4}, \qquad \xi = \int_a^x \left[\frac{w(t)}{p(t)}\right]^{1/2} dt$$

transforms

$$\frac{d}{dx}\left[p(x)\frac{d}{dx}u\right] + \big[\lambda w(x) - q(x)\big]u(x) = 0$$

into

$$\frac{d^2v}{d\xi^2} + \big[\lambda - Q(\xi)\big]v(\xi) = 0,$$

where

$$Q(\xi) = \frac{q(x(\xi))}{w(x(\xi))} + \big[p\big(x(\xi)\big)w\big(x(\xi)\big)\big]^{-1/4}\frac{d^2}{d\xi^2}(pw)^{1/4}.$$

(b) If $v_1(\xi)$ and $v_2(\xi)$ are obtained from $u_1(x)$ and $u_2(x)$, respectively, by a Liouville substitution, show that $\int_a^b w(x)u_1u_2\,dx$ is transformed into $\int_0^c v_1(\xi)v_2(\xi)\,d\xi$ with $c = \int_a^b [\frac{w}{p}]^{1/2}\,dx$.

10.2.9 The ultraspherical polynomials $C_n^{(\alpha)}(x)$ are solutions of the differential equation

$$\left\{(1-x^2)\frac{d^2}{dx^2} - (2\alpha+1)x\frac{d}{dx} + n(n+2\alpha)\right\}C_n^{(\alpha)}(x) = 0.$$

(a) Transform this differential equation into self-adjoint form.
(b) Show that the $C_n^{(\alpha)}(x)$ are orthogonal for different n. Specify the interval of integration and the weighting factor.

Note. Assume that your solutions are polynomials.

10.2.10 With $\mathcal{L}$ **not** self-adjoint,

$$\mathcal{L}u_i + \lambda_i w u_i = 0$$

and

$$\tilde{\mathcal{L}}v_j + \lambda_j w v_j = 0.$$

(a) Show that

$$\int_a^b v_j \mathcal{L} u_i \, dx = \int_a^b u_i \tilde{\mathcal{L}} v_j \, dx,$$

provided

$$u_i p_0 v_j' \Big|_a^b = v_j p_0 u_i' \Big|_a^b$$

and

$$u_i (p_1 - p_0') v_j \Big|_a^b = 0.$$

(b) Show that the orthogonality integral for the eigenfunctions u_i and v_j becomes

$$\int_a^b u_i v_j w \, dx = 0 \qquad (\lambda_i \neq \lambda_j).$$

10.2.11 In Exercise 9.5.8 the series solution of the Chebyshev equation is found to be convergent for all eigenvalues n. Therefore n is **not** quantized by the argument used for Legendre's (Exercise 9.5.5). Calculate the sum of the indicial equation $k = 0$ Chebyshev series for $n = \nu = 0.8$, 0.9, and 1.0 and for $x = 0.0(0.1)0.9$.
Note. The Chebyshev series recurrence relation is given in Exercise 5.2.16.

10.2.12 (a) Evaluate the $n = \nu = 0.9$, indicial equation $k = 0$ Chebyshev series for $x = 0.98$, 0.99, and 1.00. The series converges very slowly at $x = 1.00$. You may wish to use double precision. Upper bounds to the error in your calculation can be set by comparison with the $\nu = 1.0$ case, which corresponds to $(1 - x^2)^{1/2}$.

(b) These series solutions for eigenvalue $\nu = 0.9$ and for $\nu = 1.0$ are obviously **not** orthogonal, despite the fact that they satisfy a self-adjoint eigenvalue equation with different eigenvalues. From the behavior of the solutions in the vicinity of $x = 1.00$ try to formulate a hypothesis as to why the proof of orthogonality does not apply.

10.2.13 The Fourier expansion of the (asymmetric) square wave is given by Eq. (10.38). With $h = 2$, evaluate this series for $x = 0(\pi/18)\pi/2$, using the first (a) 10 terms, (b) 100 terms of the series.
Note. For 10 terms and $x = \pi/18$, or 10°, your Fourier representation has a sharp hump. This is the Gibbs phenomenon of Section 14.5. For 100 terms this hump has been shifted over to about 1°.

10.2.14 The **symmetric** square wave

$$f(x) = \begin{cases} 1, & |x| < \dfrac{\pi}{2} \\ -1, & \dfrac{\pi}{2} < |x| < \pi \end{cases}$$

has a Fourier expansion

$$f(x) = \frac{4}{\pi} \sum_{n=0}^{\infty} (-1)^n \frac{\cos(2n+1)x}{2n+1}.$$

Evaluate this series for $x = 0(\pi/18)\pi/2$ using the first

(a) 10 terms, (b) 100 terms of the series.

Note. As in Exercise 10.2.13, the Gibbs phenomenon appears at the discontinuity. This means that a Fourier series is not suitable for precise numerical work in the vicinity of a discontinuity.

10.3 GRAM–SCHMIDT ORTHOGONALIZATION

The Gram–Schmidt orthogonalization is a method that takes a nonorthogonal set of linearly independent vectors (see Section 3.1) or functions[8] and constructs an orthogonal set of vectors or functions over an arbitrary interval and with respect to an arbitrary weight or density factor. In the language of linear algebra, the process is equivalent to a matrix transformation relating an orthogonal set of basis vectors (functions) to a nonorthogonal set. A specific example of this matrix transformation appears in Exercise 12.2.1.

Next we apply the Gram–Schmidt procedure to a set of functions. The functions involved may be real or complex. Here for convenience they are assumed to be real. The generalization to the complex case offers no difficulty.

Before taking up orthogonalization, we should consider normalization of functions. So far no normalization has been specified. This means that

$$\int_a^b \varphi_i^2 w\,dx = N_i^2,$$

but no attention has been paid to the value of N_i. Since our basic equation (Eq. (10.8)) is linear and homogeneous, we may multiply our solution by any constant and it will still be a solution. We now demand that each solution $\varphi_i(x)$ be multiplied by N_i^{-1} so that the new (normalized) φ_i will satisfy

$$\int_a^b \varphi_i^2(x) w(x)\,dx = 1 \tag{10.39}$$

and

$$\int_a^b \varphi_i(x)\varphi_j(x) w(x)\,dx = \delta_{ij}. \tag{10.40}$$

Equation (10.39) says that we have normalized to unity. Including the property of orthogonality, we have Eq. (10.40). Functions satisfying this equation are said to be **orthonormal** (orthogonal plus unit normalization). Other normalizations are certainly possible, and indeed, by historical convention, each of the special functions of mathematical physics treated in Chapters 12 and 13 will be normalized differently.

We consider three sets of functions: an original, linearly independent given set $u_n(x)$, $n = 0, 1, 2, \ldots$; an orthogonalized set $\psi_n(x)$ to be constructed; and a final set

[8] Such a set of functions might well arise from the solutions of a PDE in which the eigenvalue was independent of one or more of the constants of separation. As an example, we have the hydrogen atom problem (Sections 10.2 and 13.2). The eigenvalue (energy) is independent of both the electron orbital angular momentum and its projection on the z-axis, m. Note, however, that the origin of the set of functions is irrelevant to the Gram–Schmidt orthogonalization procedure.

of functions $\varphi_n(x)$, which are the normalized ψ_n. The original u_n may be degenerate eigenfunctions, but this is not necessary. We shall have the following properties:

$u_n(x)$	$\psi_n(x)$	$\varphi_n(x)$
Linearly independent	*Linearly independent*	*Linearly independent*
Nonorthogonal	*Orthogonal*	*Orthogonal*
Unnormalized	Unnormalized	*Normalized* (orthonormal)

The Gram–Schmidt procedure takes the nth ψ function (ψ_n) to be $u_n(x)$ plus an unknown linear combination of the previous φ. The presence of the new $u_n(x)$ will guarantee linear independence. The requirement that $\psi_n(x)$ be orthogonal to each of the previous φ yields just enough constraints to determine each of the unknown coefficients. Then the fully determined ψ_n will be normalized to unity, yielding $\varphi_n(x)$. Then the sequence of steps is repeated for $\psi_{n+1}(x)$.

We start with $n = 0$, letting

$$\psi_0(x) = u_0(x), \tag{10.41}$$

with no "previous" φ to worry about. Then we normalize

$$\varphi_0(x) = \frac{\psi_0(x)}{[\int \psi_0^2 w\, dx]^{1/2}}. \tag{10.42}$$

For $n = 1$, let

$$\psi_1(x) = u_1(x) + a_{1,0}\varphi_0(x). \tag{10.43}$$

We demand that $\psi_1(x)$ be orthogonal to $\varphi_0(x)$. (At this stage the normalization of $\psi_1(x)$ is irrelevant.) This orthogonality leads to

$$\int \psi_1 \varphi_0 w\, dx = \int u_1 \varphi_0 w\, dx + a_{1,0} \int \varphi_0^2 w\, dx = 0. \tag{10.44}$$

Since φ_0 is normalized to unity (Eq. (10.42)), we have

$$a_{1,0} = -\int u_1 \varphi_0 w\, dx, \tag{10.45}$$

fixing the value of $a_{1,0}$. Normalizing, we define

$$\varphi_1(x) = \frac{\psi_1(x)}{(\int \psi_1^2 w\, dx)^{1/2}}. \tag{10.46}$$

Finally, we generalize so that

$$\varphi_i(x) = \frac{\psi_i(x)}{(\int \psi_i^2(x) w(x)\, dx)^{1/2}}, \tag{10.47}$$

where

$$\psi_i(x) = u_i + a_{i,0}\varphi_0 + a_{i,1}\varphi_1 + \cdots + a_{i,i-1}\varphi_{i-1}. \tag{10.48}$$

The coefficients $a_{i,j}$ are given by

$$a_{i,j} = -\int u_i \varphi_j w\, dx. \tag{10.49}$$

Equation (10.49) holds for unit normalization. If some other normalization is selected,

$$\int_a^b \left[\varphi_j(x)\right]^2 w(x)\,dx = N_j^2,$$

then Eq. (10.47) is replaced by

$$\varphi_i(x) = N_i \frac{\psi_i(x)}{(\int \psi_i^2 w\,dx)^{1/2}}. \tag{10.47a}$$

and $a_{i,j}$ becomes

$$a_{i,j} = -\frac{\int u_i \varphi_j w\,dx}{N_j^2}. \tag{10.49a}$$

Equations (10.48) and (10.49) may be rewritten in terms of projection operators, P_j. If we consider the $\varphi_n(x)$ to form a linear vector space, then the integral in Eq. (10.49) may be interpreted as the projection of u_i into the φ_j "coordinate," or the jth component of u_i. With

$$P_j u_i(x) = \left\{\int u_i(t)\varphi_j(t)w(t)\,dt\right\}\varphi_j(x),$$

Eq. (10.48) becomes

$$\psi_i(x) = \left\{1 - \sum_{j=1}^{i-1} P_j\right\} u_i(x). \tag{10.48a}$$

Subtracting off the components, $j = 1$ to $i-1$, leaves $\psi_i(x)$ orthogonal to all the $\varphi_j(x)$.

It will be noticed that although this Gram–Schmidt procedure is one possible way of constructing an orthogonal or orthonormal set, the functions $\varphi_i(x)$ are not unique. There is an infinite number of possible orthonormal sets for a given interval and a given density function.

As an illustration of the freedom involved, consider two (nonparallel) vectors **A** and **B** in the xy-plane. We may normalize **A** to unit magnitude and then form $\mathbf{B}' = a\mathbf{A} + \mathbf{B}$ so that $\mathbf{B}'$ is perpendicular to **A**. By normalizing $\mathbf{B}'$ we have completed the Gram–Schmidt orthogonalization for two vectors. But any two perpendicular unit vectors, such as $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$, could have been chosen as our orthonormal set. Again, with an infinite number of possible rotations of $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ about the z-axis, we have an infinite number of possible orthonormal sets.

Example 10.3.1 LEGENDRE POLYNOMIALS BY GRAM–SCHMIDT ORTHOGONALIZATION

Let us form an orthonormal set from the set of functions $u_n(x) = x^n$, $n = 0, 1, 2\ldots$. The interval is $-1 \le x \le 1$ and the density function is $w(x) = 1$.

In accordance with the Gram–Schmidt orthogonalization process described,

$$u_0 = 1, \qquad \text{hence} \qquad \varphi_0 = \frac{1}{\sqrt{2}}. \tag{10.50}$$

Then

$$\psi_1(x) = x + a_{1,0}\frac{1}{\sqrt{2}} \tag{10.51}$$

and

$$a_{1,0} = -\int_{-1}^{1} \frac{x}{\sqrt{2}} dx = 0 \tag{10.52}$$

by symmetry. We normalize ψ_1 to obtain

$$\varphi_1(x) = \sqrt{\frac{3}{2}}x. \tag{10.53}$$

Then we continue the Gram–Schmidt procedure with

$$\psi_2(x) = x^2 + a_{2,0}\frac{1}{\sqrt{2}} + a_{2,1}\sqrt{\frac{3}{2}}x, \tag{10.54}$$

where

$$a_{2,0} = -\int_{-1}^{1} \frac{x^2}{\sqrt{2}} dx = -\frac{\sqrt{2}}{3}, \tag{10.55}$$

$$a_{2,1} = -\int_{-1}^{1} \sqrt{\frac{3}{2}}x^3\, dx = 0, \tag{10.56}$$

again by symmetry. Therefore

$$\psi_2(x) = x^2 - \frac{1}{3}, \tag{10.57}$$

and, on normalizing to unity, we have

$$\varphi_2(x) = \sqrt{\frac{5}{2}} \cdot \frac{1}{2}\left(3x^2 - 1\right). \tag{10.58}$$

The next function, $\varphi_3(x)$, becomes

$$\varphi_3(x) = \sqrt{\frac{7}{2}} \cdot \frac{1}{2}\left(5x^3 - 3x\right). \tag{10.59}$$

Reference to Chapter 12 will show that

$$\varphi_n(x) = \sqrt{\frac{2n+1}{2}} P_n(x), \tag{10.60}$$

where $P_n(x)$ is the nth-order Legendre polynomial. Our Gram–Schmidt process provides a possible but very cumbersome method of generating the Legendre polynomials. It illustrates how a power-series expansion in $u_n(x) = x^n$, which is not orthogonal, can be converted into an orthogonal series. ■

The equations for Gram–Schmidt orthogonalization tend to be ill-conditioned because of the subtractions, Eqs. (10.48) and (10.49). A technique for avoiding this difficulty using the polynomial recurrence relation is discussed by Hamming.[9]

In Example 10.3.1 we have specified an orthogonality interval $[-1, 1]$, a unit weighting function, and a set of functions x^n to be taken one at a time in increasing order. Given all these specifications, the Gram–Schmidt procedure is unique (to within a normalization factor and an overall sign, as discussed subsequently). Our resulting orthogonal set, the Legendre polynomials, P_0 up through P_n, form a complete set for the description of polynomials of order $\leq n$ over $[-1, 1]$. This concept of completeness is taken up in detail in Section 10.4. Expansions of functions in series of Legendre polynomials are found in Section 12.3.

Orthogonal Polynomials

Example 10.3.1 has been chosen strictly to illustrate the Gram–Schmidt procedure. Although it has the advantage of introducing the Legendre polynomials, the initial functions $u_n = x^n$ are not degenerate eigenfunctions and are not solutions of Legendre's equation. They are simply a set of functions that we have here rearranged to create an orthonormal set for the given interval and given weighting function. The fact that we obtained the Legendre polynomials is not quite black magic but a direct consequence of the choice of interval and weighting function. The use of $u_n(x) = x^n$ but with other choices of interval and

Table 10.3 Orthogonal Polynomials Generated by Gram–Schmidt Orthogonalization of $u_n(x) = x^n, n = 0, 1, 2, \ldots$

Polynomials	Interval	Weighting function $w(x)$	Standard normalization
Legendre	$-1 \leq x \leq 1$	1	$\int_{-1}^{1} [P_n(x)]^2\,dx = \dfrac{2}{2n+1}$
Shifted Legendre	$0 \leq x \leq 1$	1	$\int_{0}^{1} [P_n^*(x)]^2\,dx = \dfrac{1}{2n+1}$
Chebyshev I	$-1 \leq x \leq 1$	$(1-x^2)^{-1/2}$	$\int_{-1}^{1} \dfrac{[T_n(x)]^2}{(1-x^2)^{1/2}}\,dx = \begin{cases} \pi/2, & n \neq 0 \\ \pi, & n = 0 \end{cases}$
Shifted Chebyshev I	$0 \leq x \leq 1$	$[x(1-x)]^{-1/2}$	$\int_{0}^{1} \dfrac{[T_n^*(x)]^2}{[x(1-x)]^{1/2}}dx = \begin{cases} \pi/2, & n > 0 \\ \pi, & n = 0 \end{cases}$
Chebyshev II	$-1 \leq x \leq 1$	$(1-x^2)^{1/2}$	$\int_{-1}^{1} [U_n(x)]^2(1-x^2)^{1/2}\,dx = \dfrac{\pi}{2}$
Laguerre	$0 \leq x < \infty$	e^{-x}	$\int_{0}^{\infty} [L_n(x)]^2 e^{-x}\,dx = 1$
Associated Laguerre	$0 \leq x < \infty$	$x^k e^{-x}$	$\int_{0}^{\infty} [L_n^k(x)]^2 x^k e^{-x}\,dx = \dfrac{(n+k)!}{n!}$
Hermite	$-\infty < x < \infty$	e^{-x^2}	$\int_{-\infty}^{\infty} [H_n(x)]^2 e^{-x^2}\,dx = 2^n \pi^{1/2} n!$

[9] R. W. Hamming, *Numerical Methods for Scientists and Engineers*, 2nd ed., New York: McGraw-Hill (1973). See Section 27.2 and references given there.

weighting function leads to other sets of orthogonal polynomials, as shown in Table 10.3. We consider these polynomials in detail in Chapters 12 and 13 as solutions of particular differential equations.

An examination of this orthogonalization process will reveal two arbitrary features. First, as emphasized before, it is not necessary to normalize the functions to unity. In the example just given we could have required

$$\int_{-1}^{1} \varphi_n(x)\varphi_m(x)\,dx = \frac{2}{2n+1}\delta_{nm}, \tag{10.61}$$

and the resulting set would have been the actual Legendre polynomials. Second, the sign of φ_n is always indeterminate. In the example we chose the sign by requiring the coefficient of the highest power of x in the polynomial to be positive. For the Laguerre polynomials, on the other hand, we would require the coefficient of the highest power to be $(-1)^n/n!$

Exercises

10.3.1 Rework Example 10.3.1 by replacing $\varphi_n(x)$ by the conventional Legendre polynomial, $P_n(x)$:

$$\int_{-1}^{1} \big[P_n(x)\big]^2\,dx = \frac{2}{2n+1}.$$

Using Eqs. (10.47a), and (10.49a), construct P_0, $P_1(x)$, and $P_2(x)$.

ANS. $P_0 = 1, \quad P_1 = x, \quad P_2 = \frac{3}{2}x^2 - \frac{1}{2}.$

10.3.2 Following the Gram–Schmidt procedure, construct a set of polynomials $P_n^*(x)$ orthogonal (unit weighting factor) over the range $[0, 1]$ from the set $[1, x]$. Normalize so that $P_n^*(1) = 1$.

ANS. $P_n^*(x) = 1,$
$P_1^*(x) = 2x - 1,$
$P_2^*(x) = 6x^2 - 6x + 1,$
$P_3^*(x) = 20x^3 - 30x^2 + 12x - 1.$

These are the first four **shifted** Legendre polynomials.
Note. The "*" is the standard notation for "shifted": $[0, 1]$ instead of $[-1, 1]$. It does **not** mean complex conjugate.

10.3.3 Apply the Gram–Schmidt procedure to form the first three Laguerre polynomials

$$u_n(x) = x^n, \qquad n = 0, 1, 2, \ldots, \qquad 0 \le x < \infty, \qquad w(x) = e^{-x}.$$

The conventional normalization is

$$\int_0^\infty L_m(x)L_n(x)e^{-x}\,dx = \delta_{mn}.$$

ANS. $L_0 = 1, \quad L_1 = (1 - x), \quad L_2 = \dfrac{2 - 4x + x^2}{2}.$

10.3.4 You are given

(a) a set of functions $u_n(x) = x^n, n = 0, 1, 2, \ldots,$
(b) an interval $(0, \infty)$,
(c) a weighting function $w(x) = xe^{-x}$. Use the Gram–Schmidt procedure to construct the first **three orthonormal** functions from the set $u_n(x)$ for this interval and this weighting function.

ANS. $\varphi_0(x) = 1, \quad \varphi_1(x) = (x-2)/\sqrt{2}, \quad \varphi_2(x) = \left(x^2 - 6x + 6\right)/2\sqrt{3}.$

10.3.5 Using the Gram–Schmidt orthogonalization procedure, construct the lowest three Hermite polynomials:

$$u_n(x) = x^n, \qquad n = 0, 1, 2, \ldots, \qquad -\infty < x < \infty, \qquad w(x) = e^{-x^2}.$$

For this set of polynomials the usual normalization is

$$\int_{-\infty}^{\infty} H_m(x) H_n(x) w(x)\, dx = \delta_{mn} 2^m m! \pi^{1/2}.$$

ANS. $H_0 = 1, \quad H_1 = 2x, \quad H_2 = 4x^2 - 2.$

10.3.6 Use the Gram–Schmidt orthogonalization scheme to construct the first three Chebyshev polynomials (type I):

$$u_n(x) = x^n, \qquad n = 0, 1, 2, \ldots, \qquad -1 \le x \le 1, \qquad w(x) = \left(1 - x^2\right)^{-1/2}.$$

Take the normalization

$$\int_{-1}^{1} T_m(x) T_n(x) w(x)\, dx = \delta_{mn} \begin{cases} \pi, & m = n = 0, \\ \dfrac{\pi}{2}, & m = n \ge 1. \end{cases}$$

Hint. The needed integrals are given in Exercise 8.4.3.

ANS. $T_0 = 1, \quad T_1 = x, \quad T_2 = 2x^2 - 1 \quad (T_3 = 4x^3 - 3x).$

10.3.7 Use the Gram–Schmidt orthogonalization scheme to construct the first three Chebyshev polynomials (type II):

$$u_n(x) = x^n, \qquad n = 0, 1, 2, \ldots, \qquad -1 \le x \le 1, \qquad w(x) = \left(1 - x^2\right)^{+1/2}.$$

Take the normalization to be

$$\int_{-1}^{1} U_m(x) U_n(x) w(x)\, dx = \delta_{mn} \frac{\pi}{2}.$$

Hint.

$$\int_{-1}^{1} \left(1 - x^2\right)^{1/2} x^{2n}\, dx = \frac{\pi}{2} \times \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{4 \cdot 6 \cdot 8 \cdots (2n+2)}, \qquad n = 1, 2, 3, \ldots$$
$$= \frac{\pi}{2}, \qquad n = 0.$$

ANS. $U_0 = 1, \quad U_1 = 2x, \quad U_2 = 4x^2 - 1.$

10.3.8 As a modification of Exercise 10.3.5, apply the Gram–Schmidt orthogonalization procedure to the set $u_n(x) = x^n, n = 0, 1, 2, \ldots, 0 \le x < \infty$. Take $w(x)$ to be $\exp[-x^2]$. Find the first two nonvanishing polynomials. Normalize so that the coefficient of the highest power of x is unity. In Exercise 10.3.5 the interval $(-\infty, \infty)$ led to the Hermite polynomials. These are certainly not the Hermite polynomials.

ANS. $\varphi_0 = 1, \quad \varphi_1 = x - \pi^{-1/2}$.

10.3.9 Form an orthogonal set over the interval $0 \le x < \infty$, using $u_n(x) = e^{-nx}, n = 1, 2, 3, \ldots$. Take the weighting factor, $w(x)$, to be unity. These functions are solutions of $u_n'' - n^2 u_n = 0$, which is clearly already in Sturm–Liouville (self-adjoint) form. Why doesn't the Sturm–Liouville theory guarantee the orthogonality of these functions?

10.4 COMPLETENESS OF EIGENFUNCTIONS

The third important property of an Hermitian operator is that its eigenfunctions form a complete set. This completeness means that any well-behaved (at least piecewise continuous) function $F(x)$ can be approximated by a series

$$F(x) = \sum_{n=0}^{\infty} a_n \varphi_n(x) \tag{10.62}$$

to any desired degree of accuracy.[10] More precisely, the set $\varphi_n(x)$ is called **complete**[11] if the limit of the mean square error vanishes:

$$\lim_{m\to\infty} \int_a^b \left[F(x) - \sum_{n=0}^{m} a_n \varphi_n(x) \right]^2 w(x)\, dx = 0. \tag{10.63}$$

Technically, the integral here is a Lebesgue integral. We have not required that the error vanish identically in $[a, b]$ but only that the integral of the error squared go to zero.

This convergence in the mean, Eq. (10.63), should be compared with uniform convergence (Section 5.5, Eq. (5.67)). Clearly, uniform convergence implies convergence in the mean, but the converse does not hold; convergence in the mean is less restrictive. Specifically, Eq. (10.63) is not upset by piecewise continuous functions with only a finite number of finite discontinuities. A relevant example is the Gibbs phenomenon of discontinuous Fourier series discussed in Section 14.5, which occurs for other eigenfunction series as well.

Equation (10.63) is perfectly adequate for our purposes and is far more convenient than Eq. (5.67). Indeed, since we frequently use eigenfunctions to describe discontinuous functions, convergence in the mean is all we can expect.

In Eq. (10.62) the expansion coefficients a_m may be determined by

$$a_m = \int_a^b F(x) \varphi_m^*(x) w(x)\, dx. \tag{10.64}$$

[10] If we have a finite set, as with vectors, the summation is over the number of linearly independent members of the set.

[11] Many authors use the term **closed** here.

This follows from multiplying Eq. (10.62) by $\varphi_m^*(x)w(x)$ and integrating. From the orthogonality of the eigenfunctions $\varphi_n(x)$, only the mth term survives. Here we see the value of orthogonality. Equation (10.64) may be compared with the dot or inner product of vectors, Section 1.3, and a_m interpreted as the mth projection of the function $F(x)$. Often the coefficient a_m is called a **generalized Fourier coefficient**.

For a known function $F(x)$, Eq. (10.64) gives a_m as a **definite** integral that can always be evaluated, by computer if not analytically.

In the language of linear algebra, we have a linear space, a function vector space. The linearly independent, orthonormal functions $\varphi_n(x)$ form the basis for this (infinite-dimensional) space. Equation (10.62) is a statement that the functions $\varphi_n(x)$ span this linear space. With an inner product defined by Eq. (10.64), our linear space is a **Hilbert** space.

Setting the weight function $w(x) = 1$ for simplicity, completeness in operator form for a discrete set of eigenfunctions $|\varphi_i\rangle$ becomes

$$\boxed{\sum_i |\varphi_i\rangle\langle\varphi_i| = 1.}$$

Multiplying the completeness relation by $|F\rangle$ we obtain the eigenfunction expansion

$$\boxed{|F\rangle = \sum_i |\varphi_i\rangle\langle\varphi_i|F\rangle}$$

with the generalized Fourier coefficient $a_i = \langle\varphi_i|F\rangle$. Equivalently in coordinate representation,

$$\boxed{\sum_i \varphi_i^*(y)\varphi_i(x) = \delta(x-y)}$$

implies

$$F(x) = \int F(y)\delta(x-y)\,dy = \sum_i \varphi_i(x)\int \varphi_i^*(y)F(y)\,dy.$$

Without proof, we state that the spectrum of a linear operator A that maps a Hilbert space H into itself may be divided into a discrete (or point) spectrum with eigenvectors of finite length, a continuous spectrum so that the eigenvalue equation $\mathsf{A}v = \lambda v$ with v in H does not have a unique bounded inverse $(\mathsf{A}-\lambda)^{-1}$ in a dense domain of H and a residual spectrum where $(\mathsf{A}-\lambda)^{-1}$ is unbounded in a domain not dense in H.

The question of completeness of a set of functions is often determined by comparison with a Laurent series, Section 6.5. In Section 14.1 this is done for Fourier series, thus establishing the completeness of Fourier series. For all orthogonal polynomials mentioned in Section 10.3 it is possible to find a polynomial expansion of each power of z,

$$z^n = \sum_{i=0}^{n} a_i P_i(z), \tag{10.65}$$

where $P_i(z)$ is the ith polynomial. Exercises 12.4.6, 13.1.6, 13.2.5, and 13.3.22 are specific examples of Eq. (10.65). Using Eq. (10.65), we may reexpress the Laurent expansion of $f(z)$ in terms of the polynomials, showing that the polynomial expansion exists (when it exists, it is unique, Exercise 10.4.1). The limitation of this Laurent series development is that it requires the function to be analytic. Equations (10.62) and (10.63) are more general. $F(x)$ may be only piecewise continuous. Numerous examples of the representation of such piecewise continuous functions appear in Chapter 14 (Fourier series). A proof that our Sturm–Liouville eigenfunctions form complete sets appears in Courant and Hilbert.[12]

For examples of particular eigenfunction expansions, see the following: Fourier series, Section 10.2 and Chapter 14; Bessel and Fourier–Bessel expansions, Section 11.2; Legendre series, Section 12.3; Laplace series, Section 12.6; Hermite series, Section 13.1; Laguerre series, Section 13.2; and Chebyshev series, Section 13.3.

It may also happen that the eigenfunction expansion, Eq. (10.62), is the expansion of an unknown $F(x)$ in a series of known eigenfunctions $\varphi_n(x)$ with unknown coefficients a_n. An example would be the quantum chemist's attempt to describe an (unknown) molecular wave function as a linear combination of known atomic wave functions. The unknown coefficients a_n would be determined by a variational technique — Rayleigh–Ritz, Section 17.8.

Bessel's inequality

If the set of functions $\varphi_n(x)$ does not form a complete set, possibly because we simply have not included the required infinite number of members of an infinite set, we are led to Bessel's inequality. First, consider the finite case from vector analysis. Let $\mathbf{A}$ be an n component vector,

$$\mathbf{A} = \mathbf{e}_1 a_1 + \mathbf{e}_2 a_2 + \cdots + \mathbf{e}_n a_n, \tag{10.66}$$

in which $\mathbf{e}_i$ is a unit vector and a_i is the corresponding component (projection) of $\mathbf{A}$; that is,

$$a_i = \mathbf{A} \cdot \mathbf{e}_i. \tag{10.67}$$

Then

$$\left(\mathbf{A} - \sum_i \mathbf{e}_i a_i\right)^2 \geq 0. \tag{10.68}$$

If we sum over all n components, the summation clearly, equals $\mathbf{A}$ by Eq. (10.66) and the equality holds. If, however, the summation does not include all n components, the inequality results. By expanding Eq. (10.68) and choosing the unit vectors so as to satisfy an orthogonality relation,

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}, \tag{10.69}$$

[12] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (English translation), Vol. 1, New York: Interscience (1953), reprinted, Wiley (1989), Chapter 6, Section 3.

we have

$$\mathbf{A}^2 \geq \sum_i a_i^2. \tag{10.70}$$

This is Bessel's inequality.

For real functions we consider the integral

$$\int_a^b \left[f(x) - \sum_i a_i \varphi_i(x) \right]^2 w(x)\, dx \geq 0. \tag{10.71}$$

This is the continuum analog of Eq. (10.68), letting $n \to \infty$ and replacing the summation by an integration. Again, with the weighting factor $w(x) > 0$, the integrand is nonnegative. The integral vanishes by Eq. (10.62) if we have a complete set. Otherwise it is positive. Expanding the squared term, we obtain

$$\int_a^b \left[f(x)\right]^2 w(x)\, dx - 2\sum_i a_i \int_a^b f(x)\varphi_i(x) w(x)\, dx + \sum_i a_i^2 \geq 0. \tag{10.72}$$

Applying Eq. (10.64), we have

$$\int_a^b \left[f(x)\right]^2 w(x)\, dx \geq \sum_i a_i^2. \tag{10.73}$$

Hence the sum of the squares of the expansion coefficients a_i is less than or equal to the weighted integral of $[f(x)]^2$, the equality holding if and only if the expansion is exact, that is, if the set of functions $\varphi_n(x)$ is a complete set.

In later chapters, when we consider eigenfunctions that form complete sets (such as Legendre polynomials), Eq. (10.73) with the equal sign holding will be called a **Parseval relation**.

Bessel's inequality has a variety of uses, including proof of convergence of the Fourier series.

Schwarz Inequality

The frequently used Schwarz inequality is similar to the Bessel inequality. Consider the quadratic equation with unknown x:

$$\sum_{i=1}^n (a_i x + b_i)^2 = \sum_{i=1}^n a_i^2 \left(x + \frac{b_i}{a_i} \right)^2 = 0 \tag{10.74}$$

with real a_i, b_i. If $b_i/a_i =$ constant, c, that is, independent of the index i, then the solution is $x = -c$. If b_i/a_i is not a constant in i, all terms cannot vanish simultaneously for real x. So the solution must be complex. Expanding, we find that

$$x^2 \sum_i^n a_i^2 + 2x \sum_i^n a_i b_i + \sum_i^n b_i^2 = 0, \tag{10.75}$$

and since x is complex (or $= -b_i/a_i$), the quadratic formula[13] for x leads to

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \le \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right), \tag{10.76}$$

the equality holding when b_i/a_i equals a constant, independent of i.

Once more, in terms of vectors, we have

$$(\mathbf{a}\cdot\mathbf{b})^2 \le a^2 b^2 \cos^2\theta \le a^2 b^2, \tag{10.77}$$

where θ is the angle included between $\mathbf{a}$ and $\mathbf{b}$.

The analogous Schwarz inequality for complex functions has the form

$$\left|\int_a^b f^*(x)g(x)\,dx\right|^2 \le \int_a^b f^*(x)f(x)\,dx \int_a^b g^*(x)g(x)\,dx, \tag{10.78}$$

the equality holding if and only if $g(x) = \alpha f(x)$, α being a constant. To prove this function form of the Schwarz inequality,[14] consider a complex function $\psi(x) = f(x) + \lambda g(x)$ with λ a complex constant, where $f(x)$ and $g(x)$ are any two square integrable functions (for which the integrals on the right-hand side exist). Multiplying by the complex conjugate and integrating, we obtain

$$\int_a^b \psi^*\psi\,dx \equiv \int_a^b f^* f\,dx + \lambda\int_a^b f^* g\,dx + \lambda^*\int_a^b g^* f\,dx + \lambda\lambda^*\int_a^b g^* g\,dx \ge 0. \tag{10.79}$$

The ≥ 0 appears since $\psi^*\psi$ is nonnegative, the equal $(=)$ sign holding only if $\psi(x)$ is identically zero. Noting that λ and λ^* are linearly independent, we differentiate with respect to one of them and set the derivative equal to zero to minimize $\int_a^b \psi^*\psi\,dx$:

$$\frac{\partial}{\partial\lambda^*}\int_a^b \psi^*\psi\,dx = \int_a^b g^* f\,dx + \lambda\int_a^b g^* g\,dx = 0.$$

This yields

$$\lambda = -\frac{\int_a^b g^* f\,dx}{\int_a^b g^* g\,dx}. \tag{10.80a}$$

Taking the complex conjugate, we obtain

$$\lambda^* = -\frac{\int_a^b f^* g\,dx}{\int_a^b g^* g\,dx}. \tag{10.80b}$$

Substituting these values of λ and λ^* back into Eq. (10.79), we obtain Eq. (10.78), the Schwarz inequality.

[13] With negative (or zero) discriminant.

[14] An alternate derivation is provided by the inequality $\iint [f(x)g(y) - f(y)g(x)]^*[f(x)g(y) - f(y)g(x)]\,dx\,dy \ge 0$.

In quantum mechanics $f(x)$ and $g(x)$ might each represent a state or configuration of a physical system, that is, a linear combination of wave functions. Then the Schwarz inequality gives an upper limit for the absolute value of the inner product $\int_a^b f^*(x)g(x)\,dx$. In some texts the Schwarz inequality is a key step in the derivation of the Heisenberg uncertainty principle.

The function notation of Eqs. (10.78) and (10.79) is relatively cumbersome. In advanced mathematical physics and especially in quantum mechanics it is common to use the Dirac bra-ket notation. Using this notation, we simply understand the range of integration, (a, b), and the presence of the weighting function $w(x) \geq 0$. In this notation the Schwarz inequality takes the elegant form

$$\left|\langle f|g\rangle\right|^2 \leq \langle f|f\rangle\langle g|g\rangle. \tag{78a}$$

If $g(x)$ is a normalized eigenfunction, $\varphi_i(x)$, Eq. (10.78) yields (here $w(x) = 1$)

$$a_i^* a_i \leq \int_a^b f^*(x) f(x)\,dx, \tag{10.81}$$

a result that also follows from Eq. (10.73).

For useful representations of Dirac's delta function in terms of orthogonal sets of functions and the relation between closure and completeness we refer to the relevant subsection of Section 1.15, including Exercise 1.15.16, and for coordinate versus momentum representations in quantum mechanics to Section 15.6.

Summary — Vector Spaces, Completeness

Here we summarize some properties of vector spaces, first with the vectors taken to be the familiar real vectors of Chapter 1 and then with the vectors taken to be ordinary functions. The concept of **completeness** has been developed for finite vector spaces (Chapter 1, Eq. (1.5)) and carries over into infinite vector spaces. For example, in three-dimensional Euclidean space every vector can be written in terms of a linear combination of the three coordinate unit vectors (representing a basis) involving the vector's Cartesian components as the expansion coefficients. Or a periodic function of an infinite vector space can be expanded in terms of the set of periodic functions $\sin nx, \cos nx, n = 0, 1, 2, \ldots$, that form a basis of this space. Since any periodic function with reasonable properties (spelled out in Chapter 14) can be expanded in terms of these sine and cosine functions, they are complete and form a basis of such a linear function space.

1v. We shall describe our vector space with a set of n linearly independent vectors $\mathbf{e}_i$, $i = 1, 2, \ldots, n$. If $n = 3$, then $\mathbf{e}_1 = \hat{\mathbf{x}}$, $\mathbf{e}_2 = \hat{\mathbf{y}}$, and $\mathbf{e}_3 = \hat{\mathbf{z}}$. The $n\mathbf{e}_i$ **span** the linear vector space.

1f. We shall describe our vector (function) space with a set of n linearly independent functions, $\varphi_i(x), i = 0, 1, \ldots, n-1$. The index i starts with 0 to agree with the labeling of the classical polynomials. Here $\varphi_i(x)$ is assumed to be a polynomial of degree i. The $n\varphi_i(x)$ span the linear vector (function) space.

2v. The vectors in our vector space satisfy the following relations (Section 1.2; the vector components are numbers):

a.	Vector addition is commutative	$\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$
b.	Vector addition is associative	$[\mathbf{u}+\mathbf{v}]+\mathbf{w}=\mathbf{u}+[\mathbf{v}+\mathbf{w}]$
c.	There is a null vector	$\mathbf{0}+\mathbf{v}=\mathbf{v}$
d.	Multiplication by a scalar	
	Distributive	$a[\mathbf{u}+\mathbf{v}]=a\mathbf{u}+a\mathbf{v}$
	Distributive	$(a+b)\mathbf{u}=a\mathbf{u}+b\mathbf{u}$
	Associative	$a[b\mathbf{u}]=(ab)\mathbf{u}$
e.	Multiplication	
	By unit scalar	$1\mathbf{u}=\mathbf{u}$
	By zero	$0\mathbf{u}=0$
f.	Negative vector	$(-1)\mathbf{u}=-\mathbf{u}$.

2f. The functions in our linear function space satisfy the properties listed for vectors (substitute "function" for "vector"):

$$
\begin{aligned}
f(x)+g(x)&=g(x)+f(x)\\
\big[f(x)+g(x)\big]+h(x)&=f(x)+\big[g(x)+h(x)\big]\\
0+f(x)&=f(x)\\
a\big[f(x)+g(x)\big]&=af(x)+ag(x)\\
(a+b)f(x)&=af(x)+bf(x)\\
a\big[bf(x)\big]&=(ab)f(x)\\
1\cdot f(x)&=f(x)\\
0\cdot f(x)&=0\\
(-1)\cdot f(x)&=-f(x).
\end{aligned}
$$

3v. In n-dimensional vector space an arbitrary vector $\mathbf{c}$ is described by its n components $(c_1, c_2, \ldots, c_n)$, or

$$\mathbf{c}=\sum_{i=1}^{n} c_i\mathbf{e}_i.$$

When $n\mathbf{e}_i$ (1) are linearly independent and (2) span the n-dimensional vector space, then the $\mathbf{e}_i$ form a basis and constitute a **complete** set.

3f. In n-dimensional function space a polynomial of degree $m \le n-1$ is described by

$$f(x)=\sum_{i=0}^{n-1} c_i\varphi_i(x).$$

When the $n\varphi_i(x)$ (1) are linearly independent and (2) span the n-dimensional function space, then the $\varphi_i(x)$ form a basis and constitute a **complete** set (for describing polynomials of degree $m \le n-1$).

4v. An inner product (scalar, dot product) of a vector space is defined by

$$\mathbf{c}\cdot\mathbf{d}=\sum_{i=1}^{n} c_i d_i.$$

If **c** and **d** have complex components in an orthogonal coordinate system, the inner product is defined as $\sum_{i=1}^{n} c_i^* d_i$. The inner product has the properties of

a. Distributive law of addition $\quad \mathbf{c}\cdot(\mathbf{d}+\mathbf{e}) = \mathbf{c}\cdot\mathbf{d} + \mathbf{c}\cdot\mathbf{e}$
b. Scalar multiplication $\quad \mathbf{c}\cdot a\mathbf{d} = a\mathbf{c}\cdot\mathbf{d}$
c. Complex conjugation $\quad \mathbf{c}\cdot\mathbf{d} = (\mathbf{d}\cdot\mathbf{c})^*$.

4f. An inner product of a linear space of functions is defined by

$$\langle f|g\rangle = \int_a^b f^*(x)g(x)w(x)\,dx.$$

The choice of the weighting function $w(x)$ and the interval (a, b) follows from the differential equation satisfied by $\varphi_i(x)$ and the boundary conditions — Section 10.1. In matrix terminology, Section 3.2, $|g\rangle$ is a column vector and $\langle f|$ is a row vector, the adjoint of $|f\rangle$, where both may have infinitely many components. For example, if we expand $g(x) = \sum_i g_i\varphi_i(x)$, then $|g\rangle$ has the ith component g_i in a column vector and $|f\rangle$ has f_i^* as its ith component in a row vector.

The inner product has the properties listed for vectors:

a. $\langle f|g+h\rangle = \langle f|g\rangle + \langle f|h\rangle$
b. $\langle f|ag\rangle = a\langle f|g\rangle$
c. $\langle f|g\rangle = \langle g|f\rangle^*$.

5v. Orthogonality:

$$\mathbf{e}_j\cdot\mathbf{e}_j = 0, \qquad i \neq j.$$

If the $n\mathbf{e}_i$ are not already orthogonal, the Gram–Schmidt process may be used to create an orthogonal set.

5f. Orthogonality:

$$\langle\varphi_i|\varphi_j\rangle = \int_a^b \varphi_i^*(x)\varphi_j(x)w(x)\,dx = 0, \qquad i \neq j.$$

If the $n\varphi_i(x)$ are not already orthogonal, the Gram–Schmidt process (Section 10.3) may be used to create an orthogonal set.

6v. Definition of **norm:**

$$|\mathbf{c}| = (\mathbf{c}\cdot\mathbf{c})^{1/2} = \left(\sum_{i=1}^{n} c_i^2\right)^{1/2}.$$

The basis vectors $\mathbf{e}_i$ are taken to have unit norm (length) $\mathbf{e}_i\cdot\mathbf{e}_i = 1$. The components of **c** are given by

$$c_i = \mathbf{e}_i\cdot\mathbf{c}, \qquad i = 1, 2, \ldots, n.$$

6f. Definition of **norm:**

$$\|f\| = \langle f|f\rangle^{1/2} = \left[\int_a^b |f(x)|^2 w(x)\,dx\right]^{1/2} = \left[\sum_{i=0}^{n-1} |c_i|^2\right]^{1/2},$$

Parseval's identity. $\|f\| > 0$ unless $f(x)$ is identically zero. The basis functions $\varphi_i(x)$ may be taken to have unit norm (unit normalization),

$$\|\varphi_i\| = 1.$$

The expansion coefficients of our polynomial $f(x)$ are given by

$$c_i = \langle \varphi_i | f \rangle, \qquad i = 0, 1, \ldots, n-1.$$

7v. Bessel's inequality:

$$\mathbf{c} \cdot \mathbf{c} \geq \sum_i c_i^2.$$

If the equals sign holds for all $\mathbf{c}$, it indicates that the $\mathbf{e}_i$ span the vector space; that is, they are complete.

7f. Bessel's inequality:

$$\langle f | f \rangle = \int_a^b |f(x)|^2 w(x)\, dx \geq \sum_i |c_i|^2.$$

If the equals sign holds for all allowable f, it indicates that the $\varphi_i(x)$ span the function space; that is, they are complete.

8v. Schwarz' inequality:

$$|\mathbf{c} \cdot \mathbf{d}| \leq |\mathbf{c}| \cdot |\mathbf{d}|.$$

The equals sign holds when $\mathbf{c}$ is a multiple of $\mathbf{d}$. If the angle included between $\mathbf{c}$ and $\mathbf{d}$ is θ, then $|\cos\theta| \leq 1$.

8f. Schwarz' inequality:

$$|\langle f | g \rangle| \leq \langle f | f \rangle^{1/2} \langle g | g \rangle^{1/2} = \|f\| \cdot \|g\|.$$

The equals sign holds when $f(x)$ and $g(x)$ are linearly dependent, that is, when $f(x)$ is a multiple of $g(x)$.

Now, let $n \to \infty$, forming an infinite-dimensional linear vector space, l^2.

9v. In an infinite-dimensional space our vector $\mathbf{c}$ is

$$\mathbf{c} = \sum_{i=1}^{\infty} c_i \mathbf{e}_i.$$

We require that

$$\sum_{i=1}^{\infty} c_i^2 < \infty.$$

The components of $\mathbf{c}$ are given by

$$c_i = \mathbf{e}_i \cdot \mathbf{c}, \qquad i = 1, 2, \ldots, \infty,$$

exactly as in a finite-dimensional vector space.

Then let $n \to \infty$, forming an infinite-dimensional vector (function) space L^2. Then L stands for Lebesgue, the superscript 2 for the quadratic norm, that is, the 2 in $|f(x)|^2$. Our functions need no longer be polynomials, but we do require that $f(x)$ be at least piecewise

continuous (Dirichlet conditions for Fourier series) and that $\langle f|f\rangle = \int_a^b |f(x)|^2 w(x)\,dx$ exist. This latter condition is often stated as a requirement that $f(x)$ be square integrable.

9f. Cauchy sequence (generalized Fourier expansion): Expand $f(x) = \sum_{i=0}^{\infty} f_i\varphi_i(x)$ and let

$$f_n(x) = \sum_{i=0}^{n} f_i\varphi_i(x).$$

If

$$\|f(x) - f_n(x)\| \to 0 \quad \text{as } n \to \infty$$

or

$$\lim_{n\to\infty} \int \left| f(x) - \sum_{i=0}^{n} f_i\varphi_i(x) \right|^2 w(x)\,dx = 0,$$

then we have convergence in the mean. This is analogous to the partial sum–Cauchy sequence criterion for the convergence of an infinite series, Section 5.1.

If every Cauchy sequence of allowable vectors (square integrable, piecewise continuous functions) converges to a limit vector in our linear space, the space is said to be complete. Then

$$f(x) = \sum_{i=0}^{\infty} c_i\varphi_i(x) \qquad \text{(almost everywhere)}$$

in the sense of convergence in the mean. As noted before, this is a weaker requirement than pointwise convergence (fixed value of x) or uniform convergence.

Expansion Coefficients

For a function f its expansion coefficients are defined as

$$c_i = \langle\varphi_i|f\rangle, \qquad i = 0, 1, \ldots, \infty,$$

exactly as in a finite-dimensional vector space. Hence

$$f(x) = \sum_i \langle\varphi_i|f\rangle\varphi_i(x).$$

A linear space (finite- or infinite-dimensional) that (1) has an inner product defined ($\langle f|g\rangle$) and (2) is complete is a **Hilbert space**.

Infinite-dimensional Hilbert space provides a natural mathematical frame-work for modern quantum mechanics. Away from quantum mechanics, Hilbert space retains its abstract mathematical power and beauty and has many uses.

Exercises

10.4.1 A function $f(x)$ is expanded in a series of orthonormal eigenfunctions

$$f(x) = \sum_{n=0}^{\infty} a_n \varphi_n(x).$$

Show that the series expansion is unique for a given set of $\varphi_n(x)$. The functions $\varphi_n(x)$ are being taken here as the **basis** vectors in an infinite-dimensional Hilbert space.

10.4.2 A function $f(x)$ is represented by a finite set of basis functions $\varphi_i(x)$,

$$f(x) = \sum_{i=1}^{N} c_i \varphi_i(x).$$

Show that the components c_i are unique, that no different set c_i' exists.
Note. Your basis functions are automatically linearly independent. They are not necessarily orthogonal.

10.4.3 A function $f(x)$ is approximated by a power series $\sum_{i=0}^{n-1} c_i x^i$ over the interval $[0, 1]$. Show that minimizing the mean square error leads to a set of linear equations

$$\mathsf{A}\mathbf{c} = \mathbf{b},$$

where

$$A_{ij} = \int_0^1 x^{i+j}\, dx = \frac{1}{i+j+1}, \qquad i, j = 0, 1, 2, \ldots, n-1$$

and

$$b_i = \int_0^1 x^i f(x)\, dx, \qquad i = 0, 1, 2, \ldots, n-1.$$

Note. The A_{ij} are the elements of the Hilbert matrix of order n. The determinant of this Hilbert matrix is a rapidly decreasing function of n. For $n = 5$, $\det \mathsf{A} = 3.7 \times 10^{-12}$ and the set of equations $\mathsf{A}c = \mathbf{b}$ is becoming ill-conditioned and unstable.

10.4.4 In place of the expansion of a function $F(x)$ given by

$$F(x) = \sum_{n=0}^{\infty} a_n \varphi_n(x),$$

with

$$a_n = \int_a^b F(x)\varphi_n(x)w(x)\, dx,$$

take the **finite** series approximation

$$F(x) \approx \sum_{n=0}^{m} c_n \varphi_n(x).$$

Show that the mean square error

$$\int_a^b \left[F(x) - \sum_{n=0}^{m} c_n \varphi_n(x)\right]^2 w(x)\,dx$$

is minimized by taking $c_n = a_n$.
Note. The values of the coefficients are independent of the number of terms in the finite series. This independence is a consequence of orthogonality and would not hold for a least-squares fit using powers of x.

10.4.5 From Example 10.2.2,

$$f(x) = \left\{ \begin{array}{ll} \dfrac{h}{2}, & 0 < x < \pi \\ -\dfrac{h}{2}, & -\pi < x < 0 \end{array} \right\} = \frac{2h}{\pi} \sum_{n=0}^{\infty} \frac{\sin(2n+1)x}{2n+1}.$$

(a) Show that

$$\int_{-\pi}^{\pi} \big[f(x)\big]^2\,dx = \frac{\pi}{2}h^2 = \frac{4h^2}{\pi} \sum_{n=0}^{\infty} (2n+1)^{-2}.$$

For a finite upper limit this would be Bessel's inequality. For the upper limit ∞, this is Parseval's identity.

(b) Verify that

$$\frac{\pi}{2}h^2 = \frac{4h^2}{\pi} \sum_{n=0}^{\infty} (2n+1)^{-2}$$

by evaluating the series.

Hint. The series can be expressed as the Riemann zeta function.

10.4.6 Differentiate Eq. (10.79),

$$\langle \psi | \psi \rangle = \langle f | f \rangle + \lambda \langle f | g \rangle + \lambda^* \langle g | f \rangle + \lambda \lambda^* \langle g | g \rangle,$$

with respect to λ^* and show that you get the Schwarz inequality, Eq. (10.78).

10.4.7 Derive the Schwarz inequality from the identity

$$\left[\int_a^b f(x)g(x)\,dx\right]^2 = \int_a^b \big[f(x)\big]^2 dx \int_a^b \big[g(x)\big]^2 dx - \frac{1}{2}\int_a^b \int_a^b \big[f(x)g(y) - f(y)g(x)\big]^2 dx\,dy.$$

10.4.8 If the functions $f(x)$ and $g(x)$ of the Schwarz inequality, Eq. (10.78), may be expanded in a series of eigenfunctions $\varphi_i(x)$, show that Eq. (10.78) reduces to Eq. (10.76) (with n possibly infinite).

Note the description of $f(x)$ as a vector in a function space in which $\varphi_i(x)$ corresponds to the unit vector $\mathbf{e}_1$.

10.4.9 The operator H is Hermitian and positive definite; that is, for all f:

$$\int_a^b f^* H f \, dx > 0.$$

10.4.10 A normalized wave function $\psi(x) = \sum_{n=0}^{\infty} a_n \varphi_n(x)$. The expansion coefficients a_n are known as probability amplitudes. We may define a density matrix ρ with elements $\rho_{ij} = a_i a_j^*$. Show that

$$\left(\rho^2\right)_{ij} = \rho_{ij},$$

or

$$\rho^2 = \rho.$$

This result, by definition, makes ρ a projection operator.
Hint: Use

$$\int \psi^* \psi \, dx = 1.$$

10.4.11 Show that

(a) the operator

$$|\varphi_i(x)\rangle\langle\varphi_i(t)|$$

operating on

$$f(t) = \sum_j c_j |\varphi_j(t)\rangle$$

yields

$$|\varphi_i(x)\rangle\langle\varphi_i(x)|\varphi_i(x)\rangle = 1.$$

This operator is a **projection operator** projecting $f(x)$ onto the ith coordinate, selectively picking out the ith component $c_i|\varphi_i(x)\rangle$ of $f(x)$.

(b) $\sum_i |\varphi_i(x)\rangle\langle\varphi_i(x)| = 1$.

Hint. The operator operates via the well-defined inner product.

10.5 GREEN'S FUNCTION — EIGENFUNCTION EXPANSION

A series somewhat similar to that representing $\delta(x-t)$ results when we expand the Green's function in the eigenfunctions of the corresponding homogeneous equation. In the inhomogeneous Helmholtz equation we have

$$\boxed{\nabla^2\psi(\mathbf{r})+k^2\psi(\mathbf{r})=-\rho(\mathbf{r}).} \tag{10.82}$$

The homogeneous Helmholtz equation is satisfied by its orthonormal eigenfunctions φ_n,

$$\nabla^2\varphi_n(\mathbf{r})+k_n^2\varphi_n(\mathbf{r})=0. \tag{10.83}$$

As outlined in Section 9.7, the Green's function $G(\mathbf{r}_1,\mathbf{r}_2)$ satisfies the point source equation

$$\boxed{\nabla^2 G(\mathbf{r}_1,\mathbf{r}_2)+k^2G(\mathbf{r}_1,\mathbf{r}_2)=-\delta(\mathbf{r}_1-\mathbf{r}_2)} \tag{10.84}$$

and the boundary conditions imposed on the solutions of the homogeneous equation. Because G is real, we expand the Green's function in a series of real eigenfunctions of the homogeneous equation (10.83); that is,

$$G(\mathbf{r}_1,\mathbf{r}_2)=\sum_{n=0}^{\infty}a_n(\mathbf{r}_2)\varphi_n(\mathbf{r}_1), \tag{10.85}$$

and by substituting into Eq. (10.84) we obtain

$$-\sum_{n=0}^{\infty}a_n(\mathbf{r}_2)k_n^2\varphi_n(\mathbf{r}_1)+k^2\sum_{n=0}^{\infty}a_n(\mathbf{r}_2)\varphi_n(\mathbf{r}_1)=-\sum_{n=0}^{\infty}\varphi_n(\mathbf{r}_1)\varphi_n(\mathbf{r}_2). \tag{10.86}$$

Here $\delta(\mathbf{r}_1-\mathbf{r}_2)$ has been replaced by its eigenfunction expansion, Eq. (1.190). When we employ the orthogonality of $\varphi_n(\mathbf{r}_1)$ to isolate a_n, this yields

$$\sum_{m=0}^{\infty}a_m(\mathbf{r}_2)\left(k^2-k_m^2\right)\int\varphi_n(\mathbf{r}_1)\varphi_m(\mathbf{r}_1)\,d^3r_1=-\sum_{m=0}^{\infty}\varphi_m(\mathbf{r}_2)\int\varphi_n(\mathbf{r}_1)\varphi_m(\mathbf{r}_1)\,d^3r_1,$$

or

$$a_n(\mathbf{r}_2)\left(k^2-k_n^2\right)=-\varphi_n(\mathbf{r}_2).$$

Then substituting this into Eq. (10.85), the Green's function becomes

$$\boxed{G(\mathbf{r}_1,\mathbf{r}_2)=\sum_{n=0}^{\infty}\frac{\varphi_n(\mathbf{r}_1)\varphi_n(\mathbf{r}_2)}{k_n^2-k^2},} \tag{10.87}$$

a bilinear expansion, symmetric with respect to $\mathbf{r}_1$ and $\mathbf{r}_2$, as expected. Finally, $\psi(\mathbf{r}_1)$, the desired solution of the inhomogeneous equation, is given by

$$\boxed{\psi(\mathbf{r}_1)=\int G(\mathbf{r}_1,\mathbf{r}_2)\rho(\mathbf{r}_2)\,d\tau_2.} \tag{10.88}$$

If we generalize our inhomogeneous differential equation to

$$\mathcal{L}\psi + \lambda\psi = -\rho, \tag{10.89}$$

where $\mathcal{L}$ is a Hermitian operator, we find that

$$G(\mathbf{r}_1, \mathbf{r}_2) = \sum_{n=0}^{\infty} \frac{\varphi_n(\mathbf{r}_1)\varphi_n(\mathbf{r}_2)}{\lambda_n - \lambda}, \tag{10.90}$$

where λ_n is the nth eigenvalue and φ_n is the corresponding orthonormal eigenfunction of the homogeneous differential equation

$$\mathcal{L}\psi + \lambda\psi = 0. \tag{10.91}$$

The eigenfunction expansion of the Green's function in Eq. (10.90) makes the symmetry property $G(\mathbf{r}_1, \mathbf{r}_2) = G(\mathbf{r}_2, \mathbf{r}_1)$ explicit and is often useful when comparing with solutions obtained by other means.

Green's Functions — One-Dimensional

The development of the Green's function for two- and three-dimensional systems was the topic discussed in the preceding material and in Section 9.7. Here, for simplicity, we restrict ourselves to one-dimensional cases and follow a somewhat different approach.

Defining Properties

In our one-dimensional analysis we consider first the inhomogeneous equation

$$\mathcal{L}y(x) + f(x) = 0, \tag{10.92}$$

in which $\mathcal{L}$ is the **self-adjoint** differential operator

$$\mathcal{L} = \frac{d}{dx}\left(p(x)\frac{d}{dx}\right) + q(x). \tag{10.93}$$

As in Section 10.1, $y(x)$ is required to satisfy certain boundary conditions at the endpoints a and b of our interval $[a, b]$.

We now proceed to define a rather strange and arbitrary function G over the interval $[a, b]$. At this stage the most that can be said in defense of G is that the defining properties are legitimate, or mathematically acceptable. Later, G will appear as a reasonable tool for obtaining solutions of the inhomogeneous ODE, Eq. (10.92); this role dictates its properties.

1. The interval $a \le x \le b$ is divided by a parameter t. We label $G(x) = G_1(x)$ for $a \le x < t$ and $G(x) = G_2(x)$ for $t < x \le b$.
2. The functions $G_1(x)$ and $G_2(x)$ each satisfy the homogeneous[15] equation; that is,

$$\begin{aligned} \mathcal{L}G_1(x) &= 0, \qquad a \le x < t, \\ \mathcal{L}G_2(x) &= 0, \qquad t < x \le b. \end{aligned} \tag{10.94}$$

[15] Homogeneous with respect to the unknown function. The function $f(x)$ in Eq. (10.92) is set equal to zero.

3. At $x = a$, $G_1(x)$ satisfies the boundary conditions we impose on $y(x)$, a solution of the inhomogeneous ODE, Eq. (10.92). At $x = b$, $G_2(x)$ satisfies the boundary conditions imposed on $y(x)$ at this endpoint of the interval. For convenience, the boundary conditions are taken to be homogeneous; that is, at $x = a$,

$$y(a) = 0, \qquad \text{or} \qquad y'(a) = 0, \qquad \text{or} \qquad \alpha y(a) + \beta y'(a) = 0$$

and similarly at $x = b$.

4. We demand that $G(x)$ be **continuous**,[16]

$$\lim_{x \to t_-} G_1(x) = \lim_{x \to t_+} G_2(x). \tag{10.95}$$

5. We require that $G'(x)$ be **discontinuous**, specifically that[15]

$$\frac{d}{dx} G_2(x)\Big|_t - \frac{d}{dx} G_1(x)\Big|_t = -\frac{1}{p(t)}, \tag{10.96}$$

where $p(t)$ comes from the self-adjoint operator, Eq. (10.93). Note that with the first derivative discontinuous, the second derivative does not exist.

These requirements, in effect, make G a function of two variables, $G(x, t)$. Also, we note that $G(x, t)$ depends on both the form of the differential operator $\mathcal{L}$ and the boundary conditions that $y(x)$ must satisfy. Note that we have described the properties of Green's functions for second-order differential equations. Note that for Green's functions for first-order differential equations, the discontinuities arise in G itself.

Now, assuming that we can find a function $G(x, t)$ that has these properties, we label it a Green's function and proceed to show that a solution of Eq. (10.92) is

$$y(x) = \int_a^b G(x, t) f(t)\, dt. \tag{10.97}$$

To do this we first construct the Green's function $G(x, t)$. Let $u(x)$ be a solution of the homogeneous equation that satisfies the boundary conditions at $x = a$, and let $v(x)$ be a solution that satisfies the boundary conditions at $x = b$. Then we may take[17]

$$G(x, t) = \begin{cases} c_1 u(x), & a \le x < t, \\ c_2 v(x), & t < x \le b. \end{cases} \tag{10.98}$$

Continuity at $x = t$ (Eq. (10.95)) requires

$$c_2 v(t) - c_1 u(t) = 0. \tag{10.99}$$

Finally, the discontinuity in the first derivative (Eq. (10.96)) becomes

$$c_2 v'(t) - c_1 u'(t) = -\frac{1}{p(t)}. \tag{10.100}$$

[16] Strictly speaking, this is the limit as $x \to t$.

[17] The "constants" c_1 and c_2 are independent of x, but they may (and do) depend on the other variable, t.

There will be a unique solution for our unknown coefficients c_1 and c_2 if the Wronskian determinant

$$\begin{vmatrix} u(t) & v(t) \\ u'(t) & v'(t) \end{vmatrix} = u(t)v'(t) - v(t)u'(t)$$

does not vanish. We have seen in Section 9.6 that the nonvanishing of this determinant is a necessary condition for linear independence. Let us assume $u(x)$ and $v(x)$ to be independent. (If $u(x)$ and $v(x)$ are linearly dependent, the situation becomes more complicated and is not considered here. See Courant and Hilbert in Additional Readings of Chapter 9.) For independent $u(x)$ and $v(x)$ we have the Wronskian (again from Section 9.6 or Exercise 10.1.4)

$$u(t)v'(t) - v(t)u'(t) = \frac{A}{p(t)}, \tag{10.101}$$

in which A is a constant. Equation (10.101) is sometimes called *Abel's formula*. Numerous examples have appeared in connection with Bessel and Legendre functions. Now, from Eq. (10.100), we identify

$$c_1 = -\frac{v(t)}{A}, \qquad c_2 = -\frac{u(t)}{A}. \tag{10.102}$$

Equation (10.99) is clearly satisfied. Substitution into Eq. (10.98) yields our Green's function

$$G(x,t) = \begin{cases} -\dfrac{1}{A}u(x)v(t), & a \le x < t, \\ -\dfrac{1}{A}u(t)v(x), & t < x \le b. \end{cases} \tag{10.103}$$

Note that $G(x,t) = G(t,x)$. This is the symmetry property that was proved earlier in Section 9.7. Its physical interpretation is given by the reciprocity principle (via our propagator function)—a cause at t yields the same effect at x as a cause at x produces at t. In terms of our electrostatic analogy this is obvious, the propagator function depending only on the magnitude of the distance between the two points:

$$|\mathbf{r}_1 - \mathbf{r}_2| = |\mathbf{r}_2 - \mathbf{r}_1|.$$

Green's Function Integral — Differential Equation

We have constructed $G(x,t)$, but there still remains the task of showing that the integral (Eq. (10.97)) with our new Green's function is indeed a solution of the original differential equation (10.92). This we do by direct substitution. With $G(x,t)$ given by Eq. (10.103),[18] Eq. (10.97) becomes

$$y(x) = -\frac{1}{A}\int_a^x v(x)u(t)f(t)\,dt - \frac{1}{A}\int_x^b u(x)v(t)f(t)\,dt. \tag{10.104}$$

[18] In the first integral, $a \le t \le x$. Hence $G(x,t) = G_2(x,t) = -(1/A)u(t)v(x)$. Similarly, the second integral requires $G = G_1$.

Differentiating, we obtain

$$y'(x) = -\frac{1}{A}\int_a^x v'(x)u(t)f(t)\,dt - \frac{1}{A}\int_x^b u'(x)v(t)f(t)\,dt, \tag{10.105}$$

the derivatives of the limits canceling. A second differentiation yields

$$y''(x) = -\frac{1}{A}\int_a^x v''(x)u(t)f(t)\,dt - \frac{1}{A}\int_x^b u''(x)v(t)f(t)\,dt$$
$$-\frac{1}{A}\big[u(x)v'(x) - v(x)u'(x)\big]f(x). \tag{10.106}$$

By Eqs. (10.100) and (10.102) this may be rewritten as

$$y''(x) = -\frac{v''(x)}{A}\int_a^x u(t)f(t)\,dt - \frac{u''(x)}{A}\int_x^b v(t)f(t)\,dt - \frac{f(x)}{p(x)}. \tag{10.107}$$

Now, by substituting into Eq. (10.93), we have

$$\mathcal{L}y(x) = -\frac{\mathcal{L}v(x)}{A}\int_a^x u(t)f(t)\,dt - \frac{\mathcal{L}u(x)}{A}\int_x^b v(t)f(t)\,dt - f(x). \tag{10.108}$$

Since $u(x)$ and $v(x)$ were chosen to satisfy the homogeneous equation, the $\mathcal{L}$-factors are zero and the integral terms vanish, and we see that Eq. (10.92) is satisfied.

We must also check that $y(x)$ satisfies the required boundary conditions. At point $x = a$,

$$y(a) = -\frac{u(a)}{A}\int_a^b v(t)f(t)\,dt = cu(a), \tag{10.109}$$

$$y'(a) = -\frac{u'(a)}{A}\int_a^b v(t)f(t)\,dt = cu'(a), \tag{10.110}$$

since the definite integral is a constant. We chose $u(x)$ to satisfy

$$\alpha u(a) + \beta u'(a) = 0. \tag{10.111}$$

Multiplying by the constant c, we verify that $y(x)$ also satisfies Eq. (10.111). This illustrates the utility of the **homogeneous** boundary conditions: The normalization does not matter. In quantum mechanical problems the boundary condition on the wave function is often expressed in terms of the ratio

$$\frac{\psi'(x)}{\psi(x)} = \frac{d}{dx}\ln\psi(x), \qquad \text{compared to} \qquad \frac{d}{dx}\ln u(x)\big|_{x=a} = -\frac{\alpha}{\beta},$$

Eq. (10.111). The advantage is that the wave function need not be normalized yet.

Summarizing, we have Eq. (10.97),

$$y(x) = \int_a^b G(x,t)f(t)\,dt,$$

which satisfies the differential equation (Eq. (10.92)),

$$\mathcal{L}y(x) + f(x) = 0,$$

and the boundary conditions, these boundary conditions having been built into the Green's function, $G(x,t)$.

Basically, what we have done is to use the solutions of the homogeneous equation Eq. (10.94) to construct a solution of the inhomogeneous equation. Again, Poisson's equation is an illustration. The solution (Eq. (9.148)) represents a weighted $[\rho(\mathbf{r}_2)]$ combination of solutions of the corresponding homogeneous Laplace's equation. (We followed these same steps early in this section.)

It should be noted that our $y(x)$, Eq. (10.97), is actually the **particular** solution of the differential equation, Eq. (10.92). Our boundary conditions exclude the addition of solutions of the homogeneous equation. In an actual physical problem we may well have both types of solutions. In electrostatics, for instance (compare Section 9.7), the Green's function solution of Poisson's equation gives the potential created by the given charge distribution. In addition, there may be external fields superimposed. These would be described by solutions of the homogeneous equation, Laplace's equation.

Eigenfunction, Eigenvalue Equation

The preceding analysis placed no special restrictions on our $f(x)$. Let us now assume that $f(x) = \lambda\rho(x)y(x)$.[19] Then we have

$$y(x) = \lambda \int_a^b G(x,t)\rho(t)y(t)\,dt \tag{10.112}$$

as a solution of

$$\mathcal{L}y(x) + \lambda\rho(x)y(x) = 0 \tag{10.113}$$

and its boundary conditions. Equation (10.112) is a homogeneous Fredholm integral equation of the second kind, and Eq. (10.113) is the homogeneous eigenvalue equation (with the weighting function $w(x)$ replaced by $\rho(x)$).

There is a change in the interpretation of our Green's function. It started as a propagator function, a weighting function giving the importance of the charge $\rho(\mathbf{r}_2)$ in producing the potential $\varphi(\mathbf{r}_1)$. The charge ρ was the inhomogeneous term in the inhomogeneous differential equation (10.92). Now the differential equation and the integral equation are both **homogeneous**. $G(x,t)$ has become a link relating the two equations, differential and integral.

To complete the discussion of this differential equation–integral equation equivalence, let us now show that Eq. (10.113) implies Eq. (10.112), that is, that a solution of our differential equation (10.113) with its boundary conditions satisfies the integral equation (10.112). We multiply Eq. (10.113) by $G(x,t)$, the appropriate Green's function, and integrate from $x = a$ to $x = b$ to obtain

$$\int_a^b G(x,t)\mathcal{L}y(x)\,dx + \lambda\int_a^b G(x,t)\rho(x)y(x)\,dx = 0. \tag{10.114}$$

[19] The function $\rho(x)$ is some weighting function, not a charge density.

The first integral is split in two ($x < t, x > t$), according to the construction of our Green's function, giving

$$-\int_a^t G_1(x,t)\mathcal{L}y(x)\,dx - \int_t^b G_2(x,t)\mathcal{L}y(x)\,dx = \lambda\int_a^b G(x,t)\rho(x)y(x)\,dx. \quad (10.115)$$

Note that t is the upper limit for the G_1 integrals and the lower limit for the G_2 integrals. We are going to reduce the left-hand side of Eq. (10.115) to $y(t)$. Then, with $G(x,t) = G(t,x)$, we have Eq. (10.112) (with x and t interchanged).

Applying Green's theorem to the left-hand side or, equivalently, integrating by parts, we obtain

$$\begin{aligned}
&-\int_a^t G_1(t,x)\left[\frac{d}{dx}\left(p(x)\frac{d}{dx}y(x)\right) + q(x)y(x)\right]dx \\
&\quad = -\big[G_1(x,t)p(x)y'(x)\big]\Big|_{x=a}^{x=t} + \int_a^t \left(\frac{\partial}{\partial x}G_1(x,t)\right)p(x)y'(x)\,dx \\
&\qquad -\int_a^t G_1(x,t)q(x)y(x)\,dx, \qquad (10.116)
\end{aligned}$$

with an equivalent expression for the second integral. A second integration by parts yields

$$\begin{aligned}
-\int_a^t G_1(x,t)\mathcal{L}y(x)\,dx = &-\int_a^t y(x)\mathcal{L}G_1(x,t)\,dx \\
&-\big[G_1(x,t)p(x)y'(x)\big]\Big|_{x=a}^{x=t} \\
&+\big[G_1'(x,t)p(x)y(x)\big]\Big|_{x=a}^{x=t}. \qquad (10.117)
\end{aligned}$$

The integral on the right vanishes because $\mathcal{L}G_1 = 0$. By combining the integrated terms with those from integrating G_2, we have

$$\begin{aligned}
&-p(t)\left[G_1(t,t)y'(t) - y(t)\frac{\partial}{\partial x}G_1(x,t)\Big|_{x=t} - G_2(t,t)y'(t) + y(t)\frac{\partial}{\partial x}G_2(x,t)\Big|_{x=t}\right] \\
&\quad + p(a)\left[y'(a)G_1(a,t) - y(a)\frac{\partial}{\partial x}G_1(x,t)\Big|_{x=a}\right] \\
&\quad - p(b)\left[G_2(b,t)y'(b) - y(b)\frac{\partial}{\partial x}G_2(x,t)\Big|_{x=b}\right]. \qquad (10.118)
\end{aligned}$$

Each of the last two expressions vanishes, for $G(x,t)$ and $y(x)$ satisfy the same boundary conditions. The first expression, with the help of Eqs. (10.95) and (10.96), reduces to $y(t)$. Substituting into Eq. (10.115), we have Eq. (10.112), thus completing the demonstration of the equivalence of the integral equation and the differential equation plus boundary conditions.

Example 10.5.1 LINEAR OSCILLATOR

As a simple example, consider the linear oscillator equation (for a vibrating string):

$$y''(x) + \lambda y(x) = 0. \qquad (10.119)$$

We impose the conditions $y(0) = y(1) = 0$, which correspond to a string clamped at both ends. Now, to construct our Green's function, we need solutions of the homogeneous equation $\mathcal{L}y(x) = 0$, which is $y''(x) = 0$. To satisfy the boundary conditions, we must have one solution vanish at $x = 0$, the other at $x = 1$. Such solutions (unnormalized) are

$$u(x) = x, \qquad v(x) = 1 - x. \tag{10.120}$$

We find that

$$uv' - vu' = -1 \tag{10.121}$$

or, by Eq. (10.101) with $p(x) = 1$, $A = -1$. Our Green's function becomes

$$G(x,t) = \begin{cases} x(1-t), & 0 \le x < t, \\ t(1-x), & t < x \le 1. \end{cases} \tag{10.122}$$

Hence by Eq. (10.112) our clamped vibrating string satisfies

$$y(x) = \lambda \int_0^1 G(x,t)y(t)\,dt. \tag{10.123}$$

You may show that the known solutions of Eq. (10.119),

$$y = \sin n\pi x, \qquad \lambda = n^2\pi^2,$$

do indeed satisfy Eq. (10.123). Note that our eigenvalue λ is **not** the wavelength. ■

Green's Function and the Dirac Delta Function

One more approach to the Green's function may shed additional light on our formulation and particularly on its relation to physical problems. Let us refer once more to Poisson's equation, this time for a point charge:

$$\nabla^2\varphi(\mathbf{r}) = -\frac{\rho_{\text{point}}}{\varepsilon_0}. \tag{10.124}$$

The Green's function solution of this equation was developed in Section 9.7. This time let us take a one-dimensional analog

$$\mathcal{L}y(x) + f(x)_{\text{point}} = 0. \tag{10.125}$$

Here $f(x)_{\text{point}}$ refers to a unit point "charge," or a point force. We may represent it by a number of forms, but perhaps the most convenient is

$$f(x)_{\text{point}} = \begin{cases} \dfrac{1}{2\varepsilon}, & t - \varepsilon < x < t + \varepsilon, \\ 0, & \text{elsewhere}, \end{cases} \tag{10.126}$$

which is essentially the same as Eq. (1.172). Then, integrating Eq. (10.125), we have

$$\int_{t-\varepsilon}^{t+\varepsilon} \mathcal{L}y(x)\,dx = -\int_{t-\varepsilon}^{t+\varepsilon} f(x)_{\text{point}}\,dx = -1 \tag{10.127}$$

from the definition of $f(x)$. Let us examine $\mathcal{L}y(x)$ more closely. We have

$$\int_{t-\varepsilon}^{t+\varepsilon} \frac{d}{dx}\big[p(x)y'(x)\big]\,dx + \int_{t-\varepsilon}^{t+\varepsilon} q(x)y(x)\,dx = \big|_{t-\varepsilon}^{t+\varepsilon} p(x)y'(x)\big| + \int_{t-\varepsilon}^{t+\varepsilon} q(x)y(x)\,dx = -1. \tag{10.128}$$

In the limit $\varepsilon \to 0$ we may satisfy this relation by **permitting** $y'(x)$ to have a discontinuity of $-1/p(x)$ at $x = t$, $y(x)$ itself remaining continuous.[20] These, however, are just the properties used to define our Green's function, $G(x, t)$. In addition, we note that in the limit $\varepsilon \to 0$,

$$f(x)_{\text{point}} = \delta(x - t), \tag{10.129}$$

in which $\delta(x - t)$ is our Dirac delta function, defined in this manner in Section 1.15. Hence Eq. (10.125) has become

$$\mathcal{L}G(x, t) = -\delta(x - t). \tag{10.130}$$

This is a one-dimensional version of Eq. (9.159), which we exploit for the development of Green's functions in two and three dimensions — Section 9.7. It will be recalled that we used this relation in Section 9.7 to determine our Green's functions.

Equation (10.130) could have been expected since it is actually a consequence of our differential equation, Eq. (10.92), and Green's function integral solution, Eq. (10.97). If we let $\mathcal{L}_x$ (subscript to emphasize that it operates on the x-dependence) operate on both sides of Eq. (10.97), then

$$\mathcal{L}_x y(x) = \mathcal{L}_x \int_a^b G(x, t) f(t)\,dt.$$

By Eq. (10.92) the left-hand side is just $-f(x)$. On the right $\mathcal{L}_x$, is independent of the variable of integration t, so we may write

$$-f(x) = \int_a^b \big\{\mathcal{L}_x G(x, t)\big\} f(t)\,dt.$$

By definition of Dirac's delta function, Eqs. (1.171b) and (1.183), we have Eq. (10.130).

Exercises

10.5.1 Show that

$$G(x, t) = \begin{cases} x, & 0 \le x < t, \\ t, & t < x \le 1, \end{cases}$$

is the Green's function for the operator $\mathcal{L} = d^2/dx^2$ and the boundary conditions

$$y(0) = 0, \qquad y'(1) = 0.$$

[20] The functions $p(x)$ and $q(x)$ appearing in the operator $\mathcal{L}$ are continuous functions. With $y(x)$ remaining continuous, $\int q(x)y(x)\,dx$ is certainly continuous. Hence this integral over an interval 2ε (Eq. (10.128)) vanishes as ε vanishes.

10.5.2 Find the Green's function for

(a) $\mathcal{L}y(x) = \dfrac{d^2y(x)}{dx^2} + y(x), \qquad \begin{cases} y(0) = 0, \\ y'(1) = 0. \end{cases}$

(b) $\mathcal{L}y(x) = \dfrac{d^2y(x)}{dx^2} - y(x), \quad y(x)$ finite for $-\infty < x < \infty$.

10.5.3 Find the Green's function for the operators

(a) $\mathcal{L}y(x) = \dfrac{d}{dx}\left(x\dfrac{dy(x)}{dx}\right)$.

$$\text{ANS. } G(x,t) = \begin{cases} -\ln t, & 0 \le x < t, \\ -\ln x, & t < x \le 1. \end{cases}$$

(b) $\mathcal{L}y(x) = \dfrac{d}{dx}\left(x\dfrac{dy(x)}{dx}\right) - \dfrac{n^2}{x}y(x)$, with $y(0)$ finite and $y(1) = 0$.

$$\text{ANS. } G(x,t) = \begin{cases} \dfrac{1}{2n}\left[\left(\dfrac{x}{t}\right)^n - (xt)^n\right], & 0 \le x < t, \\ \dfrac{1}{2n}\left[\left(\dfrac{t}{x}\right)^n - (xt)^n\right], & t < x \le 1. \end{cases}$$

The combination of operator and interval specified in Exercise 10.5.3(a) is pathological, in that one of the endpoints of the interval (zero) is a singular point of the operator. As a consequence, the integrated part (the surface integral of Green's theorem) does not vanish. The next four exercises explore this situation.

10.5.4 (a) Show that the particular solution of

$$\frac{d}{dx}\left[x\frac{d}{dx}y(x)\right] = -1$$

is $y_P(x) = -x$.

(b) Show that

$$y_P(x) = -x \neq \int_0^1 G(x,t)(-1)\,dt,$$

where $G(x,t)$ is the Green's function of Exercise 10.5.3(a).

10.5.5 Show that Green's theorem, Eq. (1.104) in one dimension with a Sturm–Liouville-type operator $(d/dt)p(t)(d/dt)$ replacing $\nabla \cdot \nabla$, may be rewritten as

$$\int_a^b \left[u(t)\frac{d}{dt}\left(p(t)\frac{dv(t)}{dt}\right) - v(t)\frac{d}{dt}\left(p(t)\frac{du(t)}{dt}\right)\right]dt$$
$$= \left[u(t)p(t)\frac{dv(t)}{dt} - v(t)p(t)\frac{du(t)}{dt}\right]\Bigg|_a^b.$$

10.5.6 Using the one-dimensional form of Green's theorem of Exercise 10.5.5, let

$$v(t) = y(t) \qquad \text{and} \qquad \frac{d}{dt}\left(p(t)\frac{dy(t)}{dt}\right) = -f(t),$$

$$u(t) = G(x,t) \qquad \text{and} \qquad \frac{d}{dt}\left(p(t)\frac{\partial G(x,t)}{\partial t}\right) = -\delta(x-t).$$

Show that Green's theorem yields

$$y(x) = \int_a^b G(x,t)f(t)\,dt + \left[G(x,t)p(t)\frac{dy(t)}{dt} - y(t)p(t)\frac{\partial}{\partial t}G(x,t)\right]\Bigg|_{t=a}^{t=b}.$$

10.5.7 For $p(t) = t$, $y(t) = -t$,

$$G(x,t) = \begin{cases} -\ln t, & 0 \le x < t \\ -\ln x, & t < x \le 1, \end{cases}$$

verify that the integrated part does not vanish.

10.5.8 Construct the Green's function for

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + \left(k^2x^2 - 1\right)y = 0,$$

subject to the boundary conditions

$$y(0) = 0, \qquad y(1) = 0.$$

10.5.9 Given that

$$\mathcal{L} = \left(1 - x^2\right)\frac{d^2}{dx^2} - 2x\frac{d}{dx}$$

and

$$G(\pm 1, t) \text{ remains finite,}$$

show that no Green's function can be constructed by the techniques of this section. ($u(x)$ and $v(x)$ are linearly dependent.)

10.5.10 Construct the one-dimensional Green's function for the Helmholtz equation

$$\left(\frac{d^2}{dx^2} + k^2\right)\psi(x) = g(x).$$

The boundary conditions are those for a wave advancing in the positive x-direction — assuming a time dependence e^{-iwt}.

$$\text{ANS. } G(x_1, x_2) = \frac{i}{2k}\exp\left(ik|x_1 - x_2|\right).$$

10.5.11 Construct the one-dimensional Green's function for the modified Helmholtz equation

$$\left(\frac{d^2}{dx^2} - k^2\right)\psi(x) = f(x).$$

The boundary conditions are that the Green's function must vanish for $x \to \infty$ and $x \to -\infty$.

$$\text{ANS. } G(x_1, x_2) = \frac{1}{2k}\exp\bigl(-k|x_1 - x_2|\bigr).$$

10.5.12 From the eigenfunction expansion of the Green's function show that

(a) $$\frac{2}{\pi^2}\sum_{n=1}^{\infty}\frac{\sin n\pi x \sin n\pi t}{n^2} = \begin{cases} x(1-t), & 0 \le x < t, \\ t(1-x), & t < x \le 1. \end{cases}$$

(b) $$\frac{2}{\pi^2}\sum_{n=0}^{\infty}\frac{\sin(n+\frac{1}{2})\pi x \sin(n+\frac{1}{2})\pi t}{(n+\frac{1}{2})^2} = \begin{cases} x, & 0 \le x < t, \\ t, & t < x \le 1. \end{cases}$$

Note. In Section 10.4 the Green's function of $\mathcal{L} + \lambda$ is expanded in eigenfunctions. The λ there is an adjustable parameter, not an eigenvalue.

10.5.13 In the Fredholm equation,

$$f(x) = \lambda^2 \int_a^b G(x,t)\varphi(t)\,dt,$$

$G(x,t)$ is a Green's function given by

$$G(x,t) = \sum_{n=1}^{\infty}\frac{\varphi_n(x)\varphi_n(t)}{\lambda_n^2 - \lambda^2}.$$

Show that the solution is

$$\varphi(x) = \sum_{n=1}^{\infty}\frac{\lambda_n^2 - \lambda^2}{\lambda^2}\varphi_n(x)\int_a^b f(t)\varphi_n(t)\,dt.$$

10.5.14 Show that the Green's function integral transform operator

$$\int_a^b G(x,t)[\,]\,dt$$

is equal to $-\mathcal{L}^{-1}$, in the sense that

(a) $$\mathcal{L}_x \int_a^b G(x,t)y(t)\,dt = -y(x),$$

(b) $$\int_a^b G(x,t)\mathcal{L}_t y(t)\,dt = -y(x).$$

Note. Take $\mathcal{L}y(x) + f(x) = 0$, Eq. (10.92).

10.5.15 Substitute Eq. (10.87), the eigenfunction expansion of Green's function, into Eq. (10.88) and then show that Eq. (10.88) is indeed a solution of the inhomogeneous Helmholtz equation (10.82).

10.5.16 (a) Starting with a one-dimensional inhomogeneous differential equation (Eq. (10.89)), assume that $\psi(x)$ and $\rho(x)$ may be represented by eigenfunction expansions. Without any use of the Dirac delta function or its representations, show that

$$\psi(x) = \sum_{n=0}^{\infty} \frac{\int_a^b \rho(t)\varphi_n(t)\,dt}{\lambda_n - \lambda}\varphi_n(x).$$

Note that (1) if $\rho = 0$, no solution exists unless $\lambda = \lambda_n$ and (2) if $\lambda = \lambda_n$, no solution exists unless ρ is orthogonal to φ_n. This same behavior will reappear with integral equations in Section 16.4.

(b) Interchanging summation and integration, show that you have constructed the Green's function corresponding to Eq. (10.90).

10.5.17 The eigenfunctions of the Schrödinger equation are often complex. In this case the orthogonality integral, Eq. (10.40), is replaced by

$$\int_a^b \varphi_i^*(x)\varphi_j(x)w(x)\,dx = \delta_{ij}.$$

Instead of Eq. (1.189), we have

$$\delta(\mathbf{r}_1 - \mathbf{r}_2) = \sum_{n=0}^{\infty} \varphi_n(\mathbf{r}_1)\varphi_n^*(\mathbf{r}_2).$$

Show that the Green's function, Eq. (10.87), becomes

$$G(\mathbf{r}_1, \mathbf{r}_2) = \sum_{n=0}^{\infty} \frac{\varphi_n(\mathbf{r}_1)\varphi_n^*(\mathbf{r}_2)}{k_n^2 - k^2} = G^*(\mathbf{r}_2, \mathbf{r}_1).$$

Additional Readings

Byron, F. W., Jr., and R. W. Fuller, *Mathematics of Classical and Quantum Physics*. Reading, MA: Addison-Wesley (1969).

Dennery, P., and A. Krzywicki, *Mathematics for Physicists*. Reprinted. New York: Dover (1996).

Hirsch, M., *Differential Equations, Dynamical Systems, and Linear Algebra*. San Diego: Academic Press (1974).

Miller, K. S., *Linear Differential Equations in the Real Domain*. New York: Norton (1963).

Titchmarsh, E. C., *Eigenfunction Expansions Associated with Second-Order Differential Equations*, 2nd ed., Vol. 1. London: Oxford University Press (1962), Vol. II (1958).

CHAPTER 11

BESSEL FUNCTIONS

11.1 BESSEL FUNCTIONS OF THE FIRST KIND, $J_\nu(x)$

Bessel functions appear in a wide variety of physical problems. In Section 9.3, separation of the Helmholtz, or wave, equation in circular cylindrical coordinates led to Bessel's equation. In Section 11.7 we will see that the Helmholtz equation in spherical polar coordinates also leads to a form of Bessel's equation. Bessel functions may also appear in integral form — integral representations. This may result from integral transforms (Chapter 15) or from the mathematical elegance of starting the study of Bessel functions with Hankel functions, Section 11.4.

Bessel functions and closely related functions form a rich area of mathematical analysis with many representations, many interesting and useful properties, and many interrelations. Some of the major interrelations are developed in Section 11.1 and in succeeding sections. Note that Bessel functions are not restricted to Chapter 11. The asymptotic forms are developed in Section 7.3 as well as in Section 11.6. The confluent hypergeometric representations appear in Section 13.5.

Generating Function for Integral Order

Although Bessel functions are of interest primarily as solutions of differential equations, it is instructive and convenient to develop them from a completely different approach, that of the generating function.[1] This approach also has the advantage of focusing on the functions themselves rather than on the differential equations they satisfy. Let us introduce a function of two variables,

$$g(x,t) = e^{(x/2)(t-1/t)}. \tag{11.1}$$

[1] Generating functions have already been used in Chapter 5. In Section 5.6 the generating function $(1+x)^n$ was used to derive the binomial coefficients. In Section 5.9 the generating function $x(e^x-1)^{-1}$ was used to derive the Bernoulli numbers.

Expanding this function in a Laurent series (Section 6.5), we obtain

$$\boxed{e^{(x/2)(t-1/t)} = \sum_{n=-\infty}^{\infty} J_n(x)t^n.} \tag{11.2}$$

It is instructive to compare Eq. (11.2) with the equivalent Eqs. (11.23) and (11.25).

The coefficient of t^n, $J_n(x)$, is defined to be a Bessel function of the first kind, of integral order n. Expanding the exponentials, we have a product of Maclaurin series in $xt/2$ and $-x/2t$, respectively,

$$e^{xt/2} \cdot e^{-x/2t} = \sum_{r=0}^{\infty} \left(\frac{x}{2}\right)^r \frac{t^r}{r!} \sum_{s=0}^{\infty} (-1)^s \left(\frac{x}{2}\right)^s \frac{t^{-s}}{s!}. \tag{11.3}$$

Here, the summation index r is changed to n, with $n = r - s$ and summation limits $n = -s$ to ∞, and the order of the summations is interchanged, which is justified by absolute convergence. The range of the summation over n becomes $-\infty$ to ∞, while the summation over s extends from $\max(-n, 0)$ to ∞. For a given s we get $t^n (n \geq 0)$ from $r = n + s$:

$$\left(\frac{x}{2}\right)^{n+s} \frac{t^{n+s}}{(n+s)!} (-1)^s \left(\frac{x}{2}\right)^s \frac{t^{-s}}{s!}. \tag{11.4}$$

The coefficient of t^n is then[2]

$$\boxed{J_n(x) = \sum_{s=0}^{\infty} \frac{(-1)^s}{s!(n+s)!} \left(\frac{x}{2}\right)^{n+2s} = \frac{x^n}{2^n n!} - \frac{x^{n+2}}{2^{n+2}(n+1)!} + \cdots.} \tag{11.5}$$

This series form exhibits is behavior of the Bessel function $J_n(x)$ for small x and permits numerical evaluation of $J_n(x)$. The results for J_0, J_1, and J_2 are shown in Fig. 11.1. From Section 5.3 the error in using only a finite number of terms of this alternating series in numerical evaluation is less than the first term omitted. For instance, if we want $J_n(x)$

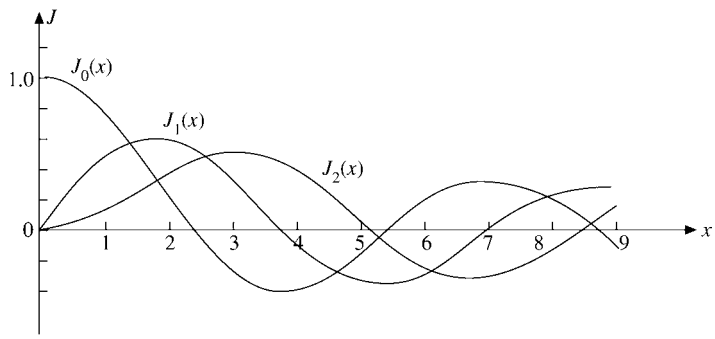

FIGURE 11.1 Bessel functions, $J_0(x)$, $J_1(x)$, and $J_2(x)$.

[2]From the steps leading to this series and from its convergence characteristics it should be clear that this series may be used with x replaced by z and with z any point in the finite complex plane.

to ±1% accuracy, the first term alone of Eq. (11.5) will suffice, provided the ratio of the second term to the first is less than 1% (in magnitude) or $x < 0.2(n+1)^{1/2}$. The Bessel functions oscillate but are **not** periodic — except in the limit as $x \to \infty$ (Section 11.6). The amplitude of $J_n(x)$ is not constant but decreases asymptotically as $x^{-1/2}$. (See Eq.(11.137) for this envelope.)

For $n < 0$, Eq. (11.5) gives

$$J_{-n}(x) = \sum_{s=0}^{\infty} \frac{(-1)^s}{s!(s-n)!}\left(\frac{x}{2}\right)^{2s-n}. \tag{11.6}$$

Since n is an integer (here), $(s-n)! \to \infty$ for $s = 0, \ldots, (n-1)$. Hence the series may be considered to start with $s = n$. Replacing s by $s+n$, we obtain

$$J_{-n}(x) = \sum_{s=0}^{\infty} \frac{(-1)^{s+n}}{s!(s+n)!}\left(\frac{x}{2}\right)^{n+2s}, \tag{11.7}$$

showing immediately that $J_n(x)$ and $J_{-n}(x)$ are not independent but are related by

$$J_{-n}(x) = (-1)^n J_n(x) \qquad \text{(integral } n\text{)}. \tag{11.8}$$

These series expressions (Eqs. (11.5) and (11.6)) may be used with n replaced by ν to **define** $J_\nu(x)$ and $J_{-\nu}(x)$ for nonintegral ν (compare Exercise 11.1.7).

Recurrence Relations

The recurrence relations for $J_n(x)$ and its derivatives may all be obtained by operating on the series, Eq. (11.5), although this requires a bit of clairvoyance (or a lot of trial and error). Verification of the known recurrence relations is straightforward, Exercise 11.1.7. Here it is convenient to obtain them from the generating function, $g(x,t)$. Differentiating both sides of Eq. (11.1) with respect to t, we find that

$$\begin{aligned}\frac{\partial}{\partial t} g(x,t) &= \frac{1}{2}x\left(1+\frac{1}{t^2}\right)e^{(x/2)(t-1/t)} \\ &= \sum_{n=-\infty}^{\infty} n J_n(x) t^{n-1},\end{aligned} \tag{11.9}$$

and substituting Eq. (11.2) for the exponential and equating the coefficients of like powers of t,[3] we obtain

$$J_{n-1}(x) + J_{n+1}(x) = \frac{2n}{x} J_n(x). \tag{11.10}$$

This is a three-term recurrence relation. Given J_0 and J_1, for example, J_2 (and any other integral order J_n) may be computed.

[3] This depends on the fact that the power-series representation is unique (Sections 5.7 and 6.5).

Differentiating Eq. (11.1) with respect to x, we have

$$\frac{\partial}{\partial x}g(x,t)=\frac{1}{2}\left(t-\frac{1}{t}\right)e^{(x/2)(t-1/t)}=\sum_{n=-\infty}^{\infty}J_n'(x)t^n. \tag{11.11}$$

Again, substituting in Eq. (11.2) and equating the coefficients of like powers of t, we obtain the result

$$J_{n-1}(x)-J_{n+1}(x)=2J_n'(x). \tag{11.12}$$

As a special case of this general recurrence relation,

$$J_0'(x)=-J_1(x). \tag{11.13}$$

Adding Eqs. (11.10) and (11.12) and dividing by 2, we have

$$J_{n-1}(x)=\frac{n}{x}J_n(x)+J_n'(x). \tag{11.14}$$

Multiplying by x^n and rearranging terms produces

$$\frac{d}{dx}\left[x^nJ_n(x)\right]=x^nJ_{n-1}(x). \tag{11.15}$$

Subtracting Eq. (11.12) from Eq. (11.10) and dividing by 2 yields

$$J_{n+1}(x)=\frac{n}{x}J_n(x)-J_n'(x). \tag{11.16}$$

Multiplying by x^{-n} and rearranging terms, we obtain

$$\frac{d}{dx}\left[x^{-n}J_n(x)\right]=-x^{-n}J_{n+1}(x). \tag{11.17}$$

Bessel's Differential Equation

Suppose we consider a set of functions $Z_\nu(x)$ that satisfies the basic recurrence relations (Eqs. (11.10) and (11.12)), but with ν not necessarily an integer and Z_ν not necessarily given by the series (Eq. (11.5)). Equation (11.14) may be rewritten ($n\to\nu$) as

$$xZ_\nu'(x)=xZ_{\nu-1}(x)-\nu Z_\nu(x). \tag{11.18}$$

On differentiating with respect to x, we have

$$xZ_\nu''(x)+(\nu+1)Z_\nu'-xZ_{\nu-1}'-Z_{\nu-1}=0. \tag{11.19}$$

Multiplying by x and then subtracting Eq. (11.18) multiplied by ν gives us

$$x^2Z_\nu''+xZ_\nu'-\nu^2Z_\nu+(\nu-1)xZ_{\nu-1}-x^2Z_{\nu-1}'=0. \tag{11.20}$$

Now we rewrite Eq. (11.16) and replace n by $\nu-1$:

$$xZ_{\nu-1}'=(\nu-1)Z_{\nu-1}-xZ_\nu. \tag{11.21}$$

Using Eq. (11.21) to eliminate $Z_{\nu-1}$ and $Z_{\nu-1}'$ from Eq. (11.20), we finally get

$$\boxed{x^2Z_\nu''+xZ_\nu'+\left(x^2-\nu^2\right)Z_\nu=0,} \tag{11.22}$$

which is **Bessel**'s ODE. Hence any functions $Z_\nu(x)$ that satisfy the recurrence relations (Eqs. (11.10) and (11.12), (11.14) and (11.16), or (11.15) and (11.17)) satisfy Bessel's equation; that is, the unknown Z_ν are Bessel functions. In particular, we have shown that the functions $J_n(x)$, defined by our generating function, satisfy Bessel's ODE. If the argument is $k\rho$ rather than x, Eq. (11.22) becomes

$$\boxed{\rho^2\frac{d^2}{d\rho^2}Z_\nu(k\rho)+\rho\frac{d}{d\rho}Z_\nu(k\rho)+\left(k^2\rho^2-\nu^2\right)Z_\nu(k\rho)=0.} \tag{11.22a}$$

Integral Representation

A particularly useful and powerful way of treating Bessel functions employs integral representations. If we return to the generating function (Eq. (11.2)), and substitute $t=e^{i\theta}$, we get

$$\begin{aligned} e^{ix\sin\theta} &= J_0(x)+2\big[J_2(x)\cos 2\theta+J_4(x)\cos 4\theta+\cdots\big] \\ &\quad +2i\big[J_1(x)\sin\theta+J_3(x)\sin 3\theta+\cdots\big], \end{aligned} \tag{11.23}$$

in which we have used the relations

$$\begin{aligned} J_1(x)e^{i\theta}+J_{-1}(x)e^{-i\theta} &= J_1(x)\left(e^{i\theta}-e^{-i\theta}\right) \\ &= 2iJ_1(x)\sin\theta, \\ J_2(x)e^{2i\theta}+J_{-2}(x)e^{-2i\theta} &= 2J_2(x)\cos 2\theta, \end{aligned} \tag{11.24}$$

and so on.

In summation notation,

$$\begin{aligned} \cos(x\sin\theta) &= J_0(x)+2\sum_{n=1}^{\infty}J_{2n}(x)\cos(2n\theta), \\ \sin(x\sin\theta) &= 2\sum_{n=1}^{\infty}J_{2n-1}(x)\sin\big[(2n-1)\theta\big], \end{aligned} \tag{11.25}$$

equating real and imaginary parts of Eq. (11.23).

By employing the orthogonality properties of cosine and sine,[4]

$$\int_0^\pi \cos n\theta\cos m\theta\,d\theta=\frac{\pi}{2}\delta_{nm}, \tag{11.26a}$$

$$\int_0^\pi \sin n\theta\sin m\theta\,d\theta=\frac{\pi}{2}\delta_{nm}, \tag{11.26b}$$

[4]They are eigenfunctions of a self-adjoint equation (linear oscillator equation) and satisfy appropriate boundary conditions (compare Sections 10.2 and 14.1).

in which n and m are **positive** integers (zero is excluded),[5] we obtain

$$\frac{1}{\pi}\int_0^{\pi}\cos(x\sin\theta)\cos n\theta\,d\theta=\begin{cases}J_n(x), & n \text{ even},\\ 0, & n \text{ odd},\end{cases} \tag{11.27}$$

$$\frac{1}{\pi}\int_0^{\pi}\sin(x\sin\theta)\sin n\theta\,d\theta=\begin{cases}0, & n \text{ even},\\ J_n(x), & n \text{ odd}.\end{cases} \tag{11.28}$$

If these two equations are added together,

$$\begin{aligned}J_n(x)&=\frac{1}{\pi}\int_0^{\pi}\left[\cos(x\sin\theta)\cos n\theta+\sin(x\sin\theta)\sin n\theta\right]d\theta\\&=\frac{1}{\pi}\int_0^{\pi}\cos(n\theta-x\sin\theta)\,d\theta, \qquad n=0,1,2,3,\ldots.\end{aligned} \tag{11.29}$$

As a special case (integrate Eq. (11.25) over $(0,\pi)$ to get)

$$J_0(x)=\frac{1}{\pi}\int_0^{\pi}\cos(x\sin\theta)\,d\theta. \tag{11.30}$$

Noting that $\cos(x\sin\theta)$ repeats itself in all four quadrants, we may write Eq. (11.30) as

$$J_0(x)=\frac{1}{2\pi}\int_0^{2\pi}\cos(x\sin\theta)\,d\theta. \tag{11.30a}$$

On the other hand, $\sin(x\sin\theta)$ reverses its sign in the third and fourth quadrants, so

$$\frac{1}{2\pi}\int_0^{2\pi}\sin(x\sin\theta)\,d\theta=0. \tag{11.30b}$$

Adding Eq. (11.30a) and i times Eq. (11.30b), we obtain the complex exponential representation

$$J_0(x)=\frac{1}{2\pi}\int_0^{2\pi}e^{ix\sin\theta}\,d\theta=\frac{1}{2\pi}\int_0^{2\pi}e^{ix\cos\theta}\,d\theta. \tag{11.30c}$$

This integral representation (Eq. (11.29)) may be obtained somewhat more directly by employing contour integration (compare Exercise 11.1.16).[6] Many other integral representations exist (compare Exercise 11.1.18).

Example 11.1.1 FRAUNHOFER DIFFRACTION, CIRCULAR APERTURE

In the theory of diffraction through a circular aperture we encounter the integral

$$\Phi\sim\int_0^{a}r\,dr\int_0^{2\pi}e^{ibr\cos\theta}\,d\theta \tag{11.31}$$

[5] Equations (11.26a) and (11.26b) hold for either m or $n=0$. If both m and $n=0$, the constant in (11.26a) becomes π; the constant in Eq. (11.26b) becomes 0.

[6] For $n=0$ a simple integration over θ from 0 to 2π will convert Eq. (11.23) into Eq. (11.30c).

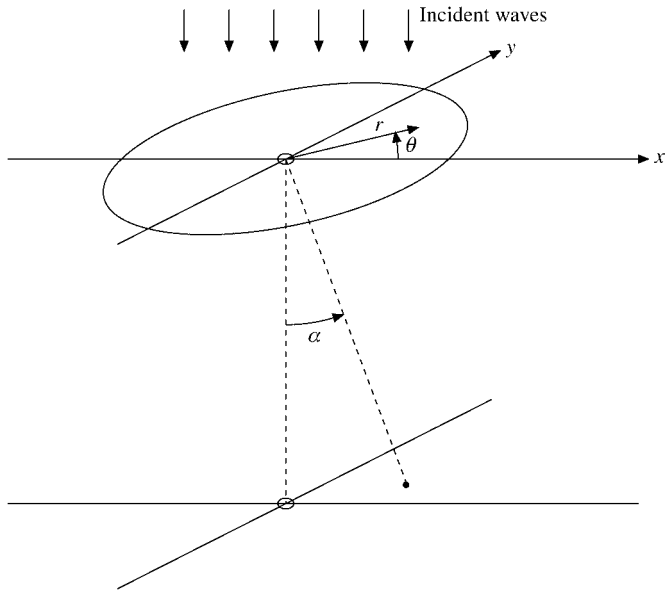

FIGURE 11.2 Fraunhofer diffraction, circular aperture.

for Φ, the amplitude of the diffracted wave.[7] Here θ is an azimuth angle in the plane of the circular aperture of radius a, and α is the angle defined by a point on a screen below the circular aperture relative to the normal through the center point. The parameter b is given by

$$b = \frac{2\pi}{\lambda}\sin\alpha, \tag{11.32}$$

with λ the wavelength of the incident wave. The other symbols are defined by Fig. 11.2. From Eq. (11.30c) we get[8]

$$\Phi \sim 2\pi \int_0^a J_0(br)r\,dr. \tag{11.33}$$

Equation (11.15) enables us to integrate Eq. (11.33) immediately to obtain

$$\Phi \sim \frac{2\pi ab}{b^2}J_1(ab) \sim \frac{\lambda a}{\sin\alpha}J_1\left(\frac{2\pi a}{\lambda}\sin\alpha\right). \tag{11.34}$$

Note here that $J_1(0) = 0$. The intensity of the light in the diffraction pattern is proportional to Φ^2 and

$$\Phi^2 \sim \left\{\frac{J_1[(2\pi a/\lambda)\sin\alpha]}{\sin\alpha}\right\}^2. \tag{11.35}$$

[7] The exponent $ibr\cos\theta$ gives the phase of the wave on the distant screen at angle α relative to the phase of the wave incident on the aperture at the point (r,θ). The imaginary exponential form of this integrand means that the integral is technically a Fourier transform, Chapter 15. In general, the Fraunhofer diffraction pattern is given by the Fourier transform of the aperture.

[8] We could also refer to Exercise 11.1.16(b).

Table 11.1 Zeros of the Bessel Functions and Their First Derivatives

Number of zero	$J_0(x)$	$J_1(x)$	$J_2(x)$	$J_3(x)$	$J_4(x)$	$J_5(x)$
1	2.4048	3.8317	5.1356	6.3802	7.5883	8.7715
2	5.5201	7.0156	8.4172	9.7610	11.0647	12.3386
3	8.6537	10.1735	11.6198	13.0152	14.3725	15.7002
4	11.7915	13.3237	14.7960	16.2235	17.6160	18.9801
5	14.9309	16.4706	17.9598	19.4094	20.8269	22.2178
	$J_0'(x)^a$	$J_1'(x)$	$J_2'(x)$	$J_3'(x)$		
1	3.8317	1.8412	3.0542	4.2012		
2	7.0156	5.3314	6.7061	8.0152		
3	10.1735	8.5363	9.9695	11.3459		

[a] $J_0'(x) = -J_1(x)$.

From Table 11.1, which lists the zeros of the Bessel functions and their first derivatives,[9] Eq. (11.35) will have a zero at

$$\frac{2\pi a}{\lambda}\sin\alpha = 3.8317\ldots, \tag{11.36}$$

or

$$\sin\alpha = \frac{3.8317\lambda}{2\pi a}. \tag{11.37}$$

For green light, $\lambda = 5.5 \times 10^{-5}$ cm. Hence, if $a = 0.5$ cm,

$$\alpha \approx \sin\alpha = 6.7 \times 10^{-5}\ \text{(radian)} \approx 14\ \text{seconds of arc}, \tag{11.38}$$

which shows that the bending or spreading of the light ray is extremely small. If this analysis had been known in the seventeenth century, the arguments against the wave theory of light would have collapsed. In mid-twentieth century this same diffraction pattern appears in the scattering of nuclear particles by atomic nuclei — a striking demonstration of the wave properties of the nuclear particles. ■

A further example of the use of Bessel functions and their roots is provided by the electromagnetic resonant cavity (Example 11.1.2) and the example and exercises of Section 11.2.

Example 11.1.2 CYLINDRICAL RESONANT CAVITY

The propagation of electromagnetic waves in hollow metallic cylinders is important in many practical devices. If the cylinder has end surfaces, it is called a **cavity**. Resonant cavities play a crucial role in many particle accelerators.

[9] Additional roots of the Bessel functions and their first derivatives may be found in C. L. Beattie, Table of first 700 zeros of Bessel functions. *Bell Syst. Tech. J.* **37**: 689 (1958), and Bell Monogr. **3055**. Roots may be accessed in Mathematica and other symbolic software and are on the Web.

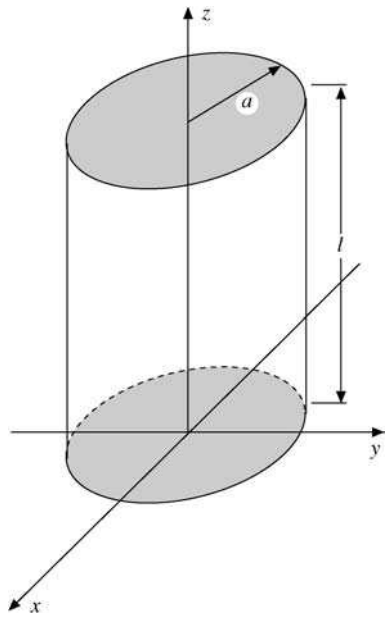

FIGURE 11.3 Cylindrical resonant cavity.

We take the z-axis along the center of the cavity with end surfaces at $z = 0$ and $z = l$ and use cylindrical coordinates suggested by the geometry. Its walls are perfect conductors, so the tangential electric field vanishes on them (as in Fig. 11.3):

$$E_z = 0 = E_\varphi \qquad \text{for } \rho = a, \qquad E_\rho = 0 = E_\varphi \qquad \text{for } z = 0, l.$$

Inside the cavity we have a vacuum, so $\varepsilon_0\mu_0 = 1/c^2$. In the interior of a resonant cavity, electromagnetic waves oscillate with harmonic time dependence $e^{-i\omega t}$, which follows from separating the time from the spatial variables in Maxwell's equations (Section 1.9), so

$$\nabla \times \nabla \times \mathbf{E} = -\frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} = \alpha^2 \mathbf{E}, \qquad \alpha = \frac{\omega}{c}.$$

With $\nabla \cdot \mathbf{E} = 0$ (vacuum, no charges) and Eq. (1.85), we obtain for the space part of the electric field

$$\nabla^2 \mathbf{E} + \alpha^2 \mathbf{E} = 0,$$

which is called the **vector Helmholtz PDE**. The z-component (E_z, space part only) satisfies the scalar Helmholtz equation,

$$\nabla^2 E_z + \alpha^2 E_z = 0. \tag{11.39}$$

The transverse electric field components $\mathbf{E}_\perp = (E_\rho, E_\varphi)$ obey the same PDE but different boundary conditions, given earlier. Once E_z is known, Maxwell's equations determine E_φ fully. See Jackson, *Electrodynamics* in Additional Readings for details.

We separate the z variable from ρ and φ, because there are no mixed derivatives $\frac{\partial^2 E_z}{\partial z\,\partial\rho}$, etc. The product solution, $E_z = v(\rho,\varphi)w(z)$, is substituted into the Helmholtz PDE for E_z using Eq. (2.35) for ∇^2 in cylindrical coordinates, and then we divide by vw, yielding

$$\frac{1}{w(z)}\frac{d^2w}{dz^2} + \frac{1}{v}\left(\frac{\partial^2 v}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial v}{\partial\rho} + \frac{1}{\rho^2}\frac{\partial^2 v}{\partial\varphi^2} + \alpha^2\right)v(\rho,\varphi) = 0.$$

This implies

$$-\frac{1}{w(z)}\frac{d^2w}{dz^2} = \frac{1}{v(\rho,\varphi)}\left(\frac{\partial^2 v}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial v}{\partial\rho} + \frac{1}{\rho^2}\frac{\partial^2 v}{\partial\varphi^2} + \alpha^2 v\right) = k^2.$$

Here, k^2 is a separation constant, because the left- and right-hand sides depend on different variables. For $w(z)$ we find the harmonic oscillator ODE with standing wave solution (not transients) that we seek,

$$w(z) = A\sin kz + B\cos kz,$$

with A, B constants. For $v(\rho,\varphi)$ we obtain

$$\frac{\partial^2 v}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial v}{\partial\rho} + \frac{1}{\rho^2}\frac{\partial^2 v}{\partial\varphi^2} + \gamma^2 v = 0, \qquad \gamma^2 = \alpha^2 - k^2.$$

In this PDE we can separate the ρ and φ variables, because there is no mixed term $\frac{\partial^2 v}{\partial\rho\,\partial\varphi}$. The product form $v = u(\rho)\Phi(\varphi)$ yields

$$\frac{\rho^2}{u(\rho)}\left(\frac{d^2u}{d\rho^2} + \frac{1}{\rho}\frac{du}{d\rho} + \gamma^2\right) = -\frac{1}{\Phi(\varphi)}\frac{d^2\Phi}{d\varphi^2} = m^2,$$

where the **separation constant m^2 must be an integer**, because the angular solution $\Phi = e^{im\varphi}$ of the ODE

$$\frac{d^2\Phi}{d\varphi^2} + m^2\Phi = 0$$

must be periodic in the azimuthal angle.

This leaves us with the radial ODE

$$\frac{d^2u}{d\rho^2} + \frac{1}{\rho}\frac{du}{d\rho} + \left(\gamma^2 - \frac{m^2}{\rho^2}\right)u = 0.$$

Dimensional arguments suggest rescaling $\rho \to r = \gamma\rho$ and dividing by γ^2, which yields

$$\frac{d^2u}{dr^2} + \frac{1}{r}\frac{du}{dr} + \left(1 - \frac{m^2}{r^2}\right)u = 0.$$

This is Bessel's ODE for $\nu = m$. We use the regular solution $J_m(\gamma\rho)$ because the (irregular) second independent solution is singular at the origin, which is unacceptable here. The complete solution is

$$E_z = J_m(\gamma\rho)e^{im\varphi}(A\sin kz + B\cos kz), \tag{11.40a}$$

where the constant γ is determined from the **boundary condition** $E_z = 0$ on the cavity surface $\rho = a$, that is, that γa be a root of the Bessel function J_m (see Table 11.1). This gives a discrete set of values $\gamma = \gamma_{mn}$, where n designates the nth root of J_m (see Table 11.1).

For the transverse magnetic or TM mode of oscillation with $H_z = 0$ Maxwell's equations imply. (See again *Resonant Cavities* in J. D. Jackson's *Electrodynamics* in Additional Readings.)

$$\mathbf{E}_\perp \sim \nabla_\perp \frac{\partial E_z}{\partial z}, \qquad \nabla_\perp = \left(\frac{\partial}{\partial \rho}, \frac{1}{\rho}\frac{\partial}{\partial \varphi} \right).$$

The form of this result suggests $E_z \sim \cos kz$, that is, setting $A = 0$ so that $\mathbf{E}_\perp \sim \sin kz = 0$ at $z = 0, l$ can be satisfied by

$$k = \frac{p\pi}{l}, \qquad p = 0, 1, 2, \ldots. \tag{11.41}$$

Thus, the **tangential** electric fields E_ρ and E_φ vanish at $z = 0$ and l. In other words, $A = 0$ corresponds to $dE_z/dz = 0$ at $z = 0$ and $z = l$ for the TM mode. Altogether then, we have

$$\gamma^2 = \frac{\omega^2}{c^2} - k^2 = \frac{\omega^2}{c^2} - \frac{p^2\pi^2}{l^2}, \tag{11.42}$$

with

$$\gamma = \gamma_{mn} = \frac{\alpha_{mn}}{a}, \tag{11.43}$$

where α_{mn} is the nth zero of J_m. The general solution

$$E_z = \sum_{m,n,p} J_m(\gamma_{mn}\rho) e^{\pm im\varphi} B_{mnp} \cos \frac{p\pi z}{l}, \tag{11.40b}$$

with constants B_{mnp}, now follows from the superposition principle.

The result of the two boundary conditions and the separation constant m^2 is that the angular frequency of our oscillation depends on three discrete parameters:

$$\omega_{mnp} = c\sqrt{\frac{\alpha_{mn}^2}{a^2} + \frac{p^2\pi^2}{l^2}}, \qquad \begin{cases} m = 0, 1, 2, \ldots, \\ n = 1, 2, 3, \ldots, \\ p = 0, 1, 2 \ldots. \end{cases} \tag{11.44}$$

These are the allowable resonant frequencies for our TM mode. The TE mode of oscillation is the topic of Exercise 11.1.26. ■

Alternate Approaches

Bessel functions are introduced here by means of a generating function, Eq. (11.2). Other approaches are possible. Listing the various possibilities, we have:

1. Generating function (magic), Eq. (11.2).
2. Series solution of Bessel's differential equation, Section 9.5.
3. Contour integrals: Some writers prefer to start with contour integral definitions of the Hankel functions, Section 7.3 and 11.4, and develop the Bessel function $J_\nu(x)$ from the Hankel functions.

4. Direct solution of physical problems: Example 11.1.1. Fraunhofer diffraction with a circular aperture, illustrates this. Incidentally, Eq. (11.31) can be treated by series expansion, if desired. Feynman[10] develops Bessel functions from a consideration of cavity resonators.

In case the generating function seems too arbitrary, it can be derived from a contour integral, Exercise 11.1.16, or from the Bessel function recurrence relations, Exercise 11.1.6. Note that the contour integral is not limited to integer ν, thus providing a starting point for developing Bessel functions.

Bessel Functions of Nonintegral Order

These different approaches are not exactly equivalent. The generating function approach is very convenient for deriving two recurrence relations, Bessel's differential equation, integral representations, addition theorems (Exercise 11.1.2), and upper and lower bounds (Exercise 11.1.1). However, you will probably have noticed that the generating function defined only Bessel functions of integral order, J_0, J_1, J_2, and so on. This is a limitation of the generating function approach that can be avoided by using the contour integral in Exercise 11.1.16 instead, thus leading to foregoing approach (3). But the Bessel function of the first kind, $J_\nu(x)$, may easily be defined for nonintegral ν by using the series (Eq. (11.5)) as a new definition.

The recurrence relations may be verified by substituting in the series form of $J_\nu(x)$ (Exercise 11.1.7). From these relations Bessel's equation follows. In fact, if ν is not an integer, there is actually an important simplification. It is found that J_ν and $J_{-\nu}$ are independent, for no relation of the form of Eq. (11.8) exists. On the other hand, for $\nu = n$, an integer, we need another solution. The development of this second solution and an investigation of its properties form the subject of Section 11.3.

Exercises

11.1.1 From the product of the generating functions $g(x,t)\cdot g(x,-t)$ show that

$$1 = \big[J_0(x)\big]^2 + 2\big[J_1(x)\big]^2 + 2\big[J_2(x)\big]^2 + \cdots$$

and therefore that $|J_0(x)| \le 1$ and $|J_n(x)| \le 1/\sqrt{2}$, $n = 1, 2, 3, \ldots$.
Hint. Use uniqueness of power series, Section 5.7.

11.1.2 Using a generating function $g(x,t) = g(u+v,t) = g(u,t)\cdot g(v,t)$, show that

(a) $$J_n(u+v) = \sum_{s=-\infty}^{\infty} J_s(u)\cdot J_{n-s}(v),$$

(b) $$J_0(u+v) = J_0(u)J_0(v) + 2\sum_{s=1}^{\infty} J_s(u)J_{-s}(v).$$

[10] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Vol. II. Reading, MA: Addison-Wesley (1964), Chapter 23.

These are addition theorems for the Bessel functions.

11.1.3 Using only the generating function

$$e^{(x/2)(t-1/t)} = \sum_{n=-\infty}^{\infty} J_n(x)t^n$$

and not the explicit series form of $J_n(x)$, show that $J_n(x)$ has odd or even parity according to whether n is odd or even, that is,[11]

$$J_n(x) = (-1)^n J_n(-x).$$

11.1.4 Derive the Jacobi–Anger expansion

$$e^{iz\cos\theta} = \sum_{m=-\infty}^{\infty} i^m J_m(z)e^{im\theta}.$$

This is an expansion of a plane wave in a series of cylindrical waves.

11.1.5 Show that

(a) $\cos x = J_0(x) + 2\sum_{n=1}^{\infty}(-1)^n J_{2n}(x)$,

(b) $\sin x = 2\sum_{n=0}^{\infty}(-1)^n J_{2n+1}(x)$.

11.1.6 To help remove the generating function from the realm of magic, show that it can be **derived** from the recurrence relation, Eq. (11.10).
Hint.

(a) Assume a generating function of the form

$$g(x,t) = \sum_{m=-\infty}^{\infty} J_m(x)t^m.$$

(b) Multiply Eq. (11.10) by t^n and sum over n.
(c) Rewrite the preceding result as

$$\left(t + \frac{1}{t}\right)g(x,t) = \frac{2t}{x}\frac{\partial g(x,t)}{\partial t}.$$

(d) Integrate and adjust the "constant" of integration (a function of x) so that the coefficient of the zeroth power, t^0, is $J_0(x)$, as given by Eq. (11.5).

11.1.7 Show, by direct differentiation, that

$$J_\nu(x) = \sum_{s=0}^{\infty} \frac{(-1)^s}{s!(s+\nu)!}\left(\frac{x}{2}\right)^{\nu+2s}$$

[11] This is easily seen from the series form (Eq. (11.5)).

satisfies the two recurrence relations

$$J_{\nu-1}(x) + J_{\nu+1}(x) = \frac{2\nu}{x} J_\nu(x),$$

$$J_{\nu-1}(x) - J_{\nu+1}(x) = 2J'_\nu(x),$$

and Bessel's differential equation

$$x^2 J''_\nu(x) + x J'_\nu(x) + \left(x^2 - \nu^2\right) J_\nu(x) = 0.$$

11.1.8 Prove that

$$\frac{\sin x}{x} = \int_0^{\pi/2} J_0(x\cos\theta)\cos\theta\, d\theta, \qquad \frac{1-\cos x}{x} = \int_0^{\pi/2} J_1(x\cos\theta)\, d\theta.$$

Hint. The definite integral

$$\int_0^{\pi/2} \cos^{2s+1}\theta\, d\theta = \frac{2\cdot 4\cdot 6\cdots(2s)}{1\cdot 3\cdot 5\cdots(2s+1)}$$

may be useful.

11.1.9 Show that

$$J_0(x) = \frac{2}{\pi}\int_0^1 \frac{\cos xt}{\sqrt{1-t^2}} dt.$$

This integral is a Fourier cosine transform (compare Section 15.3). The corresponding Fourier sine transform,

$$J_0(x) = \frac{2}{\pi}\int_1^\infty \frac{\sin xt}{\sqrt{t^2-1}} dt,$$

is established in Section 11.4 (Exercise 11.4.6) using a Hankel function integral representation.

11.1.10 Derive

$$J_n(x) = (-1)^n x^n \left(\frac{1}{x}\frac{d}{dx}\right)^n J_0(x).$$

Hint. Try mathematical induction.

11.1.11 Show that between any two consecutive zeros of $J_n(x)$ there is one and only one zero of $J_{n+1}(x)$.
Hint. Equations (11.15) and (11.17) may be useful.

11.1.12 An analysis of antenna radiation patterns for a system with a circular aperture involves the equation

$$g(u) = \int_0^1 f(r) J_0(ur) r\, dr.$$

If $f(r) = 1 - r^2$, show that

$$g(u) = \frac{2}{u^2} J_2(u).$$

11.1.13 The differential cross section in a nuclear scattering experiment is given by $d\sigma/d\Omega = |f(\theta)|^2$. An approximate treatment leads to

$$f(\theta) = \frac{-ik}{2\pi}\int_0^{2\pi}\int_0^R \exp[ik\rho\sin\theta\sin\varphi]\rho\,d\rho\,d\varphi.$$

Here θ is an angle through which the scattered particle is scattered. R is the nuclear radius. Show that

$$\frac{d\sigma}{d\Omega} = \left(\pi R^2\right)\frac{1}{\pi}\left[\frac{J_1(kR\sin\theta)}{\sin\theta}\right]^2.$$

11.1.14 A set of functions $C_n(x)$ satisfies the recurrence relations

$$C_{n-1}(x) - C_{n+1}(x) = \frac{2n}{x}C_n(x),$$
$$C_{n-1}(x) + C_{n+1}(x) = 2C_n'(x).$$

(a) What linear second-order ODE does the $C_n(x)$ satisfy?
(b) By a change of variable transform your ODE into Bessel's equation. This suggests that $C_n(x)$ may be expressed in terms of Bessel functions of transformed argument.

11.1.15 A particle (mass m) is contained in a right circular cylinder (pillbox) of radius R and height H. The particle is described by a wave function satisfying the Schrödinger wave equation

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\rho,\varphi,z) = E\psi(\rho,\varphi,z)$$

and the condition that the wave function go to zero over the surface of the pillbox. Find the lowest (zero point) permitted energy.

$$\text{ANS. } E = \frac{\hbar^2}{2m}\left[\left(\frac{z_{pq}}{R}\right)^2 + \left(\frac{n\pi}{H}\right)^2\right],$$
$$E_{\min} = \frac{\hbar^2}{2m}\left[\left(\frac{2.405}{R}\right)^2 + \left(\frac{\pi}{H}\right)^2\right],$$

where z_{pq} is the qth zero of J_p and the index p is fixed by the azimuthal dependence.

11.1.16 (a) Show by direct differentiation and substitution that

$$J_\nu(x) = \frac{1}{2\pi i}\int_C e^{(x/2)(t-1/t)}t^{-\nu-1}\,dt$$

or that the equivalent equation,

$$J_\nu(x) = \frac{1}{2\pi i}\left(\frac{x}{2}\right)^\nu\int e^{s-x^2/4s}s^{-\nu-1}\,ds,$$

satisfies Bessel's equation. C is the contour shown in Fig. 11.4. The negative real axis is the cut line.

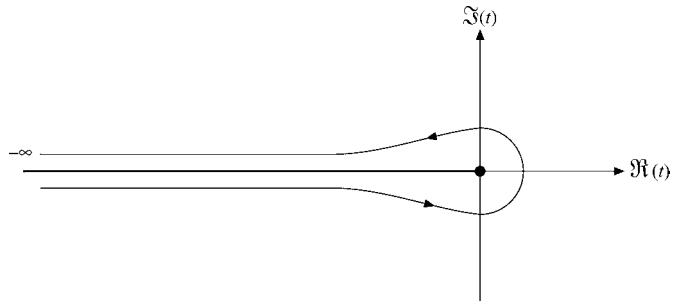

FIGURE 11.4 Bessel function contour.

Hint. Show that the total integrand (after substituting in Bessel's differential equation) may be written as a total derivative:

$$\frac{d}{dt}\left\{\exp\left[\frac{x}{2}\left(t-\frac{1}{t}\right)\right]t^{-\nu}\left[\nu+\frac{x}{2}\left(t+\frac{1}{t}\right)\right]\right\}.$$

(b) Show that the first integral (with n an integer) may be transformed into

$$J_n(x)=\frac{1}{2\pi}\int_0^{2\pi} e^{i(x\sin\theta-n\theta)}\,d\theta=\frac{i^{-n}}{2\pi}\int_0^{2\pi} e^{i(x\cos\theta+n\theta)}\,d\theta.$$

11.1.17 The contour C in Exercise 11.1.16 is deformed to the path $-\infty$ to -1, unit circle $e^{-i\pi}$ to $e^{i\pi}$, and finally -1 to $-\infty$. Show that

$$J_\nu(x)=\frac{1}{\pi}\int_0^{\pi}\cos(\nu\theta-x\sin\theta)\,d\theta-\frac{\sin\nu\pi}{\pi}\int_0^{\infty} e^{-\nu\theta-x\sinh\theta}\,d\theta.$$

This is Bessel's integral.
Hint. The negative values of the variable of integration u may be handled by using

$$u=te^{\pm ix}.$$

11.1.18 (a) Show that

$$J_\nu(x)=\frac{2}{\pi^{1/2}(\nu-\frac{1}{2})!}\left(\frac{x}{2}\right)^{\nu}\int_0^{\pi/2}\cos(x\sin\theta)\cos^{2\nu}\theta\,d\theta,$$

where $\nu>-\frac{1}{2}$.
Hint. Here is a chance to use series expansion and term-by-term integration. The formulas of Section 8.4 will prove useful.

(b) Transform the integral in part (a) into

$$J_\nu(x) = \frac{1}{\pi^{1/2}(\nu - \frac{1}{2})!}\left(\frac{x}{2}\right)^\nu \int_0^\pi \cos(x\cos\theta)\sin^{2\nu}\theta\, d\theta$$
$$= \frac{1}{\pi^{1/2}(\nu - \frac{1}{2})!}\left(\frac{x}{2}\right)^\nu \int_0^\pi e^{\pm ix\cos\theta}\sin^{2\nu}\theta\, d\theta$$
$$= \frac{1}{\pi^{1/2}(\nu - \frac{1}{2})!}\left(\frac{x}{2}\right)^\nu \int_{-1}^1 e^{\pm ipx}(1-p^2)^{\nu-1/2}\, dp.$$

These are alternate integral representations of $J_\nu(x)$.

11.1.19 (a) From

$$J_\nu(x) = \frac{1}{2\pi i}\left(\frac{x}{2}\right)^\nu \int t^{-\nu-1}e^{t-x^2/4t}\, dt$$

derive the recurrence relation

$$J'_\nu(x) = \frac{\nu}{x}J_\nu(x) - J_{\nu+1}(x).$$

(b) From

$$J_\nu(x) = \frac{1}{2\pi i}\int t^{-\nu-1}e^{(x/2)(t-1/t)}\, dt$$

derive the recurrence relation

$$J'_\nu(x) = \tfrac{1}{2}\big[J_{\nu-1}(x) - J_{\nu+1}(x)\big].$$

11.1.20 Show that the recurrence relation

$$J'_n(x) = \tfrac{1}{2}\big[J_{n-1}(x) - J_{n+1}(x)\big]$$

follows directly from differentiation of

$$J_n(x) = \frac{1}{\pi}\int_0^\pi \cos(n\theta - x\sin\theta)\, d\theta.$$

11.1.21 Evaluate

$$\int_0^\infty e^{-ax}J_0(bx)\, dx, \qquad a, b > 0.$$

Actually the results hold for $a \geq 0$, $-\infty < b < \infty$. This is a Laplace transform of J_0. *Hint*. Either an integral representation of J_0 or a series expansion will be helpful.

11.1.22 Using trigonometric forms, verify that

$$J_0(br) = \frac{1}{2\pi}\int_0^{2\pi} e^{ibr\sin\theta}\, d\theta.$$

11.1.23 (a) Plot the intensity (Φ^2 of Eq. (11.35)) as a function of $(\sin\alpha/\lambda)$ along a diameter of the circular diffraction pattern. Locate the first two minima.

(b) What fraction of the total light intensity falls within the central maximum?
Hint. $[J_1(x)]^2/x$ may be written as a derivative and the area integral of the intensity integrated by inspection.

11.1.24 The fraction of light incident on a circular aperture (normal incidence) that is transmitted is given by

$$T = 2\int_0^{2ka} J_2(x)\frac{dx}{x} - \frac{1}{2ka}\int_0^{2ka} J_2(x)\,dx.$$

Here a is the radius of the aperture and k is the wave number, $2\pi/\lambda$. Show that

$$\text{(a)}\quad T = 1 - \frac{1}{ka}\sum_{n=0}^{\infty} J_{2n+1}(2ka), \qquad \text{(b)}\quad T = 1 - \frac{1}{2ka}\int_0^{2ka} J_0(x)\,dx.$$

11.1.25 The amplitude $U(\rho, \varphi, t)$ of a vibrating circular membrane of radius a satisfies the wave equation

$$\nabla^2 U - \frac{1}{v^2}\frac{\partial^2 U}{\partial t^2} = 0.$$

Here v is the phase velocity of the wave fixed by the elastic constants and whatever damping is imposed.

(a) Show that a solution is

$$U(\rho, \varphi, t) = J_m(k\rho)\left(a_1 e^{im\varphi} + a_2 e^{-im\varphi}\right)\left(b_1 e^{i\omega t} + b_2 e^{-i\omega t}\right).$$

(b) From the Dirichlet boundary condition, $J_m(ka) = 0$, find the allowable values of the wavelength $\lambda(k = 2\pi/\lambda)$.

Note. There are other Bessel functions besides J_m, but they all diverge at $\rho = 0$. This is shown explicitly in Section 11.3. The divergent behavior is actually implicit in Eq. (11.6).

11.1.26 Example 11.1.2 describes the TM modes of electromagnetic cavity oscillation. The transverse electric (TE) modes differ, in that we work from the z component of the magnetic induction **B**:

$$\nabla^2 B_z + \alpha^2 B_z = 0$$

with boundary conditions

$$B_z(0) = B_z(l) = 0 \qquad \text{and} \qquad \left.\frac{\partial B_z}{\partial \rho}\right|_{\rho=0} = 0.$$

Show that the TE resonant frequencies are given by

$$\omega_{mnp} = c\sqrt{\frac{\beta_{mn}^2}{a^2} + \frac{p^2\pi^2}{l^2}}, \qquad p = 1, 2, 3, \ldots.$$

11.1.27 Plot the three lowest TM and the three lowest TE angular resonant frequencies, ω_{mnp}, as a function of the radius/length (a/l) ratio for $0 \le a/l \le 1.5$.
Hint. Try plotting ω^2 (in units of c^2/a^2) versus $(a/l)^2$. Why this choice?

11.1.28 A thin conducting disk of radius a carries a charge q. Show that the potential is described by

$$\varphi(r,z) = \frac{q}{4\pi\varepsilon_0 a}\int_0^\infty e^{-k|z|} J_0(kr)\frac{\sin ka}{k}dk,$$

where J_0 is the usual Bessel function and r and z are the familiar cylindrical coordinates.
Note. This is a difficult problem. One approach is through Fourier transforms such as Exercise 15.3.11. For a discussion of the physical problem see Jackson (*Classical Electrodynamics* in Additional Readings).

11.1.29 Show that

$$\int_0^a x^m J_n(x)\,dx, \qquad m \ge n \ge 0,$$

(a) is integrable in terms of Bessel functions and powers of x (such as $a^p J_q(a)$) for $m+n$ odd;
(b) may be reduced to integrated terms plus $\int_0^a J_0(x)dx$ for $m+n$ even.

11.1.30 Show that

$$\int_0^{\alpha_{0n}} \left(1 - \frac{y}{\alpha_{0n}}\right) J_0(y)\,y\,dy = \frac{1}{\alpha_{0n}}\int_0^{\alpha_{0n}} J_0(y)\,dy.$$

Here α_{0n} is the nth root of $J_0(y)$. This relation is useful (see Exercise 11.2.11): The expression on the right is easier and quicker to evaluate — and much more accurate. Taking the difference of two terms in the expression on the left leads to a large relative error.

11.1.31 The circular aperature diffraction amplitude Φ of Eq. (17.35) is proportional to $f(z) = J_1(z)/z$. The corresponding single slit diffraction amplitude is proportional to $g(z) = \sin z/z$.

(a) Calculate and plot $f(z)$ and $g(z)$ for $z = 0.0(0.2)12.0$.
(b) Locate the two lowest values of $z(z > 0)$ for which $f(z)$ takes on an extreme value. Calculate the corresponding values of $f(z)$.
(c) Locate the two lowest values of $z(z > 0)$ for which $g(z)$ takes on an extreme value. Calculate the corresponding values of $g(z)$.

11.1.32 Calculate the electrostatic potential of a charged disk $\varphi(r,z)$ from the integral form of Exercise 11.1.28. Calculate the potential for $r/a = 0.0(0.5)2.0$ and $z/a = 0.25(0.25)1.25$. Why is $z/a = 0$ omitted? Exercise 12.3.17 is a spherical harmonic version of this same problem.

11.2 ORTHOGONALITY

If Bessel's equation, Eq. (11.22a), is divided by ρ, we see that it becomes self-adjoint, and therefore, by the Sturm–Liouville theory, Section 10.2, the solutions are expected to be orthogonal — if we can arrange to have appropriate boundary conditions satisfied. To take care of the boundary conditions for a finite interval $[0, a]$, we introduce parameters a and $\alpha_{\nu m}$ into the argument of J_ν to get $J_\nu(\alpha_{\nu m}\rho/a)$. Here a is the upper limit of the cylindrical radial coordinate ρ. From Eq. (11.22a),

$$\rho\frac{d^2}{d\rho^2}J_\nu\left(\alpha_{\nu m}\frac{\rho}{a}\right)+\frac{d}{d\rho}J_\nu\left(\alpha_{\nu m}\frac{\rho}{a}\right)+\left(\frac{\alpha_{\nu m}^2\rho}{a^2}-\frac{\nu^2}{\rho}\right)J_\nu\left(\alpha_{\nu m}\frac{\rho}{a}\right)=0. \tag{11.45}$$

Changing the parameter $\alpha_{\nu m}$ to $\alpha_{\nu n}$, we find that $J_\nu(\alpha_{\nu n}\rho/a)$ satisfies

$$\rho\frac{d^2}{d\rho^2}J_\nu\left(\alpha_{\nu n}\frac{\rho}{a}\right)+\frac{d}{d\rho}J_\nu\left(\alpha_{\nu n}\frac{\rho}{a}\right)+\left(\frac{\alpha_{\nu n}^2\rho}{a^2}-\frac{\nu^2}{\rho}\right)J_\nu\left(\alpha_{\nu n}\frac{\rho}{a}\right)=0. \tag{11.45a}$$

Proceeding as in Section 10.2, we multiply Eq. (11.45) by $J_\nu(\alpha_{\nu n}\rho/a)$ and Eq. (11.45a) by $J_\nu(\alpha_{\nu m}\rho/a)$ and subtract, obtaining

$$\begin{aligned}&J_\nu\left(\alpha_{\nu n}\frac{\rho}{a}\right)\frac{d}{d\rho}\left[\rho\frac{d}{d\rho}J_\nu\left(\alpha_{\nu m}\frac{\rho}{a}\right)\right]-J_\nu\left(\alpha_{\nu m}\frac{\rho}{a}\right)\frac{d}{d\rho}\left[\rho\frac{d}{d\rho}J_\nu\left(\alpha_{\nu n}\frac{\rho}{a}\right)\right]\\&\quad=\frac{\alpha_{\nu n}^2-\alpha_{\nu m}^2}{a^2}\rho J_\nu\left(\alpha_{\nu m}\frac{\rho}{a}\right)J_\nu\left(\alpha_{\nu n}\frac{\rho}{a}\right).\end{aligned} \tag{11.46}$$

Integrating from $\rho=0$ to $\rho=a$, we obtain

$$\begin{aligned}&\int_0^a J_\nu\left(\alpha_{\nu n}\frac{\rho}{a}\right)\frac{d}{d\rho}\left[\rho\frac{d}{d\rho}J_\nu\left(\alpha_{\nu m}\frac{\rho}{a}\right)\right]d\rho-\int_0^a J_\nu\left(\alpha_{\nu m}\frac{\rho}{a}\right)\frac{d}{d\rho}\left[\rho\frac{d}{d\rho}J_\nu\left(\alpha_{\nu n}\frac{\rho}{a}\right)\right]d\rho\\&\quad=\frac{\alpha_{\nu n}^2-\alpha_{\nu m}^2}{a^2}\int_0^a J_\nu\left(\alpha_{\nu m}\frac{\rho}{a}\right)J_\nu\left(\alpha_{\nu n}\frac{\rho}{a}\right)\rho\,d\rho.\end{aligned} \tag{11.47}$$

Upon integrating by parts, we see that the left-hand side of Eq. (11.47) becomes

$$\left.\rho J_\nu\left(\alpha_{\nu n}\frac{\rho}{a}\right)\frac{d}{d\rho}J_\nu\left(\alpha_{\nu m}\frac{\rho}{a}\right)\right|_0^a-\left.\rho J_\nu\left(\alpha_{\nu m}\frac{\rho}{a}\right)\frac{d}{d\rho}J_\nu\left(\alpha_{\nu n}\frac{\rho}{a}\right)\right|_0^a. \tag{11.48}$$

For $\nu\geq 0$ the factor ρ guarantees a zero at the lower limit, $\rho=0$. Actually the lower limit on the index ν may be extended down to $\nu>-1$, Exercise 11.2.4.[12] At $\rho=a$, each expression vanishes if we choose the parameters $\alpha_{\nu n}$ and $\alpha_{\nu m}$ to be zeros, or roots of J_ν; that is, $J_\nu(\alpha_{\nu m})=0$. The subscripts now become meaningful: $\alpha_{\nu m}$ is the mth zero of J_ν.

With this choice of parameters, the left-hand side vanishes (the Sturm–Liouville boundary conditions are satisfied) and for $m\neq n$,

$$\int_0^a J_\nu\left(\alpha_{\nu m}\frac{\rho}{a}\right)J_\nu\left(\alpha_{\nu n}\frac{\rho}{a}\right)\rho\,d\rho=0. \tag{11.49}$$

This gives us orthogonality over the interval $[0, a]$.

[12] The case $\nu=-1$ reverts to $\nu=+1$, Eq. (11.8).

Normalization

The normalization integral may be developed by returning to Eq. (11.48), setting $\alpha_{\nu n} = \alpha_{\nu m} + \varepsilon$, and taking the limit $\varepsilon \to 0$ (compare Exercise 11.2.2). With the aid of the recurrence relation, Eq. (11.16), the result may be written as

$$\int_0^a \left[J_\nu\left(\alpha_{\nu m}\frac{\rho}{a}\right)\right]^2 \rho\, d\rho = \frac{a^2}{2}\left[J_{\nu+1}(\alpha_{\nu m})\right]^2. \tag{11.50}$$

Bessel Series

If we assume that the set of Bessel functions $J_\nu(\alpha_{\nu m}\rho/a))(\nu$ fixed, $m = 1, 2, 3, \ldots)$ is complete, then any well-behaved but otherwise arbitrary function $f(\rho)$ may be expanded in a Bessel series (Bessel–Fourier or Fourier–Bessel)

$$f(\rho) = \sum_{m=1}^{\infty} c_{\nu m} J_\nu\left(\alpha_{\nu m}\frac{\rho}{a}\right), \qquad 0 \le \rho \le a, \qquad \nu > -1. \tag{11.51}$$

The coefficients $c_{\nu m}$ are determined by using Eq. (11.50),

$$c_{\nu m} = \frac{2}{a^2[J_{\nu+1}(\alpha_{\nu m})]^2}\int_0^a f(\rho) J_\nu\left(\alpha_{\nu m}\frac{\rho}{a}\right)\rho\, d\rho. \tag{11.52}$$

A similar series expansion involving $J_\nu(\beta_{\nu m}\rho/a)$ with $(d/d\rho)J_\nu(\beta_{\nu m}\rho/a)|_{\rho=a} = 0$ is included in Exercises 11.2.3 and 11.2.6(b).

Example 11.2.1 ELECTROSTATIC POTENTIAL IN A HOLLOW CYLINDER

From Table 9.3 of Section 9.3 (with α replaced by k), our solution of Laplace's equation in circular cylindrical coordinates is a linear combination of

$$\psi_{km}(\rho, \varphi, z) = J_m(k\rho)[a_m \sin m\varphi + b_m \cos m\varphi]\left[c_1 e^{kz} + c_2 e^{-kz}\right]. \tag{11.53}$$

The particular linear combination is determined by the boundary conditions to be satisfied. Our cylinder here has a radius a and a height l. The top end section has a potential distribution $\psi(\rho, \varphi)$. Elsewhere on the surface the potential is zero.[13] The problem is to find the electrostatic potential

$$\psi(\rho, \varphi, z) = \sum_{k,m} \psi_{km}(\rho, \varphi, z) \tag{11.54}$$

everywhere in the interior.

For convenience, the circular cylindrical coordinates are placed as shown in Fig. 11.3. Since $\psi(\rho, \varphi, 0) = 0$, we take $c_1 = -c_2 = \frac{1}{2}$. The z dependence becomes $\sinh kz$, vanishing at $z = 0$. The requirement that $\psi = 0$ on the cylindrical sides is met by requiring the separation constant k to be

$$k = k_{mn} = \frac{\alpha_{mn}}{a}, \tag{11.55}$$

[13] If $\psi = 0$ at $z = 0, l$, but $\psi \neq 0$ for $\rho = a$, the modified Bessel functions, Section 11.5, are involved.

where the first subscript, m, gives the index of the Bessel function, whereas the second subscript identifies the particular zero of J_m.

The electrostatic potential becomes

$$\psi(\rho,\varphi,z)=\sum_{m=0}^{\infty}\sum_{n=1}^{\infty}J_m\left(\alpha_{mn}\frac{\rho}{a}\right)\cdot[a_{mn}\sin m\varphi+b_{mn}\cos m\varphi]\cdot\sinh\left(\alpha_{mn}\frac{z}{a}\right). \tag{11.56}$$

Equation (11.56) is a double series: a Bessel series in ρ and a Fourier series in φ.

At $z=l$, $\psi=\psi(\rho,\varphi)$, a known function of ρ and φ. Therefore

$$\psi(\rho,\varphi)=\sum_{m=0}^{\infty}\sum_{n=1}^{\infty}J_m\left(\alpha_{mn}\frac{\rho}{a}\right)\cdot[a_{mn}\sin m\varphi+b_{mn}\cos m\varphi]\cdot\sinh\left(\alpha_{mn}\frac{l}{a}\right). \tag{11.57}$$

The constants a_{mn} and b_{mn} are evaluated by using Eqs. (11.49) and (11.50) and the corresponding equations for $\sin\varphi$ and $\cos\varphi$ (Example 10.2.1 and Eqs. (14.2), (14.3), (14.15) to (14.17)). We find[14]

$$\left.\begin{matrix}a_{mn}\\ b_{mn}\end{matrix}\right\}=2\left[\pi a^2\sinh\left(\alpha_{mn}\frac{l}{a}\right)J_{m+1}^2(\alpha_{mn})\right]^{-1}\cdot\int_0^{2\pi}\int_0^a\psi(\rho,\varphi)J_m\left(\alpha_{mn}\frac{\rho}{a}\right)\left\{\begin{matrix}\sin m\varphi\\ \cos m\varphi\end{matrix}\right\}\rho\,d\rho\,d\varphi. \tag{11.58}$$

These are definite integrals, that is, numbers. Substituting back into Eq. (11.56), the series is specified and the potential $\psi(\rho,\varphi,z)$ is determined. ■

Continuum Form

The Bessel series, Eq. (11.51), and Exercise 11.2.6 apply to expansions over the finite interval $[0,a]$. If $a\to\infty$, then the series forms may be expected to go over into integrals. The discrete roots $\alpha_{\nu m}$ become a continuous variable α. A similar situation is encountered in the Fourier series, Section 15.2. The development of the Bessel integral from the Bessel series is left as Exercise 11.2.8.

For operations with a continuum of Bessel functions, $J_\nu(\alpha\rho)$, a key relation is the Bessel function **closure** equation,

$$\int_0^\infty J_\nu(\alpha\rho)J_\nu(\alpha'\rho)\rho\,d\rho=\frac{1}{\alpha}\delta(\alpha-\alpha'),\qquad \nu>-\frac{1}{2}. \tag{11.59}$$

This may be proved by the use of Hankel transforms, Section 15.1. An alternate approach, starting from a relation similar to Eq. (10.82), is given by Morse and Feshbach, Section 6.3. A second kind of orthogonality (varying the index) is developed for spherical Bessel functions in Section 11.7.

[14] If $m=0$, the factor 2 is omitted (compare Eq. (14.16)).

Exercises

11.2.1 Show that

$$\left(a^2-b^2\right)\int_0^P J_\nu(ax)J_\nu(bx)x\,dx = P\left[bJ_\nu(aP)J'_\nu(bP)-aJ'_\nu(aP)J_\nu(bP)\right],$$

with

$$J'_\nu(aP)=\frac{d}{d(ax)}J_\nu(ax)\Big|_{x=P},$$

$$\int_0^P \left[J_\nu(ax)\right]^2 x\,dx = \frac{P^2}{2}\left\{\left[J'_\nu(aP)\right]^2+\left(1-\frac{\nu^2}{a^2P^2}\right)\left[J_\nu(aP)\right]^2\right\}, \qquad \nu>-1.$$

These two integrals are usually called the **first and second Lommel integrals**.
Hint. We have the development of the orthogonality of the Bessel functions as an analogy.

11.2.2 Show that

$$\int_0^a \left[J_\nu\left(\alpha_{\nu m}\frac{\rho}{a}\right)\right]^2 \rho\,d\rho = \frac{a^2}{2}\left[J_{\nu+1}(\alpha_{\nu m})\right]^2, \qquad \nu>-1.$$

Here $\alpha_{\nu m}$ is the mth zero of J_ν.
Hint. With $\alpha_{\nu n}=\alpha_{\nu m}+\varepsilon$, expand $J_\nu[(\alpha_{\nu m}+\varepsilon)\rho/a]$ about $\alpha_{\nu m}\rho/a$ by a Taylor expansion.

11.2.3 (a) If $\beta_{\nu m}$ is the mth zero of $(d/d\rho)J_\nu(\beta_{\nu m}\rho/a)$, show that the Bessel functions are orthogonal over the interval $[0,a]$ with an orthogonality integral

$$\int_0^a J_\nu\left(\beta_{\nu m}\frac{\rho}{a}\right)J_\nu\left(\beta_{\nu n}\frac{\rho}{a}\right)\rho\,d\rho=0, \qquad m\neq n, \qquad \nu>-1.$$

(b) Derive the corresponding normalization integral ($m=n$).

$$\text{ANS. } \frac{a^2}{2}\left(1-\frac{\nu^2}{\beta_{\nu m}^2}\right)\left[J_\nu(\beta_{\nu m})\right]^2, \qquad \nu>-1.$$

11.2.4 Verify that the orthogonality equation, Eq. (11.49), and the normalization equation, Eq. (11.50), hold for $\nu>-1$.
Hint. Using power-series expansions, examine the behavior of Eq. (11.48) as $\rho\to 0$.

11.2.5 From Eq. (11.49) develop a proof that $J_\nu(z)$, $\nu>-1$, has no complex roots (with nonzero imaginary part).
Hint.

(a) Use the series form of $J_\nu(z)$ to exclude pure imaginary roots.
(b) Assume $\alpha_{\nu m}$ to be complex and take $\alpha_{\nu n}$ to be $\alpha^*_{\nu m}$.

11.2.6 (a) In the series expansion

$$f(\rho)=\sum_{m=1}^{\infty} c_{\nu m}J_\nu\left(\alpha_{\nu m}\frac{\rho}{a}\right), \qquad 0\le\rho\le a, \qquad \nu>-1,$$

with $J_\nu(\alpha_{\nu m}) = 0$, show that the coefficients are given by

$$c_{\nu m} = \frac{2}{a^2[J_{\nu+1}(\alpha_{\nu m})]^2} \int_0^a f(\rho) J_\nu\left(\alpha_{\nu m}\frac{\rho}{a}\right)\rho\, d\rho.$$

(b) In the series expansion

$$f(\rho) = \sum_{m=1}^{\infty} d_{\nu m} J_\nu\left(\beta_{\nu m}\frac{\rho}{a}\right), \qquad 0 \le \rho \le a, \qquad \nu > -1,$$

with $(d/d\rho)J_\nu(\beta_{\nu m}\rho/a)\,|_{\rho=a} = 0$, show that the coefficients are given by

$$d_{\nu m} = \frac{2}{a^2(1-\nu^2/\beta_{\nu m}^2)[J_\nu(\beta_{\nu m})]^2} \int_0^a f(\rho) J_\nu\left(\beta_{\nu m}\frac{\rho}{a}\right)\rho\, d\rho.$$

11.2.7 A right circular cylinder has an electrostatic potential of $\psi(\rho, \varphi)$ on both ends. The potential on the curved cylindrical surface is zero. Find the potential at all interior points. *Hint*. Choose your coordinate system and adjust your z dependence to exploit the symmetry of your potential.

11.2.8 For the continuum case, show that Eqs. (11.51) and (11.52) are replaced by

$$f(\rho) = \int_0^\infty a(\alpha) J_\nu(\alpha\rho)\, d\alpha,$$

$$a(\alpha) = \alpha \int_0^\infty f(\rho) J_\nu(\alpha\rho)\rho\, d\rho.$$

Hint. The corresponding case for sines and cosines is worked out in Section 15.2. These are Hankel transforms. A derivation for the special case $\nu = 0$ is the topic of Exercise 15.1.1.

11.2.9 A function $f(x)$ is expressed as a Bessel series:

$$f(x) = \sum_{n=1}^{\infty} a_n J_m(\alpha_{mn} x),$$

with α_{mn} the nth root of J_m. Prove the Parseval relation,

$$\int_0^1 \big[f(x)\big]^2 x\, dx = \frac{1}{2}\sum_{n=1}^{\infty} a_n^2 \big[J_{m+1}(\alpha_{mn})\big]^2.$$

11.2.10 Prove that

$$\sum_{n=1}^{\infty} (\alpha_{mn})^{-2} = \frac{1}{4(m+1)}.$$

Hint. Expand x^m in a Bessel series and apply the Parseval relation.

11.2.11 A right circular cylinder of length l has a potential

$$\psi\left(z = \pm\frac{l}{2}\right) = 100\left(1 - \frac{\rho}{a}\right),$$

where a is the radius. The potential over the curved surface (side) is zero. Using the Bessel series from Exercise 11.2.7, calculate the electrostatic potential for $\rho/a = 0.0(0.2)1.0$ and $z/l = 0.0(0.1)0.5$. Take $a/l = 0.5$.
Hint. From Exercise 11.1.30 you have

$$\int_0^{\alpha_{0n}} \left(1 - \frac{y}{\alpha_{0n}}\right) J_0(y)\, y\, dy.$$

Show that this equals

$$\frac{1}{\alpha_{0n}} \int_0^{\alpha_{0n}} J_0(y)\, dy.$$

Numerical evaluation of this latter form rather than the former is both faster and more accurate.
Note. For $\rho/a = 0.0$ and $z/l = 0.5$ the convergence is slow, 20 terms giving only 98.4 rather than 100.

Check value. For $\rho/a = 0.4$ and $z/l = 0.3$,
$\psi = 24.558$.

11.3 NEUMANN FUNCTIONS, BESSEL FUNCTIONS OF THE SECOND KIND

From the theory of ODEs it is known that Bessel's equation has two independent solutions. Indeed, for nonintegral order ν we have already found two solutions and labeled them $J_\nu(x)$ and $J_{-\nu}(x)$, using the infinite series (Eq. (11.5)). The trouble is that when ν is integral, Eq. (11.8) holds and we have but one independent solution. A second solution may be developed by the methods of Section 9.6. This yields a perfectly good second solution of Bessel's equation but is not the standard form.

Definition and Series Form

As an alternate approach, we take the particular linear combination of $J_\nu(x)$ and $J_{-\nu}(x)$

$$\boxed{N_\nu(x) = \frac{\cos\nu\pi\, J_\nu(x) - J_{-\nu}(x)}{\sin\nu\pi}.} \tag{11.60}$$

This is the Neumann function (Fig. 11.5).[15] For nonintegral ν, $N_\nu(x)$ clearly satisfies Bessel's equation, for it is a linear combination of known solutions $J_\nu(x)$ and $J_{-\nu}(x)$.

[15] In AMS-55 (see footnote 4 in Chapter 5 or Additional Readings of Chapter 8 p. for this ref.) and in most mathematics tables, this is labeled $Y_\nu(x)$.

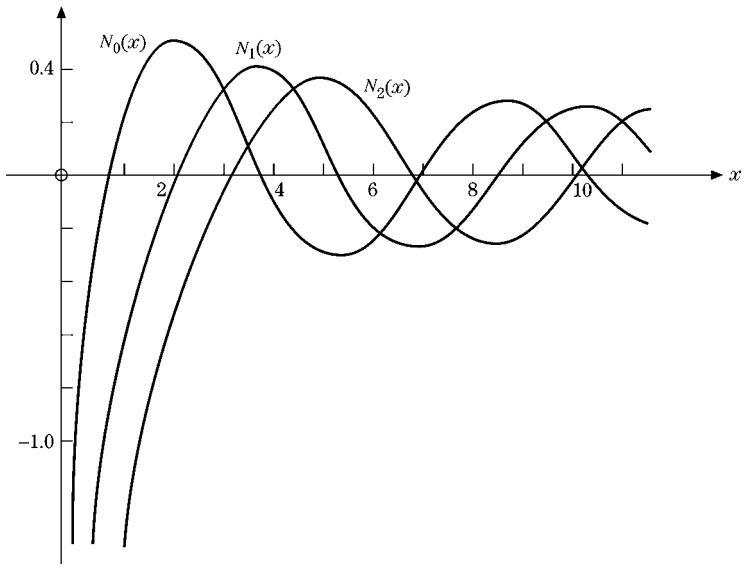

FIGURE 11.5 Neumann functions $N_0(x)$, $N_1(x)$, and $N_2(x)$.

Substituting the power-series Eq. (11.6) for $n \to \nu$ (given in Exercise 11.1.7) yields

$$N_\nu(x) = -\frac{(\nu-1)!}{\pi}\left(\frac{2}{x}\right)^\nu + \cdots,^{16} \tag{11.61}$$

for $\nu > 0$. However, for integral ν, $\nu = n$, Eq. (11.8) applies and Eq. (11.60)[16] becomes indeterminate. The definition of $N_\nu(x)$ was chosen deliberately for this indeterminate property. Again substituting the power series and evaluating $N_\nu(x)$ for $\nu \to 0$ by l'Hôpital's rule for indeterminate forms, we obtain the limiting value

$$N_0(x) = \frac{2}{\pi}(\ln x + \gamma - \ln 2) + O\left(x^2\right) \tag{11.62}$$

for $n = 0$ and $x \to 0$, using

$$\nu!(-\nu)! = \frac{\pi\nu}{\sin\pi\nu} \tag{11.63}$$

from Eq. (8.32). The first and third terms in Eq. (11.62) come from using $(d/d\nu)(x/2)^\nu = (x/2)^\nu \ln(x/2)$, while γ comes from $(d/d\nu)\nu!$ for $\nu \to 0$ using Eqs. (8.38) and (8.40). For $n > 0$ we obtain similarly

$$N_n(x) = -\frac{1}{\pi}(n-1)!\left(\frac{2}{x}\right)^n + \cdots + \frac{2}{\pi}\left(\frac{x}{2}\right)^n \frac{1}{n!}\ln\left(\frac{x}{2}\right) + \cdots. \tag{11.64}$$

Equations (11.62) and (11.64) exhibit the logarithmic dependence that was to be expected. This, of course, verifies the independence of J_n and N_n.

[16] Note that this limiting form applies to both integral and nonintegral values of the index ν.

Other Forms

As with all the other Bessel functions, $N_\nu(x)$ has integral representations. For $N_0(x)$ we have

$$N_0(x) = -\frac{2}{\pi}\int_0^\infty \cos(x\cosh t)\,dt = -\frac{2}{\pi}\int_1^\infty \frac{\cos(xt)}{(t^2-1)^{1/2}}\,dt, \qquad x>0.$$

These forms can be derived as the imaginary part of the Hankel representations of Exercise 11.4.7. The latter form is a Fourier cosine transform.

To verify that $N_\nu(x)$, our Neumann function (Fig. 11.5) or Bessel function of the second kind, actually does satisfy Bessel's equation for integral n, we may proceed as follows. L'Hôpital's rule applied to Eq. (11.60) yields

$$\begin{aligned} N_n(x) &= \left.\frac{(d/d\nu)[\cos\nu\pi\, J_\nu(x) - J_{-\nu}(x)]}{(d/d\nu)\sin\nu\pi}\right|_{\nu=n} \\ &= \frac{-\pi\sin n\pi\, J_n(x) + [\cos n\pi\,\partial J_\nu/\partial\nu - \partial J_{-\nu}/\partial\nu]|_{\nu=n}}{\pi\cos n\pi} \\ &= \frac{1}{\pi}\left[\frac{\partial J_\nu(x)}{\partial\nu} - (-1)^n\frac{\partial J_{-\nu}(x)}{\partial\nu}\right]\Bigg|_{\nu=n}. \end{aligned} \tag{11.65}$$

Differentiating Bessel's equation for $J_{\pm\nu}(x)$ with respect to ν, we have

$$x^2\frac{d^2}{dx^2}\left(\frac{\partial J_{\pm\nu}}{\partial\nu}\right) + x\frac{d}{dx}\left(\frac{\partial J_{\pm\nu}}{\partial\nu}\right) + \left(x^2-\nu^2\right)\frac{\partial J_{\pm\nu}}{\partial\nu} = 2\nu J_{\pm\nu}. \tag{11.66}$$

Multiplying the equation for $J_{-\nu}$ by $(-1)^\nu$, subtracting from the equation for J_ν (as suggested by Eq. (11.65)), and taking the limit $\nu\to n$, we obtain

$$x^2\frac{d^2}{dx^2}N_n + x\frac{d}{dx}N_n + \left(x^2-n^2\right)N_n = \frac{2n}{\pi}\left[J_n - (-1)^n J_{-n}\right]. \tag{11.67}$$

For $\nu = n$, an integer, the right-hand side vanishes by Eq. (11.8) and $N_n(x)$ is seen to be a solution of Bessel's equation. The most general solution for any ν can therefore be written as

$$y(x) = AJ_\nu(x) + BN_\nu(x). \tag{11.68}$$

It is seen from Eqs. (11.62) and (11.64) that N_n diverges, at least logarithmically. Any boundary condition that requires the solution to be finite at the origin (as in our vibrating circular membrane (Section 11.1)) automatically excludes $N_n(x)$. Conversely, in the absence of such a requirement, $N_n(x)$ must be considered.

To a certain extent the definition of the Neumann function $N_n(x)$ is arbitrary. Equations (11.62) and (11.64) contain terms of the form $a_nJ_n(x)$. Clearly, any finite value of the constant a_n would still give us a second solution of Bessel's equation. Why should a_n have the particular value implicit in Eqs. (11.62) and (11.64)? The answer involves the asymptotic dependence developed in Section 11.6. If J_n corresponds to a cosine wave, then N_n corresponds to a sine wave. This simple and convenient asymptotic phase relationship is a consequence of the particular admixture of J_n in N_n.

Recurrence Relations

Substituting Eq. (11.60) for $N_\nu(x)$ (nonintegral ν) into the recurrence relations (Eqs. (11.10) and (11.12) for $J_n(x)$, we see immediately that $N_\nu(x)$ satisfies these same recurrence relations. This actually constitutes another proof that N_ν is a solution. Note that the converse is not necessarily true. All solutions need not satisfy the same recurrence relations. An example of this sort of trouble appears in Section 11.5.

Wronskian Formulas

From Section 9.6 and Exercise 10.1.4 we have the Wronskian formula[17] for solutions of the Bessel equation,

$$u_\nu(x)v'_\nu(x) - u'_\nu(x)v_\nu(x) = \frac{A_\nu}{x}, \tag{11.69}$$

in which A_ν is a parameter that depends on the particular Bessel functions $u_\nu(x)$ and $v_\nu(x)$ being considered. A_ν is a constant in the sense that it is independent of x. Consider the special case

$$u_\nu(x) = J_\nu(x), \qquad v_\nu(x) = J_{-\nu}(x), \tag{11.70}$$

$$J_\nu J'_{-\nu} - J'_\nu J_{-\nu} = \frac{A_\nu}{x}. \tag{11.71}$$

Since A_ν is a constant, it may be identified at any convenient point, such as $x = 0$. Using the first terms in the series expansions (Eqs. (11.5) and (11.6)), we obtain

$$\begin{aligned} J_\nu &\to \frac{x^\nu}{2^\nu \nu!}, \qquad J_{-\nu} \to \frac{2^\nu x^{-\nu}}{(-\nu)!} \\ J'_\nu &\to \frac{\nu x^{\nu-1}}{2^\nu \nu!}, \qquad J'_{-\nu} \to -\frac{\nu 2^\nu x^{-\nu-1}}{(-\nu)!}. \end{aligned} \tag{11.72}$$

Substitution into Eq. (11.69) yields

$$J_\nu(x)J'_{-\nu}(x) - J'_\nu(x)J_{-\nu}(x) = \frac{-2\nu}{x\nu!(-\nu)!} = -\frac{2\sin\nu\pi}{\pi x}, \tag{11.73}$$

using Eq. (8.32). Note that A_ν vanishes for integral ν, as it must, since the nonvanishing of the Wronskian is a test of the independence of the two solutions. By Eq. (11.73), J_n and J_{-n} are clearly linearly dependent.

Using our recurrence relations, we may readily develop a large number of alternate forms, among which are

$$J_\nu J_{-\nu+1} + J_{-\nu}J_{\nu-1} = \frac{2\sin\nu\pi}{\pi x}, \tag{11.74}$$

[17] This result depends on $P(x)$ of Section 9.5 being equal to $p'(x)/p(x)$, the corresponding coefficient of the self-adjoint form of Section 10.1.

$$J_\nu J_{-\nu-1} + J_{-\nu} J_{\nu+1} = -\frac{2\sin\nu\pi}{\pi x}, \tag{11.75}$$

$$J_\nu N'_\nu - J'_\nu N_\nu = \frac{2}{\pi x}, \tag{11.76}$$

$$J_\nu N_{\nu+1} - J_{\nu+1} N_\nu = -\frac{2}{\pi x}. \tag{11.77}$$

Many more will be found in the references given at chapter's end.

You will recall that in Chapter 9 Wronskians were of great value in two respects: (1) in establishing the linear independence or linear dependence of solutions of differential equations and (2) in developing an integral form of a second solution. Here the specific forms of the Wronskians and Wronskian-derived combinations of Bessel functions are useful primarily to illustrate the general behavior of the various Bessel functions. Wronskians are of great use in checking tables of Bessel functions. In Section 10.5 Wronskians appeared in connection with Green's functions.

Example 11.3.1 COAXIAL WAVE GUIDES

We are interested in an electromagnetic wave confined between the concentric, conducting cylindrical surfaces $\rho = a$ and $\rho = b$. Most of the mathematics is worked out in Section 9.3 and Example 11.1.2. To go from the standing wave of these examples to the traveling wave here, we let $A = iB$, $A = a_{mn}$, $B = b_{mn}$ in Eq. (11.40a) and obtain

$$E_z = \sum_{m,n} b_{mn} J_m(\gamma\rho) e^{\pm im\varphi} e^{i(kz-\omega t)}. \tag{11.78}$$

Additional properties of the components of the electromagnetic wave in the simple cylindrical wave guide are explored in Exercises 11.3.8 and 11.3.9. For the coaxial wave guide one generalization is needed. The origin, $\rho = 0$, is now excluded ($0 < a \le \rho \le b$). Hence the Neumann function $N_m(\gamma\rho)$ may not be excluded. $E_z(\rho, \varphi, z, t)$ becomes

$$E_z = \sum_{m,n} \left[b_{mn} J_m(\gamma\rho) + c_{mn} N_m(\gamma\rho) \right] e^{\pm im\varphi} e^{i(kz-\omega t)}. \tag{11.79}$$

With the condition

$$H_z = 0, \tag{11.80}$$

we have the basic equations for a TM (transverse magnetic) wave.

The (tangential) electric field must vanish at the conducting surfaces (Dirichlet boundary condition), or

$$b_{mn} J_m(\gamma a) + c_{mn} N_m(\gamma a) = 0, \tag{11.81}$$

$$b_{mn} J_m(\gamma b) + c_{mn} N_m(\gamma b) = 0. \tag{11.82}$$

These transcendental equations may be solved for $\gamma(\gamma_{mn})$ and the ratio c_{mn}/b_{mn}. From Example 11.1.2,

$$k^2 = \omega^2\mu_0\varepsilon_0 - \gamma^2 = \frac{\omega^2}{c^2} - \gamma^2. \tag{11.83}$$

Since k^2 must be positive for a real wave, the minimum frequency that will be propagated (in this TM mode) is

$$\omega = \gamma c, \tag{11.84}$$

with γ fixed by the boundary conditions, Eqs. (11.81) and (11.82). This is the cutoff frequency of the wave guide.

There is also a TE (transverse electric) mode, with $E_z = 0$ and H_z given by Eq. (11.79). Then we have Neumann boundary conditions in place of Eqs. (11.81) and (11.82). Finally, for the coaxial guide (not for the plain cylindrical guide, $a = 0$), a TEM (transverse electromagnetic) mode, $E_z = H_z = 0$, is possible. This corresponds to a plane wave, as in free space.

The simpler cases (no Neumann functions, simpler boundary conditions) of a circular wave guide are included as Exercises 11.3.8 and 11.3.9.

To conclude this discussion of Neumann functions, we introduce the Neumann function $N_\nu(x)$ for the following reasons:

1. It is a second, independent solution of Bessel's equation, which completes the general solution.
2. It is required for specific physical problems such as electromagnetic waves in coaxial cables and quantum mechanical scattering theory.
3. It leads to a Green's function for the Bessel equation (Sections 9.7 and 10.5).
4. It leads directly to the two Hankel functions (Section 11.4).

■

Exercises

11.3.1 Prove that the Neumann functions N_n (with n an integer) satisfy the recurrence relations

$$N_{n-1}(x) + N_{n+1}(x) = \frac{2n}{x}N_n(x),$$
$$N_{n-1}(x) - N_{n+1}(x) = 2N_n'(x).$$

Hint. These relations may be proved by differentiating the recurrence relations for J_ν or by using the limit form of N_ν but **not** dividing everything by zero.

11.3.2 Show that

$$N_{-n}(x) = (-1)^n N_n(x).$$

11.3.3 Show that

$$N_0'(x) = -N_1(x).$$

11.3.4 If Y and Z are any two solutions of Bessel's equation, show that

$$Y_\nu(x)Z'_\nu(x) - Y'_\nu(x)Z_\nu(x) = \frac{A_\nu}{x},$$

in which A_ν may depend on ν but is independent of x. This is a special case of Exercise 10.1.4.

11.3.5 Verify the Wronskian formulas

$$J_\nu(x)J_{-\nu+1}(x) + J_{-\nu}(x)J_{\nu-1}(x) = \frac{2\sin\nu\pi}{\pi x},$$

$$J_\nu(x)N'_\nu(x) - J'_\nu(x)N_\nu(x) = \frac{2}{\pi x}.$$

11.3.6 As an alternative to letting x approach zero in the evaluation of the Wronskian constant, we may invoke uniqueness of power series (Section 5.7). The coefficient of x^{-1} in the series expansion of $u_\nu(x)v'_\nu(x) - u'_\nu(x)v_\nu(x)$ is then A_ν. Show by series expansion that the coefficients of x^0 and x^1 of $J_\nu(x)J'_{-\nu}(x) - J'_\nu(x)J_{-\nu}(x)$ are each zero.

11.3.7 (a) By differentiating and substituting into Bessel's ODE, show that

$$\int_0^\infty \cos(x\cosh t)\,dt$$

is a solution.
Hint. You can rearrange the final integral as

$$\int_0^\infty \frac{d}{dt}\left\{x\sin(x\cosh t)\sinh t\right\}dt.$$

(b) Show that

$$N_0(x) = -\frac{2}{\pi}\int_0^\infty \cos(x\cosh t)\,dt$$

is linearly independent of $J_0(x)$.

11.3.8 A cylindrical wave guide has radius r_0. Find the nonvanishing components of the electric and magnetic fields for

(a) TM_{01}, transverse magnetic wave ($H_z = H_\rho = E_\varphi = 0$),
(b) TE_{01}, transverse electric wave ($E_z = E_\rho = H_\varphi = 0$).

The subscripts 01 indicate that the longitudinal component (E_z or H_z) involves J_0 and the boundary condition is satisfied by the **first** zero of J_0 or J'_0.
Hint. All components of the wave have the same factor: $\exp i(kz - \omega t)$.

11.3.9 For a given mode of oscillation the **minimum** frequency that will be passed by a circular cylindrical wave guide (radius r_0) is

$$\nu_{\min} = \frac{c}{\lambda_c},$$

in which λ_c is fixed by the boundary condition

$$J_n\left(\frac{2\pi r_0}{\lambda_c}\right) = 0 \qquad \text{for TM}_{nm} \text{ mode,}$$

$$J_n'\left(\frac{2\pi r_0}{\lambda_c}\right) = 0 \qquad \text{for TE}_{nm} \text{ mode.}$$

The subscript n denotes the order of the Bessel function and m indicates the zero used. Find this cutoff wavelength λ_c for the three TM and three TE modes with the longest cutoff wavelengths. Explain your results in terms of the graph of J_0, J_1, and J_2 (Fig. 11.1).

11.3.10 Write a program that will compute successive roots of the Neumann function $N_n(x)$, that is α_{ns}, where $N_n(\alpha_{ns}) = 0$. Tabulate the first five roots of N_0, N_1, and N_2. Check your values for the roots against those listed in AMS-55 (see Additional Readings of Chapter 8 for the full ref.).

Check value. $\alpha_{12} = 5.42968$.

11.3.11 For the case $m = 0$, $a = 1$, and $b = 2$, the coaxial wave guide boundary conditions lead to

$$f(x) = \frac{J_0(2x)}{N_0(2x)} - \frac{J_0(x)}{N_0(x)}$$

(Fig. 11.6).

(a) Calculate $f(x)$ for $x = 0.0(0.1)10.0$ and plot $f(x)$ versus x to find the approximate location of the roots.

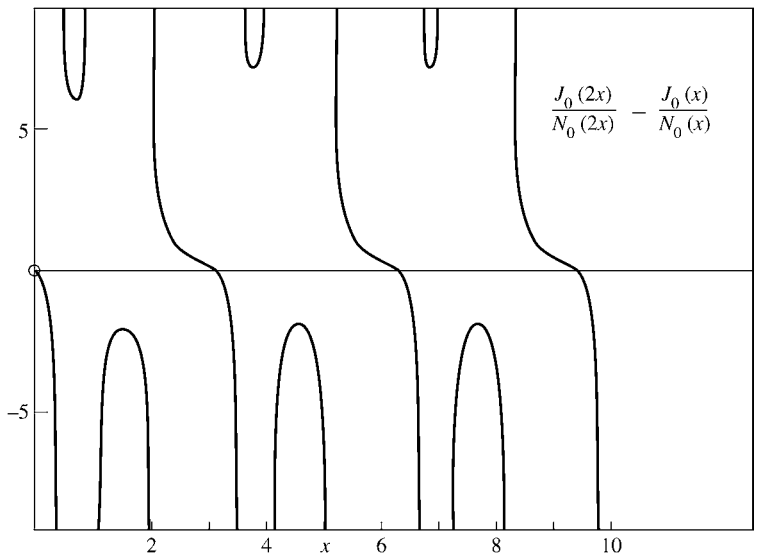

FIGURE 11.6 $f(x)$ of Exercise 11.3.11.

(b) Call a root-finding subroutine to determine the first three roots to higher precision.

ANS. 3.1230, 6.2734, 9.4182.

Note. The higher roots can be expected to appear at intervals whose length approaches n. Why? AMS-55 (see Additional Readings of Chapter 8 for the reference), gives an approximate formula for the roots. The function $g(x) = J_0(x)N_0(2x) - J_0(2x)N_0(x)$ is much better behaved than $f(x)$ previously discussed.

11.4 HANKEL FUNCTIONS

Many authors prefer to introduce the Hankel functions by means of integral representations and then to use them to define the Neumann function $N_\nu(z)$. An outline of this approach is given at the end of this section.

Definitions

Because we have already obtained the Neumann function by more elementary (and less powerful) techniques, we may use it to define the Hankel functions $H_\nu^{(1)}(x)$ and $H_\nu^{(2)}(x)$:

$$H_\nu^{(1)}(x) = J_\nu(x) + iN_\nu(x) \tag{11.85}$$

and

$$H_\nu^{(2)}(x) = J_\nu(x) - iN_\nu(x). \tag{11.86}$$

This is exactly analogous to taking

$$e^{\pm i\theta} = \cos\theta \pm i\sin\theta. \tag{11.87}$$

For real arguments, $H_\nu^{(1)}$ and $H_\nu^{(2)}$ are complex conjugates. The extent of the analogy will be seen even better when the asymptotic forms are considered (Section 11.6). Indeed, it is their asymptotic behavior that makes the Hankel functions useful.

Series expansion of $H_\nu^{(1)}(x)$ and $H_\nu^{(2)}(x)$ may be obtained by combining Eqs. (11.5) and (11.63). Often only the first term is of interest; it is given by

$$H_0^{(1)}(x) \approx i\frac{2}{\pi}\ln x + 1 + i\frac{2}{\pi}(\gamma - \ln 2) + \cdots, \tag{11.88}$$

$$H_\nu^{(1)}(x) \approx -i\frac{(\nu-1)!}{\pi}\left(\frac{2}{x}\right)^\nu + \cdots, \qquad \nu > 0, \tag{11.89}$$

$$H_0^{(2)}(x) \approx -i\frac{2}{\pi}\ln x + 1 - i\frac{2}{\pi}(\gamma - \ln 2) + \cdots, \tag{11.90}$$

$$H_\nu^{(2)}(x) \approx i\frac{(\nu-1)!}{\pi}\left(\frac{2}{x}\right)^\nu + \cdots, \qquad \nu > 0. \tag{11.91}$$

Since the Hankel functions are linear combinations (with constant coefficients) of J_ν and N_ν, they satisfy the same recurrence relations (Eqs. (11.10) and (11.12))

$$H_{\nu-1}(x) + H_{\nu+1}(x) = \frac{2\nu}{x} H_\nu(x), \tag{11.92}$$

$$H_{\nu-1}(x) - H_{\nu+1}(x) = 2H'_\nu(x), \tag{11.93}$$

for both $H_\nu^{(1)}(x)$ and $H_\nu^{(2)}(x)$.

A variety of Wronskian formulas can be developed:

$$H_\nu^{(2)} H_{\nu+1}^{(1)} - H_\nu^{(1)} H_{\nu+1}^{(2)} = \frac{4}{i\pi x}, \tag{11.94}$$

$$J_{\nu-1} H_\nu^{(1)} - J_\nu H_{\nu-1}^{(1)} = \frac{2}{i\pi x}, \tag{11.95}$$

$$J_\nu H_{\nu-1}^{(2)} - J_{\nu-1} H_\nu^{(2)} = \frac{2}{i\pi x}. \tag{11.96}$$

Example 11.4.1 CYLINDRICAL TRAVELING WAVES

As an illustration of the use of Hankel functions, consider a two-dimensional wave problem similar to the vibrating circular membrane of Exercise 11.1.25. Now imagine that the waves are generated at $r = 0$ and move outward to infinity. We replace our standing waves by traveling ones. The differential equation remains the same, but the boundary conditions change. We now demand that for large r the wave behave like

$$U \sim e^{i(kr-\omega t)} \tag{11.97}$$

to describe an outgoing wave. As before, k is the wave number. This assumes, for simplicity, that there is no azimuthal dependence, that is, no angular momentum, or $m = 0$. In Sections 7.3 and 11.6, $H_0^{(1)}(kr)$ is shown to have the asymptotic behavior (for $r \to \infty$)

$$H_0^{(1)}(kr) \sim e^{ikr}. \tag{11.98}$$

This boundary condition at infinity then determines our wave solution as

$$U(r, t) = H_0^{(1)}(kr) e^{-i\omega t}. \tag{11.99}$$

This solution diverges as $r \to 0$, which is the behavior to be expected with a source at the origin.

The choice of a two-dimensional wave problem to illustrate the Hankel function $H_0^{(1)}(z)$ is not accidental. Bessel functions may appear in a variety of ways, such as in the separation of conical coordinates. However, they enter most commonly in the radial equations from the separation of variables in the Helmholtz equation in cylindrical and in spherical polar coordinates. We have taken a degenerate form of cylindrical coordinates for this illustration. Had we used spherical polar coordinates (spherical waves), we should have encountered index $\nu = n + \frac{1}{2}$, n an integer. These special values yield the spherical Bessel functions to be discussed in Section 11.7. ■

Contour Integral Representation of the Hankel Functions

The integral representation (Schlaefli integral)

$$J_\nu(x) = \frac{1}{2\pi i}\oint_C e^{(x/2)(t-1/t)}\frac{dt}{t^{\nu+1}} \tag{11.100}$$

may easily be established as a Cauchy integral for $\nu = n$, an integer (by recognizing that the numerator is the generating function (Eq. (11.1)) and integrating around the origin). If ν is not an integer, the integrand is not single-valued and a cut line is needed in our complex plane. Choosing the negative real axis as the cut line and using the contour shown in Fig. 11.7, we can extend Eq. (11.100) to nonintegral ν. Substituting Eq. (11.100) into Bessel's ODE, we can represent the combined integrand by an exact differential that vanishes as $t \to \infty e^{\pm i\pi}$ (compare Exercise 11.1.16).

We now deform the contour so that it approaches the origin along the positive real axis, as shown in Fig. 11.8. For $x > 0$, this particular approach guarantees that the exact differential mentioned will vanish as $t \to 0$ because of the $e^{-x/2t} \to 0$ factor. Hence each of the separate portions ($\infty\, e^{-i\pi}$ to 0) and (0 to $\infty\, e^{i\pi}$) is a solution of Bessel's equation. We define

$$H_\nu^{(1)}(x) = \frac{1}{\pi i}\int_0^{\infty e^{i\pi}} e^{(x/2)(t-1/t)}\frac{dt}{t^{\nu+1}}, \tag{11.101}$$

$$H_\nu^{(2)}(x) = \frac{1}{\pi i}\int_{\infty e^{-i\pi}}^{0} e^{(x/2)(t-1/t)}\frac{dt}{t^{\nu+1}}. \tag{11.102}$$

These expressions are particularly convenient because they may be handled by the method of steepest descents (Section 7.3). $H_\nu^{(1)}(x)$ has a saddle point at $t = +i$, whereas $H_\nu^{(2)}(x)$ has a saddle point at $t = -i$.

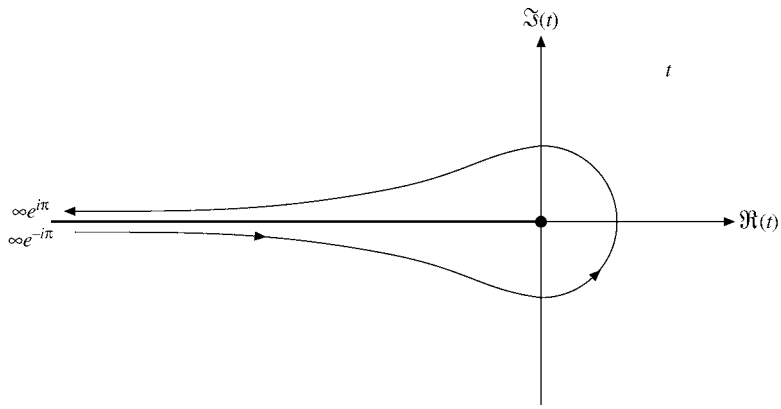

FIGURE 11.7 Bessel function contour.

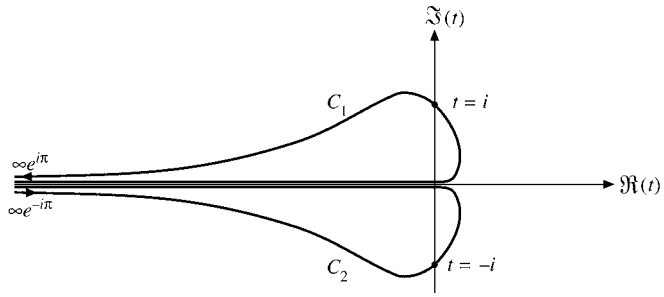

FIGURE 11.8 Hankel function contours.

The problem of relating Eqs. (11.101) and (11.102) to our earlier definition of the Hankel function (Eqs. (11.85) and (11.86)) remains. Since Eqs. (11.100) to (11.102) combined yield

$$J_\nu(x) = \frac{1}{2}\left[H_\nu^{(1)}(x) + H_\nu^{(2)}(x)\right] \tag{11.103}$$

by inspection, we need only show that

$$N_\nu(x) = \frac{1}{2i}\left[H_\nu^{(1)}(x) - H_\nu^{(2)}(x)\right]. \tag{11.104}$$

This may be accomplished by the following steps:

1. With the substitutions $t = e^{i\pi}/s$ for $H_\nu^{(1)}$ and $t = e^{-i\pi}/s$ for $H_\nu^{(2)}$, we obtain

$$H_\nu^{(1)}(x) = e^{-i\nu\pi} H_{-\nu}^{(1)}(x), \tag{11.105}$$

$$H_\nu^{(2)}(x) = e^{i\nu\pi} H_{-\nu}^{(2)}(x). \tag{11.106}$$

2. From Eqs. (11.103) ($\nu \to -\nu$), (11.105), and (11.106),

$$J_{-\nu}(x) = \frac{1}{2}\left[e^{i\nu\pi} H_\nu^{(1)}(x) + e^{-i\nu\pi} H_\nu^{(2)}(x)\right]. \tag{11.107}$$

3. Finally substitute J_ν (Eq. (11.103)) and $J_{-\nu}$ (Eq. (11.107)) into the defining equation for N_ν, Eq. (11.60). This leads to Eq. (11.104) and establishes the contour integrals Eqs. (11.101) and (11.102) as the Hankel functions.

Integral representations have appeared before: Eq. (8.35) for $\Gamma(z)$ and various representations of $J_\nu(z)$ in Section 11.1. With these integral representations of the Hankel functions, it is perhaps appropriate to ask why we are interested in integral representations. There are at least four reasons. The first is simply aesthetic appeal. Second, the integral representations help to distinguish between two linearly independent solutions. In Fig. 11.6, the contours C_1 and C_2 cross **different** saddle points (Section 7.3). For the Legendre functions the contour for $P_n(z)$ (Fig. 12.11) and that for $Q_n(z)$ encircle **different** singular points.

Third, the integral representations facilitate manipulations, analysis, and the development of relations among the various special functions. Fourth, and probably most important of all, the integral representations are extremely useful in developing asymptotic expansions. One approach, the method of steepest descents, appears in Section 7.3. A second approach, the direct expansion of an integral representation is given in Section 11.6 for the modified Bessel function $K_\nu(z)$. This same technique may be used to obtain asymptotic expansions of the confluent hypergeometric functions M and U — Exercise 13.5.13.

In conclusion, the Hankel functions are introduced here for the following reasons:

- As analogs of $e^{\pm ix}$ they are useful for describing traveling waves.
- They offer an alternate (contour integral) and a rather elegant definition of Bessel func-tions.
- $H_\nu^{(1)}$ is used to define the modified Bessel function K_ν of Section 11.5.

Exercises

11.4.1 Verify the Wronskian formulas

(a) $J_\nu(x)H_\nu^{(1)'}(x) - J_\nu'(x)H_\nu^{(1)}(x) = \frac{2i}{\pi x}$,

(b) $J_\nu(x)H_\nu^{(2)'}(x) - J_\nu'(x)H_\nu^{(2)}(x) = \frac{-2i}{\pi x}$,

(c) $N_\nu(x)H_\nu^{(1)'}(x) - N_\nu'(x)H_\nu^{(1)}(x) = \frac{-2}{\pi x}$,

(d) $N_\nu(x)H_\nu^{(2)'}(x) - N_\nu'(x)H_\nu^{(2)}(x) = \frac{-2}{\pi x}$,

(e) $H_\nu^{(1)}(x)H_\nu^{(2)'}(x) - H_\nu^{(1)'}(x)H_\nu^{(2)}(x) = \frac{-4i}{\pi x}$,

(f) $H_\nu^{(2)}(x)H_{\nu+1}^{(1)}(x) - H_\nu^{(1)}(x)H_{\nu+1}^{(2)}(x) = \frac{4}{i\pi x}$,

(g) $J_{\nu-1}(x)H_\nu^{(1)}(x) - J_\nu(x)H_{\nu-1}^{(1)}(x) = \frac{2}{i\pi x}$.

11.4.2 Show that the integral forms

(a) $$\frac{1}{i\pi}\int_{0C_1}^{\infty e^{i\pi}} e^{(x/2)(t-1/t)}\frac{dt}{t^{\nu+1}} = H_\nu^{(1)}(x),$$

(b) $$\frac{1}{i\pi}\int_{\infty e^{-i\pi}C_2}^{0} e^{(x/2)(t-1/t)}\frac{dt}{t^{\nu+1}} = H_\nu^{(2)}(x)$$

satisfy Bessel's ODE. The contours C_1 and C_2 are shown in Fig. 11.8.

11.4.3 Using the integrals and contours given in problem 11.4.2, show that

$$\frac{1}{2i}\left[H_\nu^{(1)}(x) - H_\nu^{(2)}(x)\right] = N_\nu(x).$$

11.4.4 Show that the integrals in Exercise 11.4.2 may be transformed to yield

(a) $$H_\nu^{(1)}(x) = \frac{1}{\pi i}\int_{C_3} e^{x\sinh\gamma - \nu\gamma}\,d\gamma, \qquad \text{(b)} \quad H_\nu^{(2)}(x) = \frac{1}{\pi i}\int_{C_4} e^{x\sinh\gamma - \nu\gamma}\,d\gamma$$

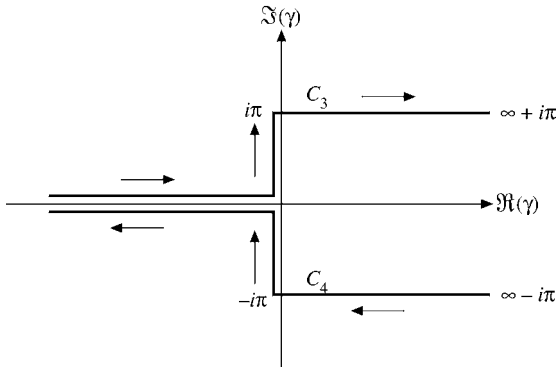

FIGURE 11.9 Hankel function contours.

(see Fig. 11.9).

11.4.5 (a) Transform $H_0^{(1)}(x)$, Eq. (11.101), into

$$H_0^{(1)}(x) = \frac{1}{i\pi}\int_C e^{ix\cosh s}\,ds,$$

where the contour C runs from $-\infty - i\pi/2$ through the origin of the s-plane to $\infty + i\pi/2$.

(b) Justify rewriting $H_0^{(1)}(x)$ as

$$H_0^{(1)}(x) = \frac{2}{i\pi}\int_0^{\infty+i\pi/2} e^{ix\cosh s}\,ds.$$

(c) Verify that this integral representation actually satisfies Bessel's differential equation. (The $i\pi/2$ in the upper limit is not essential. It serves as a convergence factor. We can replace it by $ia\pi/2$ and take the limit.)

11.4.6 From

$$H_0^{(1)}(x) = \frac{2}{i\pi}\int_0^{\infty} e^{ix\cosh s}\,ds$$

show that

(a) $J_0(x) = \dfrac{2}{\pi}\displaystyle\int_0^{\infty} \sin(x\cosh s)\,ds,$ (b) $J_0(x) = \dfrac{2}{\pi}\displaystyle\int_1^{\infty} \frac{\sin(xt)}{\sqrt{t^2-1}}dt.$

This last result is a Fourier sine transform.

11.4.7 From (see Exercises 11.4.4 and 11.4.5)

$$H_0^{(1)}(x) = \frac{2}{i\pi}\int_0^{\infty} e^{ix\cosh s}ds$$

show that

(a) $N_0(x) = -\dfrac{2}{\pi}\displaystyle\int_0^{\infty} \cos(x\cosh s)\,ds.$

(b) $$N_0(x) = -\frac{2}{\pi}\int_1^\infty \frac{\cos(xt)}{\sqrt{t^2-1)}}dt.$$

These are the integral representations in Section 11.3 (Other Forms).
This last result is a Fourier cosine transform.

11.5 MODIFIED BESSEL FUNCTIONS, $I_\nu(x)$ AND $K_\nu(x)$

The Helmholtz equation,

$$\nabla^2\psi + k^2\psi = 0,$$

separated in circular cylindrical coordinates, leads to Eq. (11.22a), the Bessel equation. Equation (11.22a) is satisfied by the Bessel and Neumann functions $J_\nu(k\rho)$ and $N_\nu(k\rho)$ and any linear combination, such as the Hankel functions $H_\nu^{(1)}(k\rho)$ and $H_\nu^{(2)}(k\rho)$. Now, the Helmholtz equation describes the space part of wave phenomena. If instead we have a diffusion problem, then the Helmholtz equation is replaced by

$$\boxed{\nabla^2\psi - k^2\psi = 0.} \tag{11.108}$$

The analog to Eq. (11.22a) is

$$\rho^2\frac{d^2}{d\rho^2}Y_\nu(k\rho) + \rho\frac{d}{d\rho}Y_\nu(k\rho) - \left(k^2\rho^2 + \nu^2\right)Y_\nu(k\rho) = 0. \tag{11.109}$$

The Helmholtz equation may be transformed into the diffusion equation by the transformation $k \to ik$. Similarly, $k \to ik$ changes Eq. (11.22a) into Eq. (11.109) and shows that

$$Y_\nu(k\rho) = Z_\nu(ik\rho).$$

The solutions of Eq. (11.109) are Bessel functions of imaginary argument. To obtain a solution that is regular at the origin, we take Z_ν as the regular Bessel function J_ν. It is customary (and convenient) to choose the normalization so that

$$Y_\nu(x) = I_\nu(x) \equiv i^{-\nu}J_\nu(ix). \tag{11.110}$$

(Here the variable $k\rho$ is being replaced by x for simplicity.) The extra $i^{-\nu}$ normalization cancels the i^ν from each term and leaves $I_\nu(x)$ real. Often this is written as

$$\boxed{I_\nu(x) = e^{-\nu\pi i/2}J_\nu\left(xe^{i\pi/2}\right).} \tag{11.111}$$

I_0 and I_1 are shown in Fig. 11.10.

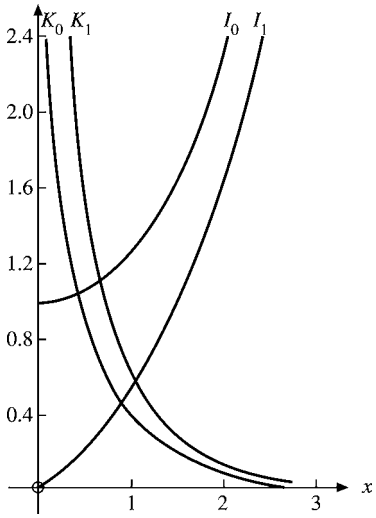

FIGURE 11.10 Modified Bessel functions.

Series Form

In terms of infinite series this is equivalent to removing the $(-1)^s$ sign in Eq. (11.5) and writing

$$I_\nu(x) = \sum_{s=0}^{\infty} \frac{1}{s!(s+\nu)!}\left(\frac{x}{2}\right)^{2s+\nu}, \qquad I_{-\nu}(x) = \sum_{s=0}^{\infty} \frac{1}{s!(s-\nu)!}\left(\frac{x}{2}\right)^{2s-\nu}. \tag{11.112}$$

For integral ν this yields

$$I_n(x) = I_{-n}(x). \tag{11.113}$$

Recurrence Relations

The recurrence relations satisfied by $I_\nu(x)$ may be developed from the series expansions, but it is perhaps easier to work from the existing recurrence relations for $J_\nu(x)$. Let us replace x by $-ix$ and rewrite Eq. (11.110) as

$$J_\nu(x) = i^\nu I_\nu(-ix). \tag{11.114}$$

Then Eq. (11.10) becomes

$$i^{\nu-1} I_{\nu-1}(-ix) + i^{\nu+1} I_{\nu+1}(-ix) = \frac{2\nu}{x} i^\nu I_\nu(-ix).$$

Replacing x by ix, we have a recurrence relation for $I_\nu(x)$,

$$I_{\nu-1}(x) - I_{\nu+1}(x) = \frac{2\nu}{x} I_\nu(x). \tag{11.115}$$

Equation (11.12) transforms to

$$I_{\nu-1}(x) + I_{\nu+1}(x) = 2I'_\nu(x). \tag{11.116}$$

These are the recurrence relations used in Exercise 11.1.14. It is worth emphasizing that although two recurrence relations, Eqs. (11.115) and (11.116) or Exercise 11.5.7, specify the second-order ODE, the converse is not true. The ODE does not uniquely fix the recurrence relations. Equations (11.115) and (11.116) and Exercise 11.5.7 provide an example.

From Eq. (11.113) it is seen that we have but one independent solution when ν is an integer, exactly as in the Bessel functions J_ν. The choice of a second, independent solution of Eq. (11.108) is essentially a matter of convenience. The second solution given here is selected on the basis of its asymptotic behavior — as shown in the next section. The confusion of choice and notation for this solution is perhaps greater than anywhere else in this field.[18] Many authors[19] choose to define a second solution in terms of the Hankel function $H_\nu^{(1)}(x)$ by

$$K_\nu(x) \equiv \frac{\pi}{2} i^{\nu+1} H_\nu^{(1)}(ix) = \frac{\pi}{2} i^{\nu+1}\big[J_\nu(ix) + iN_\nu(ix)\big]. \tag{11.117}$$

The factor $i^{\nu+1}$ makes $K_\nu(x)$ real when x is real. Using Eqs. (11.60) and (11.110), we may transform Eq. (11.117) to[20]

$$K_\nu(x) = \frac{\pi}{2}\frac{I_{-\nu}(x) - I_\nu(x)}{\sin\nu\pi}, \tag{11.118}$$

analogous to Eq. (11.60) for $N_\nu(x)$. The choice of Eq. (11.117) as a definition is somewhat unfortunate in that the function $K_\nu(x)$ does not satisfy the same recurrence relations as $I_\nu(x)$ (compare Exercises 11.5.7 and 11.5.8). To avoid this annoyance, other authors[21] have included an additional factor of $\cos\nu\pi$. This permits K_ν to satisfy the same recurrence relations as I_ν, but it has the disadvantage of making $K_\nu = 0$ for $\nu = \frac{1}{2}, \frac{3}{5}, \frac{5}{2}, \ldots$.

The series expansion of $K_\nu(x)$ follows directly from the series form of $H_\nu^{(1)}(ix)$. The lowest-order terms are (cf. Eqs. (11.61) and (11.62))

$$\begin{aligned} K_0(x) &= -\ln x - \gamma + \ln 2 + \cdots, \\ K_\nu(x) &= 2^{\nu-1}(\nu-1)!x^{-\nu} + \cdots. \end{aligned} \tag{11.119}$$

Because the modified Bessel function I_ν is related to the Bessel function J_ν, much as sinh is related to sine, I_ν and the second solution K_ν are sometimes referred to as hyperbolic Bessel functions. K_0 and K_1 are shown in Fig. 11.10.

$I_0(x)$ and $K_0(x)$ have the integral representations

$$I_0(x) = \frac{1}{\pi}\int_0^\pi \cosh(x\cos\theta)\,d\theta, \tag{11.120}$$

$$K_0(x) = \int_0^\infty \cos(x\sinh t)\,dt = \int_0^\infty \frac{\cos(xt)\,dt}{(t^2+1)^{1/2}}, \qquad x > 0. \tag{11.121}$$

[18] A discussion and comparison of notations will be found in *Math. Tables Aids Comput.* **1**: 207–308 (1944).

[19] Watson, Morse and Feshbach, Jeffreys and Jeffreys (without the $\pi/2$).

[20] For integral index n we take the limit as $\nu \to n$.

[21] Whittaker and Watson, see Additional Readings of Chapter 13.

Equation (11.120) may be derived from Eq. (11.30) for $J_0(x)$ or may be taken as a special case of Exercise 11.5.4, $\nu = 0$. The integral representation of K_0, Eq. (11.121), is a Fourier transform and may best be derived with Fourier transforms, Chapter 15, or with Green's functions Section 9.7. A variety of other forms of integral representations (including $\nu \neq 0$) appear in the exercises. These integral representations are useful in developing asymptotic forms (Section 11.6) and in connection with Fourier transforms, Chapter 15.

To put the modified Bessel functions $I_\nu(x)$ and $K_\nu(x)$ in proper perspective, we introduce them here because:

- These functions are solutions of the frequently encountered modified Bessel equation.
- They are needed for specific physical problems, such as diffusion problems.
- $K_\nu(x)$ provides a Green's function, Section 9.7.
- $K_\nu(x)$ leads to a convenient determination of asymptotic behavior (Section 11.6).

Exercises

11.5.1 Show that

$$e^{(x/2)(t+1/t)} = \sum_{n=-\infty}^{\infty} I_n(x)t^n,$$

thus generating modified Bessel functions, $I_n(x)$.

11.5.2 Verify the following identities

(a) $1 = I_0(x) + 2\sum_{n=1}^{\infty}(-1)^n I_{2n}(x),$

(b) $e^x = I_0(x) + 2\sum_{n=1}^{\infty} I_n(x),$

(c) $e^{-x} = I_0(x) + 2\sum_{n=1}^{\infty}(-1)^n I_n(x),$

(d) $\cosh x = I_0(x) + 2\sum_{n=1}^{\infty} I_{2n}(x),$

(e) $\sinh x = 2\sum_{n=1}^{\infty} I_{2n-1}(x).$

11.5.3 (a) From the generating function of Exercise 11.5.1 show that

$$I_n(x) = \frac{1}{2\pi i}\oint \exp\big[(x/2)(t+1/t)\big]\frac{dt}{t^{n+1}}.$$

(b) For $n = \nu$, not an integer, show that the preceding integral representation may be generalized to

$$I_\nu(x) = \frac{1}{2\pi i}\int_C \exp\big[(x/2)(t+1/t)\big]\frac{dt}{t^{\nu+1}}.$$

The contour C is the same as that for $J_\nu(x)$, Fig. 11.7.

11.5.4 For $\nu > -\frac{1}{2}$ show that $I_\nu(z)$ may be represented by

$$\begin{aligned} I_\nu(z) &= \frac{1}{\pi^{1/2}(\nu-\frac{1}{2})!}\left(\frac{z}{2}\right)^\nu \int_0^\pi e^{\pm z\cos\theta}\sin^{2\nu}\theta\, d\theta \\ &= \frac{1}{\pi^{1/2}(\nu-\frac{1}{2})!}\left(\frac{z}{2}\right)^\nu \int_{-1}^1 e^{\pm zp}\big(1-p^2\big)^{\nu-1/2}\, dp \\ &= \frac{2}{\pi^{1/2}(\nu-\frac{1}{2})!}\left(\frac{z}{2}\right)^\nu \int_0^{\pi/2} \cosh(z\cos\theta)\sin^{2\nu}\theta\, d\theta. \end{aligned}$$

11.5.5 A cylindrical cavity has a radius a and height l, Fig. 11.3. The ends, $z = 0$ and l, are at zero potential. The cylindrical walls, $\rho = a$, have a potential $V = V(\varphi, z)$.

(a) Show that the electrostatic potential $\Phi(\rho, \varphi, z)$ has the functional form

$$\Phi(\rho,\varphi,z) = \sum_{m=0}^{\infty}\sum_{n=1}^{\infty} I_m(k_n\rho)\sin k_n z\cdot(a_{mn}\sin m\varphi + b_{mn}\cos m\varphi),$$

where $k_n = n\pi/l$.

(b) Show that the coefficients a_{mn} and b_{mn} are given by[22]

$$\left.\begin{matrix} a_{mn} \\ b_{mn} \end{matrix}\right\} = \frac{2}{\pi l I_m(k_n a)}\int_0^{2\pi}\int_0^l V(\varphi,z)\sin k_n z\cdot\left\{\begin{matrix}\sin m\varphi \\ \cos m\varphi\end{matrix}\right\} dz\, d\varphi.$$

Hint. Expand $V(\varphi, z)$ as a double series and use the orthogonality of the trigonometric functions.

11.5.6 Verify that $K_\nu(x)$ is given by

$$K_\nu(x) = \frac{\pi}{2}\frac{I_{-\nu}(x) - I_\nu(x)}{\sin\nu\pi}$$

and from this show that

$$K_\nu(x) = K_{-\nu}(x).$$

11.5.7 Show that $K_\nu(x)$ satisfies the recurrence relations

$$\begin{aligned} K_{\nu-1}(x) - K_{\nu+1}(x) &= -\frac{2\nu}{x}K_\nu(x), \\ K_{\nu-1}(x) + K_{\nu+1}(x) &= -2K_\nu'(x). \end{aligned}$$

[22] When $m = 0$, the 2 in the coefficient is replaced by 1.

11.5.8 If $\mathcal{K}_\nu = e^{\nu\pi i} K_\nu$, show that $\mathcal{K}_\nu$ satisfies the same recurrence relations as I_ν.

11.5.9 For $\nu > -\frac{1}{2}$ show that $K_\nu(z)$ may be represented by

$$K_\nu(z) = \frac{\pi^{1/2}}{(\nu - \frac{1}{2})!}\left(\frac{z}{2}\right)^\nu \int_0^\infty e^{-z\cosh t}\sinh^{2\nu} t\, dt, \qquad -\frac{\pi}{2} < \arg z < \frac{\pi}{2}$$

$$= \frac{\pi^{1/2}}{(\nu - \frac{1}{2})!}\left(\frac{z}{2}\right)^\nu \int_1^\infty e^{-zp}(p^2-1)^{\nu-1/2}\, dp.$$

11.5.10 Show that $I_\nu(x)$ and $K_\nu(x)$ satisfy the Wronskian relation

$$I_\nu(x)K'_\nu(x) - I'_\nu(x)K_\nu(x) = -\frac{1}{x}.$$

This result is quoted in Section 9.7 in the development of a Green's function.

11.5.11 If $r = (x^2 + y^2)^{1/2}$, prove that

$$\frac{1}{r} = \frac{2}{\pi}\int_0^\infty \cos(xt)K_0(yt)\, dt.$$

This is a Fourier cosine transform of K_0.

11.5.12 (a) Verify that

$$I_0(x) = \frac{1}{\pi}\int_0^\pi \cosh(x\cos\theta)\, d\theta$$

satisfies the modified Bessel equation, $\nu = 0$.

(b) Show that this integral contains no admixture of $K_0(x)$, the irregular second solution.

(c) Verify the normalization factor $1/\pi$.

11.5.13 Verify that the integral representations

$$I_n(z) = \frac{1}{\pi}\int_0^\pi e^{z\cos t}\cos(nt)\, dt,$$

$$K_\nu(z) = \int_0^\infty e^{-z\cosh t}\cosh(\nu t)\, dt, \qquad \Re(z) > 0,$$

satisfy the modified Bessel equation by direct substitution into that equation. How can you show that the first form does not contain an admixture of K_n and that the second form does not contain an admixture of I_ν? How can you check the normalization?

11.5.14 Derive the integral representation

$$I_n(x) = \frac{1}{\pi}\int_0^\pi e^{x\cos\theta}\cos(n\theta)\, d\theta.$$

Hint. Start with the corresponding integral representation of $J_n(x)$. Equation (11.120) is a special case of this representation.

11.5.15 Show that

$$K_0(z) = \int_0^\infty e^{-z\cosh t}\,dt$$

satisfies the modified Bessel equation. How can you establish that this form is linearly independent of $I_0(z)$?

11.5.16 Show that

$$e^{ax} = I_0(a)T_0(x) + 2\sum_{n=1}^{\infty} I_n(a)T_n(x), \qquad -1 \le x \le 1.$$

$T_n(x)$ is the nth-order Chebyshev polynomial, Section 13.3.
Hint. Assume a Chebyshev series expansion. Using the orthogonality and normalization of the $T_n(x)$, solve for the coefficients of the Chebyshev series.

11.5.17 (a) Write a double precision subroutine to calculate $I_n(x)$ to 12-decimal-place accuracy for $n = 0, 1, 2, 3, \ldots$ and $0 \le x \le 1$. Check your results against the 10-place values given in AMS-55, Table 9.11, see Additional Readings of Chapter 8 for the reference.

(b) Referring to Exercise 11.5.16, calculate the coefficients in the Chebyshev expansions of $\cosh x$ and of $\sinh x$.

11.5.18 The cylindrical cavity of Exercise 11.5.5 has a potential along the cylinder walls:

$$V(z) = \begin{cases} 100\frac{z}{l}, & 0 \le \frac{z}{l} \le \frac{1}{2}, \\ 100\left(1 - \frac{z}{l}\right), & \frac{1}{2} \le \frac{z}{l} \le 1. \end{cases}$$

With the radius–height ratio $a/l = 0.5$, calculate the potential for $z/l = 0.1(0.1)0.5$ and $\rho/a = 0.0(0.2)1.0$.

Check value. For $z/l = 0.3$ and $\rho/a = 0.8$, $V = 26.396$.

11.6 ASYMPTOTIC EXPANSIONS

Frequently in physical problems there is a need to know how a given Bessel or modified Bessel function behaves for large values of the argument, that is, the asymptotic behavior. This is one occasion when computers are not very helpful. One possible approach is to develop a power-series solution of the differential equation, as in Section 9.5, but now using negative powers. This is Stokes' method, Exercise 11.6.5. The limitation is that starting from some positive value of the argument (for convergence of the series), we do not know what mixture of solutions or multiple of a given solution we have. The problem is to relate the asymptotic series (useful for large values of the variable) to the power-series or related definition (useful for small values of the variable). This relationship can be established by introducing a suitable **integral representation** and then using either the method of steepest descent, Section 7.3, or the direct expansion as developed in this section.

Expansion of an Integral Representation

As a direct approach, consider the integral representation (Exercise 11.5.9)

$$K_\nu(z) = \frac{\pi^{1/2}}{(\nu - \frac{1}{2})!}\left(\frac{z}{2}\right)^\nu \int_1^\infty e^{-zx}\left(x^2-1\right)^{\nu-1/2} dx, \qquad \nu > -\frac{1}{2}. \tag{11.122}$$

For the present let us take z to be real, although Eq. (11.122) may be established for $-\pi/2 < \arg z < \pi/2$ ($\Re(z) > 0$). We have three tasks:

1. To show that K_ν as given in Eq. (11.122) actually satisfies the modified Bessel equation (11.109).
2. To show that the regular solution I_ν is absent.
3. To show that Eq. (11.122) has the proper normalization.

1. The fact that Eq. (11.122) is a solution of the modified Bessel equation may be verified by direct substitution. We obtain

$$z^{\nu+1}\int_1^\infty \frac{d}{dx}\left[e^{-zx}\left(x^2-1\right)^{\nu+1/2}\right] dx = 0,$$

which transforms the combined integrand into the derivative of a function that vanishes at both endpoints. Hence the integral is some linear combination of I_ν and K_ν.

2. The rejection of the possibility that this solution contains I_ν constitutes Exercise 11.6.1.

3. The normalization may be verified by showing that, in the limit $z \to 0$, $K_\nu(z)$ is in agreement with Eq. (11.119). By substituting $x = 1 + t/z$,

$$\frac{\pi^{1/2}}{(\nu - \frac{1}{2})!}\left(\frac{z}{2}\right)^\nu \int_1^\infty e^{-zx}\left(x^2-1\right)^{\nu-1/2} dx$$

$$= \frac{\pi^{1/2}}{(\nu - \frac{1}{2})!}\left(\frac{z}{2}\right)^\nu e^{-z}\int_0^\infty e^{-t}\left(\frac{t^2}{z^2} + \frac{2t}{z}\right)^{\nu-1/2}\frac{dt}{z} \tag{11.123a}$$

$$= \frac{\pi^{1/2}}{(\nu - \frac{1}{2})!}\frac{e^{-z}}{2^\nu z^\nu}\int_0^\infty e^{-t}t^{2\nu-1}\left(1 + \frac{2z}{t}\right)^{\nu-1/2} dt, \tag{11.123b}$$

taking out t^2/z^2 as a factor. This substitution has changed the limits of integration to a more convenient range and has isolated the negative exponential dependence e^{-z}. The integral in Eq. (11.123b) may be evaluated for $z = 0$ to yield $(2\nu - 1)!$. Then, using the duplication formula (Section 8.4), we have

$$\lim_{z\to 0} K_\nu(z) = \frac{(\nu-1)!2^{\nu-1}}{z^\nu}, \qquad \nu > 0, \tag{11.124}$$

in agreement with Eq. (11.119), which thus checks the normalization.[23]

[23] For $\nu \to 0$ the integral diverges logarithmically, in agreement with the logarithmic divergence of $K_0(z)$ for $z \to 0$ (Section 11.5).

Now, to develop an asymptotic series for $K_\nu(z)$, we may rewrite Eq. (11.123a) as

$$K_\nu(z) = \sqrt{\frac{\pi}{2z}}\frac{e^{-z}}{(\nu-\frac{1}{2})!}\int_0^\infty e^{-t}t^{\nu-1/2}\left(1+\frac{t}{2z}\right)^{\nu-1/2}dt \tag{11.125}$$

(taking out $2t/z$ as a factor).

We expand $(1+t/2z)^{\nu-1/2}$ by the binomial theorem to obtain

$$K_\nu(z) = \sqrt{\frac{\pi}{2z}}\frac{e^{-z}}{(\nu-\frac{1}{2})!}\sum_{r=0}^{\infty}\frac{(\nu-\frac{1}{2})!}{r!(\nu-r-\frac{1}{2})!}(2z)^{-r}\int_0^\infty e^{-t}t^{\nu+r-1/2}\,dt. \tag{11.126}$$

Term-by-term integration (valid for asymptotic series) yields the desired asymptotic expansion of $K_\nu(z)$:

$$K_\nu(z) \sim \sqrt{\frac{\pi}{2z}}e^{-z}\left[1+\frac{(4\nu^2-1^2)}{1!8z}+\frac{(4\nu^2-1^2)(4\nu^2-3^2)}{2!(8z)^2}+\cdots\right]. \tag{11.127}$$

Although the integral of Eq. (11.122), integrating along the real axis, was convergent only for $-\pi/2 < \arg z < \pi/2$, Eq. (11.127) may be extended to $-3\pi/2 < \arg z < 3\pi/2$. Considered as an infinite series, Eq. (11.127) is actually divergent.[24] However, this series is asymptotic, in the sense that for large enough z, $K_\nu(z)$ may be approximated to any fixed degree of accuracy with a small number of terms. (Compare Section 5.10 for a definition and discussion of asymptotic series.)

It is convenient to rewrite Eq. (11.127) as

$$K_\nu(z) = \sqrt{\frac{\pi}{2z}}e^{-z}\big[P_\nu(iz)+iQ_\nu(iz)\big], \tag{11.128}$$

where

$$P_\nu(z) \sim 1-\frac{(\mu-1)(\mu-9)}{2!(8z)^2}+\frac{(\mu-1)(\mu-9)(\mu-25)(\mu-49)}{4!(8z)^4}-\cdots, \tag{11.129a}$$

$$Q_\nu(z) \sim \frac{\mu-1}{1!(8z)}-\frac{(\mu-1)(\mu-9)(\mu-25)}{3!(8z)^3}+\cdots, \tag{11.129b}$$

and

$$\mu = 4\nu^2.$$

It should be noted that although $P_\nu(z)$ of Eq. (11.129a) and $Q_\nu(z)$ of Eq. (11.129b) have alternating signs, the series for $P_\nu(iz)$ and $Q_\nu(iz)$ of Eq. (11.128) have all signs positive. Finally, for z large, P_ν dominates.

Then with the asymptotic form of $K_\nu(z)$, Eq. (11.128), we can obtain expansions for all other Bessel and hyperbolic Bessel functions by defining relations:

[24] Our binomial expansion is valid only for $t < 2z$ and we have integrated t out to infinity. The exponential decrease of the integrand prevents a disaster, but the resultant series is still only asymptotic, not convergent. By Table 9.3, $z = \infty$ is an essential singularity of the Bessel (and modified Bessel) equations. Fuchs' theorem does not guarantee a convergent series and we do not get a convergent series.

1. From

$$\frac{\pi}{2} i^{\nu+1} H_\nu^{(1)}(iz) = K_\nu(z) \tag{11.130}$$

we have

$$H_\nu^{(1)}(z) = \sqrt{\frac{2}{\pi z}} \exp\left\{ i\left[z - \left(\nu + \frac{1}{2}\right)\frac{\pi}{2} \right] \right\} \cdot \left[P_\nu(z) + i\, Q_\nu(z) \right], \qquad -\pi < \arg z < 2\pi. \tag{11.131}$$

2. The second Hankel function is just the complex conjugate of the first (for real argument),

$$H_\nu^{(2)}(z) = \sqrt{\frac{2}{\pi z}} \exp\left\{ -i\left[z - \left(\nu + \frac{1}{2}\right)\frac{\pi}{2} \right] \right\} \cdot \left[P_\nu(z) - i\, Q_\nu(z) \right], \qquad -2\pi < \arg z < \pi. \tag{11.132}$$

An alternate derivation of the asymptotic behavior of the Hankel functions appears in Section 7.3 as an application of the method of steepest descents.

3. Since $J_\nu(z)$ is the real part of $H_\nu^{(1)}(z)$ for real z,

$$J_\nu(z) = \sqrt{\frac{2}{\pi z}} \left\{ P_\nu(z) \cos\left[z - \left(\nu + \frac{1}{2}\right)\frac{\pi}{2} \right] - Q_\nu(z) \sin\left[z - \left(\nu + \frac{1}{2}\right)\frac{\pi}{2} \right] \right\}, \qquad -\pi < \arg z < \pi, \tag{11.133}$$

holds for real z, that is, $\arg z = 0, \pi$. Once Eq. (11.133) is established for real z, the relation is valid for complex z in the given range of argument.

4. The Neumann function is the imaginary part of $H_\nu^{(1)}(z)$ for real z, or

$$N_\nu(z) = \sqrt{\frac{2}{\pi z}} \left\{ P_\nu(z) \sin\left[z - \left(\nu + \frac{1}{2}\right)\frac{\pi}{2} \right] + Q_\nu(z) \cos\left[z - \left(\nu + \frac{1}{2}\right)\frac{\pi}{2} \right] \right\}, \qquad -\pi < \arg z < \pi. \tag{11.134}$$

Initially, this relation is established for real z, but it may be extended to the complex domain as shown.

5. Finally, the regular hyperbolic or modified Bessel function $I_\nu(z)$ is given by

$$I_\nu(z) = i^{-\nu} J_\nu(iz) \tag{11.135}$$

or

$$I_\nu(z) = \frac{e^z}{\sqrt{2\pi z}} \left[P_\nu(iz) - i\, Q_\nu(iz) \right], \qquad -\frac{\pi}{2} < \arg z < \frac{\pi}{2}. \tag{11.136}$$

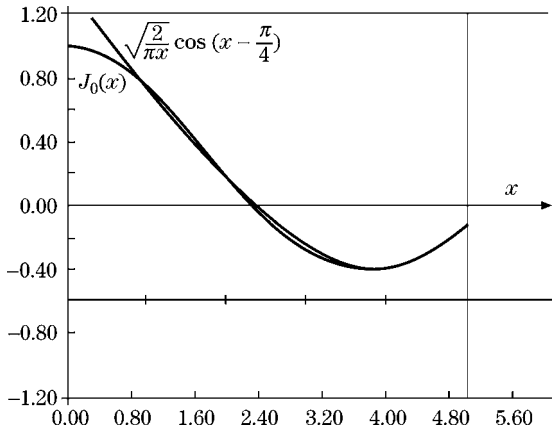

FIGURE 11.11 Asymptotic approximation of $J_0(x)$.

This completes our determination of the asymptotic expansions. However, it is perhaps worth noting the primary characteristics. Apart from the ubiquitous $z^{-1/2}$, J_ν and N_ν behave as cosine and sine, respectively. The zeros are **almost** evenly spaced at intervals of π; the spacing becomes exactly π in the limit as $z \to \infty$. The Hankel functions have been defined to behave like the imaginary exponentials, and the modified Bessel functions I_ν and K_ν go into the positive and negative exponentials. This asymptotic behavior may be sufficient to eliminate immediately one of these functions as a solution for a physical problem. We should also note that the asymptotic series $P_\nu(z)$ and $Q_\nu(z)$, Eqs. (11.129a) and (11.129b), terminate for $\nu = \pm 1/2, \pm 3/2, \ldots$ and become polynomials (in negative powers of z). For these special values of ν the asymptotic approximations become exact solutions.

It is of some interest to consider the accuracy of the asymptotic forms, taking just the first term, for example (Fig. 11.11),

$$J_n(x) \approx \sqrt{\frac{2}{\pi x}}\cos\left[x - \left(n + \frac{1}{2}\right)\left(\frac{\pi}{2}\right)\right]. \tag{11.137}$$

Clearly, the condition for the validity of Eq. (11.137) is that the sine term be negligible; that is,

$$8x \gg 4n^2 - 1. \tag{11.138}$$

For n or $\nu > 1$ the asymptotic region may be far out.

As pointed out in Section 11.3, the asymptotic forms may be used to evaluate the various Wronskian formulas (compare Exercise 11.6.3).

Exercises

11.6.1 In checking the normalization of the integral representation of $K_\nu(z)$ (Eq. (11.122)), we assumed that $I_\nu(z)$ was not present. How do we know that the integral representation (Eq. (11.122)) does not yield $K_\nu(z) + \varepsilon I_\nu(z)$ with $\varepsilon \neq 0$?

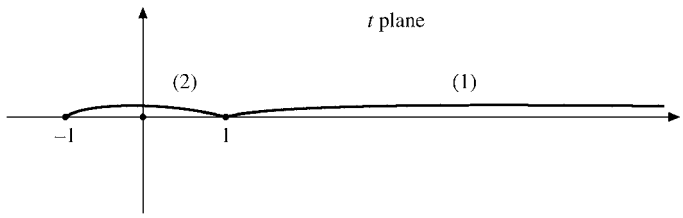

FIGURE 11.12 Modified Bessel function contours.

11.6.2 (a) Show that

$$y(z) = z^{\nu} \int e^{-zt}\left(t^2-1\right)^{\nu-1/2} dt$$

satisfies the modified Bessel equation, provided the contour is chosen so that

$$e^{-zt}\left(t^2-1\right)^{\nu+1/2}$$

has the same value at the initial and final points of the contour.

(b) Verify that the contours shown in Fig. 11.12 are suitable for this problem.

11.6.3 Use the asymptotic expansions to verify the following Wronskian formulas:

(a) $J_\nu(x)J_{-\nu-1}(x) + J_{-\nu}(x)J_{\nu+1}(x) = -2\sin\nu\pi/\pi x$,
(b) $J_\nu(x)N_{\nu+1}(x) - J_{\nu+1}(x)N_\nu(x) = -2/\pi x$,
(c) $J_\nu(x)H^{(2)}_{\nu-1}(x) - J_{\nu-1}(x)H^{(2)}_\nu(x) = 2/i\pi x$,
(d) $I_\nu(x)K'_\nu(x) - I'_\nu(x)K_\nu(x) = -1/x$,
(e) $I_\nu(x)K_{\nu+1}(x) + I_{\nu+1}(x)K_\nu(x) = 1/x$.

11.6.4 From the asymptotic form of $K_\nu(z)$, Eq. (11.127), derive the asymptotic form of $H^{(1)}_\nu(z)$, Eq. (11.131). Note particularly the phase, $(\nu+\frac{1}{2})\pi/2$.

11.6.5 Stokes' method.

(a) Replace the Bessel function in Bessel's equation by $x^{-1/2}y(x)$ and show that $y(x)$ satisfies

$$y''(x) + \left(1 - \frac{\nu^2-\frac{1}{4}}{x^2}\right)y(x) = 0.$$

(b) Develop a power-series solution with negative powers of x starting with the assumed form

$$y(x) = e^{ix}\sum_{n=0}^{\infty} a_n x^{-n}.$$

Determine the recurrence relation giving a_{n+1} in terms of a_n. Check your result against the asymptotic series, Eq. (11.131).

(c) From the results of Section 7.4 determine the initial coefficient, a_0.

11.6.6 Calculate the first 15 partial sums of $P_0(x)$ and $Q_0(x)$, Eqs. (11.129a) and (11.129b). Let x vary from 4 to 10 in unit steps. Determine the number of terms to be retained for maximum accuracy and the accuracy achieved as a function of x. Specifically, how small may x be without raising the error above 3×10^{-6}?

ANS. $x_{\min} = 6$.

11.6.7 (a) Using the asymptotic series (partial sums) $P_0(x)$ and $Q_0(x)$ determined in Exercise 11.6.6, write a function subprogram FCT(X) that will calculate $J_0(x), x$ real, for $x \geq x_{\min}$.

(b) Test your function by comparing it with the $J_0(x)$ (tables or computer library subroutine) for $x = x_{\min}(10)x_{\min} + 10$.

Note. A more accurate and perhaps simpler asymptotic form for $J_0(x)$ is given in AMS-55, Eq. (9.4.3), see Additional Readings of Chapter 8 for the reference.

11.7 SPHERICAL BESSEL FUNCTIONS

When the Helmholtz equation is separated in spherical coordinates, the radial equation has the form

$$r^2\frac{d^2R}{dr^2} + 2r\frac{dR}{dr} + \left[k^2r^2 - n(n+1)\right]R = 0. \tag{11.139}$$

This is Eq. (9.65) of Section 9.3. The parameter k enters from the original Helmholtz equation, while $n(n+1)$ is a separation constant. From the behavior of the polar angle function (Legendre's equation, Sections 9.5 and 12.5), the separation constant must have this form, with n a nonnegative integer. Equation (11.139) has the virtue of being self-adjoint, but clearly it is not Bessel's equation. However, if we substitute

$$R(kr) = \frac{Z(kr)}{(kr)^{1/2}},$$

Equation (11.139) becomes

$$r^2\frac{d^2Z}{dr^2} + r\frac{dZ}{dr} + \left[k^2r^2 - \left(n+\frac{1}{2}\right)^2\right]Z = 0, \tag{11.140}$$

which **is** Bessel's equation. Z is a Bessel function of order $n + \frac{1}{2}$ (n an integer). Because of the importance of spherical coordinates, this combination, that is,

$$\frac{Z_{n+1/2}(kr)}{(kr)^{1/2}},$$

occurs quite often.

Definitions

It is convenient to label these functions spherical Bessel functions with the following defining equations:

$$\boxed{\begin{aligned} j_n(x) &= \sqrt{\frac{\pi}{2x}}J_{n+1/2}(x), \\ n_n(x) &= \sqrt{\frac{\pi}{2x}}N_{n+1/2}(x) = (-1)^{n+1}\sqrt{\frac{\pi}{2x}}J_{-n-1/2}(x),^{25} \\ h_n^{(1)}(x) &= \sqrt{\frac{\pi}{2x}}H_{n+1/2}^{(1)}(x) = j_n(x) + in_n(x), \\ h_n^{(2)}(x) &= \sqrt{\frac{\pi}{2x}}H_{n+1/2}^{(2)}(x) = j_n(x) - in_n(x). \end{aligned}} \tag{11.141}$$

These spherical Bessel functions (Figs. 11.13 and 11.14) can be expressed in series form by using the series (Eq. (11.5)) for J_n, replacing n with $n + \frac{1}{2}$:

$$J_{n+1/2}(x) = \sum_{s=0}^{\infty} \frac{(-1)^s}{s!(s+n+\frac{1}{2})!}\left(\frac{x}{2}\right)^{2s+n+1/2}. \tag{11.142}$$

Using the Legendre duplication formula,

$$z!(z+\tfrac{1}{2})! = 2^{-2z-1}\pi^{1/2}(2z+1)!, \tag{11.143}$$

we have

$$\begin{aligned} j_n(x) &= \sqrt{\frac{\pi}{2x}}\sum_{s=0}^{\infty}\frac{(-1)^s 2^{2s+2n+1}(s+n)!}{\pi^{1/2}(2s+2n+1)!s!}\left(\frac{x}{2}\right)^{2s+n+1/2} \\ &= 2^n x^n \sum_{s=0}^{\infty}\frac{(-1)^s(s+n)!}{s!(2s+2n+1)!}x^{2s}. \end{aligned} \tag{11.144}$$

Now, $N_{n+1/2}(x) = (-1)^{n+1}J_{-n-1/2}(x)$ and from Eq. (11.5) we find that

$$J_{-n-1/2}(x) = \sum_{s=0}^{\infty}\frac{(-1)^s}{s!(s-n-\frac{1}{2})!}\left(\frac{x}{2}\right)^{2s-n-1/2}. \tag{11.145}$$

This yields

$$n_n(x) = (-1)^{n+1}\frac{2^n\pi^{1/2}}{x^{n+1}}\sum_{s=0}^{\infty}\frac{(-1)^s}{s!(s-n-\frac{1}{2})!}\left(\frac{x}{2}\right)^{2s}. \tag{11.146}$$

[25] This is possible because $\cos(n+\frac{1}{2})\pi = 0$, see Eq. (11.60).

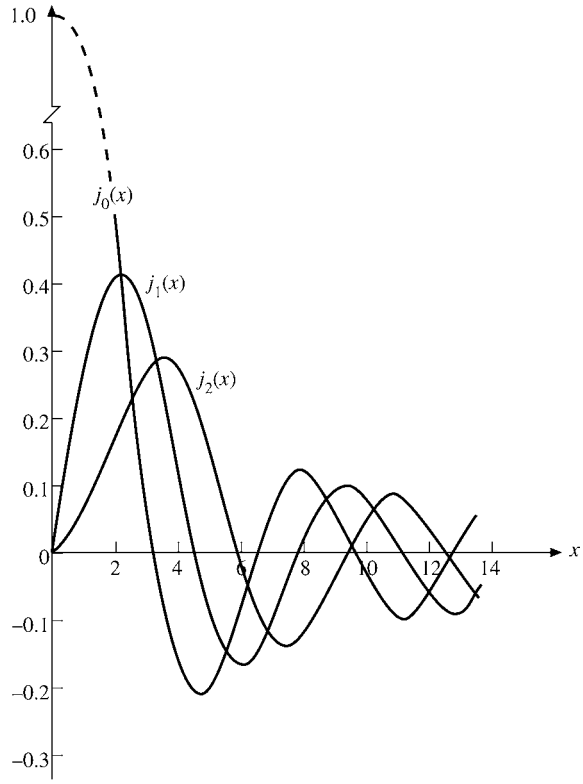

FIGURE 11.13 Spherical Bessel functions.

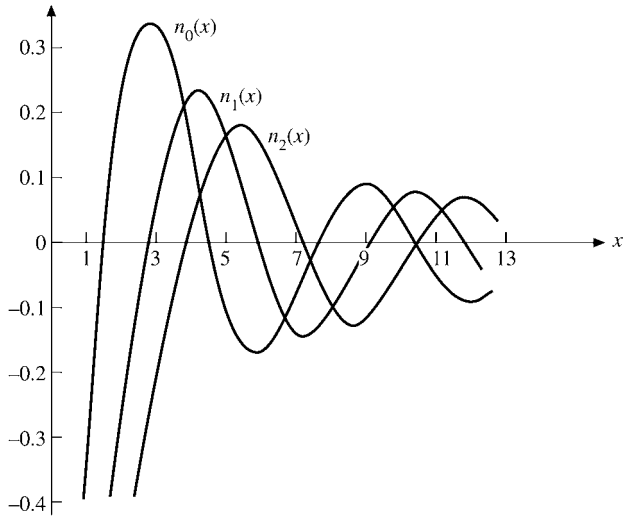

FIGURE 11.14 Spherical Neumann functions.

The Legendre duplication formula can be used again to give

$$n_n(x) = \frac{(-1)^{n+1}}{2^n x^{n+1}} \sum_{s=0}^{\infty} \frac{(-1)^s (s-n)!}{s!(2s-2n)!} x^{2s}. \tag{11.147}$$

These series forms, Eqs. (11.144) and (11.147), are useful in three ways: (1) limiting values as $x \to 0$, (2) closed-form representations for $n = 0$, and, as an extension of this, (3) an indication that the spherical Bessel functions are closely related to sine and cosine.

For the special case $n = 0$ we find from Eq. (11.144) that

$$j_0(x) = \sum_{s=0}^{\infty} \frac{(-1)^s}{(2s+1)!} x^{2s} = \frac{\sin x}{x}, \tag{11.148}$$

whereas for n_0, Eq. (11.147) yields

$$n_0(x) = -\frac{\cos x}{x}. \tag{11.149}$$

From the definition of the spherical Hankel functions (Eq. (11.141)),

$$\begin{aligned} h_0^{(1)}(x) &= \frac{1}{x}(\sin x - i\cos x) = -\frac{i}{x} e^{ix}, \\ h_0^{(2)}(x) &= \frac{1}{x}(\sin x + i\cos x) = \frac{i}{x} e^{-ix}. \end{aligned} \tag{11.150}$$

Equations (11.148) and (11.149) suggest expressing all spherical Bessel functions as combinations of sine and cosine. The appropriate combinations can be developed from the power-series solutions, Eqs. (11.144) and (11.147), but this approach is awkward. Actually the trigonometric forms are already available as the asymptotic expansion of Section 11.6. From Eqs. (11.131) and (11.129a),

$$\begin{aligned} h_n^{(1)}(x) &= \sqrt{\frac{\pi}{2z}} H_{n+1/2}^{(1)}(z) \\ &= (-i)^{n+1} \frac{e^{iz}}{z} \left\{ P_{n+1/2}(z) + i\, Q_{n+1/2}(z) \right\}. \end{aligned} \tag{11.151}$$

Now, $P_{n+1/2}$ and $Q_{n+1/2}$ are **polynomials**. This means that Eq. (11.151) is mathematically exact, not simply an asymptotic approximation. We obtain

$$\begin{aligned} h_n^{(1)}(z) &= (-i)^{n+1} \frac{e^{iz}}{z} \sum_{s=0}^{n} \frac{i^s}{s!(8z)^s} \frac{(2n+2s)!!}{(2n-2s)!!} \\ &= (-i)^{n+1} \frac{e^{iz}}{z} \sum_{s=0}^{n} \frac{i^s}{s!(2z)^s} \frac{(n+s)!}{(n-s)!}. \end{aligned} \tag{11.152}$$

Often a factor $(-i)^n = (e^{-i\pi/2})^n$ will be combined with the e^{iz} to give $e^{i(z-n\pi/2)}$. For z real, $j_n(z)$ is the real part of this, $n_n(z)$ the imaginary part, and $h_n^{(2)}(z)$ the complex conjugate. Specifically,

$$h_1^{(1)}(x) = e^{ix}\left(-\frac{1}{x} - \frac{i}{x^2}\right), \tag{11.153a}$$

$$h_2^{(1)}(x) = e^{ix}\left(\frac{i}{x} - \frac{3}{x^2} - \frac{3i}{x^3}\right), \tag{11.153b}$$

$$\begin{aligned} j_1(x) &= \frac{\sin x}{x^2} - \frac{\cos x}{x}, \\ j_2(x) &= \left(\frac{3}{x^3} - \frac{1}{x}\right)\sin x - \frac{3}{x^2}\cos x, \end{aligned} \tag{11.154}$$

$$\begin{aligned} n_1(x) &= -\frac{\cos x}{x^2} - \frac{\sin x}{x}, \\ n_2(x) &= -\left(\frac{3}{x^3} - \frac{1}{x}\right)\cos x - \frac{3}{x^2}\sin x, \end{aligned} \tag{11.155}$$

and so on.

Limiting Values

For $x \ll 1$,[26] Eqs. (11.144) and (11.147) yield

$$j_n(x) \approx \frac{2^n n!}{(2n+1)!}x^n = \frac{x^n}{(2n+1)!!}, \tag{11.156}$$

$$\begin{aligned} n_n(x) &\approx \frac{(-1)^{n+1}}{2^n}\cdot\frac{(-n)!}{(-2n)!}x^{-n-1} \\ &= -\frac{(2n)!}{2^n n!}x^{-n-1} = -(2n-1)!!x^{-n-1}. \end{aligned} \tag{11.157}$$

The transformation of factorials in the expressions for $n_n(x)$ employs Exercise 8.1.3. The limiting values of the spherical Hankel functions go as $\pm i n_n(x)$.

The asymptotic values of $j_n, n_n, h_n^{(2)}$, and $h_n^{(1)}$ may be obtained from the Bessel asymptotic forms, Section 11.6. We find

$$j_n(x) \sim \frac{1}{x}\sin\left(x - \frac{n\pi}{2}\right), \tag{11.158}$$

$$n_n(x) \sim -\frac{1}{x}\cos\left(x - \frac{n\pi}{2}\right), \tag{11.159}$$

$$h_n^{(1)}(x) \sim (-i)^{n+1}\frac{e^{ix}}{x} = -i\frac{e^{i(x-n\pi/2)}}{x}, \tag{11.160a}$$

$$h_n^{(2)}(x) \sim i^{n+1}\frac{e^{-ix}}{x} = i\frac{e^{-i(x-n\pi/2)}}{x}. \tag{11.160b}$$

[26] The condition that the second term in the series be negligible compared to the first is actually $x \ll 2[(2n+2)(2n+3)/(n+1)]^{1/2}$ for $j_n(x)$.

The condition for these spherical Bessel forms is that $x \gg n(n+1)/2$. From these asymptotic values we see that $j_n(x)$ and $n_n(x)$ are appropriate for a description of **standing spherical waves**; $h_n^{(1)}(x)$ and $h_n^{(2)}(x)$ correspond to **traveling spherical waves**. If the time dependence for the traveling waves is taken to be $e^{-i\omega t}$, then $h_n^{(1)}(x)$ yields an outgoing traveling spherical wave, $h_n^{(2)}(x)$ an incoming wave. Radiation theory in electromagnetism and scattering theory in quantum mechanics provide many applications.

Recurrence Relations

The recurrence relations to which we now turn provide a convenient way of developing the higher-order spherical Bessel functions. These recurrence relations may be derived from the series, but, as with the modified Bessel functions, it is easier to substitute into the known recurrence relations (Eqs. (11.10) and (11.12)). This gives

$$f_{n-1}(x) + f_{n+1}(x) = \frac{2n+1}{x} f_n(x), \tag{11.161}$$

$$n f_{n-1}(x) - (n+1) f_{n+1}(x) = (2n+1) f_n'(x). \tag{11.162}$$

Rearranging these relations (or substituting into Eqs. (11.15) and (11.17)), we obtain

$$\frac{d}{dx}\left[x^{n+1} f_n(x)\right] = x^{n+1} f_{n-1}(x), \tag{11.163}$$

$$\frac{d}{dx}\left[x^{-n} f_n(x)\right] = -x^{-n} f_{n+1}(x). \tag{11.164}$$

Here f_n may represent $j_n, n_n, h_n^{(1)}$, or $h_n^{(2)}$.

The specific forms, Eqs. (11.154) and (11.155), may also be readily obtained from Eq. (11.164).

By mathematical induction we may establish the Rayleigh formulas

$$j_n(x) = (-1)^n x^n \left(\frac{1}{x}\frac{d}{dx}\right)^n \left(\frac{\sin x}{x}\right), \tag{11.165}$$

$$n_n(x) = -(-1)^n x^n \left(\frac{1}{x}\frac{d}{dx}\right)^n \left(\frac{\cos x}{x}\right), \tag{11.166}$$

$$\begin{aligned} h_n^{(1)}(x) &= -i(-1)^n x^n \left(\frac{1}{x}\frac{d}{dx}\right)^n \left(\frac{e^{ix}}{x}\right), \\ h_n^{(2)}(x) &= i(-1)^n x^n \left(\frac{1}{x}\frac{d}{dx}\right)^n \left(\frac{e^{-ix}}{x}\right). \end{aligned} \tag{11.167}$$

Orthogonality

We may take the orthogonality integral for the ordinary Bessel functions (Eqs. (11.49) and (11.50)),

$$\int_0^a J_\nu\left(\alpha_{\nu p}\frac{\rho}{a}\right)J_\nu\left(\alpha_{\nu q}\frac{\rho}{a}\right)\rho\,d\rho=\frac{a^2}{2}\left[J_{\nu+1}(\alpha_{\nu p})\right]^2\delta_{pq}, \tag{11.168}$$

and substitute in the expression for j_n to obtain

$$\int_0^a j_n\left(\alpha_{np}\frac{\rho}{a}\right)j_n\left(\alpha_{nq}\frac{\rho}{a}\right)\rho^2\,d\rho=\frac{a^3}{2}\left[j_{n+1}(\alpha_{np})\right]^2\delta_{pq}. \tag{11.169}$$

Here α_{np} and α_{nq} are roots of j_n.

This represents orthogonality with respect to the roots of the Bessel functions. An illustration of this sort of orthogonality is provided in Example 11.7.1, the problem of a particle in a sphere. Equation (11.169) guarantees orthogonality of the wave functions $j_n(r)$ for fixed n. (If n varies, the accompanying spherical harmonic will provide orthogonality.)

Example 11.7.1 PARTICLE IN A SPHERE

An illustration of the use of the spherical Bessel functions is provided by the problem of a quantum mechanical particle in a sphere of radius a. Quantum theory requires that the wave function ψ, describing our particle, satisfy

$$-\frac{\hbar^2}{2m}\nabla^2\psi=E\psi, \tag{11.170}$$

and the boundary conditions (1) $\psi(r\le a)$ remains finite, (2) $\psi(a)=0$. This corresponds to a square-well potential $V=0$, $r\le a$, and $V=\infty$, $r>a$. Here $\hbar$ is Planck's constant divided by 2π, m is the mass of our particle, and E is, its energy. Let us determine the **minimum** value of the energy for which our wave equation has an acceptable solution. Equation (11.170) is Helmholtz's equation with a radial part (compare Section 9.3 for separation of variables):

$$\frac{d^2R}{dr^2}+\frac{2}{r}\frac{dR}{dr}+\left[k^2-\frac{n(n+1)}{r^2}\right]R=0, \tag{11.171}$$

with $k^2=2mE/\hbar^2$. Hence by Eq. (11.139), with $n=0$,

$$R=Aj_0(kr)+Bn_0(kr).$$

We choose the orbital angular momentum index $n=0$, for any angular dependence would raise the energy. The spherical Neumann function is rejected because of its divergent behavior at the origin. To satisfy the second boundary condition (for all angles), we require

$$ka=\frac{\sqrt{2mE}}{\hbar}a=\alpha, \tag{11.172}$$

where α is a root of j_0, that is, $j_0(\alpha) = 0$. This has the effect of limiting the allowable energies to a certain discrete set, or, in other words, application of boundary condition (2) quantizes the energy E. The smallest α is the first zero of j_0,

$$\alpha = \pi,$$

and

$$E_{\min} = \frac{\pi^2\hbar^2}{2ma^2} = \frac{h^2}{8ma^2}, \tag{11.173}$$

which means that for any finite sphere the particle energy will have a positive minimum or zero-point energy. This is an illustration of the Heisenberg uncertainty principle for Δp with $\Delta r \leq a$.

In solid-state physics, astrophysics, and other areas of physics, we may wish to know how many different solutions (energy states) correspond to energies less than or equal to some fixed energy E_0. For a cubic volume (Exercise 9.3.5) the problem is fairly simple. The considerably more difficult spherical case is worked out by R. H. Lambert, *Am. J. Phys.* **36**: 417, 1169 (1968).

The relevant orthogonality relation for the $j_n(kr)$ can be derived from the integral given in Exercise 11.7.23. ■

Another form, orthogonality with respect to the indices, may be written as

$$\int_{-\infty}^{\infty} j_m(x)\, j_n(x)\, dx = 0, \qquad m \neq n,\ m, n \geq 0. \tag{11.174}$$

The proof is left as Exercise 11.7.10. If $m = n$ (compare Exercise 11.7.11), we have

$$\int_{-\infty}^{\infty} \left[j_n(x)\right]^2 dx = \frac{\pi}{2n+1}. \tag{11.175}$$

Most physical applications of orthogonal Bessel and spherical Bessel functions involve orthogonality with varying roots and an interval $[0, a]$ and Eqs. (11.168) and (11.169) and Exercise 11.7.23 for continuous-energy eigenvalues.

The spherical Bessel functions will enter again in connection with spherical waves, but further consideration is postponed until the corresponding angular functions, the Legendre functions, have been introduced.

Exercises

11.7.1 Show that if

$$n_n(x) = \sqrt{\frac{\pi}{2x}} N_{n+1/2}(x),$$

it automatically equals

$$(-1)^{n+1}\sqrt{\frac{\pi}{2x}} J_{-n-1/2}(x).$$

11.7.2 Derive the trigonometric-polynomial forms of $j_n(z)$ and $n_n(z)$.[27]

$$j_n(z) = \frac{1}{z}\sin\left(z - \frac{n\pi}{2}\right)\sum_{s=0}^{[n/2]} \frac{(-1)^s(n+2s)!}{(2s)!(2z)^{2s}(n-2s)!} + \frac{1}{z}\cos\left(z - \frac{n\pi}{2}\right)\sum_{s=0}^{[(n-1)/2]} \frac{(-1)^s(n+2s+1)!}{(2s+1)!(2z)^{2s}(n-2s-1)!},$$

$$n_n(z) = \frac{(-1)^{n+1}}{z}\cos\left(z + \frac{n\pi}{2}\right)\sum_{s=0}^{[n/2]} \frac{(-1)^s(n+2s)!}{(2s)!(2z)^{2s}(n-2s)!} + \frac{(-1)^{n+1}}{z}\sin\left(z + \frac{n\pi}{2}\right)\sum_{s=0}^{[(n-1)/2]} \frac{(-1)^s(n+2s+1)!}{(2s+1)!(2z)^{2s+1}(n-2s-1)!}.$$

11.7.3 Use the integral representation of $J_\nu(x)$,

$$J_\nu(x) = \frac{1}{\pi^{1/2}(\nu - \frac{1}{2})!}\left(\frac{x}{2}\right)^\nu \int_{-1}^{1} e^{\pm ixp}\left(1-p^2\right)^{\nu-1/2} dp,$$

to show that the spherical Bessel functions $j_n(x)$ are expressible in terms of trigonometric functions; that is, for example,

$$j_0(x) = \frac{\sin x}{x}, \qquad j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x}.$$

11.7.4 (a) Derive the recurrence relations

$$f_{n-1}(x) + f_{n+1}(x) = \frac{2n+1}{x} f_n(x),$$
$$nf_{n-1}(x) - (n+1)f_{n+1}(x) = (2n+1)f_n'(x)$$

satisfied by the spherical Bessel functions $j_n(x), n_n(x), h_n^{(1)}(x)$, and $h_n^{(2)}(x)$.

(b) Show, from these two recurrence relations, that the spherical Bessel function $f_n(x)$ satisfies the differential equation

$$x^2 f_n''(x) + 2x f_n'(x) + \left[x^2 - n(n+1)\right] f_n(x) = 0.$$

11.7.5 Prove by mathematical induction that

$$j_n(x) = (-1)^n x^n \left(\frac{1}{x}\frac{d}{dx}\right)^n \left(\frac{\sin x}{x}\right)$$

for n an arbitrary nonnegative integer.

11.7.6 From the discussion of orthogonality of the spherical Bessel functions, show that a Wronskian relation for $j_n(x)$ and $n_n(x)$ is

$$j_n(x)n_n'(x) - j_n'(x)n_n(x) = \frac{1}{x^2}.$$

[27] The upper limit on the summation $[n/2]$ means the largest **integer** that does not exceed $n/2$.

11.7.7 Verify

$$h_n^{(1)}(x)h_n^{(2)'}(x) - h_n^{(1)'}(x)h_n^{(2)}(x) = -\frac{2i}{x^2}.$$

11.7.8 Verify Poisson's integral representation of the spherical Bessel function,

$$j_n(z) = \frac{z^n}{2^{n+1}n!}\int_0^\pi \cos(z\cos\theta)\sin^{2n+1}\theta\, d\theta.$$

11.7.9 Show that

$$\int_0^\infty J_\mu(x)J_\nu(x)\frac{dx}{x} = \frac{2}{\pi}\frac{\sin[(\mu-\nu)\pi/2]}{\mu^2-\nu^2}, \qquad \mu+\nu > -1.$$

11.7.10 Derive Eq. (11.174):

$$\int_{-\infty}^\infty j_m(x)j_n(x)\,dx = 0, \quad \begin{matrix} m \neq n \\ m, n \geq 0. \end{matrix}$$

11.7.11 Derive Eq. (11.175):

$$\int_{-\infty}^\infty \big[j_n(x)\big]^2\,dx = \frac{\pi}{2n+1}.$$

11.7.12 Set up the orthogonality integral for $j_L(kr)$ in a sphere of radius R with the boundary condition

$$j_L(kR) = 0.$$

The result is used in classifying electromagnetic radiation according to its angular momentum.

11.7.13 The Fresnel integrals (Fig. 11.15 and Exercise 5.10.2) occurring in diffraction theory are given by

$$x(t) = \sqrt{\frac{\pi}{2}}C\left(\sqrt{\frac{\pi}{2}}t\right) = \int_0^t \cos\big(v^2\big)\,dv, \qquad y(t) = \sqrt{\frac{\pi}{2}}S\left(\sqrt{\frac{\pi}{2}}t\right) = \int_0^t \sin\big(v^2\big)\,dv.$$

Show that these integrals may be expanded in series of spherical Bessel functions

$$x(s) = \frac{1}{2}\int_0^s j_{-1}(u)u^{1/2}\,du = s^{1/2}\sum_{n=0}^\infty j_{2n}(s),$$

$$y(s) = \frac{1}{2}\int_0^s j_0(u)u^{1/2}\,du = s^{1/2}\sum_{n=0}^\infty j_{2n+1}(s).$$

Hint. To establish the equality of the integral and the sum, you may wish to work with their derivatives. The spherical Bessel analogs of Eqs. (11.12) and (11.14) are helpful.

11.7.14 A hollow sphere of radius a (Helmholtz resonator) contains standing sound waves. Find the minimum frequency of oscillation in terms of the radius a and the velocity of sound v. The sound waves satisfy the wave equation

$$\nabla^2\psi = \frac{1}{v^2}\frac{\partial^2\psi}{\partial t^2}$$

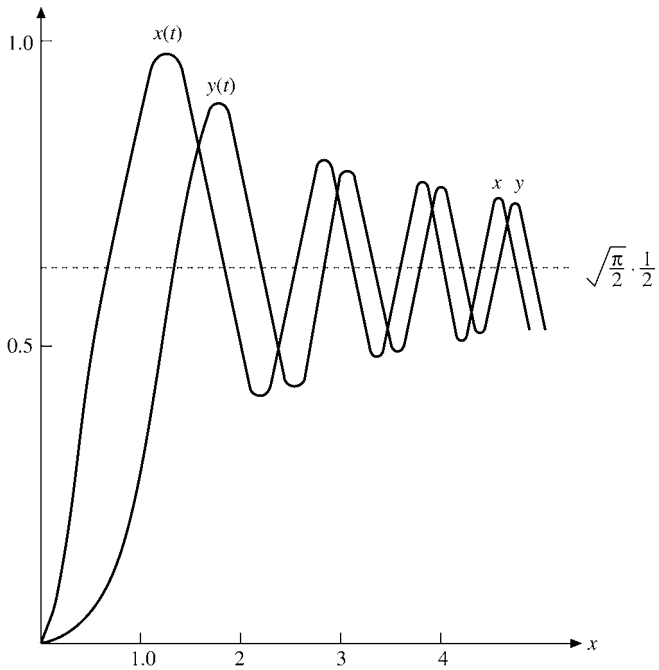

FIGURE 11.15 Fresnel integrals.

and the boundary condition

$$\frac{\partial \psi}{\partial r} = 0, \qquad r = a.$$

This is a Neumann boundary condition. Example 11.7.1 has the same PDE but with a Dirichlet boundary condition.

ANS. $\nu_{\min} = 0.3313v/a, \qquad \lambda_{\max} = 3.018a.$

11.7.15 Defining the spherical modified Bessel functions (Fig. 11.16) by

$$i_n(x) = \sqrt{\frac{\pi}{2x}} I_{n+1/2}(x), \qquad k_n(x) = \sqrt{\frac{2}{\pi x}} K_{n+1/2}(x),$$

show that

$$i_0(x) = \frac{\sinh x}{x}, \qquad k_0(x) = \frac{e^{-x}}{x}.$$

Note that the numerical factors in the definitions of i_n and k_n are not identical.

11.7.16 (a) Show that the parity of $i_n(x)$ is $(-1)^n$.
(b) Show that $k_n(x)$ has no definite parity.

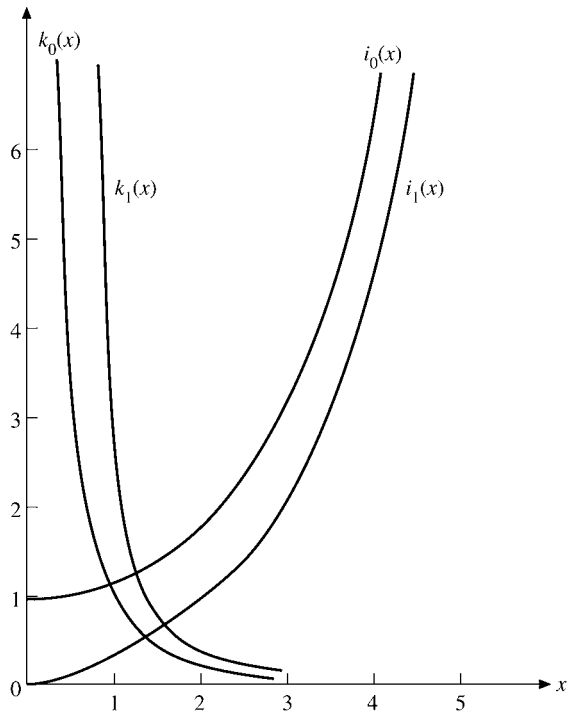

FIGURE 11.16 Spherical modified Bessel functions.

11.7.17 Show that the spherical modified Bessel functions satisfy the following relations:

(a) $i_n(x) = i^{-n} j_n(ix)$,
$k_n(x) = -i^n h_n^{(1)}(ix)$,

(b) $i_{n+1}(x) = x^n \dfrac{d}{dx}\left(x^{-n} i_n\right)$,

$k_{n+1}(x) = -x^n \dfrac{d}{dx}\left(x^{-n} k_n\right)$,

(c) $i_n(x) = x^n \left(\dfrac{1}{x}\dfrac{d}{dx}\right)^n \dfrac{\sinh x}{x}$,

$k_n(x) = (-1)^n x^n \left(\dfrac{1}{x}\dfrac{d}{dx}\right)^n \dfrac{e^{-x}}{x}$.

11.7.18 Show that the recurrence relations for $i_n(x)$ and $k_n(x)$ are

(a) $i_{n-1}(x) - i_{n+1}(x) = \dfrac{2n+1}{x} i_n(x)$,

$n i_{n-1}(x) + (n+1) i_{n+1}(x) = (2n+1) i_n'(x)$,

(b) $k_{n-1}(x) - k_{n+1}(x) = -\dfrac{2n+1}{x}k_n(x),$

$nk_{n-1}(x) + (n+1)k_{n+1}(x) = -(2n+1)k_n'(x).$

11.7.19 Derive the limiting values for the spherical modified Bessel functions

(a) $$i_n(x) \approx \frac{x^n}{(2n+1)!!}, \qquad k_n(x) \approx \frac{(2n-1)!!}{x^{n+1}}, \qquad x \ll 1.$$

(b) $$i_n(x) \sim \frac{e^x}{2x}, \qquad k_n(x) \sim \frac{e^{-x}}{x}, \qquad x \gg \frac{1}{2}n(n+1).$$

11.7.20 Show that the Wronskian of the spherical modified Bessel functions is given by

$$i_n(x)k_n'(x) - i_n'(x)k_n(x) = -\frac{1}{x^2}.$$

11.7.21 A quantum particle of mass M is trapped in a "square" well of radius a. The Schrödinger equation potential is

$$V(r) = \begin{cases} -V_0, & 0 \le r < a \\ 0, & r > a. \end{cases}$$

The particle's energy E is negative (an eigenvalue).

(a) Show that the radial part of the wave function is given by $j_l(k_1 r)$ for $0 \le r < a$ and $k_l(k_2 r)$ for $r > a$. (We require that $\psi(0)$ be finite and $\psi(\infty) \to 0$.) Here $k_1^2 = 2M(E + V_0)/\hbar^2$, $k_2^2 = -2ME/\hbar^2$, and l is the angular momentum (n in Eq. (11.139)).

(b) The boundary condition at $r = a$ is that the wave function $\psi(r)$ and its first derivative be continuous. Show that this means

$$\left.\frac{(d/dr)j_l(k_1 r)}{j_l(k_1 r)}\right|_{r=a} = \left.\frac{(d/dr)k_l(k_2 r)}{k_l(k_2 r)}\right|_{r=a}.$$

This equation determines the energy eigenvalues.

Note. This is a generalization of Example 10.1.2.

11.7.22 The quantum mechanical radial wave function for a scattered wave is given by

$$\psi_k = \frac{\sin(kr + \delta_0)}{kr},$$

where k is the wave number, $k = \sqrt{2mE/\hbar}$, and δ_0 is the scattering phase shift. Show that the normalization integral is

$$\int_0^\infty \psi_k(r)\psi_{k'}(r)r^2\,dr = \frac{\pi}{2k}\delta(k - k').$$

Hint. You can use a sine representation of the Dirac delta function. See Exercise 15.3.8.

11.7.23 Derive the spherical Bessel function closure relation

$$\frac{2a^2}{\pi}\int_0^\infty j_n(ar)\,j_n(br)r^2\,dr = \delta(a-b).$$

Note. An interesting derivation involving Fourier transforms, the Rayleigh plane-wave expansion, and spherical harmonics has been given by P. Ugincius, *Am. J. Phys.* **40**: 1690 (1972).

11.7.24 (a) Write a subroutine that will generate the spherical Bessel functions, $j_n(x)$, that is, will generate the numerical value of $j_n(x)$ given x and n.
Note. One possibility is to use the explicit known forms of j_0 and j_1 and to develop the higher index j_n, by repeated application of the recurrence relation.

(b) Check your subroutine by an independent calculation, such as Eq. (11.154). If possible, compare the machine time needed for this check with the time required for your subroutine.

11.7.25 The wave function of a particle in a sphere (Example 11.7.1) with angular momentum l is $\psi(r,\theta,\varphi) = A\,j_l(\sqrt{2ME}\,r/\hbar)Y_l^m(\theta,\varphi)$. The $Y_l^m(\theta,\varphi)$ is a spherical harmonic, described in Section 12.6. From the boundary condition $\psi(a,\theta,\varphi) = 0$ or $j_l((\sqrt{2ME})a/\hbar) = 0$ calculate the 10 lowest-energy states. Disregard the m degeneracy ($2l+1$ values of m for each choice of l). Check your results against AMS-55, Table 10.6, see Additional Readings for Chapter 8 for the reference.
Hint. You can use your spherical Bessel subroutine and a root-finding subroutine.

Check values. $j_l(\alpha_{ls}) = 0$,
$\alpha_{01} = 3.1416$
$\alpha_{11} = 4.4934$
$\alpha_{21} = 5.7635$
$\alpha_{02} = 6.2832.$

11.7.26 Let Example 11.7.1 be modified so that the potential is a finite V_0 outside ($r > a$).

(a) For $E < V_0$ show that

$$\psi_{\text{out}}(r,\theta,\varphi) \sim k_l\left(\frac{r}{\hbar}\sqrt{2M(V_0-E)}\right).$$

(b) The new boundary conditions to be satisfied at $r = a$ are

$$\psi_{\text{in}}(a,\theta,\varphi) = \psi_{\text{out}}(a,\theta,\varphi),$$
$$\frac{\partial}{\partial r}\psi_{\text{in}}(a,\theta,\varphi) = \frac{\partial}{\partial r}\psi_{\text{out}}(a,\theta,\varphi)$$

or

$$\frac{1}{\psi_{\text{in}}}\frac{\partial\psi_{\text{in}}}{\partial r}\bigg|_{r=a} = \frac{1}{\psi_{\text{out}}}\frac{\partial\psi_{\text{out}}}{\partial r}\bigg|_{r=a}.$$

For $l = 0$ show that the boundary condition at $r = a$ leads to

$$f(E) = k\left\{\cot ka - \frac{1}{ka}\right\} + k'\left\{1 + \frac{1}{k'a}\right\} = 0,$$

where $k = \sqrt{2ME}/\hbar$ and $k' = \sqrt{2M(V_0-E)}/\hbar$.

(c) With $a = 4\pi\varepsilon_0\hbar^2/Me^2$ (Bohr radius) and $V_0 = 4Me^4/2\hbar^2$, compute the possible bound states ($0 < E < V_0$).
Hint. Call a root-finding subroutine after you know the approximate location of the roots of

$$f(E) = 0 \qquad (0 \le E \le V_0).$$

(d) Show that when $a = 4\pi\varepsilon_0\hbar^2/Me^2$ the minimum value of V_0 for which a bound state exists is $V_0 = 2.4674Me^4/2\hbar^2$.

11.7.27 In some nuclear stripping reactions the differential cross section is proportional to $j_l(x)^2$, where l is the angular momentum. The location of the maximum on the curve of experimental data permits a determination of l, if the location of the (first) maximum of $j_l(x)$ is known. Compute the location of the first maximum of $j_1(x)$, $j_2(x)$, and $j_3(x)$. *Note*. For better accuracy look for the first zero of $j_l'(x)$. Why is this more accurate than direct location of the maximum?

Additional Readings

Jackson, J. D., *Classical Electrodynamics*, 3rd ed., New York: J. Wiley (1999).

McBride, E. B., *Obtaining Generating Functions*. New York: Springer-Verlag (1971). An introduction to methods of obtaining generating functions.

Watson, G. N., *A Treatise on the Theory of Bessel Functions*, 2nd ed. Cambridge, UK: Cambridge University Press (1952). This is the definitive text on Bessel functions and their properties. Although difficult reading, it is invaluable as the ultimate reference.

Watson, G. N., *A Treatise on the Theory of Bessel Functions*, 1st ed. Cambridge, UK: Cambridge University Press (1922). See also the references listed at the end of Chapter 13.

This page intentionally left blank

CHAPTER 12

LEGENDRE FUNCTIONS

12.1 GENERATING FUNCTION

Legendre polynomials appear in many different mathematical and physical situations. (1) They may originate as solutions of the Legendre ODE which we have already encountered in the separation of variables (Section 9.3) for Laplace's equation, Helmholtz's equation, and similar ODEs in spherical polar coordinates. (2) They enter as a consequence of a Rodrigues' formula (Section 12.4). (3) They arise as a consequence of demanding a complete, orthogonal set of functions over the interval $[-1, 1]$ (Gram–Schmidt orthogonalization, Section 10.3). (4) In quantum mechanics they (really the spherical harmonics, Sections 12.6 and 12.7) represent angular momentum eigenfunctions. (5) They are generated by a generating function. We introduce Legendre polynomials here by way of a generating function.

Physical Basis — Electrostatics

As with Bessel functions, it is convenient to introduce the Legendre polynomials by means of a generating function, which here appears in a physical context. Consider an electric charge q placed on the z-axis at $z = a$. As shown in Fig. 12.1, the electrostatic potential of charge q is

$$\varphi = \frac{1}{4\pi\varepsilon_0} \cdot \frac{q}{r_1} \qquad \text{(SI units)}. \tag{12.1}$$

We want to express the electrostatic potential in terms of the spherical polar coordinates r and θ (the coordinate φ is absent because of symmetry about the z-axis). Using the law of cosines in Fig. 12.1, we obtain

$$\varphi = \frac{q}{4\pi\varepsilon_0}\left(r^2 + a^2 - 2ar\cos\theta\right)^{-1/2}. \tag{12.2}$$

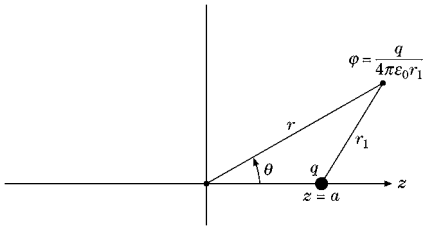

FIGURE 12.1 Electrostatic potential. Charge q displaced from origin.

Legendre Polynomials

Consider the case of $r > a$ or, more precisely, $r^2 > |a^2 - 2ar\cos\theta|$. The radical in Eq. (12.2) may be expanded in a binomial series and then rearranged in powers of (a/r). The Legendre polynomial $P_n(\cos\theta)$ (see Fig. 12.2) is defined as the coefficient of the nth power in

$$\varphi = \frac{q}{4\pi\varepsilon_0 r}\sum_{n=0}^{\infty} P_n(\cos\theta)\left(\frac{a}{r}\right)^n. \tag{12.3}$$

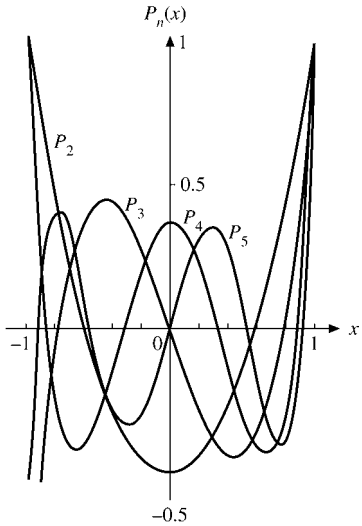

FIGURE 12.2 Legendre polynomials $P_2(x)$, $P_3(x)$, $P_4(x)$, and $P_5(x)$.

Dropping the factor $q/4\pi\varepsilon_0 r$ and using x and t instead of $\cos\theta$ and a/r, respectively, we have

$$\boxed{g(t,x)=\left(1-2xt+t^2\right)^{-1/2}=\sum_{n=0}^{\infty}P_n(x)t^n, \qquad |t|<1.} \tag{12.4}$$

Equation (12.4) is our generating function formula. In the next section it is shown that $|P_n(\cos\theta)|\le 1$, which means that the series expansion (Eq. (12.4)) is convergent for $|t|<1$.[1] Indeed, the series is convergent for $|t|=1$ except for $|x|=1$.

In physical applications Eq. (12.4) often appears in the vector form (see Section 9.7)

$$\frac{1}{|\mathbf{r}_1-\mathbf{r}_2|}=\frac{1}{r_>}\sum_{n=0}^{\infty}\left(\frac{r_<}{r_>}\right)^n P_n(\cos\theta), \tag{12.4a}$$

where

$$\left.\begin{aligned} r_> &= |\mathbf{r}_1| \\ r_< &= |\mathbf{r}_2| \end{aligned}\right\} \qquad \text{for } |\mathbf{r}_1|>|\mathbf{r}_2|, \tag{12.4b}$$

and

$$\left.\begin{aligned} r_> &= |\mathbf{r}_2| \\ r_< &= |\mathbf{r}_1| \end{aligned}\right\} \qquad \text{for } |\mathbf{r}_2|>|\mathbf{r}_1|. \tag{12.4c}$$

Using the binomial theorem (Section 5.6) and Exercise 8.1.15, we expand the generating function as (compare Eq. (12.33))

$$\begin{aligned}\left(1-2xt+t^2\right)^{-1/2} &= \sum_{n=0}^{\infty}\frac{(2n)!}{2^{2n}(n!)^2}\left(2xt-t^2\right)^n \\ &= 1+\sum_{n=1}^{\infty}\frac{(2n-1)!!}{(2n)!!}\left(2xt-t^2\right)^n. \end{aligned} \tag{12.5}$$

For the first few Legendre polynomials, say, P_0, P_1, and P_2, we need the coefficients of t^0, t^1, and t^2. These powers of t appear only in the terms $n=0$, 1, and 2, and hence we may limit our attention to the first three terms of the infinite series:

$$\begin{aligned}&\frac{0!}{2^0(0!)^2}\left(2xt-t^2\right)^0+\frac{2!}{2^2(1!)^2}\left(2xt-t^2\right)^1+\frac{4!}{2^4(2!)^2}\left(2xt-t^2\right)^2 \\ &\quad=1t^0+xt^1+\left(\frac{3}{2}x^2-\frac{1}{2}\right)t^2+\mathcal{O}\left(t^3\right).\end{aligned}$$

Then, from Eq. (12.4) (and uniqueness of power series),

$$P_0(x)=1, \qquad P_1(x)=x, \qquad P_2(x)=\frac{3}{2}x^2-\frac{1}{2}.$$

We repeat this limited development in a vector framework later in this section.

[1] Note that the series in Eq. (12.3) is convergent for $r>a$, even though the binomial expansion involved is valid only for $r>(a^2+2ar)^{1/2}$ and $\cos\theta=-1$, or $r>a(1+\sqrt{2})$.

In employing a general treatment, we find that the binomial expansion of the $(2xt - t^2)^n$ factor yields the double series

$$\left(1 - 2xt + t^2\right)^{-1/2} = \sum_{n=0}^{\infty} \frac{(2n)!}{2^{2n}(n!)^2} t^n \sum_{k=0}^{n} (-1)^k \frac{n!}{k!(n-k)!} (2x)^{n-k} t^k$$

$$= \sum_{n=0}^{\infty} \sum_{k=0}^{n} (-1)^k \frac{(2n)!}{2^{2n} n! k! (n-k)!} \cdot (2x)^{n-k} t^{n+k}. \tag{12.6}$$

From Eq. (5.64) of Section 5.4 (rearranging the order of summation), Eq. (12.6) becomes

$$\left(1 - 2xt + t^2\right)^{-1/2} = \sum_{n=0}^{\infty} \sum_{k=0}^{[n/2]} (-1)^k \frac{(2n-2k)!}{2^{2n-2k} k!(n-k)!(n-2k)!} \cdot (2x)^{n-2k} t^n, \tag{12.7}$$

with the t^n independent of the index k.[2] Now, equating our two power series (Eqs. (12.4) and (12.7)) term by term, we have[3]

$$P_n(x) = \sum_{k=0}^{[n/2]} (-1)^k \frac{(2n-2k)!}{2^n k!(n-k)!(n-2k)!} x^{n-2k}. \tag{12.8}$$

Hence, for n even, P_n has only even powers of x and even parity (see Eq. (12.37)), and odd powers and odd parity for odd n.

Linear Electric Multipoles

Returning to the electric charge on the z-axis, we demonstrate the usefulness and power of the generating function by adding a charge $-q$ at $z = -a$, as shown in Fig. 12.3. The

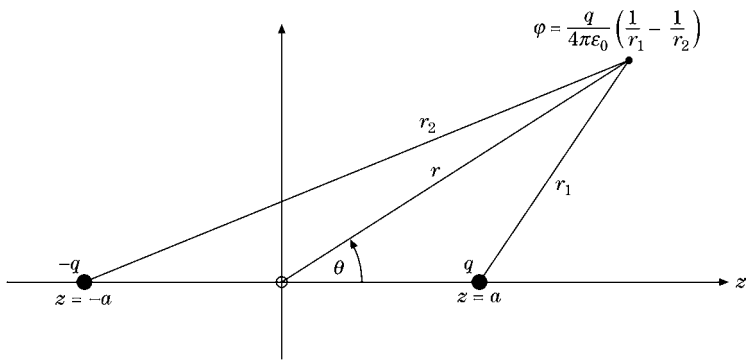

FIGURE 12.3 Electric dipole.

[2] $[n/2] = n/2$ for n even, $(n-1)/2$ for n odd.

[3] Equation (12.8) starts with x^n. By changing the index, we can transform it into a series that starts with x^0 for n even and x^1 for n odd. These ascending series are given as hypergeometric functions in Eqs. (13.138) and (13.139), Section 13.4.

potential becomes

$$\varphi = \frac{q}{4\pi\varepsilon_0}\left(\frac{1}{r_1} - \frac{1}{r_2}\right), \tag{12.9}$$

and by using the law of cosines, we have

$$\varphi = \frac{q}{4\pi\varepsilon_0 r}\left\{\left[1 - 2\left(\frac{a}{r}\right)\cos\theta + \left(\frac{a}{r}\right)^2\right]^{-1/2} - \left[1 + 2\left(\frac{a}{r}\right)\cos\theta + \left(\frac{a}{r}\right)^2\right]^{-1/2}\right\}, \qquad (r > a).$$

Clearly, the second radical is like the first, except that a has been replaced by $-a$. Then, using Eq. (12.4), we obtain

$$\varphi = \frac{q}{4\pi\varepsilon_0 r}\left[\sum_{n=0}^{\infty} P_n(\cos\theta)\left(\frac{a}{r}\right)^n - \sum_{n=0}^{\infty} P_n(\cos\theta)(-1)^n\left(\frac{a}{r}\right)^n\right]$$

$$= \frac{2q}{4\pi\varepsilon_0 r}\left[P_1(\cos\theta)\left(\frac{a}{r}\right) + P_3(\cos\theta)\left(\frac{a}{r}\right)^3 + \cdots\right]. \tag{12.10}$$

The first term (and dominant term for $r \gg a$) is

$$\varphi = \frac{2aq}{4\pi\varepsilon_0}\cdot\frac{P_1(\cos\theta)}{r^2}, \tag{12.11}$$

which is the **electric dipole** potential, and $2aq$ is the dipole moment (Fig. 12.3). This analysis may be extended by placing additional charges on the z-axis so that the P_1 term, as well as the P_0 (monopole) term, is canceled. For instance, charges of q at $z = a$ and $z = -a$, $-2q$ at $z = 0$ give rise to a potential whose series expansion starts with $P_2(\cos\theta)$. This is a linear electric quadrupole. Two linear quadrupoles may be placed so that the quadrupole term is canceled but the P_3, the octupole term, survives.

Vector Expansion

We consider the electrostatic potential produced by a distributed charge $\rho(\mathbf{r}_2)$:

$$\varphi(\mathbf{r}_1) = \frac{1}{4\pi\varepsilon_0}\int \frac{\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d^3 r_2. \tag{12.12a}$$

This expression has already appeared in Sections 1.16 and 9.7. Taking the denominator of the integrand, using first the law of cosines and then a binomial expansion, yields (see Fig. 1.42)

$$\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} = \left(r_1^2 - 2\mathbf{r}_1\cdot\mathbf{r_2} + r_2^2\right)^{-1/2} \tag{12.12b}$$

$$= \frac{1}{r_1}\left[1 + \left(-\frac{2\mathbf{r}_1\cdot\mathbf{r}_2}{r_1^2} + \frac{r_2^2}{r_1^2}\right)\right]^{-1/2}, \qquad \text{for } r_1 > r_2$$

$$= \frac{1}{r_1}\left[1 + \frac{\mathbf{r}_1\cdot\mathbf{r}_2}{r_1^2} - \frac{1}{2}\frac{r_2^2}{r_1^2} + \frac{3}{2}\frac{(\mathbf{r}_1\cdot\mathbf{r}_2)^2}{r_1^4} + \mathcal{O}\left(\frac{r_2}{r_1}\right)^3\right].$$

(For $r_1 = 1, r_2 = t$, and $\mathbf{r}_1 \cdot \mathbf{r}_2 = xt$, Eq. (12.12b) reduces to the generating function, Eq. (12.4).)

The first term in the square bracket, 1, yields a potential

$$\varphi_0(\mathbf{r}_1) = \frac{1}{4\pi\varepsilon_0}\frac{1}{r_1}\int \rho(\mathbf{r}_2)\, d^3r_2. \tag{12.12c}$$

The integral is just the total charge. This part of the total potential is an electric **monopole**.

The second term yields

$$\varphi_1(\mathbf{r}_1) = \frac{1}{4\pi\varepsilon_0}\frac{\mathbf{r}_1\cdot}{r_1^3}\int \mathbf{r}_2\rho(\mathbf{r}_2)\, d^3r_2, \tag{12.12d}$$

where the integral is the dipole moment whose charge density $\rho(\mathbf{r}_2)$ is weighted by a moment arm $\mathbf{r}_2$. We have an electric dipole potential. For atomic or nuclear states of definite parity, $\rho(\mathbf{r}_2)$ is an even function and the dipole integral is identically zero.

The last two terms, both of order $(r_2/r_1)^2$, may be handled by using Cartesian coordinates:

$$(\mathbf{r}_1 \cdot \mathbf{r}_2)^2 = \sum_{i=1}^{3} x_{1i}x_{2i} \sum_{j=1}^{3} x_{1j}x_{2j}.$$

Rearranging variables to take the x_1 components outside the integral yields

$$\varphi_2(\mathbf{r}_1) = \frac{1}{4\pi\varepsilon_0}\frac{1}{2r_1^5}\sum_{i,j=1}^{3} x_{1i}x_{1j}\int \left[3x_{2i}x_{2j} - \delta_{ij}r_2^2\right]\rho(\mathbf{r}_2)\, d^3r_2. \tag{12.12e}$$

This is the electric **quadrupole** term. We note that the square bracket in the integrand forms a symmetric, zero-trace tensor.

A general electrostatic multipole expansion can also be developed by using Eq. (12.12a) for the potential $\varphi(r_1)$ and replacing $1/(4\pi|\mathbf{r}_1 - \mathbf{r}_2|)$ by Green's function, Eq. (9.187). This yields the potential $\varphi(\mathbf{r}_1)$ as a (double) series of the spherical harmonics $Y_l^m(\theta_1, \varphi_1)$ and $Y_l^m(\theta_2, \varphi_2)$.

Before leaving multipole fields, perhaps we should emphasize three points.

- First, an electric (or magnetic) multipole is isolated and well defined only if all lower-order multipoles vanish. For instance, the potential of one charge q at $z = a$ was expanded in a series of Legendre polynomials. Although we refer to the $P_1(\cos\theta)$ term in this expansion as a dipole term, it should be remembered that this term exists only because of our choice of coordinates. We also have a monopole, $P_0(\cos\theta)$.
- Second, in physical systems we do not encounter pure multipoles. As an example, the potential of the finite dipole (q at $z = a, -q$ at $z = -a$) contained a $P_3(\cos\theta)$ term. These higher-order terms may be eliminated by shrinking the multipole to a point multipole, in this case keeping the product qa constant ($a \to 0, q \to \infty$) to maintain the same dipole moment.

- Third, the multipole theory is not restricted to electrical phenomena. Planetary configurations are described in terms of mass multipoles, Sections 12.3 and 12.6. Gravitational radiation depends on the time behavior of mass quadrupoles. (The gravitational radiation field is a **tensor** field. The radiation quanta, gravitons, carry two units of angular momentum.)

It might also be noted that a multipole expansion is actually a decomposition into the irreducible representations of the rotation group (Section 4.2).

Extension to Ultraspherical Polynomials

The generating function used here, $g(t, x)$, is actually a special case of a more general generating function,

$$\boxed{\frac{1}{(1-2xt+t^2)^{\alpha}} = \sum_{n=0}^{\infty} C_n^{(\alpha)}(x)t^n.} \tag{12.13}$$

The coefficients $C_n^{(\alpha)}(x)$ are the ultraspherical polynomials (proportional to the Gegenbauer polynomials). For $\alpha = 1/2$ this equation reduces to Eq. (12.4); that is, $C_n^{(1/2)}(x) = P_n(x)$. The cases $a = 0$ and $\alpha = 1$ are considered in Chapter 13 in connection with the Chebyshev polynomials.

Exercises

12.1.1 Develop the electrostatic potential for the array of charges shown. This is a linear electric quadrupole (Fig. 12.4).

12.1.2 Calculate the electrostatic potential of the array of charges shown in Fig. 12.5. Here is an example of two equal but oppositely directed dipoles. The dipole contributions cancel. The octupole terms do not cancel.

12.1.3 Show that the electrostatic potential produced by a charge q at $z = a$ for $r < a$ is

$$\varphi(\mathbf{r}) = \frac{q}{4\pi\varepsilon_0 a}\sum_{n=0}^{\infty}\left(\frac{r}{a}\right)^n P_n(\cos\theta).$$

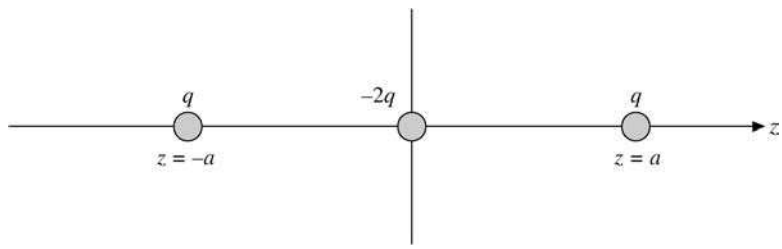

FIGURE 12.4 Linear electric quadrupole.

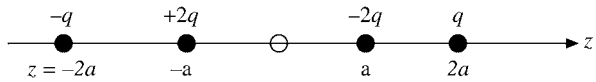

FIGURE 12.5 Linear electric octupole.

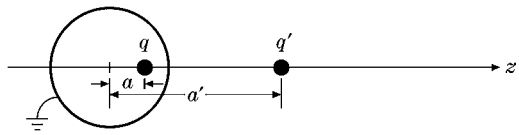

FIGURE 12.6

12.1.4 Using $\mathbf{E} = -\nabla\varphi$, determine the components of the electric field corresponding to the (pure) electric dipole potential

$$\varphi(\mathbf{r}) = \frac{2aqP_1(\cos\theta)}{4\pi\varepsilon_0 r^2}.$$

Here it is assumed that $r \gg a$.

$$\text{ANS. } E_r = +\frac{4aq\cos\theta}{4\pi\varepsilon_0 r^3}, \quad E_\theta = +\frac{2aq\sin\theta}{4\pi\varepsilon_0 r^3}, \quad E_\varphi = 0.$$

12.1.5 A point electric dipole of strength $p^{(1)}$ is placed at $z = a$; a second point electric dipole of equal but opposite strength is at the origin. Keeping the product $p^{(1)}a$ constant, let $a \to 0$. Show that this results in a point electric quadrupole.
Hint. Exercise 12.2.5 (when proved) will be helpful.

12.1.6 A point charge q is in the interior of a hollow conducting sphere of radius r_0. The charge q is displaced a distance a from the center of the sphere. If the conducting sphere is grounded, show that the potential in the interior produced by q and the distributed induced charge is the same as that produced by q and its image charge q'. The image charge is at a distance $a' = r_0^2/a$ from the center, collinear with q and the origin (Fig. 12.6).
Hint. Calculate the electrostatic potential for $a < r_0 < a'$. Show that the potential vanishes for $r = r_0$ if we take $q' = -qr_0/a$.

12.1.7 Prove that

$$P_n(\cos\theta) = (-1)^n \frac{r^{n+1}}{n!}\frac{\partial^n}{\partial z^n}\left(\frac{1}{r}\right).$$

Hint. Compare the Legendre polynomial expansion of the generating function ($a \to \Delta z$, Fig. 12.1) with a Taylor series expansion of $1/r$, where z dependence of r changes from z to $z - \Delta z$ (Fig. 12.7).

12.1.8 By differentiation and direct substitution of the series form, Eq. (12.8), show that $P_n(x)$ satisfies the Legendre ODE. Note that there is no restriction upon x. We may have any x, $-\infty < x < \infty$, and indeed any z in the entire finite complex plane.

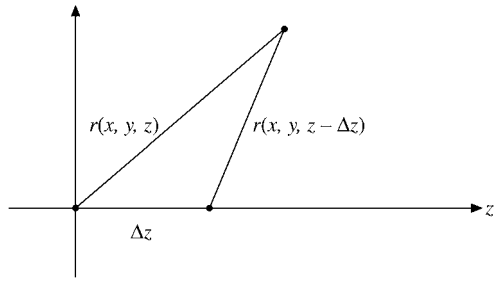

FIGURE 12.7

12.1.9 The Chebyshev polynomials (type II) are generated by (Eq. (13.93), Section 13.3)

$$\frac{1}{1-2xt+t^2}=\sum_{n=0}^{\infty}U_n(x)t^n.$$

Using the techniques of Section 5.4 for transforming series, develop a series representation of $U_n(x)$.

ANS. $U_n(x)=\sum_{k=0}^{[n/2]}(-1)^k\frac{(n-k)!}{k!(n-2k)!}(2x)^{n-2k}$.

12.2 RECURRENCE RELATIONS AND SPECIAL PROPERTIES

Recurrence Relations

The Legendre polynomial generating function provides a convenient way of deriving the recurrence relations[4] and some special properties. If our generating function (Eq. (12.4)) is differentiated with respect to t, we obtain

$$\frac{\partial g(t,x)}{\partial t}=\frac{x-t}{(1-2xt+t^2)^{3/2}}=\sum_{n=0}^{\infty}nP_n(x)t^{n-1}. \tag{12.14}$$

By substituting Eq. (12.4) into this and rearranging terms, we have

$$\left(1-2xt+t^2\right)\sum_{n=0}^{\infty}nP_n(x)t^{n-1}+(t-x)\sum_{n=0}^{\infty}P_n(x)t^n=0. \tag{12.15}$$

The left-hand side is a power series in t. Since this power series vanishes for all values of t, the coefficient of each power of t is equal to zero; that is, our power series is unique (Section 5.7). These coefficients are found by separating the individual summations and

[4]We can also apply the explicit series form Eq. (12.8) directly.

using distinctive summation indices:

$$\sum_{m=0}^{\infty} m P_m(x) t^{m-1} - \sum_{n=0}^{\infty} 2nx P_n(x) t^n + \sum_{s=0}^{\infty} s P_s(x) t^{s+1}$$
$$+ \sum_{s=0}^{\infty} P_s(x) t^{s+1} - \sum_{n=0}^{\infty} x P_n(x) t^n = 0. \tag{12.16}$$

Now, letting $m = n + 1$, $s = n - 1$, we find

$$(2n+1)x P_n(x) = (n+1) P_{n+1}(x) + n P_{n-1}(x), \qquad n = 1, 2, 3, \ldots. \tag{12.17}$$

This is another three-term recurrence relation, similar to (but not identical with) the recurrence relation for Bessel functions. With this recurrence relation we may easily construct the higher Legendre polynomials. If we take $n = 1$ and insert the easily found values of $P_0(x)$ and $P_1(x)$ (Exercise 12.1.7 or Eq. (12.8)), we obtain

$$3x P_1(x) = 2P_2(x) + P_0(x), \tag{12.18}$$

or

$$P_2(x) = \frac{1}{2}\left(3x^2 - 1\right). \tag{12.19}$$

This process may be continued indefinitely, the first few Legendre polynomials are listed in Table 12.1.

As cumbersome as it may appear at first, this technique is actually more efficient for a digital computer than is direct evaluation of the series (Eq. (12.8)). For greater stability (to avoid undue accumulation and magnification of round-off error), Eq. (12.17) is rewritten as

$$P_{n+1}(x) = 2x P_n(x) - P_{n-1}(x) - \frac{1}{n+1}\left[x P_n(x) - P_{n-1}(x)\right]. \tag{12.17a}$$

One starts with $P_0(x) = 1$, $P_1(x) = x$, and computes the **numerical** values of all the $P_n(x)$ for a given value of x up to the desired $P_N(x)$. The values of $P_n(x), 0 \le n < N$, are available as a fringe benefit.

Table 12.1 Legendre Polynomials

$P_0(x) = 1$
$P_1(x) = x$
$P_2(x) = \frac{1}{2}(3x^2 - 1)$
$P_3(x) = \frac{1}{2}(5x^3 - 3x)$
$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3)$
$P_5(x) = \frac{1}{8}(63x^5 - 70x^3 + 15x)$
$P_6(x) = \frac{1}{16}(231x^6 - 315x^4 + 105x^2 - 5)$
$P_7(x) = \frac{1}{16}(429x^7 - 693x^5 + 315x^3 - 35x)$
$P_8(x) = \frac{1}{128}(6435x^8 - 12012x^6 + 6930x^4 - 1260x^2 + 35)$

Differential Equations

More information about the behavior of the Legendre polynomials can be obtained if we now differentiate Eq. (12.4) with respect to x. This gives

$$\frac{\partial g(t,x)}{\partial x} = \frac{t}{(1-2xt+t^2)^{3/2}} = \sum_{n=0}^{\infty} P_n'(x)t^n, \tag{12.20}$$

or

$$\left(1-2xt+t^2\right)\sum_{n=0}^{\infty} P_n'(x)t^n - t\sum_{n=0}^{\infty} P_n(x)t^n = 0. \tag{12.21}$$

As before, the coefficient of each power of t is set equal to zero and we obtain

$$P_{n+1}'(x) + P_{n-1}'(x) = 2xP_n'(x) + P_n(x). \tag{12.22}$$

A more useful relation may be found by differentiating Eq. (12.17) with respect to x and multiplying by 2. To this we add $(2n+1)$ times Eq. (12.22), canceling the P_n' term. The result is

$$P_{n+1}'(x) - P_{n-1}'(x) = (2n+1)P_n(x). \tag{12.23}$$

From Eqs. (12.22) and (12.23) numerous additional equations may be developed,[5] including

$$P_{n+1}'(x) = (n+1)P_n(x) + xP_n'(x), \tag{12.24}$$

$$P_{n-1}'(x) = -nP_n(x) + xP_n'(x), \tag{12.25}$$

$$\left(1-x^2\right)P_n'(x) = nP_{n-1}(x) - nxP_n(x), \tag{12.26}$$

$$\left(1-x^2\right)P_n'(x) = (n+1)xP_n(x) - (n+1)P_{n+1}(x). \tag{12.27}$$

By differentiating Eq. (12.26) and using Eq. (12.25) to eliminate $P_{n-1}'(x)$, we find that $P_n(x)$ satisfies the linear second-order ODE

$$\boxed{\left(1-x^2\right)P_n''(x) - 2xP_n'(x) + n(n+1)P_n(x) = 0.} \tag{12.28}$$

The previous equations, Eqs. (12.22) to (12.27), are all first-order ODEs, but with polynomials of two different indices. The price for having all indices alike is a second-order

[5] Using the equation number in parentheses to denote the left-hand side of the equation, we may write the derivatives as

$$2\cdot\tfrac{d}{dx}(12.17) + (2n+1)\cdot(12.22) \Rightarrow (12.23),$$

$$\tfrac{1}{2}\{(12.22) + (12.23)\} \Rightarrow (12.24),$$

$$\tfrac{1}{2}\{(12.22) - (12.23)\} \Rightarrow (12.25),$$

$$(12.24)_{n\to n-1} + x\cdot(12.25) \Rightarrow (12.26),$$

$$\tfrac{d}{dx}(12.26) + n\cdot(12.25) \Rightarrow (12.28).$$

differential equation. Equation (12.28) is **Legendre's** ODE. We now see that the polynomials $P_n(x)$ generated by the power series for $(1-2xt+t^2)^{-1/2}$ satisfy Legendre's equation, which, of course, is why they are called Legendre polynomials.

In Eq. (12.28) differentiation is with respect to x ($x=\cos\theta$). Frequently, we encounter Legendre's equation expressed in terms of differentiation with respect to θ:

$$\boxed{\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{dP_n(\cos\theta)}{d\theta}\right)+n(n+1)P_n(\cos\theta)=0.} \tag{12.29}$$

Special Values

Our generating function provides still more information about the Legendre polynomials. If we set $x=1$, Eq. (12.4) becomes

$$\frac{1}{(1-2t+t^2)^{1/2}}=\frac{1}{1-t}=\sum_{n=0}^{\infty}t^n, \tag{12.30}$$

using a binomial expansion or the geometric series, Example 5.1.1. But Eq. (12.4) for $x=1$ defines

$$\frac{1}{(1-2t+t^2)^{1/2}}=\sum_{n=0}^{\infty}P_n(1)t^n.$$

Comparing the two series expansions (uniqueness of power series, Section 5.7), we have

$$P_n(1)=1. \tag{12.31}$$

If we let $x=-1$ in Eq. (12.4) and use

$$\frac{1}{(1+2t+t^2)^{1/2}}=\frac{1}{1+t},$$

this shows that

$$P_n(-1)=(-1)^n. \tag{12.32}$$

For obtaining these results, we find that the generating function is more convenient than the explicit series form, Eq. (12.8).

If we take $x=0$ in Eq. (12.4), using the binomial expansion

$$\left(1+t^2\right)^{-1/2}=1-\frac{1}{2}t^2+\frac{3}{8}t^4+\cdots+(-1)^n\frac{1\cdot3\cdots(2n-1)}{2^n n!}t^{2n}+\cdots, \tag{12.33}$$

we have[6]

$$P_{2n}(0)=(-1)^n\frac{1\cdot3\cdots(2n-1)}{2^n n!}=(-1)^n\frac{(2n-1)!!}{(2n)!!}=\frac{(-1)^n(2n)!}{2^{2n}(n!)^2} \tag{12.34}$$

$$P_{2n+1}(0)=0, \qquad n=0,1,2\ldots. \tag{12.35}$$

These results also follow from Eq. (12.8) by inspection.

[6]The double factorial notation is defined in Section 8.1:

$$(2n)!!=2\cdot4\cdot6\cdots(2n), \qquad (2n-1)!!=1\cdot3\cdot5\cdots(2n-1), \qquad (-1)!!=1.$$

Parity

Some of these results are special cases of the parity property of the Legendre polynomials. We refer once more to Eqs. (12.4) and (12.8). If we replace x by $-x$ and t by $-t$, the generating function is unchanged. Hence

$$g(t,x) = g(-t,-x) = \left[1-2(-t)(-x)+(-t)^2\right]^{-1/2}$$
$$= \sum_{n=0}^{\infty} P_n(-x)(-t)^n = \sum_{n=0}^{\infty} P_n(x)t^n. \tag{12.36}$$

Comparing these two series, we have

$$\boxed{P_n(-x) = (-1)^n P_n(x);} \tag{12.37}$$

that is, the polynomial functions are odd or even (with respect to $x = 0, \theta = \pi/2$) according to whether the index n is odd or even. This is the parity,[7] or reflection, property that plays such an important role in quantum mechanics. For central forces the index n is a measure of the orbital angular momentum, thus linking parity and orbital angular momentum.

This parity property is confirmed by the series solution and for the special values tabulated in Table 12.1. It might also be noted that Eq. (12.37) may be predicted by inspection of Eq. (12.17), the recurrence relation. Specifically, if $P_{n-1}(x)$ and $xP_n(x)$ are even, then $P_{n+1}(x)$ must be even.

Upper and Lower Bounds for $P_n(\cos\theta)$

Finally, in addition to these results, our generating function enables us to set an upper limit on $|P_n(\cos\theta)|$. We have

$$\left(1-2t\cos\theta+t^2\right)^{-1/2} = \left(1-te^{i\theta}\right)^{-1/2}\left(1-te^{-i\theta}\right)^{-1/2}$$
$$= \left(1+\tfrac{1}{2}te^{i\theta}+\tfrac{3}{8}t^2e^{2i\theta}+\cdots\right)$$
$$\cdot\left(1+\tfrac{1}{2}te^{-i\theta}+\tfrac{3}{8}t^2e^{-2i\theta}+\cdots\right), \tag{12.38}$$

with all coefficients **positive**. Our Legendre polynomial, $P_n(\cos\theta)$, still the coefficient of t^n, may now be written as a sum of terms of the form

$$\tfrac{1}{2}a_m\left(e^{im\theta}+e^{-im\theta}\right) = a_m\cos m\theta \tag{12.39a}$$

with all the a_m **positive** and m and n both even or odd so that

$$P_n(\cos\theta) = \sum_{m=0 \text{ or } 1}^{n} a_m\cos m\theta. \tag{12.39b}$$

[7] In spherical polar coordinates the inversion of the point (r,θ,φ) through the origin is accomplished by the transformation $[r\to r, \theta\to\pi-\theta$, and $\varphi\to\varphi\pm\pi]$. Then, $\cos\theta\to\cos(\pi-\theta) = -\cos\theta$, corresponding to $x\to -x$ (compare Exercise 2.5.8).

This series, Eq. (12.39b), is clearly a maximum when $\theta = 0$ and $\cos m\theta = 1$. But for $x = \cos\theta = 1$, Eq. (12.31) shows that $P_n(1) = 1$. Therefore

$$|P_n(\cos\theta)| \le P_n(1) = 1. \tag{12.39c}$$

A fringe benefit of Eq. (12.39b) is that it shows that our Legendre polynomial is a linear combination of $\cos m\theta$. This means that the Legendre polynomials form a complete set for any functions that may be expanded by a Fourier cosine series (Section 14.1) over the interval $[0, \pi]$.

- In this section various useful properties of the Legendre polynomials are derived from the generating function, Eq. (12.4).
- The explicit series representation, Eq. (12.8), offers an alternate and sometimes superior approach.

Exercises

12.2.1 Given the series

$$\alpha_0 + \alpha_2\cos^2\theta + \alpha_4\cos^4\theta + \alpha_6\cos^6\theta = a_0P_0 + a_2P_2 + a_4P_4 + a_6P_6,$$

express the coefficients α_i as a column vector $\boldsymbol{\alpha}$ and the coefficients a_i as a column vector $\mathbf{a}$ and determine the matrices A and B such that

$$\mathsf{A}\boldsymbol{\alpha} = \mathbf{a} \qquad \text{and} \qquad \mathsf{B}\mathbf{a} = \boldsymbol{\alpha}.$$

Check your computation by showing that $\mathsf{AB} = \mathbf{1}$ (unit matrix). Repeat for the odd case

$$\alpha_1\cos\theta + \alpha_3\cos^3\theta + \alpha_5\cos^5\theta + \alpha_7\cos^7\theta = a_1P_1 + a_3P_3 + a_5P_5 + a_7P_7.$$

Note. $P_n(\cos\theta)$ and $\cos^n\theta$ are tabulated in terms of each other in AMS-55 (see Additional Readings of Chapter 8 for the complete reference).

12.2.2 By differentiating the generating function $g(t, x)$ with respect to t, multiplying by $2t$, and then adding $g(t, x)$, show that

$$\frac{1 - t^2}{(1 - 2tx + t^2)^{3/2}} = \sum_{n=0}^{\infty}(2n + 1)P_n(x)t^n.$$

This result is useful in calculating the charge induced on a grounded metal sphere by a point charge q.

12.2.3 (a) Derive Eq. (12.27),

$$\left(1 - x^2\right)P_n'(x) = (n + 1)xP_n(x) - (n + 1)P_{n+1}(x).$$

(b) Write out the relation of Eq. (12.27) to preceding equations in symbolic form analogous to the symbolic forms for Eqs. (12.23) to (12.26).

12.2.4 A point electric octupole may be constructed by placing a point electric quadrupole (pole strength $p^{(2)}$ in the z-direction) at $z = a$ and an equal but opposite point electric quadrupole at $z = 0$ and then letting $a \to 0$, subject to $p^{(2)}a =$ constant. Find the electrostatic potential corresponding to a point electric octupole. Show from the construction of the point electric octupole that the corresponding potential may be obtained by differentiating the point quadrupole potential.

12.2.5 Operating in **spherical polar coordinates**, show that

$$\frac{\partial}{\partial z}\left[\frac{P_n(\cos\theta)}{r^{n+1}}\right] = -(n+1)\frac{P_{n+1}(\cos\theta)}{r^{n+2}}.$$

This is the key step in the mathematical argument that the derivative of one multipole leads to the next higher multipole.
Hint. Compare Exercise 2.5.12.

12.2.6 From

$$P_L(\cos\theta) = \frac{1}{L!}\frac{\partial^L}{\partial t^L}\left(1 - 2t\cos\theta + t^2\right)^{-1/2}\Big|_{t=0}$$

show that

$$P_L(1) = 1, \qquad P_L(-1) = (-1)^L.$$

12.2.7 Prove that

$$P_n'(1) = \frac{d}{dx}P_n(x)\Big|_{x=1} = \frac{1}{2}n(n+1).$$

12.2.8 Show that $P_n(\cos\theta) = (-1)^n P_n(-\cos\theta)$ by use of the recurrence relation relating P_n, P_{n+1}, and P_{n-1} and your knowledge of P_0 and P_1.

12.2.9 From Eq. (12.38) write out the coefficient of t^2 in terms of $\cos n\theta$, $n \le 2$. This coefficient is $P_2(\cos\theta)$.

12.2.10 Write a program that will generate the coefficients a_s in the polynomial form of the Legendre polynomial

$$P_n(x) = \sum_{s=0}^{n} a_s x^s.$$

12.2.11 (a) Calculate $P_{10}(x)$ over the range $[0, 1]$ and plot your results.
(b) Calculate precise (at least to five decimal places) values of the five positive roots of $P_{10}(x)$. Compare your values with the values listed in AMS-55, Table 25.4. (For the complete reference, see Additional Readings of Chapter 8.)

12.2.12 (a) Calculate the **largest** root of $P_n(x)$ for $n = 2(1)50$.
(b) Develop an approximation for the largest root from the hypergeometric representation of $P_n(x)$ (Section 13.4) and compare your values from part (a) with your hypergeometric approximation. Compare also with the values listed in AMS-55, Table 25.4. (For the complete reference, see Additional Readings of Chapter 8.)

12.2.13 (a) From Exercise 12.2.1 and AMS-55, Table 22.9, develop the 6×6 matrix B that will transform a series of even-order Legendre polynomials through $P_{10}(x)$ into a power series $\sum_{n=0}^{5} \alpha_{2n} x^{2n}$.

(b) Calculate A as B^{-1}. Check the elements of A against the values listed in AMS-55, Table 22.9. (For the complete reference, see Additional Readings of Chapter 8.)

(c) By using matrix multiplication, transform some even power series $\sum_{n=0}^{5} \alpha_{2n} x^{2n}$ into a Legendre series.

12.2.14 Write a subroutine that will transform a finite power series $\sum_{n=0}^{N} a_n x^n$ into a Legendre series $\sum_{n=0}^{N} b_n P_n(x)$. Use the recurrence relation, Eq. (12.17), and follow the technique outlined in Section 13.3 for a Chebyshev series.

12.3 ORTHOGONALITY

Legendre's ODE (12.28) may be written in the form

$$\frac{d}{dx}\left[\left(1-x^2\right)P_n'(x)\right] + n(n+1)P_n(x) = 0, \tag{12.40}$$

showing clearly that it is self-adjoint. Subject to satisfying certain boundary conditions, then, it is known that the solutions $P_n(x)$ will be orthogonal. Upon comparing Eq. (12.40) with Eqs. (10.6) and (10.8) we see that the weight function $w(x) = 1$, $\mathcal{L} = (d/dx)(1-x^2)(d/dx)$, $p(x) = 1 - x^2$, and the eigenvalue $\lambda = n(n+1)$. The integration limits on x are ± 1, where $p(\pm 1) = 0$. Then for $m \neq n$, Eq. (10.34) becomes

$$\int_{-1}^{1} P_n(x) P_m(x)\, dx = 0,^{8} \tag{12.41}$$

$$\int_{0}^{\pi} P_n(\cos\theta) P_m(\cos\theta) \sin\theta\, d\theta = 0, \tag{12.42}$$

showing that $P_n(x)$ and $P_m(x)$ are orthogonal for the interval $[-1, 1]$. This orthogonality may also be demonstrated by using Rodrigues' definition of $P_n(x)$ (compare Section 12.4, Exercise 12.4.2).

We shall need to evaluate the integral (Eq. (12.41)) when $n = m$. Certainly it is no longer zero. From our generating function,

$$\left(1 - 2tx + t^2\right)^{-1} = \left[\sum_{n=0}^{\infty} P_n(x) t^n\right]^2. \tag{12.43}$$

Integrating from $x = -1$ to $x = +1$, we have

$$\int_{-1}^{1} \frac{dx}{1 - 2tx + t^2} = \sum_{n=0}^{\infty} t^{2n} \int_{-1}^{1} \left[P_n(x)\right]^2 dx; \tag{12.44}$$

[8] In Section 10.4 such integrals are interpreted as inner products in a linear vector (function) space. Alternate notations are

$$\int_{-1}^{1} \left[P_n(x)\right]^* P_m(x)\, dx \equiv \langle P_n | P_m \rangle \equiv (P_n, P_m).$$

The $\langle\,\rangle$ form, popularized by Dirac, is common in the physics literature. The $(\,)$ form is more common in the mathematics literature.

the cross terms in the series vanish by means of Eq. (12.42). Using $y = 1 - 2tx + t^2$, $dy = -2t\,dx$, we obtain

$$\int_{-1}^{1} \frac{dx}{1-2tx+t^2} = \frac{1}{2t}\int_{(1-t)^2}^{(1+t)^2} \frac{dy}{y} = \frac{1}{t}\ln\left(\frac{1+t}{1-t}\right). \tag{12.45}$$

Expanding this in a power series (Exercise 5.4.1) gives us

$$\frac{1}{t}\ln\left(\frac{1+t}{1-t}\right) = 2\sum_{n=0}^{\infty}\frac{t^{2n}}{2n+1}. \tag{12.46}$$

Comparing power-series coefficients of Eqs. (12.44) and (12.46), we must have

$$\int_{-1}^{1}\left[P_n(x)\right]^2 dx = \frac{2}{2n+1}. \tag{12.47}$$

Combining Eq. (12.42) with Eq. (12.47) we have the orthonormality condition

$$\int_{-1}^{1} P_m(x)\,P_n(x)\,dx = \frac{2\delta_{mn}}{2n+1}. \tag{12.48}$$

We shall return to this result in Section 12.6 when we construct the orthonormal spherical harmonics.

Expansion of Functions, Legendre Series

In addition to orthogonality, the Sturm–Liouville theory implies that the Legendre polynomials form a complete set. Let us assume, then, that the series

$$\sum_{n=0}^{\infty} a_n P_n(x) = f(x) \tag{12.49}$$

converges in the mean (Section 10.4) in the interval $[-1, 1]$. This demands that $f(x)$ and $f'(x)$ be at least sectionally continuous in this interval. The coefficients a_n are found by multiplying the series by $P_m(x)$ and integrating term by term. Using the orthogonality property expressed in Eqs. (12.42) and (12.48), we obtain

$$\frac{2}{2m+1}a_m = \int_{-1}^{1} f(x)\,P_m(x)\,dx. \tag{12.50}$$

We replace the variable of integration x by t and the index m by n. Then, substituting into Eq. (12.49), we have

$$f(x) = \sum_{n=0}^{\infty}\frac{2n+1}{2}\left(\int_{-1}^{1} f(t)\,P_n(t)\,dt\right)P_n(x). \tag{12.51}$$

This expansion in a series of Legendre polynomials is usually referred to as a Legendre series.[9] Its properties are quite similar to the more familiar Fourier series (Chapter 14). In

[9]Note that Eq. (12.50) gives a_m as a **definite** integral, that is, a number for a given $f(x)$.

particular, we can use the orthogonality property (Eq. (12.48)) to show that the series is unique.

On a more abstract (and more powerful) level, Eq. (12.51) gives the representation of $f(x)$ in the vector space of Legendre polynomials (a Hilbert space, Section 10.4).

From the viewpoint of integral transforms (Chapter 15), Eq. (12.50) may be considered a finite Legendre transform of $f(x)$. Equation (12.51) is then the inverse transform. It may also be interpreted in terms of the **projection operators** of quantum theory. We may take $\mathcal{P}_m$ in

$$[\mathcal{P}_m f](x) \equiv P_m(x)\frac{2m+1}{2}\int_{-1}^{1} P_m(t)\big[f(t)\big]\,dt$$

as an (integral) operator, ready to operate on $f(t)$. (The $f(t)$ would go in the square bracket as a factor in the integrand.) Then, from Eq. (12.50),

$$[\mathcal{P}_m f](x) = a_m P_m(x).^{10}$$

The operator $\mathcal{P}_m$ projects out the mth component of the function f.

Equation (12.3), which leads directly to the generating function definition of Legendre polynomials, is a Legendre expansion of $1/r_1$. This Legendre expansion of $1/r_1$ or $1/r_{12}$ appears in several exercises of Section 12.8. Going beyond a Coulomb field, the $1/r_{12}$ is often replaced by a potential $V(|\mathbf{r}_1 - \mathbf{r}_2|)$, and the solution of the problem is again effected by a Legendre expansion.

The Legendre series, Eq. (12.49), has been treated as a **known** function $f(x)$ that we arbitrarily chose to expand in a series of Legendre polynomials. Sometimes the origin and nature of the Legendre series are different. In the next examples we consider **unknown** functions we know can be represented by a Legendre series because of the differential equation the unknown functions satisfy. As before, the problem is to determine the unknown coefficients in the series expansion. Here, however, the coefficients are not found by Eq. (12.50). Rather, they are determined by demanding that the Legendre series match a known solution at a boundary. These are boundary value problems.

Example 12.3.1 EARTH'S GRAVITATIONAL FIELD

An example of a Legendre series is provided by the description of the Earth's gravitational potential U (for exterior points), neglecting azimuthal effects. With

$$R = \text{equatorial radius} = 6378.1 \pm 0.1 \text{ km}$$
$$\frac{GM}{R} = 62.494 \pm 0.001 \text{ km}^2/\text{s}^2,$$

we write

$$U(r,\theta) = \frac{GM}{R}\left[\frac{R}{r} - \sum_{n=2}^{\infty} a_n \left(\frac{R}{r}\right)^{n+1} P_n(\cos\theta)\right], \tag{12.52}$$

[10] The dependent variables are arbitrary. Here x came from the x in $\mathcal{P}_m$.

a Legendre series. Artificial satellite motions have shown that

$$a_2 = (1{,}082{,}635 \pm 11) \times 10^{-9},$$
$$a_3 = (-2{,}531 \pm 7) \times 10^{-9},$$
$$a_4 = (-1{,}600 \pm 12) \times 10^{-9}.$$

This is the famous pear-shaped deformation of the Earth. Other coefficients have been computed through $n = 20$. Note that P_1 is omitted because the origin from which r is measured is the Earth's center of mass (P_1 would represent a displacement).

More recent satellite data permit a determination of the longitudinal dependence of the Earth's gravitational field. Such dependence may be described by a Laplace series (Section 12.6). ■

Example 12.3.2 SPHERE IN A UNIFORM FIELD

Another illustration of the use of Legendre polynomials is provided by the problem of a neutral conducting sphere (radius r_0) placed in a (previously) uniform electric field (Fig. 12.8). The problem is to find the new, perturbed, electrostatic potential. If we call the electrostatic potential[11] V, it satisfies

$$\nabla^2 V = 0, \tag{12.53}$$

Laplace's equation. We select spherical polar coordinates because of the spherical shape of the conductor. (This will simplify the application of the boundary condition at the surface of the conductor.) Separating variables and glancing at Table 9.2, we can write the unknown potential $V(r, \theta)$ in the region outside the sphere as a linear combination of solutions:

$$V(r, \theta) = \sum_{n=0}^{\infty} a_n r^n P_n(\cos\theta) + \sum_{n=0}^{\infty} b_n \frac{P_n(\cos\theta)}{r^{n+1}}. \tag{12.54}$$

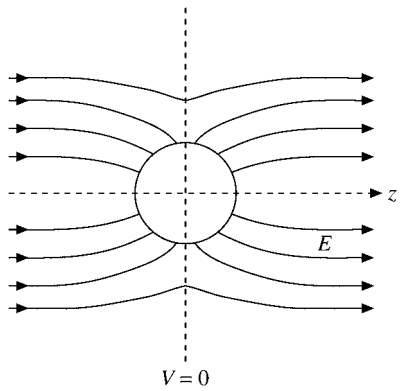

FIGURE 12.8 Conducting sphere in a uniform field.

[11] It should be emphasized that this is not a presentation of a Legendre-series expansion of a known $V(\cos\theta)$. Here we are back to **boundary value** problems of PDEs.

No φ-dependence appears because of the axial symmetry of our problem. (The center of the conducting sphere is taken as the origin and the z-axis is oriented parallel to the original uniform field.)

It might be noted here that n is an integer, because only for integral n is the θ dependence well behaved at $\cos\theta = \pm 1$. For nonintegral n the solutions of Legendre's equation diverge at the ends of the interval $[-1, 1]$, the poles $\theta = 0, \pi$ of the sphere (compare Example 5.2.4 and Exercises 5.2.15 and 9.5.5). It is for this same reason that the second solution of Legendre's equation, Q_n, is also excluded.

Now we turn to our (Dirichlet) boundary conditions to determine the unknown a_n and b_n of our series solution, Eq. (12.54). If the original, unperturbed electrostatic field is E_0, we require, as one boundary condition,

$$V(r \to \infty) = -E_0 z = -E_0 r\cos\theta = -E_0 r P_1(\cos\theta). \tag{12.55}$$

Since our Legendre series is unique, we may equate coefficients of $P_n(\cos\theta)$ in Eq. (12.54) $(r \to \infty)$ and Eq. (12.55) to obtain

$$a_n = 0, \qquad n > 1 \qquad \text{and} \qquad n = 0, \qquad a_1 = -E_0. \tag{12.56}$$

If $a_n \neq 0$ for $n > 1$, these terms would dominate at large r and the boundary condition (Eq. (12.55)) could not be satisfied.

As a second boundary condition, we may choose the conducting sphere and the plane $\theta = \pi/2$ to be at zero potential, which means that Eq. (12.54) now becomes

$$V(r = r_0) = \frac{b_0}{r_0} + \left(\frac{b_1}{r_0^2} - E_0 r_0\right) P_1(\cos\theta) + \sum_{n=2}^{\infty} b_n \frac{P_n(\cos\theta)}{r_0^{n+1}} = 0. \tag{12.57}$$

In order that this may hold for all values of θ, each coefficient of $P_n(\cos\theta)$ must vanish.[12] Hence

$$b_0 = 0,^{13} \qquad b_n = 0, \qquad n \geq 2, \tag{12.58}$$

whereas

$$b_1 = E_0 r_0^3. \tag{12.59}$$

The electrostatic potential (outside the sphere) is then

$$V = -E_0 r P_1(\cos\theta) + \frac{E_0 r_0^3}{r^2} P_1(\cos\theta) = -E_0 r P_1(\cos\theta)\left(1 - \frac{r_0^3}{r^3}\right). \tag{12.60}$$

In Section 1.16 it was shown that a solution of Laplace's equation that satisfied the boundary conditions over the entire boundary was unique. The electrostatic potential V, as given by Eq. (12.60), is a solution of Laplace's equation. It satisfies our boundary conditions and therefore is the solution of Laplace's equation for this problem.

[12] Again, this is equivalent to saying that a series expansion in Legendre polynomials (or any complete orthogonal set) is unique.

[13] The coefficient of P_0 is b_0/r_0. We set $b_0 = 0$ because there is no net charge on the sphere. If there is a net charge q, then $b_0 \neq 0$.

It may further be shown (Exercise 12.3.13) that there is an induced surface charge density

$$\sigma = -\varepsilon_0 \frac{\partial V}{\partial r}\bigg|_{r=r_0} = 3\varepsilon_0 E_0 \cos\theta \tag{12.61}$$

on the surface of the sphere and an induced electric dipole moment (Exercise 12.3.13)

$$P = 4\pi r_0^3 \varepsilon_0 E_0. \tag{12.62}$$

■

Example 12.3.3 ELECTROSTATIC POTENTIAL OF A RING OF CHARGE

As a further example, consider the electrostatic potential produced by a conducting ring carrying a total electric charge q (Fig. 12.9). From electrostatics (and Section 1.14) the potential ψ satisfies Laplace's equation. Separating variables in spherical polar coordinates (compare Table 9.2), we obtain

$$\psi(r,\theta) = \sum_{n=0}^{\infty} c_n \frac{a^n}{r^{n+1}} P_n(\cos\theta), \qquad r > a. \tag{12.63a}$$

Here a is the radius of the ring that is assumed to be in the $\theta = \pi/2$ plane. There is no φ (azimuthal) dependence because of the cylindrical symmetry of the system. The terms with positive exponent in the radial dependence have been rejected because the potential must have an asymptotic behavior,

$$\psi \sim \frac{q}{4\pi\varepsilon_0} \cdot \frac{1}{r}, \qquad r \gg a. \tag{12.63b}$$

The problem is to determine the coefficients c_n in Eq. (12.63a). This may be done by evaluating $\psi(r,\theta)$ at $\theta = 0, r = z$, and comparing with an independent calculation of the

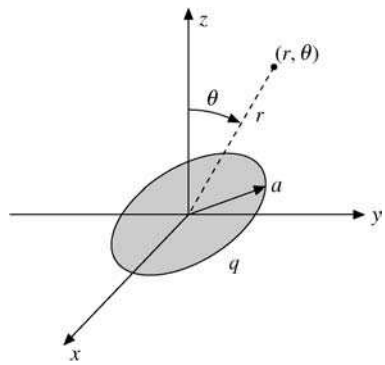

FIGURE 12.9 Charged, conducting ring.

potential from Coulomb's law. In effect, we are using a boundary condition along the z-axis. From Coulomb's law (with all charge equidistant),

$$\psi(r,\theta) = \frac{q}{4\pi\varepsilon_0}\cdot\frac{1}{(z^2+a^2)^{1/2}}, \qquad \begin{cases}\theta = 0\\ r = z,\end{cases}$$

$$= \frac{q}{4\pi\varepsilon_0 z}\sum_{s=0}^{\infty}(-1)^s\frac{(2s)!}{2^{2s}(s!)^2}\left(\frac{a}{z}\right)^{2s}, \qquad z > a. \tag{12.63c}$$

The last step uses the result of Exercise 8.1.15. Now, Eq. (12.63a) evaluated at $\theta = 0, r = z$ (with $P_n(1) = 1$), yields

$$\psi(r,\theta) = \sum_{n=0}^{\infty} c_n\frac{a^n}{z^{n+1}}, \qquad r = z. \tag{12.63d}$$

Comparing Eqs. (12.63c) and (12.63d), we get $c_n = 0$ for n odd. Setting $n = 2s$, we have

$$c_{2s} = \frac{q}{4\pi\varepsilon_0}(-1)^s\frac{(2s)!}{2^{2s}(s!)^2}, \tag{12.63e}$$

and our electrostatic potential $\psi(r,\theta)$ is given by

$$\psi(r,\theta) = \frac{q}{4\pi\varepsilon_0 r}\sum_{s=0}^{\infty}(-1)^s\frac{(2s)!}{2^{2s}(s!)^2}\left(\frac{a}{r}\right)^{2s}P_{2s}(\cos\theta), \qquad r > a. \tag{12.63f}$$

The magnetic analog of this problem appears in Example 12.5.3. ■

Exercises

12.3.1 You have constructed a set of orthogonal functions by the Gram–Schmidt process (Section 10.3), taking $u_n(x) = x^n$, $n = 0, 1, 2, \ldots$, in increasing order with $w(x) = 1$ and an interval $-1 \le x \le 1$. Prove that the nth such function constructed is proportional to $P_n(x)$.
Hint. Use mathematical induction.

12.3.2 Expand the 1-D Dirac delta function in a series of Legendre polynomials using the interval $-1 \le x \le 1$.

12.3.3 Verify the Dirac delta function expansions

$$\delta(1-x) = \sum_{n=0}^{\infty}\frac{2n+1}{2}P_n(x)$$

$$\delta(1+x) = \sum_{n=0}^{\infty}(-1)^n\frac{2n+1}{2}P_n(x).$$

These expressions appear in a resolution of the Rayleigh plane-wave expansion (Exercise 12.4.7) into incoming and outgoing spherical waves.
Note. Assume that the **entire** Dirac delta function is covered when integrating over $[-1, 1]$.

12.3.4 Neutrons (mass 1) are being scattered by a nucleus of mass A ($A > 1$). In the center-of-mass system the scattering is isotropic. Then, in the laboratory system the average of the cosine of the angle of deflection of the neutron is

$$\langle \cos\psi \rangle = \frac{1}{2}\int_0^\pi \frac{A\cos\theta + 1}{(A^2 + 2A\cos\theta + 1)^{1/2}} \sin\theta\, d\theta.$$

Show, by expansion of the denominator, that $\langle \cos\psi \rangle = 2/3A$.

12.3.5 A particular function $f(x)$ defined over the interval $[-1, 1]$ is expanded in a Legendre series over this same interval. Show that the expansion is unique.

12.3.6 A function $f(x)$ is expanded in a Legendre series $f(x) = \sum_{n=0}^{\infty} a_n P_n(x)$. Show that

$$\int_{-1}^{1} \big[f(x)\big]^2 dx = \sum_{n=0}^{\infty} \frac{2a_n^2}{2n+1}.$$

This is the Legendre form of the Fourier series Parseval identity, Exercise 14.4.2. It also illustrates Bessel's inequality, Eq. (10.72), becoming an equality for a complete set.

12.3.7 Derive the recurrence relation

$$\big(1 - x^2\big) P_n'(x) = n P_{n-1}(x) - n x P_n(x)$$

from the Legendre polynomial generating function.

12.3.8 Evaluate $\int_0^1 P_n(x)\, dx$.

$$\begin{aligned} \textit{ANS. } n = 2s; &\qquad 1 \text{ for } s = 0,\ 0 \text{ for } s > 0, \\ n = 2s+1; &\qquad P_{2s}(0)/(2s+2) = (-1)^s (2s-1)!!/1(2s+2)!! \end{aligned}$$

Hint. Use a recurrence relation to replace $P_n(x)$ by derivatives and then integrate by inspection. Alternatively, you can integrate the generating function.

12.3.9 (a) For

$$f(x) = \begin{cases} +1, & 0 < x < 1 \\ -1, & -1 < x < 0, \end{cases}$$

show that

$$\int_{-1}^{1} \big[f(x)\big]^2 dx = 2\sum_{n=0}^{\infty} (4n+3)\left[\frac{(2n-1)!!}{(2n+2)!!}\right]^2.$$

(b) By testing the series, prove that the series is convergent.

12.3.10 Prove that

$$\begin{aligned} \int_{-1}^{1} x\big(1 - x^2\big) P_n' P_m'\, dx &= 0, \qquad \text{unless } m = n \pm 1, \\ &= \frac{2n(n^2-1)}{4n^2-1}\delta_{m,n-1}, \qquad \text{if } m < n. \\ &= \frac{2n(n+2)(n+1)}{(2n+1)(2n+3)}\delta_{m,n+1}, \qquad \text{if } m > n. \end{aligned}$$

12.3.11 The amplitude of a scattered wave is given by

$$f(\theta) = \frac{1}{k}\sum_{l=0}^{\infty}(2l+1)\exp[i\delta_l]\sin\delta_l P_l(\cos\theta).$$

Here θ is the angle of scattering, l is the angular momentum eigenvalue, $\hbar k$ is the incident momentum, and δ_l is the phase shift produced by the central potential that is doing the scattering. The total cross section is $\sigma_{\text{tot}} = \int |f(\theta)|^2\, d\Omega$. Show that

$$\sigma_{\text{tot}} = \frac{4\pi}{k^2}\sum_{l=0}^{\infty}(2l+1)\sin^2\delta_l.$$

12.3.12 The coincidence counting rate, $W(\theta)$, in a gamma–gamma angular correlation experiment has the form

$$W(\theta) = \sum_{n=0}^{\infty} a_{2n} P_{2n}(\cos\theta).$$

Show that data in the range $\pi/2 \le \theta \le \pi$ can, in principle, define the function $W(\theta)$ (and permit a determination of the coefficients a_{2n}). This means that although data in the range $0 \le \theta < \pi/2$ may be useful as a check, they are not essential.

12.3.13 A conducting sphere of radius r_0 is placed in an initially uniform electric field, $\mathbf{E}_0$. Show the following:

(a) The induced surface charge density is

$$\sigma = 3\varepsilon_0 E_0 \cos\theta.$$

(b) The induced electric dipole moment is

$$P = 4\pi r_0^3 \varepsilon_0 E_0.$$

The induced electric dipole moment can be calculated either from the surface charge [part (a)] or by noting that the final electric field $\mathbf{E}$ is the result of superimposing a dipole field on the original uniform field.

12.3.14 A charge q is displaced a distance a along the z-axis from the center of a spherical cavity of radius R.

(a) Show that the electric field averaged over the volume $a \le r \le R$ is zero.
(b) Show that the electric field averaged over the volume $0 \le r \le a$ is

$$\mathbf{E} = \hat{\mathbf{z}}E_z = -\hat{\mathbf{z}}\frac{q}{4\pi\varepsilon_0 a^2} \qquad (\text{SI units}) = -\hat{\mathbf{z}}\frac{nqa}{3\varepsilon_0},$$

where n is the number of such displaced charges per unit volume. This is a basic calculation in the polarization of a dielectric.
Hint. $\mathbf{E} = -\nabla\varphi$.

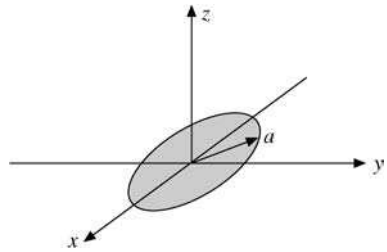

FIGURE 12.10 Charged, conducting disk.

12.3.15 Determine the electrostatic potential (Legendre expansion) of a circular ring of electric charge for $r < a$.

12.3.16 Calculate the electric field produced by the charged conducting ring of Example 12.3.3 for

(a) $r > a$, (b) $r < a$.

12.3.17 As an extension of Example 12.3.3, find the potential $\psi(r, \theta)$ produced by a charged conducting disk, Fig. 12.10, for $r > a$, the radius of the disk. The charge density σ (on each side of the disk) is

$$\sigma(\rho) = \frac{q}{4\pi a(a^2 - \rho^2)^{1/2}}, \qquad \rho^2 = x^2 + y^2.$$

Hint. The definite integral you get can be evaluated as a beta function, Section 8.4. For more details see Section 5.03 of Smythe in Additional Readings.

$$\text{ANS. } \psi(r, \theta) = \frac{q}{4\pi\varepsilon_0 r} \sum_{l=0}^{\infty} (-1)^l \frac{1}{2l+1} \left(\frac{a}{r}\right)^{2l} P_{2l}(\cos\theta).$$

12.3.18 From the result of Exercise 12.3.17 calculate the potential of the disk. Since you are violating the condition $r > a$, justify your calculation.
Hint. You may run into the series given in Exercise 5.2.9.

12.3.19 The hemisphere defined by $r = a, 0 \le \theta < \pi/2$, has an electrostatic potential $+V_0$. The hemisphere $r = a, \pi/2 < \theta \le \pi$ has an electrostatic potential $-V_0$. Show that the potential at interior points is

$$V = V_0 \sum_{n=0}^{\infty} \frac{4n+3}{2n+2} \left(\frac{r}{a}\right)^{2n+1} P_{2n}(0) P_{2n+1}(\cos\theta)$$

$$= V_0 \sum_{n=0}^{\infty} (-1)^n \frac{(4n+3)(2n-1)!!}{(2n+2)!!} \left(\frac{r}{a}\right)^{2n+1} P_{2n+1}(\cos\theta).$$

Hint. You need Exercise 12.3.8.

12.3.20 A conducting sphere of radius a is divided into two electrically separate hemispheres by a thin insulating barrier at its equator. The top hemisphere is maintained at a potential V_0, the bottom hemisphere at $-V_0$.

(a) Show that the electrostatic potential **exterior** to the two hemispheres is

$$V(r,\theta) = V_0 \sum_{s=0}^{\infty} (-1)^s (4s+3) \frac{(2s-1)!!}{(2s+2)!!} \left(\frac{a}{r}\right)^{2s+2} P_{2s+1}(\cos\theta).$$

(b) Calculate the electric charge density σ on the outside surface. Note that your series diverges at $\cos\theta = \pm 1$, as you expect from the infinite capacitance of this system (zero thickness for the insulating barrier).

$$\text{ANS. } \sigma = \varepsilon_0 E_n = -\varepsilon_0 \frac{\partial V}{\partial r}\bigg|_{r=a}$$

$$= \varepsilon_0 V_0 \sum_{s=0}^{\infty} (-1)^s (4s+3) \frac{(2s-1)!!}{(2s)!!} P_{2s+1}(\cos\theta).$$

12.3.21 In the notation of Section 10.4, $\varphi_s(x) = \sqrt{(2s+1)/2}\,P_s(x)$, a Legendre polynomial is renormalized to unity. Explain how $|\varphi_s\rangle\langle\varphi_s|$ acts as a projection operator. In particular, show that if $|f\rangle = \sum_n a'_n |\varphi_n\rangle$, then

$$|\varphi_s\rangle\langle\varphi_s|f\rangle = a'_s|\varphi_s\rangle.$$

12.3.22 Expand x^8 as a Legendre series. Determine the Legendre coefficients from Eq. (12.50),

$$a_m = \frac{2m+1}{2} \int_{-1}^{1} x^8 P_m(x)\,dx.$$

Check your values against AMS-55, Table 22.9. (For the complete reference, see Additional Readings in Chapter 8). This illustrates the expansion of a simple function $f(x)$. Actually if $f(x)$ is expressed as a power series, the technique of Exercise 12.2.14 is both faster and more accurate.
Hint. Gaussian quadrature can be used to evaluate the integral.

12.3.23 Calculate and tabulate the electrostatic potential created by a ring of charge, Example 12.3.3, for $r/a = 1.5(0.5)5.0$ and $\theta = 0°(15°)90°$. Carry terms through $P_{22}(\cos\theta)$. *Note*. The convergence of your series will be slow for $r/a = 1.5$. Truncating the series at P_{22} limits you to about four-significant-figure accuracy.

Check value. For $r/a = 2.5$ and $\theta = 60°$, $\psi = 0.40272(q/4\pi\varepsilon_0 r)$.

12.3.24 Calculate and tabulate the electrostatic potential created by a charged disk, Exercise 12.3.17, for $r/a = 1.5(0.5)5.0$ and $\theta = 0°(15°)90°$. Carry terms through $P_{22}(\cos\theta)$.

Check value. For $r/a = 2.0$ and $\theta = 15°$, $\psi = 0.46638(q/4\pi\varepsilon_0 r)$.

12.3.25 Calculate the first five (nonvanishing) coefficients in the Legendre series expansion of $f(x) = 1 - |x|$ using Eq. (12.51)—numerical integration. Actually these coefficients can be obtained in closed form. Compare your coefficients with those obtained from Exercise 13.3.28.

ANS. $a_0 = 0.5000$, $a_2 = -0.6250$, $a_4 = 0.1875$, $a_6 = -0.1016$, $a_8 = 0.0664$.

12.3.26 Calculate and tabulate the exterior electrostatic potential created by the two charged hemispheres of Exercise 12.3.20, for $r/a = 1.5(0.5)5.0$ and $\theta = 0°(15°)90°$. Carry terms through $P_{23}(\cos\theta)$.

Check value. For $r/a = 2.0$ and $\theta = 45°$, $V = 0.27066V_0$.

12.3.27 (a) Given $f(x) = 2.0$, $|x| < 0.5$; $f(x) = 0$, $0.5 < |x| < 1.0$, expand $f(x)$ in a Legendre series and calculate the coefficients a_n through a_{80} (analytically).

(b) Evaluate $\sum_{n=0}^{80} a_n P_n(x)$ for $x = 0.400(0.005)0.600$. Plot your results.

Note. This illustrates the Gibbs phenomenon of Section 14.5 and the danger of trying to calculate with a series expansion in the vicinity of a discontinuity.

12.4 ALTERNATE DEFINITIONS OF LEGENDRE POLYNOMIALS

Rodrigues' Formula

The series form of the Legendre polynomials (Eq. (12.8)) of Section 12.1 may be transformed as follows. From Eq. (12.8),

$$P_n(x) = \sum_{r=0}^{[n/2]} (-1)^r \frac{(2n-2r)!}{2^n r!(n-2r)!(n-r)!} x^{n-2r}. \tag{12.64}$$

For n an integer,

$$\begin{aligned} P_n(x) &= \sum_{r=0}^{[n/2]} (-1)^r \frac{1}{2^n r!(n-r)!} \left(\frac{d}{dx}\right)^n x^{2n-2r} \\ &= \frac{1}{2^n n!} \left(\frac{d}{dx}\right)^n \sum_{r=0}^{n} \frac{(-1)^r n!}{r!(n-r)!} x^{2n-2r}. \end{aligned} \tag{12.64a}$$

Note the extension of the upper limit. The reader is asked to show in Exercise 12.4.1 that the additional terms $[n/2]+1$ to n in the summation contribute nothing. However, the effect of these extra terms is to permit the replacement of the new summation by $(x^2-1)^n$ (binomial theorem once again) to obtain

$$P_n(x) = \frac{1}{2^n n!} \left(\frac{d}{dx}\right)^n (x^2-1)^n. \tag{12.65}$$

This is Rodrigues' formula. It is useful in proving many of the properties of the Legendre polynomials, such as orthogonality. A related application is seen in Exercise 12.4.3. The Rodrigues definition is extended in Section 12.5 to define the associated Legendre functions. In Section 12.7 it is used to identify the orbital angular momentum eigenfunctions.

Schlaefli Integral

Rodrigues' formula provides a means of developing an integral representation of $P_n(z)$. Using Cauchy's integral formula (Section 6.4)

$$f(z) = \frac{1}{2\pi i}\oint \frac{f(t)}{t-z}dt \tag{12.66}$$

with

$$f(z) = \left(z^2-1\right)^n, \tag{12.67}$$

we have

$$\left(z^2-1\right)^n = \frac{1}{2\pi i}\oint \frac{(t^2-1)^n}{t-z}dt. \tag{12.68}$$

Differentiating n times with respect to z and multiplying by $1/2^n n!$ gives

$$P_n(z) = \frac{1}{2^n n!}\frac{d^n}{dz^n}\left(z^2-1\right)^n = \frac{2^{-n}}{2\pi i}\oint \frac{(t^2-1)^n}{(t-z)^{n+1}}dt, \tag{12.69}$$

with the contour enclosing the point $t = z$.

This is the Schlaefli integral. Margenau and Murphy[14] use this to derive the recurrence relations we obtained from the generating function.

The Schlaefli integral may readily be shown to satisfy Legendre's equation by differentiation and direct substitution (Fig. 12.11). We obtain

$$\left(1-z^2\right)\frac{d^2P_n}{dz^2} - 2z\frac{dP_n}{dz} + n(n+1)P_n = \frac{n+1}{2^n 2\pi i}\oint \frac{d}{dt}\left[\frac{(t^2-1)^{n+1}}{(t-z)^{n+2}}\right]dt. \tag{12.70}$$

For integral n our function $(t^2-1)^{n+1}/(t-z)^{n+2}$ is single-valued, and the integral around the closed path vanishes. The Schlaefli integral may also be used to define $P_\nu(z)$ for non-integral ν integrating around the points $t = z$, $t = 1$, but not crossing the cut line -1 to $-\infty$. We could equally well encircle the points $t = z$ and $t = -1$, but this would lead to

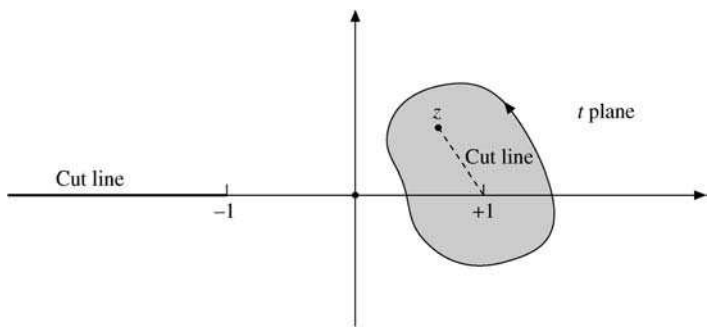

FIGURE 12.11 Schlaefli integral contour.

[14]H. Margenau and G. M. Murphy, *The Mathematics of Physics and Chemistry*, 2nd ed., Princeton, NJ: Van Nostrand (1956), Section 3.5.

nothing new. A contour about $t = +1$ and $t = -1$ will lead to a second solution, $Q_\nu(z)$, Section 12.10.

Exercises

12.4.1 Show that **each** term in the summation

$$\sum_{r=[n/2]+1}^{n} \left(\frac{d}{dx}\right)^n \frac{(-1)^r n!}{r!(n-r)!} x^{2n-2r}$$

vanishes (r and n integral).

12.4.2 Using Rodrigues' formula, show that the $P_n(x)$ are orthogonal and that

$$\int_{-1}^{1} \big[P_n(x)\big]^2 dx = \frac{2}{2n+1}.$$

Hint. Use Rodrigues' formula and integrate by parts.

12.4.3 Show that $\int_{-1}^{1} x^m P_n(x) dx = 0$ when $m < n$.
Hint. Use Rodrigues' formula or expand x^m in Legendre polynomials.

12.4.4 Show that

$$\int_{-1}^{1} x^n P_n(x)\, dx = \frac{2^{n+1} n! n!}{(2n+1)!}.$$

Note. You are expected to use Rodrigues' formula and integrate by parts, but also see if you can get the result from Eq. (12.8) by inspection.

12.4.5 Show that

$$\int_{-1}^{1} x^{2r} P_{2n}(x)\, dx = \frac{2^{2n+1}(2r)!(r+n!)}{(2r+2n+1)!(r-n)!}, \qquad r \geq n.$$

12.4.6 As a generalization of Exercises 12.4.4 and 12.4.5, show that the Legendre expansions of x^s are

(a) $$x^{2r} = \sum_{n=0}^{r} \frac{2^{2n}(4n+1)(2r)!(r+n)!}{(2r+2n+1)!(r-n)!} P_{2n}(x), \qquad s = 2r,$$

(b) $$x^{2r+1} = \sum_{n=0}^{r} \frac{2^{2n+1}(4n+3)(2r+1)!(r+n+1)!}{(2r+2n+3)!(r-n)!} P_{2n+1}(x), \qquad s = 2r+1.$$

12.4.7 A plane wave may be expanded in a series of spherical waves by the Rayleigh equation,

$$e^{ikr\cos\gamma} = \sum_{n=0}^{\infty} a_n j_n(kr) P_n(\cos\gamma).$$

Show that $a_n = i^n(2n+1)$.

Hint.

1. Use the orthogonality of the P_n to solve for $a_n j_n(kr)$.
2. Differentiate n times with respect to (kr) and set $r = 0$ to eliminate the r-dependence.
3. Evaluate the remaining integral by Exercise 12.4.4.

Note. This problem may also be treated by noting that both sides of the equation satisfy the Heemholtz equation. The equality can be established by showing that the solutions have the same behavior at the origin and also behave alike at large distances. A "by inspection" type of solution is developed in Section 9.7 using Green's functions.

12.4.8 Verify the Rayleigh equation of Exercise 12.4.7 by starting with the following steps:

(a) Differentiate with respect to (kr) to establish

$$\sum_n a_n j_n'(kr) P_n(\cos\gamma) = i \sum_n a_n j_n(kr) \cos\gamma\, P_n(\cos\gamma).$$

(b) Use a recurrence relation to replace $\cos\gamma\, P_n(\cos\gamma)$ by a linear combination of P_{n-1} and P_{n+1}.

(c) Use a recurrence relation to replace j_n' by a linear combination of j_{n-1} and j_{n+1}.

12.4.9 From Exercise 12.4.7 show that

$$j_n(kr) = \frac{1}{2i^n} \int_{-1}^{1} e^{ikr\mu} P_n(\mu)\, d\mu.$$

This means that (apart from a constant factor) the spherical Bessel function $j_n(kr)$ is the Fourier transform of the Legendre polynomial $P_n(\mu)$.

12.4.10 The Legendre polynomials and the spherical Bessel functions are related by

$$j_n(z) = \frac{1}{2}(-i)^n \int_0^{\pi} e^{iz\cos\theta} P_n(\cos\theta) \sin\theta\, d\theta, \qquad n = 0, 1, 2, \ldots.$$

Verify this relation by transforming the right-hand side into

$$\frac{z^n}{2^{n+1} n!} \int_0^{\pi} \cos(z\cos\theta) \sin^{2n+1}\theta\, d\theta$$

and using Exercise 11.7.8.

12.4.11 By direct evaluation of the Schlaefli integral show that $P_n(1) = 1$.

12.4.12 Explain why the contour of the Schlaefli integral, Eq. (12.69), is chosen to enclose the points $t = z$ and $t = 1$ when $n \to \nu$, not an integer.

12.4.13 In numerical work (for example, the Gauss–Legendre quadrature) it is useful to establish that $P_n(x)$ has n real zeros in the interior of $[-1, 1]$. Show that this is so.
Hint. Rolle's theorem shows that the first derivative of $(x^2 - 1)^{2n}$ has one zero in the interior of $[-1, 1]$. Extend this argument to the second, third, and ultimately the nth derivative.

12.5 ASSOCIATED LEGENDRE FUNCTIONS

When Helmholtz's equation is separated in spherical polar coordinates (Section 9.3), one of the separated ODEs is the associated Legendre equation

$$\boxed{\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{dP_n^m(\cos\theta)}{d\theta}\right)+\left[n(n+1)-\frac{m^2}{\sin^2\theta}\right]P_n^m(\cos\theta)=0.} \tag{12.71}$$

With $x=\cos\theta$, this becomes

$$\boxed{\left(1-x^2\right)\frac{d^2}{dx^2}P_n^m(x)-2x\frac{d}{dx}P_n^m(x)+\left[n(n+1)-\frac{m^2}{1-x^2}\right]P_n^m(x)=0.} \tag{12.72}$$

If the azimuthal separation constant $m^2=0$, we have Legendre's equation, Eq. (12.28). The regular solutions $P_n^m(x)$ (with m not necessarily zero, but an integer) are

$$v\equiv P_n^m(x)=\left(1-x^2\right)^{m/2}\frac{d^m}{dx^m}P_n(x) \tag{12.73a}$$

with $m\geq 0$ an integer.

One way of developing the solution of the associated Legendre equation is to start with the regular Legendre equation and convert it into the associated Legendre equation by using multiple differentiation. These multiple differentiations are suggested by Eq. (12.73a), the generation of associated Legendre polynomials, and spherical harmonics of Section 12.6 more generally, in Section 4.3 using raising or lowering operators of Eq. (4.69) repeatedly. For their derivative form see Exercise 12.6.8. We take Legendre's equation

$$\left(1-x^2\right)P_n''-2xP_n'+n(n+1)P_n=0, \tag{12.74}$$

and with the help of Leibniz' formula[15] differentiate m times. The result is

$$\left(1-x^2\right)u''-2x(m+1)u'+(n-m)(n+m+1)u=0, \tag{12.75}$$

where

$$u\equiv\frac{d^m}{dx^m}P_n(x). \tag{12.76}$$

Equation (12.74) is not self-adjoint. To put it into self-adjoint form and convert the weighting function to 1, we replace $u(x)$ by

$$v(x)=\left(1-x^2\right)^{m/2}u(x)=\left(1-x^2\right)^{m/2}\frac{d^mP_n(x)}{dx^m}. \tag{12.73b}$$

[15] Leibniz' formula for the nth derivative of a product is

$$\frac{d^n}{dx^n}\left[A(x)B(x)\right]=\sum_{s=0}^{n}\binom{n}{s}\frac{d^{n-s}}{dx^{n-s}}A(x)\frac{d^s}{dx^s}B(x),\qquad\binom{n}{s}=\frac{n!}{(n-s)!s!},$$

a binomial coefficient.

Solving for u and differentiating, we obtain

$$u' = \left(v' + \frac{mxv}{1-x^2}\right)\left(1-x^2\right)^{-m/2}, \tag{12.77}$$

$$u'' = \left[v'' + \frac{2mxv'}{1-x^2} + \frac{mv}{1-x^2} + \frac{m(m+2)x^2v}{(1-x^2)^2}\right]\cdot\left(1-x^2\right)^{-m/2}. \tag{12.78}$$

Substituting into Eq. (12.74), we find that the new function v satisfies the self-adjoint ODE

$$\left(1-x^2\right)v'' - 2xv' + \left[n(n+1) - \frac{m^2}{1-x^2}\right]v = 0, \tag{12.79}$$

which is the associated Legendre equation; it reduces to Legendre's equation when m is set equal to zero. Expressed in spherical polar coordinates, the associated Legendre equation is

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{dv}{d\theta}\right) + \left[n(n+1) - \frac{m^2}{\sin^2\theta}\right]v = 0. \tag{12.80}$$

Associated Legendre Polynomials

The regular solutions, relabeled $P_n^m(x)$, are

$$v \equiv P_n^m(x) = \left(1-x^2\right)^{m/2}\frac{d^m}{dx^m}P_n(x). \tag{12.73c}$$

These are the associated Legendre functions.[16] Since the highest power of x in $P_n(x)$ is x^n, we must have $m \le n$ (or the m-fold differentiation will drive our function to zero). In quantum mechanics the requirement that $m \le n$ has the physical interpretation that the expectation value of the square of the z component of the angular momentum is less than or equal to the expectation value of the square of the angular momentum vector $\mathbf{L}$,

$$\left\langle L_z^2\right\rangle \le \left\langle L^2\right\rangle \equiv \int \psi_{lm}^*\mathbf{L}^2\psi_{lm}\,d^3r.$$

From the form of Eq. (12.73c) we might expect m to be nonnegative. However, if $P_n(x)$ is expressed by Rodrigues' formula, this limitation on m is relaxed and we may have $-n \le m \le n$, negative as well as positive values of m being permitted. These limits are consistent with those obtained by means of raising and lowering operators in Chapter 4. In particular, $|m| > n$ is ruled out. This also follows from Eq. (12.73c). Using Leibniz' differentiation formula once again, we can show (Exercise 12.5.1) that $P_n^m(x)$ and $P_n^{-m}(x)$ are related by

$$P_n^{-m}(x) = (-1)^m\frac{(n-m)!}{(n+m)!}P_n^m(x). \tag{12.81}$$

[16] Occasionally (as in AMS-55; for the complete reference, see the Additional Readings of Chapter 8), one finds the associated Legendre functions defined with an additional factor of $(-1)^m$. This $(-1)^m$ seems an unnecessary complication at this point. It will be included in the definition of the spherical harmonics $Y_n^m(\theta,\varphi)$ in Section 12.6. Our definition agrees with Jackson's *Electrodynamics* (see Additional Readings of Chapter 11 for this reference). Note also that the upper index m is not an exponent.

From our definition of the associated Legendre functions $P_n^m(x)$,

$$P_n^0(x) = P_n(x). \tag{12.82}$$

A generating function for the associated Legendre functions is obtained, via Eq. (12.71), from that of the ordinary Legendre polynomials:

$$\frac{(2m)!(1-x^2)^{m/2}}{2^m m!(1-2tx+t^2)^{m+1/2}} = \sum_{s=0}^{\infty} P_{s+m}^m(x)t^s. \tag{12.83}$$

If we drop the factor $(1-x^2)^{m/2} = \sin^m\theta$ from this formula and define the **polynomials** $\mathcal{P}_{s+m}^m(x) = P_{s+m}^m(x)(1-x^2)^{-m/2}$, then we obtain a practical form of the generating function,

$$g_m(x,t) \equiv \frac{(2m)!}{2^m m!(1-2tx+t^2)^{m+1/2}} = \sum_{s=0}^{\infty} \mathcal{P}_{s+m}^m(x)t^s. \tag{12.84}$$

We can derive a recursion relation for associated Legendre polynomials that is analogous to Eqs. (12.14) and (12.17) by differentiation as follows:

$$\left(1-2tx+t^2\right)\frac{\partial g_m}{\partial t} = (2m+1)(x-t)g_m(x,t).$$

Substituting the defining expansions for associated Legendre polynomials we get

$$\left(1-2tx+t^2\right)\sum_s s\mathcal{P}_{s+m}^m(x)t^{s-1} = (2m+1)\sum_s\left[x\mathcal{P}_{s+m}^m t^s - \mathcal{P}_{s+m}^m t^{s+1}\right].$$

Comparing coefficients of powers of t in these power series, we obtain the recurrence relation

$$(s+1)\mathcal{P}_{s+m+1}^m - (2m+1+2s)x\mathcal{P}_{s+m}^m + (s+2m)\mathcal{P}_{s+m-1}^m = 0. \tag{12.85}$$

For $m=0$ and $s=n$ this relation is Eq. (12.17).

Before we can use this relation we need to initialize it, that is, relate the associated Legendre polynomials to ordinary Legendre polynomials. We can use $\mathcal{P}_m^m = (2m-1)!!$ from Eq. (12.73c). Also, since $|m| \le n$, we may set $P_n^{n+1} = 0$ and use this to obtain starting values for various recursive processes. We observe that

$$\left(1-2xt+t^2\right)g_1(x,t) = \left(1-2xt+t^2\right)^{-1/2} = \sum_s P_s(x)t^s, \tag{12.86}$$

so upon inserting Eq. (12.84) we get the recursion

$$\mathcal{P}_{s+1}^1 - 2x\mathcal{P}_s^1 + \mathcal{P}_{s-1}^1 = P_s(x). \tag{12.87}$$

More generally, we also have the identity

$$\left(1-2xt+t^2\right)g_{m+1}(x,t) = (2m+1)g_m(x,t), \tag{12.88}$$

from which we extract the recursion

$$\mathcal{P}_{s+m+1}^{m+1} - 2x\mathcal{P}_{s+m}^{m+1} + \mathcal{P}_{s+m-1}^{m+1} = (2m+1)\mathcal{P}_{s+m}^m(x), \tag{12.89}$$

which relates the associated Legendre polynomials with superindex $m+1$ to those with m. For $m=0$ we recover the initial recursion Eq. (12.87).

Table 12.2 Associated Legendre Functions

$P_1^1(x) = (1-x^2)^{1/2} = \sin\theta$
$P_2^1(x) = 3x(1-x^2)^{1/2} = 3\cos\theta\sin\theta$
$P_2^2(x) = 3(1-x^2) = 3\sin^2\theta$
$P_3^1(x) = \frac{3}{2}(5x^2-1)(1-x^2)^{1/2} = \frac{3}{2}(5\cos^2\theta - 1)\sin\theta$
$P_3^2(x) = 15x(1-x^2) = 15\cos\theta\sin^2\theta$
$P_3^3(x) = 15(1-x^2)^{3/2} = 15\sin^3\theta$
$P_4^1(x) = \frac{5}{2}(7x^3-3x)(1-x^2)^{1/2} = \frac{5}{2}(7\cos^3\theta - 3\cos\theta)\sin\theta$
$P_4^2(x) = \frac{15}{2}(7x^2-1)(1-x^2) = \frac{15}{2}(7\cos^2\theta - 1)\sin^2\theta$
$P_4^3(x) = 105x(1-x^2)^{3/2} = 105\cos\theta\sin^3\theta$
$P_4^4(x) = 105(1-x^2)^2 = 105\sin^4\theta$

Example 12.5.1 LOWEST ASSOCIATED LEGENDRE POLYNOMIALS

Now we are ready to derive the entries of Table 12.2. For $m = 1$ and $s = 0$, Eq. (12.87) yields $\mathcal{P}_1^1 = 1$, because $\mathcal{P}_0^1 = 0 = \mathcal{P}_{-1}^1$ do not occur in the definition, Eq. (12.84), of the associated Legendre polynomials. Multiplying by $(1-x^2)^{1/2} = \sin\theta$ we get the first line of Table 12.2. For $s = 1$ we find, from Eq. (12.87),

$$\mathcal{P}_2^1(x) = P_1 + 2x\mathcal{P}_1^1 = x + 2x = 3x,$$

from which the second line of Table 12.2, $3\cos\theta\sin\theta$, follows upon multiplying by $\sin\theta$. For $s = 2$ we get

$$\mathcal{P}_3^1(x) = P_2 + 2x\mathcal{P}_2^1 - \mathcal{P}_1^1 = \frac{1}{2}\left(3x^2-1\right) + 6x^2 - 1 = \frac{15}{2}x^2 - \frac{3}{2},$$

in agreement with line 4 of Table 12.2. To get line 3 we use Eq. (12.88). For $m = 1,\ s = 0$, this gives $\mathcal{P}_2^2(x) = 3\mathcal{P}_1^1(x) = 3$, and multiplying by $1 - x^2 = \sin^2\theta$ reproduces line 3 of Table 12.2. For lines 5, 8, 9, Eq. (12.84) may be used, which we leave as an exercise. More generally, we use Eq. (12.89) instead of Eq. (12.87) to get a starting value of $\mathcal{P}_m^m$. Then Eq. (12.85) reduces to a two-term formula for $\mathcal{P}_m^m$, giving $(2m-1)!!$. Note that, if $m = 0$, this is $(-1)!! = 1$. ■

Example 12.5.2 SPECIAL VALUES

For $x = 1$ we use

$$\left(1 - 2t + t^2\right)^{-m-1/2} = (1-t)^{-2m-1} = \sum_{s=0}^{\infty} \binom{-2m-1}{s} t^s$$

in Eq. (12.84) and find

$$\mathcal{P}_{s+m}^m(1) = \frac{(2m)!}{2^m m!}\binom{-2m-1}{s}, \tag{12.90}$$

where

$$\binom{-m}{s} = 1$$

for $s = 0$ and

$$\binom{-m}{s} = \frac{(-m)(-m-1)\cdots(1-s-m)}{s!}$$

for $s \geq 1$. For $m = 1$, $s = 0$ we have $\mathcal{P}_1^1(1) = \binom{-3}{0} = 1$; for $s = 1$, $\mathcal{P}_2^1(1) = -\binom{-3}{1} = 3$; for $s = 2$, $\mathcal{P}_3^1(1) = \binom{-3}{2} = \frac{(-3)(-4)}{2} = 6 = \frac{3}{2}(5-1)$, which all agree with Table 12.2. For $x = 0$ we can also use the binomial expansion, which we leave as an exercise. ■

Recurrence Relations

As expected and already seen, the associated Legendre functions satisfy recurrence relations. Because of the existence of two indices instead of just one, we have a wide variety of recurrence relations:

$$P_n^{m+1} - \frac{2mx}{(1-x^2)^{1/2}} P_n^m + \left[n(n+1) - m(m-1)\right] P_n^{m-1} = 0, \tag{12.91}$$

$$(2n+1)x\,P_n^m = (n+m)\,P_{n-1}^m + (n-m+1)\,P_{n+1}^m, \tag{12.92}$$

$$\begin{aligned}(2n+1)\left(1-x^2\right)^{1/2} P_n^m &= P_{n+1}^{m+1} - P_{n-1}^{m+1} \\ &= (n+m)(n+m-1)\,P_{n-1}^{m-1} \\ &\quad - (n-m+1)(n-m+2)\,P_{n+1}^{m-1},\end{aligned} \tag{12.93}$$

$$\left(1-x^2\right)^{1/2} P_n^{m'} = \frac{1}{2} P_n^{m+1} - \frac{1}{2}(n+m)(n-m+1)\,P_n^{m-1}. \tag{12.94}$$

These relations, and many other similar ones, may be verified by use of the generating function (Eq. (12.4)), by substitution of the series solution of the associated Legendre equation (12.79) or reduction to the Legendre polynomial recurrence relations, using Eq. (12.73c). As an example of the last method, consider Eq. (12.93). It is similar to Eq. (12.23):

$$(2n+1)\,P_n(x) = P'_{n+1}(x) - P'_{n-1}(x). \tag{12.95}$$

Let us differentiate this Legendre polynomial recurrence relation m times to obtain

$$\begin{aligned}(2n+1)\frac{d^m}{dx^m} P_n(x) &= \frac{d^m}{dx^m} P'_{n+1}(x) - \frac{d^m}{dx^m} P'_{n-1}(x) \\ &= \frac{d^{m+1}}{dx^{m+1}} P_{n+1}(x) - \frac{d^{m+1}}{dx^{m+1}} P_{n-1}(x).\end{aligned} \tag{12.96}$$

Now multiplying by $(1-x^2)^{(m+1)/2}$ and using the definition of $P_n(x)$, we obtain the first part of Eq. (12.93).

Parity

The parity relation satisfied by the associated Legendre functions may be determined by examination of the defining equation (12.73c). As $x \to -x$, we already know that $P_n(x)$ contributes a $(-1)^n$. The m-fold differentiation yields a factor of $(-1)^m$. Hence we have

$$P_n^m(-x) = (-1)^{n+m} P_n^m(x). \tag{12.97}$$

A glance at Table 12.2 verifies this for $1 \le m \le n \le 4$.

Also, from the definition in Eq. (12.73c),

$$P_n^m(\pm 1) = 0, \qquad \text{for } m \neq 0. \tag{12.98}$$

Orthogonality

The orthogonality of the $P_n^m(x)$ follows from the ODE, just as for the $P_n(x)$ (Section 12.3), if m is the same for both functions. However, it is instructive to demonstrate the orthogonality by another method, a method that will also provide the normalization constant.

Using the definition in Eq. (12.73c) and Rodrigues' formula (Eq. (12.65)) for $P_n(x)$, we find

$$\int_{-1}^{1} P_p^m(x) P_q^m(x)\, dx = \frac{(-1)^m}{2^{p+q}\, p!q!} \int_{-1}^{1} X^m \left(\frac{d^{p+m}}{dx^{p+m}} X^p \right) \frac{d^{q+m}}{dx^{q+m}} X^q\, dx. \tag{12.99}$$

The function X is given by $X \equiv (x^2 - 1)$. If $p \neq q$, let us assume that $p < q$. Notice that the superscript m is the same for both functions. This is an essential condition. The technique is to integrate repeatedly by parts; all the integrated parts will vanish as long as there is a factor $X = x^2 - 1$. Let us integrate $q + m$ times to obtain

$$\int_{-1}^{1} P_p^m(x) P_q^m(x)\, dx = \frac{(-1)^m (-1)^{q+m}}{2^{p+q}\, p!q!} \int_{-1}^{1} X^q \frac{d^{q+m}}{dx^{q+m}} \left(X^m \frac{d^{p+m}}{dx^{p+m}} X^p \right) dx. \tag{12.100}$$

The integrand on the right-hand side is now expanded by Leibniz' formula to give

$$\begin{aligned} & X^q \frac{d^{q+m}}{dx^{q+m}} \left(X^m \frac{d^{p+m}}{dx^{p+m}} X^p \right) \\ & \quad = X^q \sum_{i=0}^{q+m} \frac{(q+m)!}{i!(q+m-i)!} \left(\frac{d^{q+m-i}}{dx^{q+m-i}} X^m \right) \frac{d^{p+m+i}}{dx^{p+m+i}} X^p. \end{aligned} \tag{12.101}$$

Since the term X^m contains no power of x greater than x^{2m}, we must have

$$q + m - i \le 2m \tag{12.102}$$

or the derivative will vanish. Similarly,

$$p + m + i \le 2p. \tag{12.103}$$

Adding both inequalities yields

$$q \le p, \tag{12.104}$$

which contradicts our assumption that $p < q$. Hence, there is no solution for i and the integral vanishes. The same result obviously will follow if $p > q$.

For the remaining case, $p = q$, we have the single term corresponding to $i = q - m$. Putting Eq. (12.101) into Eq. (12.100), we have

$$\int_{-1}^{1}\left[P_q^m(x)\right]^2 dx = \frac{(-1)^{q+2m}(q+m)!}{2^{2q}q!q!(2m)!(q-m)!}\int_{-1}^{1} X^q\left(\frac{d^{2m}}{dx^{2m}}X^m\right)\left(\frac{d^{2q}}{dx^{2q}}X^q\right)dx. \tag{12.105}$$

Since

$$X^m = \left(x^2-1\right)^m = x^{2m} - mx^{2m-2} + \cdots, \tag{12.106}$$

$$\frac{d^{2m}}{dx^{2m}}X^m = (2m)!, \tag{12.107}$$

Eq. (12.105) reduces to

$$\int_{-1}^{1}\left[P_q^m(x)\right]^2 dx = \frac{(-1)^{q+2m}(2q)!(q+m)!}{2^{2q}q!q!(q-m)!}\int_{-1}^{1} X^q\, dx. \tag{12.108}$$

The integral on the right is just

$$(-1)^q\int_0^{\pi}\sin^{2q+1}\theta\, d\theta = \frac{(-1)^q 2^{2q+1}q!q!}{(2q+1)!} \tag{12.109}$$

(compare Exercise 8.4.9). Combining Eqs. (12.108) and (12.109), we have the **orthogonality integral**,

$$\int_{-1}^{1} P_p^m(x)P_q^m(x)\,dx = \frac{2}{2q+1}\cdot\frac{(q+m)!}{(q-m)!}\delta_{pq}, \tag{12.110}$$

or, in spherical polar coordinates,

$$\int_0^{\pi} P_p^m(\cos\theta)P_q^m(\cos\theta)\sin\theta\, d\theta = \frac{2}{2q+1}\cdot\frac{(q+m)!}{(q-m)!}\delta_{pq}. \tag{12.111}$$

The orthogonality of the Legendre polynomials is a special case of this result, obtained by setting m equal to zero; that is, for $m = 0$, Eq. (12.110) reduces to Eqs. (12.47) and (12.48). In both Eqs. (12.110) and (12.111), our Sturm–Liouville theory of Chapter 10 could provide the Kronecker delta. A special calculation, such as the analysis here, is required for the normalization constant.

The orthogonality of the associated Legendre functions over the same interval and with the same weighting factor as the Legendre polynomials does not contradict the uniqueness of the Gram–Schmidt construction of the Legendre polynomials, Example 10.3.1. Table 12.2 suggests (and Section 12.4 verifies) that $\int_{-1}^{1} P_p^m(x)P_q^m(x)\,dx$ may be written as

$$\int_{-1}^{1} P_p^m(x)P_q^m(x)\left(1-x^2\right)^m dx,$$

where we defined earlier

$$\mathcal{P}_p^m(x)\left(1-x^2\right)^{m/2} = P_p^m(x).$$

The functions $P_p^m(x)$ may be constructed by the Gram–Schmidt procedure with the weighting function $w(x) = (1 - x^2)^m$.

It is possible to develop an orthogonality relation for associated Legendre functions of the same lower index but different upper index. We find

$$\int_{-1}^{1} P_n^m(x) P_n^k(x) \left(1 - x^2\right)^{-1} dx = \frac{(n+m)!}{m(n-m!)} \delta_{m,k}. \tag{12.112}$$

Note that a new weighting factor, $(1 - x^2)^{-1}$, has been introduced. This relation is a mathematical curiosity. In physical problems with spherical symmetry solutions of Eqs. (12.80) and (9.64) appear in conjunction with those of Eq. (9.61), and orthogonality of the azimuthal dependence makes the two upper indices equal and always leads to Eq. (12.111).

Example 12.5.3 MAGNETIC INDUCTION FIELD OF A CURRENT LOOP

Like the other ODEs of mathematical physics, the associated Legendre equation is likely to pop up quite unexpectedly. As an illustration, consider the magnetic induction field **B** and magnetic vector potential **A** created by a single circular current loop in the equatorial plane (Fig. 12.12).

We know from electromagnetic theory that the contribution of current element $I\, d\boldsymbol{\lambda}$ to the magnetic vector potential is

$$d\mathbf{A} = \frac{\mu_0}{4\pi} \frac{I\, d\boldsymbol{\lambda}}{r}. \tag{12.113}$$

(This follows from Exercise 1.14.4 and Section 9.7) Equation (12.113), plus the symmetry of our system, shows that **A** has only a $\hat{\boldsymbol{\varphi}}$ component and that the component is independent

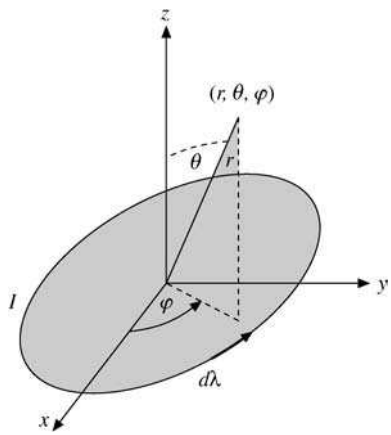

FIGURE 12.12 Circular current loop.

of φ,[17]

$$\mathbf{A} = \hat{\boldsymbol{\varphi}} A_\varphi(r, \theta). \tag{12.114}$$

By Maxwell's equations,

$$\nabla \times \mathbf{H} = \mathbf{J}, \qquad \frac{\partial \mathbf{D}}{\partial t} = 0 \qquad \text{(SI units)}. \tag{12.115}$$

Since

$$\mu_0 \mathbf{H} = \mathbf{B} = \nabla \times \mathbf{A}, \tag{12.116}$$

we have

$$\nabla \times (\nabla \times \mathbf{A}) = \mu_0 \mathbf{J}, \tag{12.117}$$

where $\mathbf{J}$ is the current density. In our problem $\mathbf{J}$ is zero everywhere except in the current loop. Therefore, away from the loop,

$$\nabla \times \nabla \times \hat{\boldsymbol{\varphi}} A_\varphi(r, \theta) = 0, \tag{12.118}$$

using Eq. (12.114).

From the expression for the curl in spherical polar coordinates (Section 2.5), we obtain (Example 2.5.2)

$$\begin{aligned} \nabla \times \left[\nabla \times \hat{\boldsymbol{\varphi}} A_\varphi(r, \theta)\right] &= \hat{\boldsymbol{\varphi}} \left[-\frac{\partial^2 A_\varphi}{\partial r^2} - \frac{2}{r}\frac{\partial A_\varphi}{\partial r} - \frac{1}{r^2}\frac{\partial^2 A_\varphi}{\partial \theta^2} - \frac{1}{r^2}\frac{\partial}{\partial \theta}(\cot\theta A_\varphi) \right] \\ &= 0. \end{aligned} \tag{12.119}$$

Letting $A_\varphi(r, \theta) = R(r)\Theta(\theta)$ and separating variables, we have

$$r^2 \frac{d^2 R}{dr^2} + 2\frac{dR}{dr} - n(n+1)R = 0, \tag{12.120}$$

$$\frac{d^2\Theta}{d\theta^2} + \cot\theta \frac{d\Theta}{d\theta} + n(n+1)\Theta - \frac{\Theta}{\sin^2\theta} = 0. \tag{12.121}$$

The second equation is the associated Legendre equation (12.80) with $m = 1$, and we may immediately write

$$\Theta(\theta) = P_n^1(\cos\theta). \tag{12.122}$$

The separation constant $n(n+1)$, n a nonnegative integer, was chosen to keep this solution well behaved.

By trial, letting $R(r) = r^\alpha$, we find that $\alpha = n$, or $-n-1$. The first possibility is discarded, for our solution must vanish as $r \to \infty$. Hence

$$A_{\varphi n} = \frac{b_n}{r^{n+1}} P_n^1(\cos\theta) = c_n \left(\frac{a}{r}\right)^{n+1} P_n^1(\cos\theta) \tag{12.123}$$

[17] Pair off corresponding current elements $I\, d\boldsymbol{\lambda}(\varphi_1)$ and $I\, d\boldsymbol{\lambda}(\varphi_2)$, where $\varphi - \varphi_1 = \varphi_2 - \varphi$.

and

$$A_\varphi(r,\theta)=\sum_{n=1}^{\infty}c_n\left(\frac{a}{r}\right)^{n+1}P_n^1(\cos\theta)\qquad(r>a). \tag{12.124}$$

Here a is the radius of the current loop.

Since A_φ must be invariant to reflection in the equatorial plane, by the symmetry of our problem,

$$A_\varphi(r,\cos\theta)=A_\varphi(r,-\cos\theta), \tag{12.125}$$

the parity property of $P_n^m(\cos\theta)$ (Eq. (12.97)) shows that $c_n=0$ for n even.

To complete the evaluation of the constants, we may use Eq. (12.124) to calculate B_z along the z-axis ($B_z=B_r(r,\theta=0)$) and compare with the expression obtained from the Biot–Savart law. This is the same technique as used in Example 12.3.3. We have (compare Eq. (2.47))

$$B_r=\nabla\times\mathbf{A}\big|_r=\frac{1}{r\sin\theta}\left[\frac{\partial}{\partial\theta}(\sin\theta A_\varphi)\right]=\frac{\cot\theta}{r}A_\varphi+\frac{1}{r}\frac{\partial A_\varphi}{\partial\theta}. \tag{12.126}$$

Using

$$\frac{\partial P_n^1(\cos\theta)}{\partial\theta}=-\sin\theta\frac{dP_n^1(\cos\theta)}{d(\cos\theta)}=-\frac{1}{2}P_n^2+\frac{n(n+1)}{2}P_n^0 \tag{12.127}$$

(Eq. (12.94)) and then Eq. (12.91) with $m=1$,

$$P_n^2(\cos\theta)-\frac{2\cos\theta}{\sin\theta}P_n^1(\cos\theta)+n(n+1)P_n(\cos\theta)=0, \tag{12.128}$$

we obtain

$$B_r(r,\theta)=\sum_{n=1}^{\infty}c_n n(n+1)\frac{a^{n+1}}{r^{n+2}}P_n(\cos\theta),\qquad r>a, \tag{12.129}$$

(for all θ). In particular, for $\theta=0$,

$$B_r(r,0)=\sum_{n=1}^{\infty}c_n n(n+1)\frac{a^{n+1}}{r^{n+2}}. \tag{12.130}$$

We may also obtain

$$B_\theta(r,\theta)=-\frac{1}{r}\frac{\partial(rA_\varphi)}{\partial r}=\sum_{n=1}^{\infty}c_n n\frac{a^{n+1}}{r^{n+2}}P_n^1(\cos\theta),\qquad r>a, \tag{12.131}$$

The Biot–Savart law states that

$$d\mathbf{B}=\frac{\mu_0}{4\pi}I\frac{d\boldsymbol{\lambda}\times\hat{\mathbf{r}}}{r^2}\qquad\text{(SI units)}. \tag{12.132}$$

We now integrate over the perimeter of our loop (radius a). The geometry is shown in Fig. 12.13. The resulting magnetic induction field is $\hat{\mathbf{z}}B_z$, along the z-axis, with

$$B_z=\frac{\mu_0 I}{2}a^2\left(a^2+z^2\right)^{-3/2}=\frac{\mu_0 I}{2}\frac{a^2}{z^3}\left(1+\frac{a^2}{z^2}\right)^{-3/2}. \tag{12.133}$$

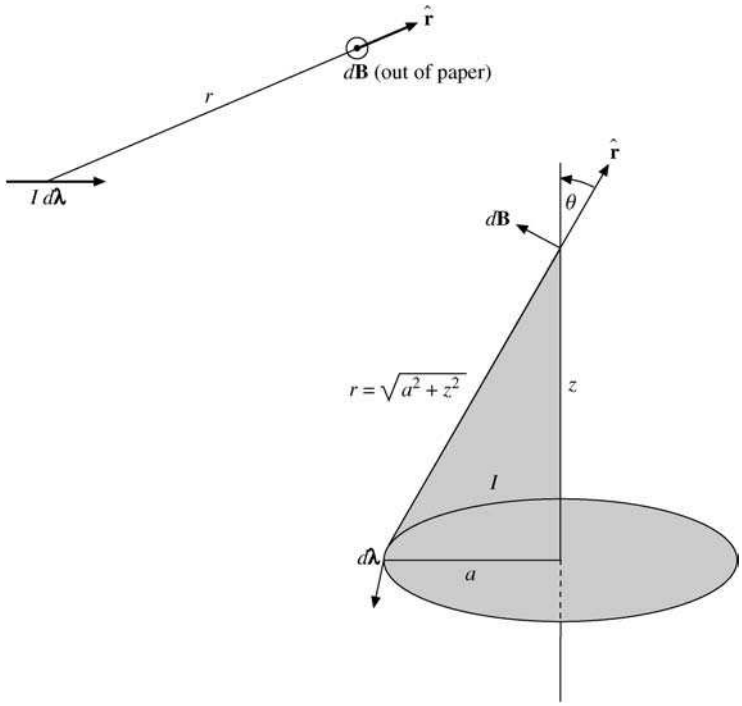

FIGURE 12.13 Biot–Savart law applied to a circular loop.

Expanding by the binomial theorem, we obtain

$$B_z = \frac{\mu_0 I}{2}\frac{a^2}{z^3}\left[1 - \frac{3}{2}\left(\frac{a}{z}\right)^2 + \frac{15}{8}\left(\frac{a}{z}\right)^4 - \cdots\right]$$
$$= \frac{\mu_0 I}{2}\frac{a^2}{z^3}\sum_{s=0}^{\infty}(-1)^s\frac{(2s+1)!!}{(2s)!!}\left(\frac{a}{z}\right)^{2s}, \qquad z > a. \tag{12.134}$$

Equating Eqs. (12.130) and (12.134) term by term (with $r = z$),[18] we find

$$c_1 = \frac{\mu_0 I}{4}, \qquad c_3 = -\frac{\mu_0 I}{16}, \qquad c_2 = c_4 = \cdots = 0.$$
$$c_n = (-1)^{(n-1)/2}\frac{\mu_0 I}{2n(n+1)}\cdot\frac{(n/2)!}{[(n-1)/2]!(\frac{1}{2})!}, \qquad n \text{ odd}. \tag{12.135}$$

Equivalently, we may write

$$c_{2n+1} = (-1)^n\frac{\mu_0 I}{2^{2n+2}}\cdot\frac{(2n)!}{n!(n+1)!} = (-1)^n\frac{\mu_0 I}{2}\cdot\frac{(2n-1)!!}{(2n+2)!!} \tag{12.136}$$

[18]The descending power series is also unique.

and

$$A_\varphi(r,\theta) = \left(\frac{a}{r}\right)^2 \sum_{n=0}^{\infty} c_{2n+1}\left(\frac{a}{r}\right)^{2n} P_{2n+1}^1(\cos\theta), \tag{12.137}$$

$$B_r(r,\theta) = \frac{a^2}{r^3}\sum_{n=0}^{\infty} c_{2n+1}(2n+1)(2n+2)\left(\frac{a}{r}\right)^{2n} P_{2n+1}(\cos\theta), \tag{12.138}$$

$$B_\theta(r,\theta) = \frac{a^2}{r^3}\sum_{n=0}^{\infty} c_{2n+1}(2n+1)\left(\frac{a}{r}\right)^{2n} P_{2n+1}^1(\cos\theta). \tag{12.139}$$

These fields may be described in closed form by the use of elliptic integrals. Exercise 5.8.4 is an illustration of this approach. A third possibility is direct integration of Eq. (12.113) by expanding the denominator of the integral for A_φ in Exercise 5.8.4 as a Legendre polynomial generating function. The current is specified by Dirac delta functions. These methods have the advantage of yielding the constants c_n directly.

A comparison of magnetic current loop dipole fields and finite electric dipole fields may be of interest. For the magnetic current loop dipole, the preceding analysis gives

$$B_r(r,\theta) = \frac{\mu_0 I}{2}\frac{a^2}{r^3}\left[P_1 - \frac{3}{2}\left(\frac{a}{r}\right)^2 P_3 + \cdots\right], \tag{12.140}$$

$$B_\theta(r,\theta) = \frac{\mu_0 I}{4}\frac{a^2}{r^3}\left[P_1^1 - \frac{3}{4}\left(\frac{a}{r}\right)^2 P_3^1 + \cdots\right]. \tag{12.141}$$

From the finite electric dipole potential of Section 12.1 we have

$$E_r(r,\theta) = \frac{qa}{\pi\varepsilon_0 r^3}\left[P_1 + 2\left(\frac{a}{r}\right)^2 P_3 + \cdots\right], \tag{12.142}$$

$$E_\theta(r,\theta) = \frac{qa}{2\pi\varepsilon_0 r^3}\left[P_1^1 + \left(\frac{a}{r}\right)^2 P_3^1 + \cdots\right]. \tag{12.143}$$

The two fields agree in form as far as the leading term is concerned ($r^{-3}P_1$), and this is the basis for calling them both dipole fields.

As with electric multipoles, it is sometimes convenient to discuss **point** magnetic multipoles (see Fig. 12.14). For the dipole case, Eqs. (12.140) and (12.141), the point dipole is formed by taking the limit $a \to 0$, $I \to \infty$, with Ia^2 held constant. With $\mathbf{n}$ a unit vector normal to the current loop (positive sense by right-hand rule, Section 1.10), the magnetic moment $\mathbf{m}$ is given by $\mathbf{m} = \mathbf{n}I\pi a^2$. ■

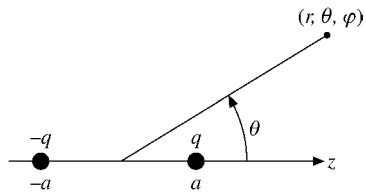

FIGURE 12.14 Electric dipole.

Exercises

12.5.1 Prove that

$$P_n^{-m}(x) = (-1)^m \frac{(n-m)!}{(n+m)!} P_n^m(x),$$

where $P_n^m(x)$ is defined by

$$P_n^m(x) = \frac{1}{2^n n!}\left(1-x^2\right)^{m/2} \frac{d^{n+m}}{dx^{n+m}}\left(x^2-1\right)^n.$$

Hint. One approach is to apply Leibniz' formula to $(x+1)^n(x-1)^n$.

12.5.2 Show that

$$P_{2n}^1(0) = 0,$$

$$P_{2n+1}^1(0) = (-1)^n \frac{(2n+1)!}{(2^n n!)^2} = (-1)^n \frac{(2n+1)!!}{(2n)!!},$$

by each of these three methods:

(a) use of recurrence relations,
(b) expansion of the generating function,
(c) Rodrigues' formula.

12.5.3 Evaluate $P_n^m(0)$.

$$\text{ANS. } P_n^m(0) = \begin{cases} (-1)^{(n-m)/2} \dfrac{(n+m)!}{2^n((n-m)/2)!((n+m)/2!)}, & n+m \text{ even}, \\ 0, & n+m \text{ odd}. \end{cases}$$

$$\text{Also, } P_n^m(0) = (-1)^{(n-m)/2} \frac{(n+m-1)!!}{(n-m)!!}, n+m \text{ even}.$$

12.5.4 Show that

$$P_n^n(\cos\theta) = (2n-1)!! \sin^n\theta, \qquad n = 0, 1, 2, \ldots.$$

12.5.5 Derive the associated Legendre recurrence relation

$$P_n^{m+1}(x) - \frac{2mx}{(1-x^2)^{1/2}} P_n^m(x) + \left[n(n+1) - m(m-1)\right] P_n^{m-1}(x) = 0.$$

12.5.6 Develop a recurrence relation that will yield $P_n^1(x)$ as

$$P_n^1(x) = f_1(x,n) P_n(x) + f_2(x,n) P_{n-1}(x).$$

Follow either (a) or (b).

(a) Derive a recurrence relation of the preceding form. Give $f_1(x,n)$ and $f_2(x,n)$ explicitly.
(b) Find the appropriate recurrence relation in print.
    (1) Give the source.

(2) Verify the recurrence relation.

$$\text{ANS. } P_n^1(x) = -\frac{nx}{(1-x^2)^{1/2}}P_n + \frac{n}{(1-x^2)^{1/2}}P_{n-1}.$$

12.5.7 Show that

$$\sin\theta\frac{d}{d\cos\theta}P_n(\cos\theta) = P_n^1(\cos\theta).$$

12.5.8 Show that

(a) $$\int_0^\pi \left(\frac{dP_n^m}{d\theta}\frac{dP_{n'}^m}{d\theta} + \frac{m^2P_n^mP_{n'}^m}{\sin^2\theta}\right)\sin\theta\,d\theta = \frac{2n(n+1)}{2n+1}\frac{(n+m)!}{(n-m)!}\delta_{nn'},$$

(b) $$\int_0^\pi \left(\frac{P_n^1}{\sin\theta}\frac{dP_{n'}^1}{d\theta} + \frac{P_{n'}^1}{\sin\theta}\frac{dP_n^1}{d\theta}\right)\sin\theta\,d\theta = 0.$$

These integrals occur in the theory of scattering of electromagnetic waves by spheres.

12.5.9 As a repeat of Exercise 12.3.10, show, using associated Legendre functions, that

$$\int_{-1}^1 x\left(1-x^2\right)P_n'(x)P_m'(x)\,dx = \frac{n+1}{2n+1}\cdot\frac{2}{2n-1}\cdot\frac{n!}{(n-2)!}\delta_{m,n-1} + \frac{n}{2n+1}\cdot\frac{2}{2n+3}\cdot\frac{(n+2)!}{n!}\delta_{m,n+1}.$$

12.5.10 Evaluate

$$\int_0^\pi \sin^2\theta\,P_n^1(\cos\theta)\,d\theta.$$

12.5.11 The associated Legendre polynomial $P_n^m(x)$ satisfies the self-adjoint ODE

$$\left(1-x^2\right)\frac{d^2P_n^m(x)}{dx^2} - 2x\frac{dP_n^m(x)}{dx} + \left[n(n+1) - \frac{m^2}{1-x^2}\right]P_n^m(x) = 0.$$

From the differential equations for $P_n^m(x)$ and $P_n^k(x)$ show that

$$\int_{-1}^1 P_n^m(x)P_n^k(x)\frac{dx}{1-x^2} = 0,$$

for $k \neq m$.

12.5.12 Determine the vector potential of a magnetic quadrupole by differentiating the magnetic dipole potential.

$$\text{ANS. } \mathbf{A}_{MQ} = \frac{\mu_0}{2}\left(Ia^2\right)(dz)\hat{\boldsymbol{\varphi}}\frac{P_2^1(\cos\theta)}{r^3} + \text{higher-order terms.}$$

$$\mathbf{B}_{MQ} = \mu_0\left(Ia^2\right)(dz)\left[\hat{\mathbf{r}}\frac{3P_2(\cos\theta)}{r^4} + \hat{\boldsymbol{\theta}}\frac{P_2^1(\cos\theta)}{r^4}\right].$$

This corresponds to placing a current loop of radius a at $z \to dz$ and an oppositely directed current loop at $z \to -dz$ and letting $a \to 0$ subject to $(dz)a$ (dipole strength) equal constant.
Another approach to this problem would be to integrate $d\mathbf{A}$ (Eq. (12.113), to expand the denominator in a series of Legendre polynomials, and to use the Legendre polynomial addition theorem (Section 12.8).

12.5.13 A single loop of wire of radius a carries a constant current I.

(a) Find the magnetic induction $\mathbf{B}$ for $r < a,\ \theta = \pi/2$.
(b) Calculate the integral of the magnetic flux $(\mathbf{B} \cdot d\boldsymbol{\sigma})$ over the area of the current loop, that is

$$\int_0^a \int_0^{2\pi} B_z\left(r, \theta = \frac{\pi}{2}\right) d\varphi\, r\, dr.$$

ANS. ∞.

The Earth is within such a ring current, in which I approximates millions of amperes arising from the drift of charged particles in the Van Allen belt.

12.5.14 (a) Show that in the point dipole limit the magnetic induction field of the current loop becomes

$$B_r(r, \theta) = \frac{\mu_0}{2\pi} \frac{m}{r^3} P_1(\cos\theta),$$
$$B_\theta(r, \theta) = \frac{\mu_0}{2\pi} \frac{m}{r^3} P_1^1(\cos\theta)$$

with $m = I\pi a^2$.
(b) Compare these results with the magnetic induction of the point magnetic dipole of Exercise 1.8.17. Take $\mathbf{m} = \hat{\mathbf{z}}m$.

12.5.15 A uniformly charged spherical shell is rotating with constant angular velocity.

(a) Calculate the magnetic induction $\mathbf{B}$ along the axis of rotation outside the sphere.
(b) Using the vector potential series of Section 12.5, find $\mathbf{A}$ and then $\mathbf{B}$ for all space outside the sphere.

12.5.16 In the liquid drop model of the nucleus, the spherical nucleus is subjected to small deformations. Consider a sphere of radius r_0 that is deformed so that its new surface is given by

$$r = r_0\big[1 + \alpha_2 P_2(\cos\theta)\big].$$

Find the area of the deformed sphere through terms of order α_2^2.
Hint.

$$dA = \left[r^2 + \left(\frac{dr}{d\theta}\right)^2\right]^{1/2} r \sin\theta\, d\theta\, d\varphi.$$

ANS. $A = 4\pi r_0^2\big[1 + \frac{4}{5}\alpha_2^2 + \mathcal{O}\big(\alpha_2^3\big)\big]$.

Note. The area element dA follows from noting that the line element ds for fixed φ is given by

$$ds = \left(r^2\,d\theta^2 + dr^2\right)^{1/2} = \left(r^2 + \left(\frac{dr}{d\theta}\right)^2\right)^{1/2} d\theta.$$

12.5.17 A nuclear particle is in a spherical square well potential $V(r,\theta,\varphi) = 0$ for $0 \le r < a$ and ∞ for $r > a$. The particle is described by a wave function $\psi(r,\theta,\varphi)$ which satisfies the wave equation

$$-\frac{\hbar^2}{2M}\nabla^2\psi + V_0\psi = E\psi, \qquad r < a,$$

and the boundary condition

$$\psi(r = a) = 0.$$

Show that for the energy E to be a minimum there must be no angular dependence in the wave function; that is, $\psi = \psi(r)$.
Hint. The problem centers on the boundary condition on the radial function.

12.5.18 (a) Write a subroutine to calculate the numerical value of the associated Legendre function $P_N^1(x)$ for given values of N and x.
Hint. With the known forms of P_1^1 and P_2^1 you can use the recurrence relation Eq. (12.92) to generate P_N^1, $N > 2$.

(b) Check your subroutine by having it calculate $P_N^1(x)$ for $x = 0.0(0.5)\ 1.0$ and $N = 1(1)10$. Check these numerical values against the known values of $P_N^1(0)$ and $P_N^1(1)$ and against the tabulated values of $P_N^1(0.5)$.

12.5.19 Calculate the magnetic vector potential of a current loop, Example 12.5.1. Tabulate your results for $r/a = 1.5(0.5)5.0$ and $\theta = 0°(15°)90°$. Include terms in the series expansion, Eq. (12.137), until the absolute values of the terms drop below the leading term by a factor of 10^5 or more.
Note. This associated Legendre expansion can be checked by comparison with the elliptic integral solution, Exercise 5.8.4.

Check value. For $r/a = 4.0$ and $\theta = 20°$,
$A_\varphi/\mu_0 I = 4.9398 \times 10^{-3}$.

12.6 SPHERICAL HARMONICS

In the separation of variables of (1) Laplace's equation, (2) Helmholtz's or the space-dependence of the electromagnetic wave equation, and (3) the Schrödinger wave equation for central force fields,

$$\nabla^2\psi + k^2 f(r)\psi = 0, \tag{12.144}$$

the angular dependence, coming entirely from the Laplacian operator, is[19]

$$\frac{\Phi(\varphi)}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right)+\frac{\Theta(\theta)}{\sin^2\theta}\frac{d^2\Phi(\varphi)}{d\varphi^2}+n(n+1)\Theta(\theta)\Phi(\varphi)=0. \tag{12.145}$$

Azimuthal Dependence — Orthogonality

The separated azimuthal equation is

$$\frac{1}{\Phi(\varphi)}\frac{d^2\Phi(\varphi)}{d\varphi^2}=-m^2, \tag{12.146}$$

with solutions

$$\Phi(\varphi)=e^{-im\varphi},\ e^{im\varphi}, \tag{12.147}$$

with m integer, which satisfy the orthogonal condition

$$\int_0^{2\pi} e^{-im_1\varphi}e^{im_2\varphi}\,d\varphi=2\pi\delta_{m_1m_2}. \tag{12.148}$$

Notice that it is the product $\Phi^*_{m_1}(\varphi)\Phi_{m_2}(\varphi)$ that is taken and that * is used to indicate the complex conjugate function. This choice is not required, but it is convenient for quantum mechanical calculations. We could have used

$$\Phi=\sin m\varphi,\ \cos m\varphi \tag{12.149}$$

and the conditions of orthogonality that form the basis for Fourier series (Chapter 14). For applications such as describing the Earth's gravitational or magnetic field, $\sin m\varphi$ and $\cos m\varphi$ would be the preferred choice (see Example 12.6.1).

In electrostatics and most other physical problems we require m to be an integer in order that $\Phi(\varphi)$ be a single-valued function of the azimuth angle. In quantum mechanics the question is much more involved: Compare the footnote in Section 9.3.

By means of Eq. (12.148),

$$\Phi_m=\frac{1}{\sqrt{2\pi}}e^{im\varphi} \tag{12.150}$$

is orthonormal (orthogonal and normalized) with respect to integration over the azimuth angle φ.

[19] For a separation constant of the form $n(n+1)$ with n an integer, a Legendre-equation-series solution becomes a polynomial. Otherwise both series solutions diverge, Exercise 9.5.5.

Polar Angle Dependence

Splitting off the azimuthal dependence, the polar angle dependence (θ) leads to the associated Legendre equation (12.80), which is satisfied by the associated Legendre functions; that is, $\Theta(\theta) = P_n^m(\cos\theta)$. To include negative values of m, we use Rodrigues' formula, Eq. (12.65), in the definition of $P_n^m(\cos\theta)$. This leads to

$$P_n^m(\cos\theta) = \frac{1}{2^n n!}\left(1-x^2\right)^{m/2}\frac{d^{m+n}}{dx^{m+n}}\left(x^2-1\right)^n, \qquad -n \le m \le n. \tag{12.151}$$

$P_n^m(\cos\theta)$ and $P_n^{-m}(\cos\theta)$ are related as indicated in Exercise 12.5.1. An advantage of this approach over simply defining $P_n^m(\cos\theta)$ for $0 \le m \le n$ and requiring that $P_n^{-m} = P_n^m$ is that the recurrence relations valid for $0 \le m \le n$ remain valid for $-n \le m < 0$.

Normalizing the associated Legendre function by Eq. (12.110), we obtain the orthonormal functions

$$\sqrt{\frac{2n+1}{2}\frac{(n-m)!}{(n+m)!}}\,P_n^m(\cos\theta), \qquad -n \le m \le n, \tag{12.152}$$

which are orthonormal with respect to the polar angle θ.

Spherical Harmonics

The function $\Phi_m(\varphi)$ (Eq. (12.150)) is orthonormal with respect to the azimuthal angle φ. We take the product of $\Phi_m(\varphi)$ and the orthonormal function in polar angle from Eq. (12.152) and define

$$Y_n^m(\theta,\varphi) \equiv (-1)^m\sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}}\,P_n^m(\cos\theta)e^{im\varphi} \tag{12.153}$$

to obtain functions of two angles (and two indices) that are orthonormal over the spherical surface. These $Y_n^m(\theta,\varphi)$ are spherical harmonics, of which the first few are plotted in Fig. 12.15. The complete orthogonality integral becomes

$$\int_{\varphi=0}^{2\pi}\int_{\theta=0}^{\pi} Y_{n_1}^{m_1*}(\theta,\varphi)Y_{n_2}^{m_2}(\theta,\varphi)\sin\theta\,d\theta\,d\varphi = \delta_{n_1n_2}\delta_{m_1m_2}. \tag{12.154}$$

The extra $(-1)^m$ included in the defining equation of $Y_n^m(\theta,\varphi)$ deserves some comment. It is clearly legitimate, since Eq. (12.144) is linear and homogeneous. It is not necessary, but in moving on to certain quantum mechanical calculations, particularly in the quantum theory of angular momentum (Section 12.7), it is most convenient. The factor $(-1)^m$ is a phase factor, often called the Condon–Shortley phase, after the authors of a classic text on atomic spectroscopy. The effect of this $(-1)^m$ (Eq. (12.153)) and the $(-1)^m$ of Eq. (12.73c) for $P_n^{-m}(\cos\theta)$ is to introduce an alternation of sign among the positive m spherical harmonics. This is shown in Table 12.3.

The functions $Y_n^m(\theta,\varphi)$ acquired the name **spherical harmonics** first because they are defined over the surface of a sphere with θ the polar angle and φ the azimuth. The **harmonic** was included because solutions of Laplace's equation were called harmonic functions and $Y_n^m(\cos,\varphi)$ is the angular part of such a solution.

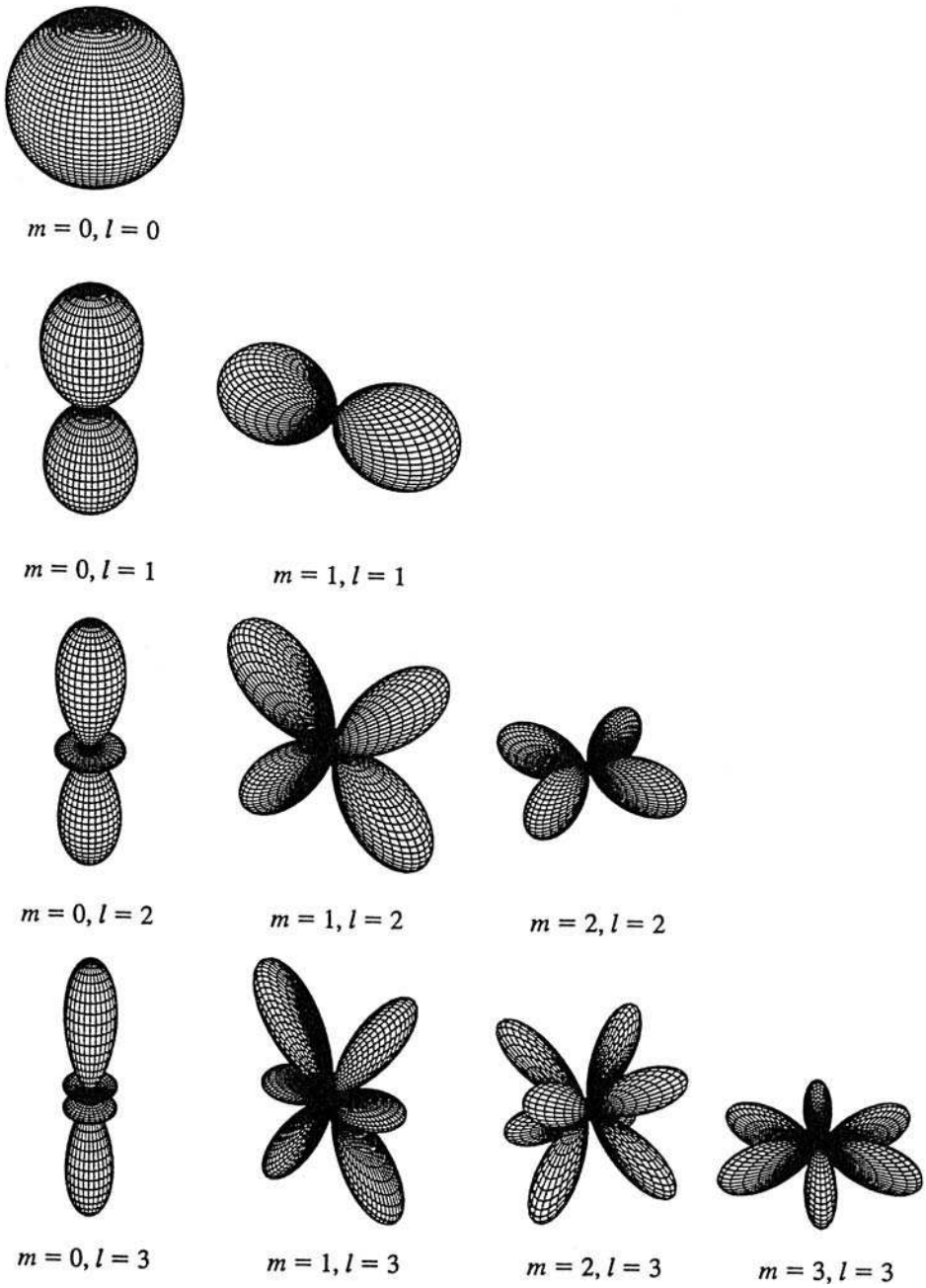

FIGURE 12.15 $[\Re Y_l^m(\theta, \varphi)]^2$ for $0 \leq l \leq 3, m = 0, \ldots, 3$.

Table 12.3 Spherical Harmonics (Condon–Shortley Phase)

$$Y_0^0(\theta,\varphi)=\frac{1}{\sqrt{4\pi}}$$

$$Y_1^1(\theta,\varphi)=-\sqrt{\frac{3}{8\pi}}\sin\theta e^{i\varphi}$$

$$Y_1^0(\theta,\varphi)=\sqrt{\frac{3}{4\pi}}\cos\theta$$

$$Y_1^{-1}(\theta,\varphi)=+\sqrt{\frac{3}{8\pi}}\sin\theta e^{-i\varphi}$$

$$Y_2^2(\theta,\varphi)=\sqrt{\frac{5}{96\pi}}3\sin^2\theta e^{2i\varphi}$$

$$Y_2^1(\theta,\varphi)=-\sqrt{\frac{5}{24\pi}}3\sin\theta\cos\theta e^{i\varphi}$$

$$Y_2^0(\theta,\varphi)=\sqrt{\frac{5}{4\pi}}\left(\frac{3}{2}\cos^2\theta-\frac{1}{2}\right)$$

$$Y_2^{-1}(\theta,\varphi)=+\sqrt{\frac{5}{24\pi}}3\sin\theta\cos\theta e^{-i\varphi}$$

$$Y_2^{-2}(\theta,\varphi)=\sqrt{\frac{5}{96\pi}}3\sin^2\theta e^{-2i\varphi}$$

In the framework of quantum mechanics Eq. (12.145) becomes an orbital angular momentum equation and the solution $Y_L^M(\theta,\varphi)$ (n replaced by L, m replaced by M) is an angular momentum eigenfunction, L being the angular momentum quantum number and M the z-axis projection of L. These relationships are developed in more detail in Sections 4.3 and 12.7.

Laplace Series, Expansion Theorem

Part of the importance of spherical harmonics lies in the completeness property, a consequence of the Sturm–Liouville form of Laplace's equation. This property, in this case, means that any function $f(\theta,\varphi)$ (with sufficient continuity properties) evaluated over the surface of the sphere can be expanded in a uniformly convergent double series of spherical harmonics[20] (Laplace's series):

$$f(\theta,\varphi)=\sum_{m,n}a_{mn}Y_n^m(\theta,\varphi). \tag{12.155}$$

If $f(\theta,\varphi)$ is known, the coefficients can be immediately found by the use of the orthogonality integral.

[20]For a proof of this fundamental theorem see E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics*, New York: Chelsea (1955), Chapter VII. If $f(\theta,\varphi)$ is discontinuous we may still have convergence in the mean, Section 10.4.

Table 12.4 Gravity Field Coefficients, Eq. (12.156)

Coefficient[a]	Earth	Moon	Mars
C_{20}	1.083×10^{-3}	$(0.200 \pm 0.002) \times 10^{-3}$	$(1.96 \pm 0.01) \times 10^{-3}$
C_{22}	0.16×10^{-5}	$(2.4 \pm 0.5) \times 10^{-5}$	$(-5 \pm 1) \times 10^{-5}$
S_{22}	-0.09×10^{-5}	$(0.5 \pm 0.6) \times 10^{-5}$	$(3 \pm 1) \times 10^{-5}$

[a] C_{20} represents an equatorial bulge, whereas C_{22} and S_{22} represent an azimuthal dependence of the gravitational field.

Example 12.6.1 LAPLACE SERIES — GRAVITY FIELDS

The gravity fields of the Earth, the Moon, and Mars have been described by a Laplace series with real eigenfunctions:

$$U(r,\theta,\varphi) = \frac{GM}{R}\left[\frac{R}{r} - \sum_{n=2}^{\infty}\sum_{m=0}^{n}\left(\frac{R}{r}\right)^{n+1}\left\{C_{nm}Y_{mn}^{e}(\theta,\varphi) + S_{nm}Y_{mn}^{o}(\theta,\varphi)\right\}\right]. \quad (12.156)$$

Here M is the mass of the body and R is the equatorial radius. The real functions Y_{mn}^{e} and Y_{mn}^{o} are defined by

$$Y_{mn}^{e}(\theta,\varphi) = P_n^m(\cos\theta)\cos m\varphi, \qquad Y_{mn}^{o}(\theta,\varphi) = P_n^m(\cos\theta)\sin m\varphi.$$

For applications such as this, the real trigonometric forms are preferred to the imaginary exponential form of $Y_L^M(\theta,\varphi)$. Satellite measurements have led to the numerical values shown in Table 12.4. ■

Exercises

12.6.1 Show that the parity of $Y_L^M(\theta,\varphi)$ is $(-1)^L$. Note the disappearance of any M dependence.
Hint. For the parity operation in spherical polar coordinates see Exercise 2.5.8 and footnote 7 in Section 12.2.

12.6.2 Prove that

$$Y_L^M(0,\varphi) = \left(\frac{2L+1}{4\pi}\right)^{1/2}\delta_{M,0}.$$

12.6.3 In the theory of Coulomb excitation of nuclei we encounter $Y_L^M(\pi/2,0)$. Show that

$$Y_L^M\left(\frac{\pi}{2},0\right) = \left(\frac{2L+1}{4\pi}\right)^{1/2}\frac{[(L-M)!(L+M)!]^{1/2}}{(L-M)!!(L+M)!!}(-1)^{(L+M)/2} \qquad \text{for } L+M \text{ even},$$
$$= 0 \qquad \text{for } L+M \text{ odd}.$$

Here

$$(2n)!! = 2n(2n-2)\cdots 6\cdot 4\cdot 2,$$
$$(2n+1)!! = (2n+1)(2n-1)\cdots 5\cdot 3\cdot 1.$$

12.6.4 (a) Express the elements of the quadrupole moment tensor $x_i x_j$ as a linear combination of the spherical harmonics Y_2^m (and Y_0^0).
Note. The tensor $x_i x_j$ is reducible. The Y_0^0 indicates the presence of a scalar component.

(b) The quadrupole moment tensor is usually defined as

$$Q_{ij} = \int \left(3x_i x_j - r^2\delta_{ij}\right)\rho(\mathbf{r})\, d\tau,$$

with $\rho(\mathbf{r})$ the charge density. Express the components of $(3x_i x_j - r^2\delta_{ij})$ in terms of $r^2 Y_2^M$.

(c) What is the significance of the $-r^2\delta_{ij}$ term?
Hint. Compare Sections 2.9 and 4.4.

12.6.5 The orthogonal azimuthal functions yield a useful representation of the Dirac delta function. Show that

$$\delta(\varphi_1 - \varphi_2) = \frac{1}{2\pi}\sum_{m=-\infty}^{\infty} \exp\big[im(\varphi_1 - \varphi_2)\big].$$

12.6.6 Derive the spherical harmonic closure relation

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{+l} Y_l^m(\theta_1,\varphi_1)Y_l^{m*}(\theta_2,\varphi_2) = \frac{1}{\sin\theta_1}\delta(\theta_1-\theta_2)\delta(\varphi_1-\varphi_2)$$
$$= \delta(\cos\theta_1 - \cos\theta_2)\delta(\varphi_1-\varphi_2).$$

12.6.7 The quantum mechanical angular momentum operators $L_x \pm iL_y$ are given by

$$L_x + iL_y = e^{i\varphi}\left(\frac{\partial}{\partial\theta} + i\cot\theta\frac{\partial}{\partial\varphi}\right),$$
$$L_x - iL_y = -e^{-i\varphi}\left(\frac{\partial}{\partial\theta} - i\cot\theta\frac{\partial}{\partial\varphi}\right).$$

Show that

(a) $(L_x + iL_y)Y_L^M(\theta,\varphi) = \sqrt{(L-M)(L+M+1)}\,Y_L^{M+1}(\theta,\varphi)$,

(b) $(L_x - iL_y)Y_L^M(\theta,\varphi) = \sqrt{(L+M)(L-M+1)}\,Y_L^{M-1}(\theta,\varphi)$.

12.6.8 With $L_\pm$ given by

$$L_\pm = L_x \pm iL_y = \pm e^{\pm i\varphi}\left[\frac{\partial}{\partial\theta} \pm i\cot\theta\frac{\partial}{\partial\varphi}\right],$$

show that

(a) $Y_l^m = \sqrt{\dfrac{(l+m)!}{(2l)!(l-m)!}}(L_-)^{l-m}Y_l^l$,

(b) $$Y_l^m = \sqrt{\frac{(l-m)!}{(2l)!(l+m)!}}(L_+)^{l+m}Y_l^{-l}.$$

12.6.9 In some circumstances it is desirable to replace the imaginary exponential of our spherical harmonic by sine or cosine. Morse and Feshbach (see the General References at book's end) define

$$Y_{mn}^e = P_n^m(\cos\theta)\cos m\varphi,$$
$$Y_{mn}^o = P_n^m(\cos\theta)\sin m\varphi,$$

where

$$\int_0^{2\pi}\int_0^{\pi}\left[Y_{mn}^{e \text{ or } o}(\theta,\varphi)\right]^2\sin\theta\,d\theta\,d\varphi = \frac{4\pi}{2(2n+1)}\frac{(n+m)!}{(n-m)!}, \qquad n = 1, 2, \ldots$$
$$= 4\pi \qquad \text{for } n = 0 \ (Y_{00}^o \text{ is undefined}).$$

These spherical harmonics are often named according to the patterns of their positive and negative regions on the surface of a sphere — zonal harmonics for $m = 0$, sectoral harmonics for $m = n$, and tesseral harmonics for $0 < m < n$. For Y_{mn}^e, $n = 4$, $m = 0, 2, 4$, indicate on a diagram of a hemisphere (one diagram for each spherical harmonic) the regions in which the spherical harmonic is positive.

12.6.10 A function $f(r, \theta, \varphi)$ may be expressed as a Laplace series

$$f(r,\theta,\varphi) = \sum_{l,m} a_{lm} r^l Y_l^m(\theta,\varphi).$$

With $\langle\,\rangle_{\text{sphere}}$ used to mean the average over a sphere (centered on the origin), show that

$$\left\langle f(r,\theta,\varphi)\right\rangle_{\text{sphere}} = f(0,0,0).$$

12.7 ORBITAL ANGULAR MOMENTUM OPERATORS

Now we return to the specific orbital angular momentum operators L_x, L_y, and L_z of quantum mechanics introduced in Section 4.3. Equation (4.68) becomes

$$L_z\psi_{LM}(\theta,\varphi) = M\psi_{LM}(\theta,\varphi),$$

and we want to show that

$$\psi_{LM}(\theta,\varphi) = Y_L^M(\theta,\varphi)$$

are the eigenfunctions $|LM\rangle$ of L^2 and L_z of Section 4.3 in spherical polar coordinates, the spherical harmonics. The explicit form of $L_z = -i\,\partial/\partial\varphi$ from Exercise 2.5.13 indicates that ψ_{LM} has a φ dependence of $\exp(iM\varphi)$ — with M an integer to keep ψ_{LM} single-valued. And if M is an integer, then L is an integer also.

To determine the θ dependence of $\psi_{LM}(\theta,\varphi)$, we proceed in two main steps: (1) the determination of $\psi_{LL}(\theta,\varphi)$ and (2) the development of $\psi_{LM}(\theta,\varphi)$ in terms of ψ_{LL} with the phase fixed by ψ_{L0}. Let

$$\psi_{LM}(\theta,\varphi) = \Theta_{LM}(\theta)e^{iM\varphi}. \tag{12.157}$$

From $L_+\psi_{LL} = 0$, L being the largest M, using the form of L_+ given in Exercises 2.5.14 and 12.6.7, we have

$$e^{i(L+1)\varphi}\left[\frac{d}{d\theta} - L\cot\theta\right]\Theta_{LL}(\theta) = 0, \tag{12.158}$$

and thus

$$\psi_{LL}(\theta,\varphi) = c_L \sin^L\theta e^{iL\varphi}. \tag{12.159}$$

Normalizing, we obtain

$$c_L^* c_L \int_0^{2\pi}\int_0^{\pi} \sin^{2L+1}\theta\, d\theta\, d\varphi = 1. \tag{12.160}$$

The θ integral may be evaluated as a beta function (Exercise 8.4.9) and

$$|c_L| = \sqrt{\frac{(2L+1)!!}{4\pi(2L)!!}} = \frac{\sqrt{(2L)!}}{2^L L!}\sqrt{\frac{2L+1}{4\pi}}. \tag{12.161}$$

This completes our first step.

To obtain the ψ_{LM}, $M \neq \pm L$, we return to the ladder operators. From Eqs. (4.83) and (4.84) and as shown in Exercise 12.7.2 (J_+ replaced by L_+ and J_- replaced by L_-),

$$\begin{aligned}\psi_{LM}(\theta,\varphi) &= \sqrt{\frac{(L+M)!}{(2L)!(L-M)!}}(L_-)^{L-M}\psi_{LL}(\theta,\varphi),\\ \psi_{LM}(\theta,\varphi) &= \sqrt{\frac{(L-M)!}{(2L)!(L+M)!}}(L_+)^{L+M}\psi_{L,-L}(\theta,\varphi).\end{aligned} \tag{12.162}$$

Again, note that the relative phases are set by the ladder operators. L_+ and L_- operating on $\Theta_{LM}(\theta)e^{iM\varphi}$ may be written as

$$\begin{aligned}L_+\Theta_{LM}(\theta)e^{iM\varphi} &= e^{i(M+1)\varphi}\left[\frac{d}{d\theta} - M\cot\theta\right]\Theta_{LM}(\theta)\\ &= e^{i(M+1)\varphi}\sin^{1+M}\theta\frac{d}{d(\cos\theta)}\sin^{-M}\Theta_{LM}(\theta),\\ L_-\Theta_{LM}(\theta)e^{iM\varphi} &= -e^{i(M-1)\varphi}\left[\frac{d}{d\theta} + M\cot\theta\right]\Theta_{LM}(\theta)\\ &= e^{i(M-1)\varphi}\sin^{1-M}\theta\frac{d}{d(\cos\theta)}\sin^{M}\theta\Theta_{LM}(\theta).\end{aligned} \tag{12.163}$$

Repeating these operations n times yields

$$\begin{aligned}(L_+)^n\Theta_{LM}(\theta)e^{iM\varphi} &= (-1)^n e^{i(M+n)\varphi}\sin^{n+M}\theta\frac{d^n\sin^{-M}\theta\Theta_{LM}(\theta)}{d(\cos\theta)^n},\\ (L_-)^n\Theta_{LM}(\theta)e^{iM\varphi} &= e^{i(M-n)\varphi}\sin^{n-M}\theta\frac{d^n\sin^{M}\theta\Theta_{LM}(\theta)}{d(\cos\theta)^n}.\end{aligned} \tag{12.164}$$

From Eq. (12.163),

$$\psi_{LM}(\theta,\varphi) = c_L \sqrt{\frac{(L+M)!}{(2L)!(L-M)!}} e^{iM\varphi} \sin^{-M}\theta \frac{d^{L-M}}{d(\cos\theta)^{L-M}} \sin^{2L}\theta, \tag{12.165}$$

and for $M = -L$:

$$\begin{aligned}\psi_{L,-L}(\theta,\varphi) &= \frac{c_L}{(2L)!} e^{-iL\varphi} \sin^L\theta \frac{d^{2L}}{d(\cos\theta)^{2L}} \sin^{2L}\theta \\ &= (-1)^L c_L \sin^L\theta e^{-iL\varphi}.\end{aligned} \tag{12.166}$$

Note the characteristic $(-1)^L$ phase of $\psi_{L,-L}$ relative to $\psi_{L,L}$. This $(-1)^L$ enters from

$$\sin^{2L}\theta = \left(1-x^2\right)^L = (-1)^L\left(x^2-1\right)^L. \tag{12.167}$$

Combining Eqs. (12.163), (12.163), and (12.166), we obtain

$$\psi_{LM}(\theta,\varphi) = (-1)^L c_L \sqrt{\frac{(L-M)!}{(2L)!(L+M)!}} (-1)^{L+M} e^{iM\varphi} \sin^M\theta \frac{d^{L+M}\sin^{2L}\theta}{d(\cos\theta)^{L+M}}. \tag{12.168}$$

Equations (12.165) and (12.168) agree if

$$\psi_{L0}(\theta,\varphi) = c_L \frac{1}{\sqrt{(2L)!}} \frac{d^L}{(d\cos\theta)^L} \sin^{2L}\theta. \tag{12.169}$$

Using Rodrigues' formula, Eq. (12.65), we have

$$\begin{aligned}\psi_{L0}(\theta,\varphi) &= (-1)^L c_L \frac{2^L L!}{\sqrt{(2L)!}} P_L(\cos\theta) \\ &= (-1)^L \frac{c_L}{|c_L|} \sqrt{\frac{2L+1}{4\pi}} P_L(\cos\theta).\end{aligned} \tag{12.170}$$

The last equality follows from Eq. (12.161). We now demand that $\psi_{L0}(0,0)$ be real and positive. Therefore

$$c_L = (-1)^L |c_L| = (-1)^L \frac{\sqrt{(2L)!}}{2^L L!} \sqrt{\frac{2L+1}{4\pi}}. \tag{12.171}$$

With $(-1)^L c_L/|c_L| = 1$, $\psi_{L0}(\theta,\varphi)$ in Eq. (12.170) may be identified with the spherical harmonic $Y_L^0(\theta,\varphi)$ of Section 12.6.

When we substitute the value of $(-1)^L c_L$ into Eq. (12.168),

$$\begin{aligned}\psi_{LM}(\theta,\varphi) &= \frac{\sqrt{(2L)!}}{2^L L!}\sqrt{\frac{2L+1}{4\pi}}\sqrt{\frac{(L-M)!}{(2L)!(L+M)!}}(-1)^{L+M}\\ &\quad\cdot e^{iM\varphi}\sin^M\theta\frac{d^{L+M}}{d(\cos\theta)^{L+M}}\sin^{2L}\theta\\ &= \sqrt{\frac{2L+1}{4\pi}}\sqrt{\frac{(L-M)!}{(L+M)!}}e^{iM\varphi}(-1)^M\\ &\quad\cdot\left\{\frac{1}{2^L L!}\left(1-x^2\right)^{M/2}\frac{d^{L+M}}{dx^{L+M}}\left(x^2-1\right)^L\right\}, \qquad x=\cos\theta, \qquad M\geq 0.\end{aligned} \tag{12.172}$$

The expression in the curly bracket is identified as the associated Legendre function (Eq. (12.151), and we have

$$\begin{aligned}\psi_{LM}(\theta,\varphi) &= Y_L^M(\theta,\varphi)\\ &= (-1)^M\sqrt{\frac{2L+1}{4\pi}\cdot\frac{(L-M)!}{(L+M)!}}\cdot P_L^M(\cos\theta)e^{iM\varphi}, \qquad M\geq 0,\end{aligned} \tag{12.173}$$

in complete agreement with Section 12.6. Then by Eq. (12.73c), Y_L^M for negative superscript is given by

$$Y_L^{-M}(\theta,\varphi) = (-1)^M\left[Y_L^M(\theta,\varphi)\right]^*. \tag{12.174}$$

- Our angular momentum eigenfunctions $\psi_{LM}(\theta,\varphi)$ are identified with the spherical harmonics. The phase factor $(-1)^M$ is associated with the positive values of M and is seen to be a consequence of the ladder operators.
- Our development of spherical harmonics here may be considered a portion of Lie algebra — related to group theory, Section 4.3.

Exercises

12.7.1 Using the known forms of L_+ and L_- (Exercises 2.5.14 and 12.6.7), show that

$$\int\left[Y_L^M\right]^* L_-\left(L_+Y_L^M\right)d\Omega = \int\left(L_+Y_L^M\right)^*\left(L_+Y_L^M\right)d\Omega.$$

12.7.2 Derive the relations

(a) $$\psi_{LM}(\theta,\varphi) = \sqrt{\frac{(L+M)!}{(2L)!(L-M)!}}(L_-)^{L-M}\psi_{LL}(\theta,\varphi),$$

(b) $$\psi_{LM}(\theta,\varphi) = \sqrt{\frac{(L-M)!}{(2L)!(L+M)!}}(L_+)^{L+M}\psi_{L,-L}(\theta,\varphi).$$

Hint. Equations (4.83) and (4.84) may be helpful.

12.7.3 Derive the multiple operator equations

$$(L_+)^n \Theta_{LM}(\theta) e^{iM\varphi} = (-1)^n e^{i(M+n)\varphi} \sin^{n+M}\theta \frac{d^n \sin^{-M}\theta \Theta_{LM}(\theta)}{d(\cos\theta)^n},$$

$$(L_-)^n \Theta_{LM}(\theta) e^{iM\varphi} = e^{i(M-n)\varphi} \sin^{n-M}\theta \frac{d^n \sin^{M}\theta \Theta_{LM}(\theta)}{d(\cos\theta)^n}.$$

Hint. Try mathematical induction.

12.7.4 Show, using $(L_-)^n$, that

$$Y_L^{-M}(\theta,\varphi) = (-1)^M Y_L^{*M}(\theta,\varphi).$$

12.7.5 Verify by explicit calculation that

(a) $L_+ Y_1^0(\theta,\varphi) = -\sqrt{\frac{3}{4\pi}} \sin\theta e^{i\varphi} = \sqrt{2} Y_1^1(\theta,\varphi),$

(b) $L_- Y_1^0(\theta,\varphi) = +\sqrt{\frac{3}{4\pi}} \sin\theta e^{-i\varphi} = \sqrt{2} Y_1^{-1}(\theta,\varphi).$

The signs (Condon–Shortley phase) are a consequence of the ladder operators L_+ and L_-.

12.8 THE ADDITION THEOREM FOR SPHERICAL HARMONICS

Trigonometric Identity

In the following discussion, (θ_1, φ_1) and (θ_2, φ_2) denote two different directions in our spherical coordinate system (x_1, y_1, z_1), separated by an angle γ (Fig. 12.16). The polar angles θ_1, θ_2 are measured from the z_1-axis. These angles satisfy the trigonometric identity

$$\cos\gamma = \cos\theta_1 \cos\theta_2 + \sin\theta_1 \sin\theta_2 \cos(\varphi_1 - \varphi_2), \tag{12.175}$$

which is perhaps most easily proved by vector methods (compare Chapter 1).

The addition theorem, then, asserts that

$$\boxed{P_n(\cos\gamma) = \frac{4\pi}{2n+1} \sum_{m=-n}^{n} (-1)^m Y_n^m(\theta_1,\varphi_1) Y_n^{-m}(\theta_2,\varphi_2),} \tag{12.176}$$

or equivalently,

$$P_n(\cos\gamma) = \frac{4\pi}{2n+1} \sum_{m=-n}^{n} Y_n^m(\theta_1,\varphi_1) \left[Y_n^m(\theta_2,\varphi_2)\right]^*.^{21} \tag{12.177}$$

[21] The asterisk for complex conjugation may go on either spherical harmonic.

In terms of the associated Legendre functions, the addition theorem is

$$P_n(\cos\gamma) = P_n(\cos\theta_1)P_n(\cos\theta_2) + 2\sum_{m=1}^{n}\frac{(n-m)!}{(n+m)!}P_n^m(\cos\theta_1)P_n^m(\cos\theta_2)\cos m(\varphi_1-\varphi_2). \tag{12.178}$$

Equation (12.175) is a special case of Eq. (12.178), $n = 1$.

Derivation of Addition Theorem

We now derive Eq. (12.177). Let (γ, ξ) be the angles that specify the direction (θ_1, φ_1) in a coordinate system (x_2, y_2, z_2) whose axis is aligned with (θ_2, φ_2). (Actually, the choice of the 0 azimuth angle ξ in Fig. 12.16 is irrelevant.) First, we expand $Y_n^m(\theta_1, \varphi_1)$ in spherical harmonics in the (γ, ξ) angular variables:

$$Y_n^m(\theta_1, \varphi_1) = \sum_{\sigma=-n}^{n} a_{n\sigma}^m Y_n^\sigma(\gamma, \xi). \tag{12.179}$$

We write no summation over n in Eq. (12.179) because the angular momentum n of Y_n^m is conserved (see Section 4.3); as a spherical harmonic, $Y_n^m(\theta_1, \varphi_1)$ is an eigenfunction of $\mathbf{L}^2$ with eigenvalue $n(n+1)$.

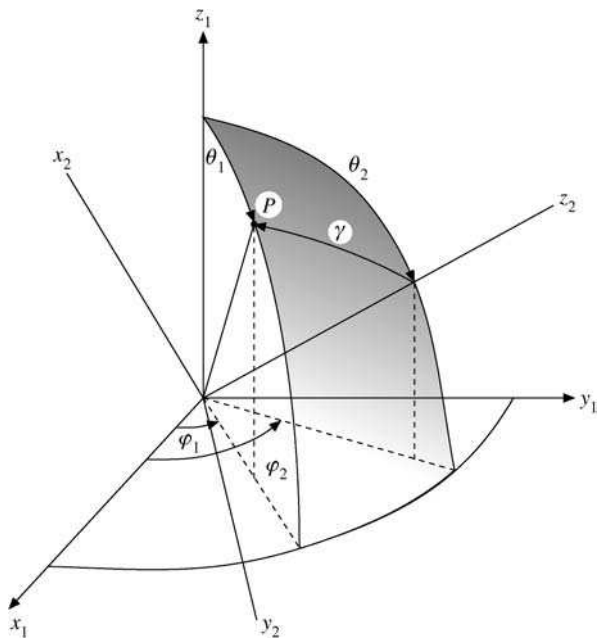

FIGURE 12.16 Two directions separated by an angle γ.

We need for our proof only the coefficient a_{n0}^m, which we get by multiplying Eq. (12.179) by $[Y_n^0(\gamma,\xi)]^*$ and integrating over the sphere:

$$a_{n0}^m = \int Y_n^m(\theta_1,\varphi_1)\big[Y_n^0(\gamma,\xi)\big]^* d\Omega_{\gamma,\xi}. \tag{12.180}$$

Similarly, we expand $P_n(\cos\gamma)$ in terms of spherical harmonics $Y_n^m(\theta_1,\varphi_1)$:

$$P_n(\cos\gamma) = \left(\frac{4\pi}{2n+1}\right)^{1/2} Y_n^0(\gamma,\xi) = \sum_{m=-n}^{n} b_{nm} Y_n^m(\theta_1,\varphi_1), \tag{12.181}$$

where the b_{nm} will, of course, depend on θ_2,φ_2, that is, on the orientation of the z_2-axis. Multiplying by $[Y_n^m(\theta_1,\varphi_1)]^*$ and integrating with respect to θ_1 and φ_1 over the sphere, we have

$$b_{nm} = \int P_n(\cos\gamma) Y_n^{m*}(\theta_1,\varphi_1)\, d\Omega_{\theta_1,\varphi_1}. \tag{12.182}$$

In terms of spherical harmonics Eq. (12.182) becomes

$$\left(\frac{4\pi}{2n+1}\right)^{1/2} \int Y_n^0(\gamma,\psi)\big[Y_n^m(\theta_1,\varphi_1)\big]^* d\Omega = b_{nm}. \tag{12.183}$$

Note that the subscripts have been dropped from the solid angle element $d\Omega$. Since the range of integration is over all solid angles, the choice of polar axis is irrelevant. Then comparing Eqs. (12.180) and (12.183), we see that

$$b_{nm}^* = a_{n0}^m \left(\frac{4\pi}{2n+1}\right)^{1/2}. \tag{12.184}$$

Now we evaluate $Y_n^m(\theta_2,\varphi_2)$ using the expansion of Eq. (12.179) and noting that the values of (γ,ξ) corresponding to $(\theta_1,\varphi_1) = (\theta_2,\varphi_2)$ are $(0,0)$. The result is

$$Y_n^m(\theta_2,\varphi_2) = a_{n0}^m Y_n^0(0,0) = a_{n0}^m \left(\frac{2n+1}{4\pi}\right)^{1/2}, \tag{12.185}$$

all terms with nonzero σ vanishing. Substituting this back into Eq. (12.184), we obtain

$$b_{nm} = \frac{4\pi}{2n+1}\big[Y_n^m(\theta_2,\varphi_2)\big]^*. \tag{12.186}$$

Finally, substituting this expression for b_{nm} into the summation, Eq. (12.181) yields Eq. (12.177), thus proving our addition theorem.

Those familiar with group theory will find a much more elegant proof of Eq. (12.177) by using the rotation group.[22] This is Exercise 4.4.5.

One application of the addition theorem is in the construction of a Green's function for the three-dimensional Laplace equation in spherical polar coordinates. If the source is on

[22] Compare M. E. Rose, *Elementary Theory of Angular Momentum*, New York: Wiley (1957).

the polar axis at the point $(r = a, \theta = 0, \varphi = 0)$, then, by Eq. (12.4a),

$$\frac{1}{R} = \frac{1}{|\mathbf{r} - \hat{\mathbf{z}}a|} = \sum_{n=0}^{\infty} P_n(\cos\gamma)\frac{a^n}{r^{n+1}}, \qquad r > a$$
$$= \sum_{n=0}^{\infty} P_n(\cos\gamma)\frac{r^n}{a^{n+1}}, \qquad r < a. \tag{12.187}$$

Rotating our coordinate system to put the source at (a, θ_2, φ_2) and the point of observation at (r, θ_1, φ_1), we obtain

$$G(r, \theta_1, \varphi_1, a, \theta_2, \varphi_2) = \frac{1}{R}$$
$$= \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \frac{4\pi}{2n+1}\left[Y_n^m(\theta_1, \varphi_1)\right]^* Y_n^m(\theta_2, \varphi_2)\frac{a^n}{r^{n+1}}, \qquad r > a,$$
$$= \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \frac{4\pi}{2n+1}\left[Y_n^m(\theta_1, \varphi_1)\right]^* Y_n^m(\theta_2, \varphi_2)\frac{r^n}{a^{n+1}}, \qquad r < a. \tag{12.188}$$

In Section 9.7 this argument is reversed to provide another derivation of the Legendre polynomial addition theorem.

Exercises

12.8.1 In proving the addition theorem, we assumed that $Y_n^k(\theta_1, \varphi_1)$ could be expanded in a series of $Y_n^m(\theta_2, \varphi_2)$, in which m varied from $-n$ to $+n$ but n was held fixed. What arguments can you develop to justify summing only over the upper index, m, and **not** over the lower index, n?
Hints. One possibility is to examine the homogeneity of the Y_n^m, that is, Y_n^m may be expressed entirely in terms of the form $\cos^{n-p}\theta \sin^p\theta$, or $x^{n-p-s}y^p z^s/r^n$. Another possibility is to examine the behavior of the Legendre equation under rotation of the coordinate system.

12.8.2 An atomic electron with angular momentum L and magnetic quantum number M has a wave function

$$\psi(r, \theta, \varphi) = f(r)Y_L^M(\theta, \varphi).$$

Show that the sum of the electron densities in a given complete shell is spherically symmetric; that is, $\sum_{M=-L}^{L} \psi^*(r, \theta, \varphi)\psi(r, \theta, \varphi)$ is independent of θ and φ.

12.8.3 The potential of an electron at point $\mathbf{r}_e$ in the field of Z protons at points $\mathbf{r}_p$ is

$$\Phi = -\frac{e^2}{4\pi\varepsilon_0}\sum_{p=1}^{Z}\frac{1}{|\mathbf{r}_e - \mathbf{r}_p|}.$$

Show that this may be written as

$$\Phi = -\frac{e^2}{4\pi\varepsilon_0 r_e}\sum_{p=1}^{Z}\sum_{L,M}\left(\frac{r_p}{r_e}\right)^L \frac{4\pi}{2L+1}\left[Y_L^M(\theta_p,\varphi_p)\right]^* Y_L^M(\theta_e,\varphi_e),$$

where $r_e > r_p$. How should Φ be written for $r_e < r_p$?

12.8.4 Two protons are **uniformly** distributed within the same spherical volume. If the coordinates of one element of charge are (r_1,θ_1,φ_1) and the coordinates of the other are (r_2,θ_2,φ_2) and r_{12} is the distance between them, the element of energy of repulsion will be given by

$$d\psi = \rho^2\frac{d\tau_1\,d\tau_2}{r_{12}} = \rho^2\frac{r_1^2\,dr_1\sin\theta_1\,d\theta_1\,d\varphi_1\,r_2^2\,dr_2\sin\theta_2\,d\theta_2\,d\varphi_2}{r_{12}}.$$

Here

$$\rho = \frac{\text{charge}}{\text{volume}} = \frac{3e}{4\pi R^3}, \qquad \text{charge density},$$
$$r_{12}^2 = r_1^2 + r_2^2 - 2r_1r_2\cos\gamma.$$

Calculate the total electrostatic energy (of repulsion) of the two protons. This calculation is used in accounting for the mass difference in "mirror" nuclei, such as O^{15} and N^{15}.

ANS. $\dfrac{6}{5}\dfrac{e^2}{R}$.

This is **double** that required to create a uniformly charged sphere because we have two separate cloud charges interacting, not one charge interacting with itself (with permutation of pairs **not** considered).

12.8.5 Each of the two $1S$ electrons in helium may be described by a hydrogenic wave function

$$\psi(\mathbf{r}) = \left(\frac{Z^3}{\pi a_0^3}\right)^{1/2} e^{-Zr/a_0}$$

in the absence of the other electron. Here Z, the atomic number, is 2. The symbol a_0 is the Bohr radius, $\hbar^2/me^2$. Find the mutual potential energy of the two electrons, given by

$$\int \psi^*(\mathbf{r}_1)\psi^*(\mathbf{r}_2)\frac{e^2}{r_{12}}\psi(\mathbf{r}_1)\psi(\mathbf{r}_2)\,d^3r_1\,d^3r_2.$$

ANS. $\dfrac{5e^2Z}{8a_0}$.

Note. $d^3r_1 = r^2\,dr_1\sin\theta_1\,d\theta_1\,d\varphi_1 \equiv d\tau_1, \qquad r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|.$

12.8.6 The probability of finding a $1S$ hydrogen electron in a volume element $r^2\,dr\sin\theta\,d\theta\,d\varphi$ is

$$\frac{1}{\pi a_0^3}\exp[-2r/a_0]r^2\,dr\sin\theta\,d\theta\,d\varphi.$$

Find the corresponding electrostatic potential. Calculate the potential from

$$V(\mathbf{r}_1) = \frac{q}{4\pi\varepsilon_0}\int \frac{\rho(\mathbf{r}_2)}{r_{12}} d^3r_2,$$

with r_1 **not** on the z-axis. Expand r_{12}. Apply the Legendre polynomial addition theorem and show that the angular dependence of $V(\mathbf{r}_1)$ drops out.

$$\text{ANS. } V(\mathbf{r}_1) = \frac{q}{4\pi\varepsilon_0}\left\{\frac{1}{2r_1}\gamma\left(3, \frac{2r_1}{a_0}\right) + \frac{1}{a_0}\Gamma\left(2, \frac{2r_1}{a_0}\right)\right\}.$$

12.8.7 A hydrogen electron in a $2P$ orbit has a charge distribution

$$\rho = \frac{q}{64\pi a_0^5} r^2 e^{-r/a_0}\sin^2\theta,$$

where a_0 is the Bohr radius, $\hbar^2/me^2$. Find the electrostatic potential corresponding to this charge distribution.

12.8.8 The electric current density produced by a $2P$ electron in a hydrogen atom is

$$\mathbf{J} = \hat{\boldsymbol{\varphi}}\frac{q\hbar}{32ma_0^5} e^{-r/a_0} r\sin\theta.$$

Using

$$\mathbf{A}(\mathbf{r}_1) = \frac{\mu_0}{4\pi}\int \frac{\mathbf{J}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d^3r_2,$$

find the magnetic vector potential produced by this hydrogen electron.
Hint. Resolve into Cartesian components. Use the addition theorem to eliminate γ, the angle included between $\mathbf{r}_1$ and $\mathbf{r}_2$.

12.8.9 (a) As a Laplace series and as an example of Eq. (1.190) (now with complex functions), show that

$$\delta(\Omega_1 - \Omega_2) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} Y_n^{m*}(\theta_2, \varphi_2) Y_n^m(\theta_1, \varphi_1).$$

(b) Show also that this **same** Dirac delta function may be written as

$$\delta(\Omega_1 - \Omega_2) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} P_n(\cos\gamma).$$

Now, if you can justify equating the summations over n **term by term**, you have an alternate derivation of the spherical harmonic addition theorem.

12.9 INTEGRALS OF PRODUCTS OF THREE SPHERICAL HARMONICS

Frequently in quantum mechanics we encounter integrals of the general form

$$\langle Y_{L_1}^{M_1} | Y_{L_2}^{M_2} | Y_{L_3}^{M_3} \rangle = \int_0^{2\pi} \int_0^{\pi} [Y_{L_1}^{M_1}]^* Y_{L_2}^{M_2} Y_{L_3}^{M_3} \sin\theta \, d\theta \, d\varphi$$

$$= \sqrt{\frac{(2L_2+1)(2L_3+1)}{4\pi(2L_1+1)}} C(L_2 L_3 L_1|000) C(L_2 L_3 L_1|M_2 M_3 M_1), \tag{12.189}$$

in which all spherical harmonics depend on θ, φ. The first factor in the integrand may come from the wave function of a final state and the third factor from an initial state, whereas the middle factor may represent an operator that is being evaluated or whose "matrix element" is being determined.

By using group theoretical methods, as in the quantum theory of angular momentum, we may give a general expression for the forms listed. The analysis involves the vector–addition or Clebsch–Gordan coefficients from Section 4.4, which are tabulated. Three general restrictions appear.

1. The integral vanishes unless the triangle condition of the L's (angular momentum) is zero, $|L_1 - L_3| \le L_2 \le L_1 + L_3$.
2. The integral vanishes unless $M_2 + M_3 = M_1$. Here we have the theoretical foundation of the vector model of atomic spectroscopy.
3. Finally, the integral vanishes unless the product $[Y_{L_1}^{M_1}]^* Y_{L_2}^{M_2} Y_{L_3}^{M_3}$ is even, that is, unless $L_1 + L_2 + L_3$ is an even integer. This is a parity conservation law.

The key to the determination of the integral in Eq. (12.189) is the expansion of the product of two spherical harmonics depending on the same angles (in contrast to the addition theorem), which are coupled by Clebsch–Gordan coefficients to angular momentum L, M, which, from its rotational transformation properties, must be proportional to $Y_L^M(\theta, \varphi)$; that is,

$$\sum_{M_1, M_2} C(L_2 L_3 L_1|M_2 M_3 M_1) Y_{L_2}^{M_2}(\theta, \varphi) Y_{L_3}^{M_3}(\theta, \varphi) \sim Y_{L_1}^{M_1}(\theta, \varphi).$$

For details we refer to Edmonds.[23]

Let us outline some of the steps of this general and powerful approach using Section 4.4. The Wigner–Eckart theorem applied to the matrix element in Eq. (12.189) yields

$$\langle Y_{L_1}^{M_1} | Y_{L_2}^{M_2} | Y_{L_3}^{M_3} \rangle = (-1)^{L_2 - L_3 + L_1} C(L_2 L_3 L_1 | M_2 M_3 M_1) \cdot \frac{\langle Y_{L_1} \| Y_{L_2} \| Y_{L_3} \rangle}{\sqrt{(2L_1+1)}}, \tag{12.190}$$

[23] E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra*, Cambridge, UK: Cambridge University Press (1951); M. E. Rose, *Elementary Theory of Angular Momentum*, New York: Wiley (1957); A. Edmonds, *Angular Momentum in Quantum Mechanics*, Princeton, NJ: Princeton University Press (1957); E. P. Wigner, *Group Theory and Its Applications to Quantum Mechanics* (translated by J. J. Griffin), New York: Academic Press (1959).

where the double bars denote the reduced matrix element, which no longer depends on the M_i. Selection rules (1) and (2) mentioned earlier follow directly from the Clebsch–Gordan coefficient in Eq. (12.190). Next we use Eq. (12.190) for $M_1 = M_2 = M_3 = 0$ in conjunction with Eq. (12.153) for $m = 0$, which yields

$$\begin{aligned}\langle Y^0_{L_1} | Y^0_{L_2} | Y^0_{L_3} \rangle &= (-1)^{L_2+L_3+L_1} \\ &\quad \frac{}{\sqrt{2L_1+1}} C(L_2L_3L_1|000) \cdot \langle Y_{L_1} \| Y_{L_2} \| Y_{L_3} \rangle \\ &= \sqrt{\frac{(2L_1+1)(2L_2+1)(2L_3+1)}{4\pi}} \\ &\quad \cdot \frac{1}{2} \cdot \int_{-1}^{1} P_{L_1}(x) P_{L_2}(x) P_{L_3}(x)\, dx, \end{aligned} \tag{12.191}$$

where $x = \cos\theta$. By elementary methods it can be shown that

$$\int_{-1}^{1} P_{L_1}(x) P_{L_2}(x) P_{L_3}(x)\, dx = \frac{2}{2L_1+1} C(L_2L_3L_1|000)^2. \tag{12.192}$$

Substituting Eq. (12.192) into (12.191) we obtain

$$\langle Y_{L_1} \| Y_{L_2} \| Y_{L_3} \rangle = (-1)^{L_2-L_3+L_1} C(L_2L_3L_1|000) \sqrt{\frac{(2L_2+1)(2L_3+1)}{4\pi}}. \tag{12.193}$$

The aforementioned parity selection rule (3) above follows from Eq. (12.193) in conjunction with the phase relation

$$C(L_2L_3L_1|-M_2,-M_3,-M_1) = (-1)^{L_2+L_3-L_1} C(L_2L_3L_1|M_2M_3M_1). \tag{12.194}$$

Note that the vector-addition coefficients are developed in terms of the Condon–Shortley phase convention,[23] in which the $(-1)^m$ of Eq. (12.153) is associated with the positive m.

It is possible to evaluate many of the commonly encountered integrals of this form with the techniques already developed. The integration over azimuth may be carried out by inspection:

$$\int_0^{2\pi} e^{-iM_1\varphi} e^{iM_2\varphi} e^{iM_3\varphi}\, d\varphi = 2\pi\, \delta_{M_2+M_3-M_1,0}. \tag{12.195}$$

Physically this corresponds to the conservation of the z component of angular momentum.

Application of Recurrence Relations

A glance at Table 12.3 will show that the θ-dependence of $Y^{M_2}_{L_2}$, that is, $P^{M_2}_{L_2}(\theta)$, can be expressed in terms of $\cos\theta$ and $\sin\theta$. However, a factor of $\cos\theta$ or $\sin\theta$ may be combined with the $Y^{M_3}_{L_3}$ factor by using the associated Legendre polynomial recurrence relations. For

instance, from Eqs. (12.92) and (12.93) we get

$$\cos\theta Y_L^M = +\left[\frac{(L-M+1)(L+M+1)}{(2L+1)(2L+3)}\right]^{1/2} Y_{L+1}^M + \left[\frac{(L-M)(L+M)}{(2L-1)(2L+1)}\right]^{1/2} Y_{L-1}^M \tag{12.196}$$

$$e^{i\varphi}\sin\theta Y_L^M = -\left[\frac{(L+M+1)(L+M+2)}{(2L+1)(2L+3)}\right]^{1/2} Y_{L+1}^{M+1} + \left[\frac{(L-M)(L-M-1)}{(2L-1)(2L+1)}\right]^{1/2} Y_{L-1}^{M+1} \tag{12.197}$$

$$e^{-i\varphi}\sin\theta Y_L^M = +\left[\frac{(L-M+1)(L-M+2)}{(2L+1)(2L+3)}\right]^{1/2} Y_{L+1}^{M-1} - \left[\frac{(L+M)(L+M-1)}{(2L-1)(2L+1)}\right]^{1/2} Y_{L-1}^{M-1}. \tag{12.198}$$

Using these equations, we obtain

$$\int Y_{L_1}^{M_1 *}\cos\theta Y_L^M\, d\Omega = \left[\frac{(L-M+1)(L+M+1)}{(2L+1)(2L+3)}\right]^{1/2} \delta_{M_1,M}\delta_{L_1,L+1} + \left[\frac{(L-M)(L+M)}{(2L-1)(2L+1)}\right]^{1/2} \delta_{M_1,M}\delta_{L_1,L-1}. \tag{12.199}$$

The occurrence of the Kronecker delta $(L_1, L \pm 1)$ is an aspect of the conservation of angular momentum. Physically, this integral arises in a consideration of ordinary atomic electromagnetic radiation (electric dipole). It leads to the familiar selection rule that transitions to an atomic level with orbital angular momentum quantum number L_1 can originate only from atomic levels with quantum numbers $L_1 - 1$ or $L_1 + 1$. The application to expressions such as

$$\text{quadrupole moment} \sim \int Y_L^{M*}(\theta,\varphi) P_2(\cos\theta) Y_L^M(\theta,\varphi)\, d\Omega$$

is more involved but perfectly straightforward.

Exercises

12.9.1 Verify

(a) $$\int Y_L^M(\theta,\varphi) Y_0^0(\theta,\varphi) Y_L^{M*}(\theta,\varphi)\, d\Omega = \frac{1}{\sqrt{4\pi}},$$

(b) $$\int Y_L^M Y_1^0 Y_{L+1}^{M*}\, d\Omega = \sqrt{\frac{3}{4\pi}}\sqrt{\frac{(L+M+1)(L-M+1)}{(2L+1)(2L+3)}},$$

(c) $$\int Y_L^M Y_1^1 Y_{L+1}^{M+1*}\,d\Omega = \sqrt{\frac{3}{8\pi}}\sqrt{\frac{(L+M+1)(L+M+2)}{(2L+1)(2L+3)}},$$

(d) $$\int Y_L^M Y_1^1 Y_{L-1}^{M+1*}\,d\Omega = -\sqrt{\frac{3}{8\pi}}\sqrt{\frac{(L-M)(L-M-1)}{(2L-1)(2L+1)}}.$$

These integrals were used in an investigation of the angular correlation of internal conversion electrons.

12.9.2 Show that

(a) $$\int_{-1}^{1} x P_L(x) P_N(x)\,dx = \begin{cases} \dfrac{2(L+1)}{(2L+1)(2L+3)}, & N = L+1, \\ \dfrac{2L}{(2L-1)(2L+1)}, & N = L-1, \end{cases}$$

(b) $$\int_{-1}^{1} x^2 P_L(x) P_N(x)\,dx = \begin{cases} \dfrac{2(L+1)(L+2)}{(2L+1)(2L+3)(2L+5)}, & N = L+2, \\ \dfrac{2(2L^2+2L-1)}{(2L-1)(2L+1)(2L+3)}, & N = L, \\ \dfrac{2L(L-1)}{(2L-3)(2L-1)(2L+1)}, & N = L-2. \end{cases}$$

12.9.3 Since $xP_n(x)$ is a polynomial (degree $n+1$), it may be represented by the Legendre series

$$xP_n(x) = \sum_{s=0}^{\infty} a_s P_s(x).$$

(a) Show that $a_s = 0$ for $s < n-1$ and $s > n+1$.
(b) Calculate a_{n-1}, a_n, and a_{n+1} and show that you have reproduced the recurrence relation, Eq. 12.17.

Note. This argument may be put in a general form to demonstrate the existence of a three-term recurrence relation for any of our complete sets of orthogonal polynomials:

$$x\varphi_n = a_{n+1}\varphi_{n+1} + a_n\varphi_n + a_{n-1}\varphi_{n-1}.$$

12.9.4 Show that Eq. (12.199) is a special case of Eq. (12.190) and derive the reduced matrix element $\langle Y_{L_1} \| Y_1 \| Y_L \rangle$.

ANS. $$\langle Y_{L_1} \| Y_1 \| Y_L \rangle = (-1)^{L_1+1-L} C(1LL_1|000)\frac{\sqrt{3(2L+1)}}{4\pi}.$$

12.10 LEGENDRE FUNCTIONS OF THE SECOND KIND

In all the analysis so far in this chapter we have been dealing with one solution of Legendre's equation, the solution $P_n(\cos\theta)$, which is regular (finite) at the two singular points of

the differential equation, $\cos\theta = \pm 1$. From the general theory of differential equations it is known that a second solution exists. We develop this second solution, Q_n, with nonnegative integer n (because Q_n in applications will occur in conjunction with P_n), by a series solution of Legendre's equation. Later a closed form will be obtained.

Series Solutions of Legendre's Equation

To solve

$$\frac{d}{dx}\left[(1-x^2)\frac{dy}{dx}\right] + n(n+1)y = 0 \tag{12.200}$$

we proceed as in Chapter 9, letting[24]

$$y = \sum_{\lambda=0}^{\infty} a_\lambda x^{k+\lambda}, \tag{12.201}$$

with

$$y' = \sum_{\lambda=0}^{\infty} (k+\lambda) a_\lambda x^{k+\lambda-1}, \tag{12.202}$$

$$y'' = \sum_{\lambda=0}^{\infty} (k+\lambda)(k+\lambda-1) a_\lambda x^{k+\lambda-2}. \tag{12.203}$$

Substitution into the original differential equation gives

$$\sum_{\lambda=0}^{\infty} (k+\lambda)(k+\lambda-1) a_\lambda x^{k+\lambda-2} + \sum_{\lambda=0}^{\infty} \big[n(n+1) - 2(k+\lambda) - (k+\lambda)(k+\lambda-1)\big] a_\lambda x^{k+\lambda} = 0. \tag{12.204}$$

The **indicial equation** is

$$k(k-1) = 0, \tag{12.205}$$

with solutions $k = 0, 1$. We try first $k = 0$ with $a_0 = 1, a_1 = 0$. Then our series is described by the recurrence relation

$$(\lambda+2)(\lambda+1)a_{\lambda+2} + \big[n(n+1) - 2\lambda - \lambda(\lambda-1)\big] a_\lambda = 0, \tag{12.206}$$

which becomes

$$a_{\lambda+2} = -\frac{(n+\lambda+1)(n-\lambda)}{(\lambda+1)(\lambda+2)} a_\lambda. \tag{12.207}$$

[24] Note that x may be replaced by the complex variable z.

Labeling this series, from Eq. (12.201), $y(x) = p_n(x)$, we have

$$p_n(x) = 1 - \frac{n(n+1)}{2!}x^2 + \frac{(n-2)n(n+1)(n+3)}{4!}x^4 + \cdots. \tag{12.208}$$

The second solution of the indicial equation, $k = 1$, with $a_0 = 0, a_1 = 1$, leads to the recurrence relation

$$a_{\lambda+2} = -\frac{(n+\lambda+2)(n-\lambda-1)}{(\lambda+2)(\lambda+3)}a_\lambda. \tag{12.209}$$

Labeling this series, from Eq. (12.201), $y(x) = q_n(x)$, we obtain

$$q_n(x) = x - \frac{(n-1)(n+2)}{3!}x^3 + \frac{(n-3)(n-1)(n+2)(n+4)}{5!}x^5 - \cdots. \tag{12.210}$$

Our general solution of Eq. (12.200), then, is

$$y_n(x) = A_n p_n(x) + B_n q_n(x), \tag{12.211}$$

provided we have convergence. From Gauss' test, Section 5.2 (see Example 5.2.4), we do not have convergence at $x = \pm 1$. To get out of this difficulty, we set the separation constant n equal to an integer (Exercise 9.5.5) and convert the infinite series into a polynomial.

For n a positive even integer (or zero), series p_n terminates, and with a proper choice of a normalizing factor (selected to obtain agreement with the definition of $P_n(x)$ in Section 12.1)

$$\begin{aligned} P_n(x) &= (-1)^{n/2}\frac{n!}{2^n[(n/2)!]^2}p_n(x) = (-1)^s\frac{(2s)!}{2^{2s}(s!)^2}p_{2s}(x) \\ &= (-1)^s\frac{(2s-1)!!}{(2s)!!}p_{2s}(x), \qquad \text{for } n = 2s. \end{aligned} \tag{12.212}$$

If n is a positive odd integer, series q_n terminates after a finite number of terms, and we write

$$\begin{aligned} P_n(x) &= (-1)^{n-1)/2}\frac{n!}{2^{n-1}\{[n-1)/2]!\}^2}q_n(x) \\ &= (-1)^s\frac{(2s+1)!}{2^{2s}(s!)^2}q_{2s+1}(x) = (-1)^s\frac{(2s+1)!!}{(2s)!!}q_{2s+1}(x), \qquad \text{for } n = 2s+1. \end{aligned} \tag{12.213}$$

Note that these expressions hold for all real values of x, $-\infty < x < \infty$, and for complex values in the finite complex plane. The constants that multiply p_n and q_n are chosen to make P_n agree with Legendre polynomials given by the generating function.

Equations (12.208) and (12.210) may still be used with $n = \nu$, not an integer, but now the series no longer terminates, and the range of convergence becomes $-1 < x < 1$. The endpoints, $x = \pm 1$, are not included.

It is sometimes convenient to reverse the order of the terms in the series. This may be done by putting

$$\begin{aligned} s &= \frac{n}{2} - \lambda \qquad \text{in the first form of } P_n(x), \qquad n \text{ even}, \\ s &= \frac{n-1}{2} - \lambda \qquad \text{in the second form of } P_n(x), \qquad n \text{ odd}, \end{aligned}$$

so that Eqs. (12.212) and (12.213) become

$$P_n(x) = \sum_{s=0}^{[n/2]} (-1)^s \frac{(2n-2s)!}{2^n s!(n-s)!(n-2s)!} x^{n-2s}, \tag{12.214}$$

where the upper limit $s = n/2$ (for n even) or $(n-1)/2$ (for n odd). This reproduces Eq. (12.8) of Section 12.1, which is obtained directly from the generating function. This agreement with Eq. (12.8) is the reason for the particular choice of normalization in Eqs. (12.212) and (12.213).

$Q_n(x)$ Functions of the Second Kind

It will be noticed that we have used only p_n for n even and q_n for n odd (because they terminated for this choice of n). We may now define a second solution of Legendre's equation (Fig. 12.17) by

$$\begin{aligned} Q_n(x) &= (-1)^{n/2} \frac{[n/2]!^2 2^n}{n!} q_n(x) \\ &= (-1)^s \frac{(2s)!!}{(2s-1)!!} q_{2s}(x), \qquad \text{for } n \text{ even}, \ n = 2s, \end{aligned} \tag{12.215}$$

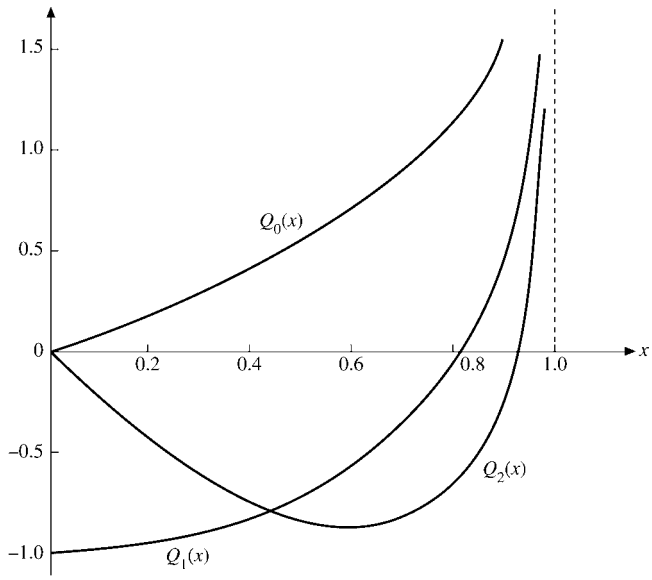

FIGURE 12.17 Second Legendre function, $Q_n(x)$, $0 \le x < 1$.

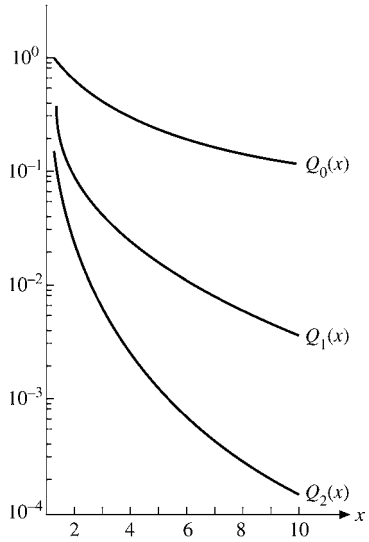

FIGURE 12.18 Second Legendre function, $Q_n(x)$, $x > 1$.

$$Q_n(x) = (-1)^{(n+1)/2}\frac{\{[(n-1)/2]!\}^2 2^{n-1}}{n!}p_n(x)$$
$$= (-1)^{s+1}\frac{(2s)!!}{(2s+1)!!}p_{2s+1}(x), \qquad \text{for } n \text{ odd},\ n = 2s+1. \tag{12.216}$$

This choice of normalizing factors forces Q_n to satisfy the same recurrence relations as P_n. This may be verified by substituting Eqs. (12.215) and (12.216) into Eqs. (12.17) and (12.26). Inspection of the (series) recurrence relations (Eqs. (12.207) and (12.209)), that is, by the Cauchy ratio test, shows that $Q_n(x)$ will converge for $-1 < x < 1$. If $|x| \geq 1$, these series forms of our second solution **diverge**. A solution in a series of negative powers of x can be developed for the region $|x| > 1$ (Fig. 12.18), but we proceed to a closed-form solution that can be used over the entire complex plane (apart from the singular points $x = \pm 1$ and with care on cut lines).

Closed-Form Solutions

Frequently, a closed form of the second solution, $Q_n(z)$, is desirable. This may be obtained by the method discussed in Section 9.6. We write

$$Q_n(z) = P_n(z)\left\{A_n + B_n\int^z \frac{dx}{(1-x^2)[P_n(x)]^2}\right\}, \tag{12.217}$$

in which the constant A_n replaces the evaluation of the integral at the arbitrary lower limit. Both constants, A_n and B_n, may be determined for special cases.

For $n = 0$, Eq. (12.217) yields

$$Q_0(z) = P_0(z)\left\{A_0 + B_0\int^z \frac{dx}{(1-x^2)[P_0(x)]^2}\right\} = A_0 + B_0\frac{1}{2}\ln\frac{1+z}{1-z}$$
$$= A_0 + B_0\left(z + \frac{z^3}{3} + \frac{z^5}{5} + \cdots + \frac{z^{2s+1}}{2s+1} + \cdots\right), \tag{12.218}$$

the last expression following from a Maclaurin expansion of the logarithm. Comparing this with the series solution (Eq. (12.210)), we obtain

$$Q_0(z) = q_0(z) = z + \frac{z^3}{3} + \frac{z^5}{5} + \cdots + \frac{z^{2s+1}}{2s+1} + \cdots, \tag{12.219}$$

we have $A_0 = 0$, $B_0 = 1$. Similar results follow for $n = 1$. We obtain

$$Q_1(z) = z\left[A_1 + B_1\int^z \frac{dx}{(1-x^2)x^2}\right]$$
$$= A_1 z + B_1 z\left(\frac{1}{2}\ln\frac{1+z}{1-z} - \frac{1}{z}\right). \tag{12.220}$$

Expanding in a power series and comparing with $Q_1(z) = -p_1(z)$, we have $A_1 = 0$, $B_1 = 1$. Therefore we may write

$$Q_0(z) = \frac{1}{2}\ln\frac{1+z}{1-z}, \qquad Q_1(z) = \frac{1}{2}z\ln\frac{1+z}{1-z} - 1, \qquad |z| < 1. \tag{12.221}$$

Perhaps the best way of determining the higher-order $Q_n(z)$ is to use the recurrence relation (Eq. (12.17)), which may be verified for both $x^2 < 1$ and $x^2 > 1$ by substituting in the series forms. This recurrence relation technique yields

$$Q_2(z) = \frac{1}{2}P_2(z)\ln\frac{1+z}{1-z} - \frac{3}{2}P_1(z). \tag{12.222}$$

Repeated application of the recurrence formula leads to

$$Q_n(z) = \frac{1}{2}P_n(z)\ln\frac{1+z}{1-z} - \frac{2n-1}{1\cdot n}P_{n-1}(z) - \frac{2n-5}{3(n-1)}P_{n-3}(z) - \cdots. \tag{12.223}$$

From the form $\ln[(1+z)/(1-z)]$ it will be seen that for real z these expressions hold in the range $-1 < x < 1$. If we wish to have closed forms valid outside this range, we need only replace

$$\ln\frac{1+x}{1-x} \qquad \text{by} \qquad \ln\frac{z+1}{z-1}.$$

When using the latter form, valid for large z, we take the line interval $-1 \le x \le 1$ as a cut line. Values of $Q_n(x)$ on the cut line are customarily assigned by the relation

$$Q_n(x) = \frac{1}{2}\big[Q_n(x+i0) + Q_n(x-i0)\big], \tag{12.224}$$

the arithmetic average of approaches from the positive imaginary side and from the negative imaginary side. Note that for $z \to x > 1$, $z - 1 \to (1 - x)e^{\pm i\pi}$. The result is that for **all** z, except on the real axis, $-1 \le x \le 1$, we have

$$Q_0(z) = \frac{1}{2}\ln\frac{z+1}{z-1}, \tag{12.225}$$

$$Q_1(z) = \frac{1}{2}z\ln\frac{z+1}{z-1} - 1, \tag{12.226}$$

and so on.

For convenient reference some special values of $Q_n(z)$ are given.

1. $Q_n(1) = \infty$, from the logarithmic term (Eq. (12.223)).
2. $Q_n(\infty) = 0$. This is best obtained from a representation of $Q_n(x)$ as a series of negative powers of x, Exercise 12.10.4.
3. $Q_n(-z) = (-1)^{n+1}Q_n(z)$. This follows from the series form. It may also be derived by using $Q_0(z)$, $Q_1(z)$ and the recurrence relation (Eq. (12.17)).
4. $Q_n(0) = 0$, for n even, by (3).
5. $$Q_n(0) = (-1)^{(n+1)/2}\frac{\{[(n-1)/2]!\}^2}{n!}2^{n-1}$$
$$= (-1)^{s+1}\frac{(2s)!!}{(2s+1)!!}, \qquad \text{for } n \text{ odd},\ n = 2s+1.$$

This last result comes from the series form (Eq. (12.216)) with $p_n(0) = 1$.

Exercises

12.10.1 Derive the parity relation for $Q_n(x)$.

12.10.2 From Eqs. (12.212) and (12.213) show that

(a) $$P_{2n}(x) = \frac{(-1)^n}{2^{2n-1}}\sum_{s=0}^{n}(-1)^s\frac{(2n+2s-1)!}{(2s)!(n+s-1)!(n-s)!}x^{2s}.$$

(b) $$P_{2n+1}(x) = \frac{(-1)^n}{2^{2n}}\sum_{s=0}^{n}(-1)^s\frac{(2n+2s+1)!}{(2s+1)!(n+s)!(n-s)!}x^{2s+1}.$$

Check the normalization by showing that one term of each series agrees with the corresponding term of Eq. (12.8).

12.10.3 Show that

(a) $$Q_{2n}(x) = (-1)^n 2^{2n}\sum_{s=0}^{n}(-1)^s\frac{(n+s)!(n-s)!}{(2s+1)!(2n-2s)!}x^{2s+1}$$
$$+ 2^{2n}\sum_{s=n+1}^{\infty}\frac{(n+s)!(2s-2n)!}{(2s+1)!(s-n)!}x^{2s+1}, \qquad |x| < 1.$$

(b) $$Q_{2n+1}(x) = (-1)^{n+1}2^{2n}\sum_{s=0}^{n}(-1)^s\frac{(n+s)!(n-s)!}{(2s)!(2n-2s+1)!}x^{2s} + 2^{2n+1}\sum_{s=n+1}^{\infty}\frac{(n+s)!(2s-2n-2)!}{(2s)!(s-n-1)!}x^{2s}, \qquad |x|<1.$$

12.10.4 (a) Starting with the assumed form

$$Q_n(x) = \sum_{\lambda=0}^{\infty} b_{-\lambda}x^{k-\lambda},$$

show that

$$Q_n(x) = b_0 x^{-n-1}\sum_{s=0}^{\infty}\frac{(n+s)!(n+2s)!(2n+1)!}{s!(n!)^2(2n+2s+1)!}x^{-2s}.$$

(b) The standard choice of b_0 is

$$b_0 = \frac{2^n(n!)^2}{(2n+1)!}.$$

Show that this choice of b_o brings this negative power-series form of $Q_n(x)$ into agreement with the closed-form solutions.

12.10.5 Verify that the Legendre functions of the second kind, $Q_n(x)$, satisfy the same recurrence relations as $P_n(x)$, both for $|x|<1$ and for $|x|>1$:

$$(2n+1)xQ_n(x) = (n+1)Q_{n+1}(x) + nQ_{n-1}(x),$$
$$(2n+1)Q_n(x) = Q'_{n+1}(x) - Q'_{n-1}(x).$$

12.10.6 (a) Using the recurrence relations, prove (independent of the Wronskian relation) that

$$n\big[P_n(x)Q_{n-1}(x) - P_{n-1}(x)Q_n(x)\big] = P_1(x)Q_0(x) - P_0(x)Q_1(x).$$

(b) By direct substitution show that the right-hand side of this equation equals 1.

12.10.7 (a) Write a subroutine that will generate $Q_n(x)$ and Q_0 through Q_{n-1} based on the recurrence relation for these Legendre functions of the second kind. Take x to be within $(-1, 1)$ — excluding the endpoints.
Hint. Take $Q_0(x)$ and $Q_1(x)$ to be known.

(b) Test your subroutine for accuracy by computing $Q_{10}(x)$ and comparing with the values tabulated in AMS-55 (for a complete reference, see Additional Readings in Chapter 8).

12.11 VECTOR SPHERICAL HARMONICS

Most of our attention in this chapter has been directed toward solving the equations of scalar fields, such as the electrostatic potential. This was done primarily because the scalar fields are easier to handle than vector fields. However, with scalar field problems under firm control, more and more attention is being paid to vector field problems.

Maxwell's equations for the vacuum, where the external current and charge densities vanish, lead to the wave (or vector Helmholtz) equation for the vector potential **A**. In a partial wave expansion of **A** in spherical polar coordinates we want to use angular eigenfunctions that are vectors. To this end we write the coordinate unit vectors $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$ in spherical notation (see Section 4.4),

$$\hat{\mathbf{e}}_{+1} = -\frac{\hat{\mathbf{x}} + i\hat{\mathbf{y}}}{\sqrt{2}}, \qquad \hat{\mathbf{e}}_0 = \hat{\mathbf{z}}, \qquad \hat{\mathbf{e}}_{-1} = \frac{\hat{\mathbf{x}} - i\hat{\mathbf{y}}}{\sqrt{2}}, \tag{12.227}$$

so that $\hat{\mathbf{e}}_m$ form a spherical tensor of rank 1. If we couple the spherical harmonics with the $\hat{\mathbf{e}}_m$ to total angular momentum **J** using the relevant Clebsch–Gordan coefficients, we are led to the vector spherical harmonics:

$$\mathbf{Y}_{JLM_J}(\theta, \varphi) = \sum_{m,M} C(L1J|MmM_J) Y_L^M(\theta, \varphi) \hat{\mathbf{e}}_m. \tag{12.228}$$

It is obvious that they obey the orthogonality relations

$$\int \mathbf{Y}^*_{JLM_J}(\theta, \varphi) \cdot \mathbf{Y}_{J'L'M'_J}(\theta, \varphi)\, d\Omega = \delta_{JJ'}\delta_{LL'}\delta_{M_J M'_M}. \tag{12.229}$$

Given J, the selection rules of angular momentum coupling tell us that L can only take on the values $J+1$, J, and $J-1$. If we look up the Clebsch–Gordan coefficients and invert Eq. (12.228) we get

$$\hat{\mathbf{r}} Y_L^M(\theta, \varphi) = -\left[\frac{L+1}{2L+1}\right]^{1/2} \mathbf{Y}_{LL+1M} + \left[\frac{L}{2L+1}\right]^{1/2} \mathbf{Y}_{LL-1M}, \tag{12.230}$$

displaying the vector character of the **Y** and the orbital angular momentum contents, $L+1$ and $L-1$, of $\hat{\mathbf{r}} Y_L^M$.

Under the parity operations (coordinate inversion) the vector spherical harmonics transform as

$$\begin{aligned} \mathbf{Y}_{LL+1M}(\theta', \varphi') &= (-1)^{L+1} \mathbf{Y}_{LL+1M}(\theta, \varphi), \\ \mathbf{Y}_{LL-1M}(\theta', \varphi') &= (-1)^{L+1} \mathbf{Y}_{LL-1M}(\theta, \varphi), \\ \mathbf{Y}_{LLM}(\theta', \varphi') &= (-1)^{L} \mathbf{Y}_{LLM}(\theta, \varphi), \end{aligned} \tag{12.231}$$

where

$$\theta' = \pi - \theta \qquad \varphi' = \pi + \varphi. \tag{12.232}$$

The vector spherical harmonics are useful in a further development of the gradient (Eq. (2.46)), divergence (Eq. (2.47)) and curl (Eq. (2.49)) operators in spherical polar co-

ordinates:

$$\nabla\big[F(r)Y_L^M(\theta,\varphi)\big] = -\left[\frac{L+1}{2L+1}\right]^{1/2}\left[\frac{d}{dr}-\frac{L}{r}\right]F\mathbf{Y}_{LL+1M} + \left[\frac{L}{2L+1}\right]^{1/2}\left[\frac{d}{dr}+\frac{L+1}{r}\right]F\mathbf{Y}_{LL-1M}, \tag{12.233}$$

$$\nabla\cdot\big[F(r)\mathbf{Y}_{LL+1M}(\theta,\varphi)\big] = -\left(\frac{L+1}{2L+1}\right)^{1/2}\left[\frac{dF}{dr}+\frac{L+2}{r}F\right]Y_L^M(\theta,\varphi), \tag{12.234}$$

$$\nabla\cdot\big[F(r)\mathbf{Y}_{LL-1M}(\theta,\varphi)\big] = \left(\frac{L}{2L+1}\right)^{1/2}\left[\frac{dF}{dr}-\frac{L-1}{r}F\right]Y_L^M(\theta,\varphi), \tag{12.235}$$

$$\nabla\cdot\big[F(r)\mathbf{Y}_{LLM}(\theta,\varphi)\big] = 0, \tag{12.236}$$

$$\nabla\times\big[F(r)\mathbf{Y}_{LL+1M}\big] = i\left[\frac{L}{2L+1}\right]^{1/2}\left[\frac{dF}{dr}+\frac{L+2}{r}F\right]\mathbf{Y}_{LLM}, \tag{12.237}$$

$$\nabla\times\big[F(r)\mathbf{Y}_{LLM}\big] = i\left(\frac{L}{2L+1}\right)^{1/2}\left[\frac{dF}{dr}-\frac{L}{r}F\right]\mathbf{Y}_{LL+1M} + i\left(\frac{L+1}{2L+1}\right)^{1/2}\left[\frac{dF}{dr}+\frac{L+1}{r}F\right]\mathbf{Y}_{LL-1M}, \tag{12.238}$$

$$\nabla\times\big[F(r)\mathbf{Y}_{LL-1M}\big] = i\left[\frac{L+1}{2L+1}\right]^{1/2}\left[\frac{dF}{dr}-\frac{L-1}{r}F\right]\mathbf{Y}_{LLM}. \tag{12.239}$$

If we substitute Eq. (12.230) into the radial component $\hat{\mathbf{r}}\partial/\partial r$ of the gradient operator, for example, we obtain both dF/dr terms in Eq. (12.233). For a complete derivation of Eqs. (12.233) to (12.239) we refer to the literature.[25] These relations play an important role in the partial wave expansion of classical and quantum electrodynamics.

The definitions of the vector spherical harmonics given here are dictated by convenience, primarily in quantum mechanical calculations, in which the angular momentum is a significant parameter. Further examples of the usefulness and power of the vector spherical harmonics will be found in Blatt and Weisskopf,[25] in Morse and Feshbach (see General References book's end), and in Jackson's *Classical Electrodynamics*, 3rd ed., New York: J. Wiley & Sons (1998), which use vector spherical harmonics in a description of multipole radiation and related electromagnetic problems.

- Vector spherical harmonics are developed from coupling L units of orbital angular momentum and 1 unit of spin angular momentum. An extension, coupling L units of orbital angular momentum and 2 units of spin angular momentum to form **tensor** spherical harmonics, is presented by Mathews.[26]

[25] E. H. Hill, Theory of vector spherical harmonics, *Am. J. Phys.* **22**: 211 (1954); also J. M. Blatt and V. Weisskopf, *Theoretical Nuclear Physics*, New York: Wiley (1952). Note that Hill assigns phases in accordance with the Condon–Shortley phase convention (Section 4.4). In Hill's notation $\mathbf{X}_{LM}=\mathbf{Y}_{LLM}$, $\mathbf{V}_{LM}=\mathbf{Y}_{LL+1M}$, $\mathbf{W}_{LM}=\mathbf{Y}_{LL-1M}$.

[26] J. Mathews, Gravitational multipole radiation, in *In Memoriam* (H.P. Robertson, ed.), Philadelphia: Society for Industrial and Applied Mathematics (1963).

- The major application of tensor spherical harmonics is in the investigation of gravitational radiation.

Exercises

12.11.1 Construct the $l = 0, m = 0$ and $l = 1, m = 0$ vector spherical harmonics.

ANS. $\mathbf{Y}_{010} = -\hat{\mathbf{r}}(4\pi)^{-1/2}$
$\mathbf{Y}_{000} = 0$
$\mathbf{Y}_{120} = -\hat{\mathbf{r}}(2\pi)^{-1/2}\cos\theta - \hat{\boldsymbol{\theta}}(8\pi)^{-1/2}\sin\theta$
$\mathbf{Y}_{110} = \hat{\boldsymbol{\varphi}} i(3/8\pi)^{1/2}\sin\theta$
$\mathbf{Y}_{100} = \hat{\mathbf{r}}(4\pi)^{-1/2}\cos\theta - \hat{\boldsymbol{\theta}}(4\pi)^{-1/2}\sin\theta.$

12.11.2 Verify that the parity of $\mathbf{Y}_{LL+1M}$ is $(-1)^{L+1}$, that of $\mathbf{Y}_{LLM}$ is $(-1)^L$, and that of $\mathbf{Y}_{LL-1M}$ is $(-1)^{L+1}$. What happened to the M-dependence of the parity?
Hint. $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\varphi}}$ have odd parity; $\hat{\boldsymbol{\theta}}$ has even parity (compare Exercise 2.5.8).

12.11.3 Verify the orthonormality of the vector spherical harmonics $\mathbf{Y}_{JLM_J}$.

12.11.4 In Jackson's *Classical Electrodynamics*, 3rd ed., (see Additional Readings of Chapter 11 for the reference) defines $\mathbf{Y}_{LLM}$ by the equation

$$\mathbf{Y}_{LLM}(\theta,\varphi) = \frac{1}{\sqrt{L(L+1)}}\mathbf{L}Y_L^M(\theta,\varphi),$$

in which the angular momentum operator $\mathbf{L}$ is given by

$$\mathbf{L} = -i(\mathbf{r}\times\boldsymbol{\nabla}).$$

Show that this definition agrees with Eq. (12.228).

12.11.5 Show that

$$\sum_{M=-L}^{L}\mathbf{Y}^*_{LLM}(\theta,\varphi)\cdot\mathbf{Y}_{LLM}(\theta,\varphi) = \frac{2L+1}{4\pi}.$$

Hint. One way is to use Exercise 12.11.4 with $\mathbf{L}$ expanded in Cartesian coordinates using the raising and lowering operators of Section 4.3.

12.11.6 Show that

$$\int \mathbf{Y}_{LLM}\cdot(\hat{\mathbf{r}}\times\mathbf{Y}_{LLM})\,d\Omega = 0.$$

The integrand represents an interference term in electromagnetic radiation that contributes to angular distributions but not to total intensity.

Additional Readings

Hobson, E. W., *The Theory of Spherical and Ellipsoidal Harmonics*. New York: Chelsea (1955). This is a very complete reference and the classic text on Legendre polynomials and all related functions.

Smythe, W. R., *Static and Dynamic Electricity*, 3rd ed. New York: McGraw-Hill (1989).

See also the references listed in Sections 4.4 and 12.9 and at the end of Chapter 13.

CHAPTER 13

MORE SPECIAL FUNCTIONS

In this chapter we shall study four sets of orthogonal polynomials, Hermite, Laguerre, and Chebyshev[1] of first and second kinds. Although these four sets are of less importance in mathematical physics than are the Bessel and Legendre functions of Chapters 11 and 12, they are used and therefore deserve attention. For example, Hermite polynomials occur in solutions of the simple harmonic oscillator of quantum mechanics and Laguerre polynomials in wave functions of the hydrogen atom. Because the general mathematical techniques duplicate those of the preceding two chapters, the development of these functions is only outlined. Detailed proofs, along the lines of Chapters 11 and 12, are left to the reader. We express these polynomials and other functions in terms of hypergeometric and confluent hypergeometric functions. To conclude the chapter, we give an introduction to Mathieu functions, which arise as solutions of ODEs and PDEs with elliptical boundary conditions.

13.1 HERMITE FUNCTIONS

Generating Functions — Hermite Polynomials

The Hermite polynomials (Fig. 13.1), $H_n(x)$, may be defined by the generating function[2]

$$\boxed{g(x,t) = e^{-t^2+2tx} = \sum_{n=0}^{\infty} H_n(x)\frac{t^n}{n!}.} \tag{13.1}$$

[1] This is the spelling choice of AMS-55 (for the complete reference see footnote 4 in Chapter 5). However, a variety of names, such as Tschebyscheff, is encountered.

[2] A derivation of this Hermite-generating function is outlined in Exercise 13.1.1.

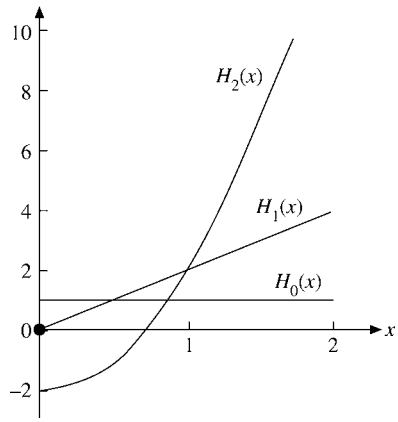

FIGURE 13.1 Hermite polynomials.

Recurrence Relations

Note the absence of a superscript, which distinguishes Hermite polynomials from the unrelated Hankel functions. From the generating function we find that the Hermite polynomials satisfy the recurrence relations

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x) \tag{13.2}$$

and

$$H_n'(x) = 2nH_{n-1}(x). \tag{13.3}$$

Equation (13.2) is obtained by differentiating the generating function with respect to t:

$$\frac{\partial g}{\partial t} = (-2t + 2x)e^{-t^2+2tx} = \sum_{n=0}^{\infty} H_{n+1}(x)\frac{t^n}{n!}$$

$$= -2\sum_{n=0}^{\infty} H_n(x)\frac{t^{n+1}}{n!} + 2x\sum_{n=0}^{\infty} H_n(x)\frac{t^n}{n!},$$

which can be rewritten as

$$\sum_{n=0}^{\infty} \frac{t^n}{n!}\left[H_{n+1}(x) - 2xH_n(x) + 2nH_{n-1}(x)\right] = 0.$$

Because each coefficient of this power series vanishes, Eq. (13.2) is established. Similarly, differentiation with respect to x leads to

$$\frac{\partial g}{\partial x} = 2te^{-t^2+2tx} = \sum_{n=0}^{\infty} H_n'(x)\frac{t^n}{n!} = 2\sum_{n=0}^{\infty} H_n(x)\frac{t^{n+1}}{n!},$$

which yields Eq. (13.3) upon shifting the summation index n in the last sum $n + 1 \to n$.

Table 13.1 Hermite Polynomials

$H_0(x) = 1$
$H_1(x) = 2x$
$H_2(x) = 4x^2 - 2$
$H_3(x) = 8x^3 - 12x$
$H_4(x) = 16x^4 - 48x^2 + 12$
$H_5(x) = 32x^5 - 160x^3 + 120x$
$H_6(x) = 64x^6 - 480x^4 + 720x^2 - 120$

The Maclaurin expansion of the generating function

$$e^{-t^2+2tx} = \sum_{n=0}^{\infty} \frac{(2tx - t^2)^n}{n!} = 1 + \left(2tx - t^2\right) + \cdots \tag{13.4}$$

gives $H_0(x) = 1$ and $H_1(x) = 2x$, and then the recursion Eq. (13.2) permits the construction of any $H_n(x)$ desired (integral n). For convenient reference the first several Hermite polynomials are listed in Table 13.1.

Special values of the Hermite polynomials follow from the generating function for $x = 0$:

$$e^{-t^2} = \sum_{n=0}^{\infty} \frac{(-t^2)^n}{n!} = \sum_{n=0}^{\infty} H_n(0) \frac{t^n}{n!},$$

that is,

$$H_{2n}(0) = (-1)^n \frac{(2n)!}{n!}, \qquad H_{2n+1}(0) = 0, \qquad n = 0, 1, \ldots. \tag{13.5}$$

We also obtain from the generating function the important parity relation

$$H_n(x) = (-1)^n H_n(-x) \tag{13.6}$$

by noting that Eq. (13.1) yields

$$g(-x, -t) = \sum_{n=0}^{\infty} H_n(-x) \frac{(-t)^n}{n!} = g(x, t) = \sum_{n=0}^{\infty} H_n(x) \frac{t^n}{n!}.$$

Alternate Representations

The Rodrigues representation of $H_n(x)$ is

$$\boxed{H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2}.} \tag{13.7}$$

Let us show this using mathematical induction as follows.

Example 13.1.1 RODRIGUES REPRESENTATION

We rewrite the generating function as $g(x,t) = e^{x^2}e^{-(t-x)^2}$ and note that

$$\frac{\partial}{\partial t}e^{-(t-x)^2} = -\frac{\partial}{\partial x}e^{-(t-x)^2}.$$

This yields

$$\left.\frac{\partial g}{\partial t}\right|_{t=0} = (2x-2t)g\Big|_{t=0} = 2x = H_1(x) = -e^{x^2}\,\frac{d}{dx}e^{-x^2},$$

which is the initial $n=1$ case. Assuming the case n of Eq. (13.7) as valid, we now use the operator identity $\frac{d}{dx}e^{x^2} = 2xe^{x^2} + e^{x^2}\frac{d}{dx}$ in

$$\begin{aligned}(-1)^{n+1}e^{x^2}\frac{d^{n+1}}{dx^{n+1}}e^{-x^2} &= (-1)^{n+1}\left[\frac{d}{dx}e^{x^2} - 2xe^{x^2}\right]\frac{d^n}{dx^n}e^{-x^2}\\ &= -\frac{d}{dx}H_n(x) + 2xH_n(x) = H_{n+1}(x)\end{aligned}$$

to establish the $n+1$ case, with the last equality following from Eqs.(13.2) and (13.3).

More directly, differentiation of the generating function n times with respect to t and then setting t equal to zero yields

$$H_n(x) = \frac{\partial^n}{\partial t^n}\left(e^{-t^2+2tx}\right)\Big|_{t=0} = (-1)^n e^{x^2}\frac{\partial^n}{\partial x^n}e^{-(t-x)^2}\Big|_{t=0} = (-1)^n e^{x^2}\frac{d^n}{dx^n}e^{-x^2}.$$

■

A second representation may be obtained by using the calculus of residues (Section 7.1). If we multiply Eq. (13.1) by t^{-m-1} and integrate around the origin in the complex t-plane, only the term with $H_m(x)$ will survive:

$$H_m(x) = \frac{m!}{2\pi i}\oint t^{-m-1}e^{-t^2+2tx}\,dt. \tag{13.8}$$

Also, from the Maclaurin expansion, Eq. (13.4), we can derive our Hermite polynomial $H_n(x)$ in series form: Using the binomial expansion of $(2x-t)^\nu$ and the index $N = s+\nu$,

$$\begin{aligned}e^{-t^2+2tx} &= \sum_{\nu=0}^{\infty}\frac{t^\nu}{\nu!}(2x-t)^\nu = \sum_{\nu=0}^{\infty}\frac{t^\nu}{\nu!}\sum_{s=0}^{\nu}\binom{\nu}{s}(2x)^{\nu-s}(-t)^s\\ &= \sum_{N=0}^{\infty}\frac{t^N}{N!}\sum_{s=0}^{[N/2]}(2x)^{N-2s}(-1)^s\frac{N!}{(N-s)!}\binom{N-s}{s},\end{aligned}$$

where $[N/2]$ is the largest integer less than or equal to $N/2$. Writing the binomial coefficient in terms of factorials and using Eq. (13.1) we obtain

$$H_N(x) = \sum_{s=0}^{[N/2]}(2x)^{N-2s}(-1)^s\frac{N!}{s!(N-2s)!}.$$

More explicitly, replacing $N \to n$, we have

$$\begin{aligned} H_n(x) &= (2x)^n - \frac{2n!}{(n-2)!2!}(2x)^{n-2} + \frac{4n!}{(n-4)!4!}(2x)^{n-4}1\cdot 3\cdots \\ &= \sum_{s=0}^{[n/2]} (-2)^s (2x)^{n-2s} \binom{n}{2s} 1\cdot 3\cdot 5\cdots(2s-1) \\ &= \sum_{s=0}^{[n/2]} (-1)^s (2x)^{n-2s} \frac{n!}{(n-2s)!s!}. \end{aligned} \tag{13.9}$$

This series terminates for integral n and yields our Hermite polynomial.

Orthogonality

If we substitute the recursion Eq. (13.3) into Eq. (13.2) we can eliminate the index $n-1$, obtaining

$$H_{n+1}(x) = 2xH_n(x) - H_n'(x),$$

which was used already in Example 13.1.1. If we differentiate this recursion relation and substitute Eq. (13.3) for the index $n+1$ we find

$$H_{n+1}'(x) = 2(n+1)H_n(x) = 2H_n(x) + 2xH_n'(x) - H_n''(x),$$

which can be rearranged to the second-order ODE for Hermite polynomials. Thus, the recurrence relations (Eqs. (13.2) and (13.3)) lead to the second-order ODE

$$\boxed{H_n''(x) - 2xH_n'(x) + 2nH_n(x) = 0,} \tag{13.10}$$

which is clearly **not** self-adjoint.

To put the ODE in self-adjoint form, following Section 10.1, we multiply by $\exp(-x^2)$, Exercise 10.1.2. This leads to the orthogonality integral

$$\int_{-\infty}^{\infty} H_m(x)H_n(x)e^{-x^2}\,dx = 0, \qquad m \neq n, \tag{13.11}$$

with the weighting function $\exp(-x^2)$, a consequence of putting the ODE into self-adjoint form. The interval $(-\infty, \infty)$ is chosen to obtain the Hermitian operator boundary conditions, Section 10.1. It is sometimes convenient to absorb the weighting function into the Hermite polynomials. We may define

$$\varphi_n(x) = e^{-x^2/2}H_n(x), \tag{13.12}$$

with $\varphi_n(x)$ no longer a polynomial.

Substitution into Eq. (13.10) yields the differential equation for $\varphi_n(x)$,

$$\varphi_n''(x) + \left(2n + 1 - x^2\right)\varphi_n(x) = 0. \tag{13.13}$$

This is the differential equation for a quantum mechanical, simple harmonic oscillator, which is perhaps the most important physics application of the Hermite polynomials.

Equation (13.13) is self-adjoint, and the solutions $\varphi_n(x)$ are orthogonal for the interval $(-\infty < x < \infty)$ with a unit weighting function.

The problem of normalizing these functions remains. Proceeding as in Section 12.3, we multiply Eq. (13.1) by itself and then by e^{-x^2}. This yields

$$e^{-x^2}e^{-s^2+2sx}e^{-t^2+2tx} = \sum_{m,n=0}^{\infty} e^{-x^2}H_m(x)H_n(x)\frac{s^m t^n}{m!n!}.$$

When we integrate this relation over x from $-\infty$ to $+\infty$, the cross terms of the double sum drop out because of the orthogonality property:[3]

$$\begin{aligned}\sum_{n=0}^{\infty}\frac{(st)^n}{n!n!}\int_{-\infty}^{\infty}e^{-x^2}\big[H_n(x)\big]^2\,dx &= \int_{-\infty}^{\infty}e^{-x^2-s^2+2sx-t^2+2tx}\,dx\\ &= \int_{-\infty}^{\infty}e^{-(x-s-t)^2}e^{2st}\,dx\\ &= \pi^{1/2}e^{2st} = \pi^{1/2}\sum_{n=0}^{\infty}\frac{2^n(st)^n}{n!},\end{aligned} \tag{13.14}$$

using Eqs. (8.6) and (8.8). By equating coefficients of like powers of st, we obtain

$$\int_{-\infty}^{\infty}e^{-x^2}\big[H_n(x)\big]^2\,dx = 2^n\pi^{1/2}n!. \tag{13.15}$$

Quantum Mechanical Simple Harmonic Oscillator

The following development of Hermite polynomials via simple harmonic oscillator wave functions $\phi_n(x)$ is analogous to the use of the raising and lowering operators for angular momentum operators presented in Section 4.3. This means that we derive the eigenvalues $n + 1/2$ and eigenfunctions (the $H_n(x)$) without assuming the development that led to Eq. (13.13). The key aspect of the eigenvalue Eq. (13.13), $(\frac{d^2}{dx^2} - x^2)\varphi_n(x) = -(2n+1)\varphi_n(x)$, is that the Hamiltonian

$$-2\mathcal{H} \equiv \frac{d^2}{dx^2} - x^2 = \left(\frac{d}{dx} - x\right)\left(\frac{d}{dx} + x\right) + \left[x, \frac{d}{dx}\right] \tag{13.16}$$

almost factorizes. Using naively $a^2 - b^2 = (a-b)(a+b)$, the basic commutator $[p_x, x] = \hbar/i$ of quantum mechanics (with momentum $p_x = (\hbar/i)d/dx$) enters as a correction in Eq. (13.16). (Because p_x is Hermitian, d/dx is anti-Hermitian, $(d/dx)^\dagger = -d/dx$.) This commutator can be evaluated as follows. Imagine the differential operator d/dx acts on a wave function $\varphi(x)$ to the right, as in Eq. (13.13), so

$$\frac{d}{dx}(x\varphi) = x\frac{d}{dx}\varphi + \varphi, \tag{13.17}$$

[3]The cross terms $(m \neq n)$ may be left in, if desired. Then, when the coefficients of $s^\alpha t^\beta$ are equated, the orthogonality will be apparent.

by the product rule. Dropping the wave function φ from Eq. (13.17), we rewrite Eq. (13.17) as

$$\frac{d}{dx}x - x\frac{d}{dx} \equiv \left[\frac{d}{dx}, x\right] = 1, \tag{13.18}$$

a constant, and then verify Eq. (13.16) directly by expanding the product of operators. The product form of Eq. (13.16), up to the constant commutator, suggests introducing the non-Hermitian operators

$$\boxed{\hat{a}^\dagger \equiv \frac{1}{\sqrt{2}}\left(x - \frac{d}{dx}\right), \qquad \hat{a} \equiv \frac{1}{\sqrt{2}}\left(x + \frac{d}{dx}\right),} \tag{13.19}$$

with $(\hat{a})^\dagger = \hat{a}^\dagger$, which are adjoints of each other. They obey the commutation relations

$$\left[\hat{a}, \hat{a}^\dagger\right] = \left[\frac{d}{dx}, x\right] = 1, \qquad [\hat{a}, \hat{a}] = 0 = \left[\hat{a}^\dagger, \hat{a}^\dagger\right], \tag{13.20}$$

which are characteristic of these operators and straightforward to derive from Eq. (13.18) and

$$\left[\frac{d}{dx}, \frac{d}{dx}\right] = 0 = [x, x] \qquad \text{and} \qquad \left[x, \frac{d}{dx}\right] = -\left[\frac{d}{dx}, x\right].$$

Returning to Eq. (13.16) and using Eq. (13.19) we rewrite the Hamiltonian as

$$\mathcal{H} = \hat{a}^\dagger\hat{a} + \frac{1}{2} = \hat{a}^\dagger\hat{a} + \frac{1}{2}\left(\hat{a}^\dagger\hat{a} + \hat{a}\hat{a}^\dagger\right) = \frac{1}{2}\left(\hat{a}^\dagger\hat{a} + \hat{a}\hat{a}^\dagger\right) \tag{13.21}$$

and introduce the Hermitian **number operator** $N = \hat{a}^\dagger\hat{a}$ so that $\mathcal{H} = N + 1/2$. Let $|n\rangle$ be an eigenfunction of $\mathcal{H}$,

$$\mathcal{H}|n\rangle = \lambda_n|n\rangle,$$

whose eigenvalue λ_n is unknown at this point. Now we prove the key property that N has nonnegative integer eigenvalues

$$\boxed{N|n\rangle = \left(\lambda_n - \frac{1}{2}\right)|n\rangle = n|n\rangle, \qquad n = 0, 1, 2\ldots,} \tag{13.22}$$

that is, $\lambda_n = n + 1/2$. Since $\hat{a}|n\rangle$ is complex conjugate to $\langle n|\hat{a}^\dagger$, the normalization integral $\langle n|\hat{a}^\dagger\hat{a}|n\rangle \geq 0$ and is finite. From

$$\left(\hat{a}|n\rangle\right)^\dagger\hat{a}|n\rangle = \langle n|\hat{a}^\dagger\hat{a}|n\rangle = \left(\lambda_n - \frac{1}{2}\right) \geq 0 \tag{13.23}$$

we see that N has nonnegative eigenvalues.

We now show that if $\hat{a}|n\rangle$ is nonzero it is an eigenfunction with eigenvalue $\lambda_{n-1} = \lambda_n - 1$. After normalizing $\hat{a}|n\rangle$, this state is designated $|n-1\rangle$. This is proved by the commutation relations

$$\boxed{\left[N, \hat{a}^\dagger\right] = \hat{a}^\dagger, \qquad [N, \hat{a}] = -\hat{a},} \tag{13.24}$$

which follow from Eq. (13.20). These commutation relations characterize N as the number operator. To see this, we determine the eigenvalue of N for the states $\hat{a}^\dagger|n\rangle$ and $\hat{a}|n\rangle$. Using $\hat{a}\hat{a}^\dagger = N + [\hat{a}, \hat{a}^\dagger] = N + 1$, we find that

$$\begin{aligned} N\big(\hat{a}^\dagger|n\rangle\big) &= \hat{a}^\dagger\big(\hat{a}\hat{a}^\dagger\big)|n\rangle = \hat{a}^\dagger\big(\big[\hat{a}, \hat{a}^\dagger\big] + N\big)|n\rangle \\ &= \hat{a}^\dagger(N+1)|n\rangle = \left(\lambda_n + \frac{1}{2}\right)\hat{a}^\dagger|n\rangle = (n+1)\hat{a}^\dagger|n\rangle, \qquad (13.25) \\ N\big(\hat{a}|n\rangle\big) &= \big(\hat{a}\hat{a}^\dagger - 1\big)\hat{a}|n\rangle = \hat{a}(N-1)|n\rangle = (n-1)\hat{a}|n\rangle. \end{aligned}$$

In other words, N acting on $\hat{a}^\dagger|n\rangle$ shows that $\hat{a}^\dagger$ has raised the eigenvalue n corresponding to $|n\rangle$ by one unit, whence its name **raising**, or **creation**, **operator**. Applying $\hat{a}^\dagger$ repeatedly, we can reach all higher states. There is no upper limit to the sequence of eigenvalues. Similarly, $\hat{a}$ lowers the eigenvalue n by one unit; hence it is a **lowering** (or **annihilation**) **operator** because. Therefore,

$$\hat{a}^\dagger|n\rangle \sim |n+1\rangle, \qquad \hat{a}|n\rangle \sim |n-1\rangle. \qquad (13.26)$$

Applying $\hat{a}$ repeatedly, we can reach the lowest, or ground, state $|0\rangle$ with eigenvalue λ_0. We cannot step lower because $\lambda_0 \geq 1/2$. Therefore $\hat{a}|0\rangle \equiv 0$, suggesting we construct $\psi_0 = |0\rangle$ from the (factored) **first-order** ODE

$$\sqrt{2}\hat{a}\psi_0 = \left(\frac{d}{dx} + x\right)\psi_0 = 0. \qquad (13.27)$$

Integrating

$$\frac{\psi_0'}{\psi_0} = -x, \qquad (13.28)$$

we obtain

$$\ln\psi_0 = -\frac{1}{2}x^2 + \ln c_0, \qquad (13.29)$$

where c_0 is an integration constant. The solution,

$$\psi_0(x) = c_0 e^{-x^2/2}, \qquad (13.30)$$

is a Gaussian that can be normalized, with $c_0 = \pi^{-1/4}$ using the error integral, Eqs. (8.6) and (8.8). Substituting ψ_0 into Eq. (13.13) we find

$$\mathcal{H}|0\rangle = \left(\hat{a}^\dagger\hat{a} + \frac{1}{2}\right)|0\rangle = \frac{1}{2}|0\rangle, \qquad (13.31)$$

so its energy eigenvalue is $\lambda_0 = 1/2$ and its number eigenvalue is $n = 0$, confirming the notation $|0\rangle$. Applying $\hat{a}^\dagger$ repeatedly to $\psi_0 = |0\rangle$, all other eigenvalues are confirmed to be $\lambda_n = n + 1/2$, proving Eq. (13.13). The normalizations needed for Eq. (13.26) follow from Eqs. (13.25) and (13.23) and

$$\langle n|\hat{a}\hat{a}^\dagger|n\rangle = \langle n|\hat{a}^\dagger\hat{a} + 1|n\rangle = n + 1, \qquad (13.32)$$

showing

$$\boxed{\sqrt{n+1}|n+1\rangle = \hat{a}^\dagger|n\rangle, \qquad \sqrt{n}|n-1\rangle = \hat{a}|n\rangle.} \tag{13.33}$$

Thus, the excited-state wave functions, ψ_1, ψ_2, and so on, are generated by the raising operator

$$|1\rangle = \hat{a}^\dagger|0\rangle = \frac{1}{\sqrt{2}}\left(x - \frac{d}{dx}\right)\psi_0(x) = \frac{x\sqrt{2}}{\pi^{1/4}}e^{-x^2/2}, \tag{13.34}$$

yielding (and leading to upcoming Eq. (13.38))

$$\psi_n(x) = N_n H_n(x) e^{-x^2/2}, \qquad N_n \equiv \pi^{-1/4}\left(2^n n!\right)^{-1/2}, \tag{13.35}$$

where H_n are the Hermite polynomials (Fig. 13.2).

As shown, the Hermite polynomials are used in analyzing the quantum mechanical simple harmonic oscillator. For a potential energy $V = \frac{1}{2}Kz^2 = \frac{1}{2}m\omega^2 z^2$ (force $\mathbf{F} = -\nabla V = -Kz\hat{\mathbf{z}}$), the Schrödinger wave equation is

$$-\frac{\hbar^2}{2m}\nabla^2\Psi(z) + \frac{1}{2}Kz^2\Psi(z) = E\Psi(z). \tag{13.36}$$

Our oscillating particle has mass m and total energy E. By use of the abbreviations

$$\begin{gathered} x = \alpha z \qquad \text{with} \qquad \alpha^4 = \frac{mK}{\hbar^2} = \frac{m^2\omega^2}{\hbar^2}, \\ \lambda = \frac{2E}{\hbar}\left(\frac{m}{K}\right)^{1/2} = \frac{2E}{\hbar\omega}, \end{gathered} \tag{13.37}$$

in which ω is the angular frequency of the corresponding classical oscillator, Eq. (13.36) becomes (with $\Psi(z) = \Psi(x/\alpha) = \psi(x)$)

$$\frac{d^2\psi(x)}{dx^2} + \left(\lambda - x^2\right)\psi(x) = 0. \tag{13.38}$$

This is Eq. (13.13) with $\lambda = 2n + 1$. Hence (Fig. 13.2),

$$\boxed{\psi_n(x) = 2^{-n/2}\pi^{-1/4}(n!)^{-1/2}e^{-x^2/2}H_n(x),} \qquad \text{(normalized)}. \tag{13.39}$$

Alternatively, the requirement that n be an integer is dictated by the boundary conditions of the quantum mechanical system,

$$\lim_{z\to\pm\infty}\Psi(z) = 0.$$

Specifically, if $n \to \nu$, not an integer, a power-series solution of Eq. (13.13) (Exercise 9.5.6) shows that $H_\nu(x)$ will behave as $x^\nu e^{x^2}$ for large x. The functions $\psi_\nu(x)$ and $\Psi_\nu(z)$ will therefore blow up at infinity, and it will be impossible to normalize the wave function $\Psi(z)$. With this requirement, the energy E becomes

$$\boxed{E = \left(n + \frac{1}{2}\right)\hbar\omega.} \tag{13.40}$$

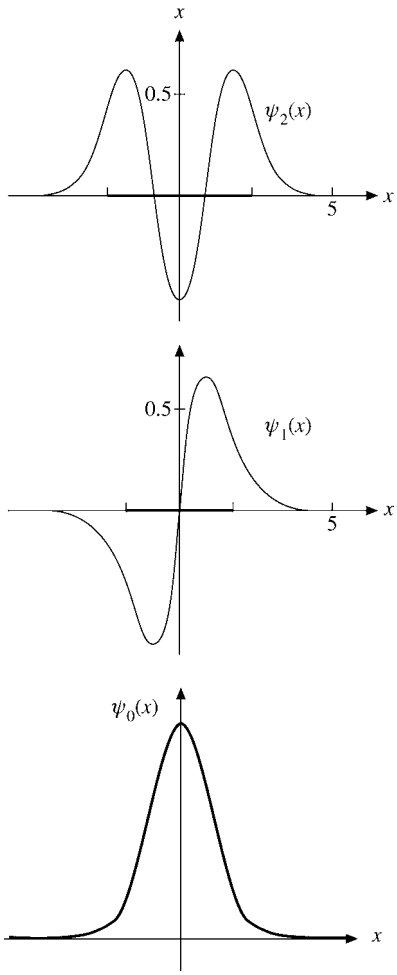

FIGURE 13.2 Quantum mechanical oscillator wave functions: The heavy bar on the x-axis indicates the allowed range of the classical oscillator with the same total energy.

As n ranges over integral values ($n \geq 0$), we see that the energy is quantized and that there is a minimum or zero point energy

$$E_{\min} = \frac{1}{2}\hbar\omega. \tag{13.41}$$

This zero point energy is an aspect of the uncertainty principle, a genuine quantum phenomenon.

In quantum mechanical problems, particularly in molecular spectroscopy, a number of integrals of the form

$$\int_{-\infty}^{\infty} x^r e^{-x^2} H_n(x) H_m(x)\, dx$$

are needed. Examples for $r = 1$ and $r = 2$ (with $n = m$) are included in the exercises at the end of this section. A large number of other examples are contained in Wilson, Decius, and Cross.[4]

In the dynamics and spectroscopy of molecules in the Born–Oppenheimer approximation, the motion of a molecule is separated into electronic, vibrational and rotational motion. Each vibrating atom contributes to a matrix element two Hermite polynomials, its initial state and another one to its final state. Thus, integrals of products of Hermite polynomials are needed.

Example 13.1.2 THREEFOLD HERMITE FORMULA

We want to calculate the following integral involving $m = 3$ Hermite polynomials:

$$\boxed{I_3 \equiv \int_{-\infty}^{\infty} e^{-x^2} H_{N_1}(x) H_{N_2}(x) H_{N_3}(x)\, dx,} \tag{13.42}$$

where $N_i \geq 0$ are integers. The formula (due to E. C. Titchmarsh, *J. Lond. Math. Soc.* **23**: 15 (1948), see Gradshteyn and Ryzhik, p. 838, in the Additional Readings) generalizes the $m = 2$ case needed for the orthogonality and normalization of Hermite polynomials. To derive it, we start with the product of three generating functions of Hermite polynomials, multiply by e^{-x^2}, and integrate over x in order to generate I_3:

$$\begin{aligned} Z_3 \equiv \int_{-\infty}^{\infty} e^{-x^2} \prod_{j=1}^{3} e^{2xt_j - t_j^2}\, dx &= \int_{-\infty}^{\infty} e^{-(\sum_{j=1}^{3} t_j - x)^2 + 2(t_1t_2 + t_1t_3 + t_2t_3)}\, dx \\ &= \sqrt{\pi} e^{2(t_1t_2 + t_1t_3 + t_2t_3)}. \end{aligned} \tag{13.43}$$

The last equality follows from substituting $y = x - \sum_j t_j$ and using the error integral $\int_{-\infty}^{\infty} e^{-y^2} dy = \sqrt{\pi}$, Eqs. (8.6) and (8.8). Expanding the generating functions in terms of Hermite polynomials we obtain

$$\begin{aligned} Z_3 &= \sum_{N_1, N_2, N_3 = 0}^{\infty} \frac{t_1^{N_1} t_2^{N_2} t_3^{N_3}}{N_1! N_2! N_3!} \int_{-\infty}^{\infty} e^{-x^2} H_{N_1}(x) H_{N_2}(x) H_{N_3}(x)\, dx \\ &= \sqrt{\pi} \sum_{N=0}^{\infty} \frac{2^N}{N!} (t_1t_2 + t_1t_3 + t_2t_3)^N \\ &= \sqrt{\pi} \sum_{N=0}^{\infty} \frac{2^N}{N!} \sum_{0 \leq n_i \leq N,\ \sum_i n_i = N} \frac{N!}{n_1! n_2! n_3!} (t_1t_2)^{n_1} (t_1t_3)^{n_2} (t_2t_3)^{n_3}, \end{aligned}$$

[4] E. B. Wilson, Jr., J. C. Decius, and P. C. Cross, *Molecular Vibrations*, New York: McGraw-Hill (1955), reprinted Dover (1980).

using the polynomial expansion

$$\left(\sum_{j=1}^{m} a_j\right)^N = \sum_{0\le n_i\le m} \frac{N!}{n_1!\cdots n_m!} a_1^{n_1}\cdots a_m^{n_m}.$$

The powers of the foregoing $t_j t_k$ become

$$(t_1t_2)^{n_1}(t_1t_3)^{n_2}(t_2t_3)^{n_3} = t_1^{N_1}t_2^{N_2}t_3^{N_3};$$

$$N_1 = n_1 + n_2, \qquad N_2 = n_1 + n_3, \qquad N_3 = n_2 + n_3.$$

That is, from

$$2N = 2(n_1 + n_2 + n_3) = N_1 + N_2 + N_3$$

there follows

$$2N = 2n_1 + 2N_3 = 2n_2 + 2N_2 = 2n_3 + 2N_1,$$

so we obtain

$$n_1 = N - N_3, \qquad n_2 = N - N_2, \qquad n_3 = N - N_1.$$

The n_i are all fixed (making this case special and easy) because the N_i are fixed, and $2N = \sum_{i=1}^{3} N_i$, with $N \ge 0$ an integer by parity. Hence, upon comparing the foregoing like $t_1t_2t_3$ powers,

$$\boxed{I_3 = \frac{\sqrt{\pi}\,2^N N_1!N_2!N_3!}{(N-N_1)!(N-N_2)!(N-N_3)!},} \tag{13.44}$$

which is the desired formula. If we **order** $N_1 \ge N_2 \ge N_3 \ge 0$, then $n_1 \ge n_2 \ge n_3 \ge 0$ follows, being equivalent to $N - N_3 \ge N - N_2 \ge N - N_1 \ge 0$, which occur in the denominators of the factorials of I_3. ■

Example 13.1.3 DIRECT EXPANSION OF PRODUCTS OF HERMITE POLYNOMIALS

In an alternative approach, we now start again from the generating function identity

$$\sum_{N_1,N_2=0}^{\infty} H_{N_1}(x)H_{N_2}(x)\frac{t_1^{N_1}}{N_1!}\frac{t_2^{N_2}}{N_2!} = e^{2x(t_1+t_2)-t_1^2-t_2^2} = e^{2x(t_1+t_2)-(t_1+t_2)^2}\cdot e^{2t_1t_2}$$

$$= \sum_{N=0}^{\infty} H_N(x)\frac{(t_1+t_2)^N}{N!}\sum_{\nu=0}^{\infty}\frac{(2t_1t_2)^\nu}{\nu!}.$$

Using the binomial expansion and then comparing like powers of $t_1 t_2$ we extract an identity due to E. Feldheim (*J. Lond. Math. Soc.* **13**: 22 (1938)):

$$H_{N_1}(x)H_{N_2}(x) = \sum_{\nu=0}^{\min(N_1,N_2)} H_{N_1+N_2-2\nu}\frac{N_1!N_2!2^\nu}{\nu!(N_1+N_2-2\nu)!}\binom{N_1+N_2-2\nu}{N_1-\nu}$$

$$= \sum_{0\le\nu\le\min(N_1,N_2)} H_{N_1+N_2-2\nu}2^\nu\nu!\binom{N_1}{\nu}\binom{N_2}{\nu}. \tag{13.45}$$

For $\nu = 0$ the coefficient of $H_{N_1+N_2}$ is obviously unity. Special cases, such as

$$H_1^2 = H_2+2, \qquad H_1H_2 = H_3+4H_1, \qquad H_2^2 = H_4+8H_2+8, \qquad H_1H_3 = H_4+6H_2,$$

can be derived from Table 13.1 and agree with the general twofold product formula.

This compact formula can be generalized to products of m Hermite polynomials, and this in turn yields a new closed form result for the integral I_m.

Let us begin with a new result for I_4 containing a product of four Hermite polynomials. Inserting the Feldheim identity for $H_{N_1}H_{N_2}$ and $H_{N_3}H_{N_4}$ and using orthogonality

$$\int_{-\infty}^{\infty} e^{-x^2}H_{N_1}H_{N_2}\,dx = \sqrt{\pi}2^{N_1}N_1!\delta_{N_1N_2}$$

for the remaining product of two Hermite polynomials yields

$$\begin{aligned} I_4 &= \int_{-\infty}^{\infty} e^{-x^2}H_{N_1}H_{N_2}H_{N_3}H_{N_4}\,dx \\ &= \sum_{0\le\mu\le\min(N_1,N_2);0\le\nu\le\min(N_3,N_4)} 2^\mu\mu! \\ &\quad\cdot\binom{N_1}{\mu}\binom{N_2}{\mu}2^\nu\nu!\binom{N_3}{\nu}\binom{N_4}{\nu}\int_{-\infty}^{\infty} e^{-x^2}H_{N_1+N_2-2\mu}H_{N_3+N_4-2\nu}\,dx \\ &= \sum_{\nu=0}^{N_4}\frac{\sqrt{\pi}2^M(N_3+N_4-2\nu)!N_1!N_2!N_3!N_4!}{(M-N_3-N_4-\nu)!(M-N_1+\nu)!(M-N_2+\nu)!(N_3-\nu)!(N_4-\nu)!\nu!}. \end{aligned} \tag{13.46}$$

Here we use the notation $M = (N_1+N_2+N_3+N_4)/2$ and write the binomial coefficients explicitly, so

$$\begin{aligned} \tfrac{1}{2}(N_1+N_2-N_3-N_4) &= M-N_3-N_4, \\ \tfrac{1}{2}(N_1-N_2+N_3+N_4) &= M-N_2, \\ \tfrac{1}{2}(N_2-N_1+N_3+N_4) &= M-N_1. \end{aligned}$$

From orthogonality we have $\mu = (N_1+N_2-N_3-N_4)/2+\nu$. The upper limit of ν is $\min(N_3, N_4, M-N_1, M-N_2) = \min(N_4, M-N_1)$ and the lower limit is $\max(0, N_3+N_4-M) = 0$, if we order $N_1 \ge N_2 \ge N_3 \ge N_4$.

Now we return to the product expansion of m Hermite polynomials and the corresponding new result from it for I_m. We prove a generalized Feldheim identity,

$$H_{N_1}(x)\cdots H_{N_m}(x)=\sum_{\nu_1,\ldots,\nu_{m-1}} H_M(x)a_{\nu_1,\ldots,\nu_{m-1}}, \tag{13.47}$$

where

$$M=\sum_{i=1}^{m-1}(N_i-2\nu_i)+N_m,$$

by mathematical induction. Multiplying this by $H_{N_{m+1}}$ and using the Feldheim identity, we end up with the same formula for $m+1$ Hermite polynomials, including the recursion relation

$$a_{\nu_1,\ldots,\nu_m}=a_{\nu_1,\ldots,\nu_{m-1}}2^{\nu_m}\nu_m!\binom{N_{m+1}}{\nu_m}\binom{\sum_{i=1}^{m-1}(N_i-2\nu_i)+N_{m+1}}{\nu_m}.$$

Its solution is

$$a_{\nu_1,\ldots,\nu_{m-1}}=\prod_{i=1}^{m-1}\binom{N_{i+1}}{\nu_i}\binom{\sum_{j=1}^{i-1}(N_j-2\nu_j)+N_i}{\nu_i}2^{\nu_i}\nu_i!. \tag{13.48}$$

The limits of the summation indices are

$$0\le\nu_1\le\min(N_1,N_2),\qquad 0\le\nu_2\le\min(N_3,N_1+N_2-2\nu_1),\ldots,$$
$$0\le\nu_{m-1}\le\min\left(N_m,\sum_{i=1}^{m-2}(N_i-2\nu_i)+N_{m-1}\right). \tag{13.49}$$

We now apply this generalized Feldheim identity, with indices ordered as $N_1\ge N_2\ge\cdots\ge N_m$, to I_m, grouping $H_{N_2}\cdots H_{N_m}$ together and using orthogonality for the remaining product of two Hermite polynomials $H_{N_1}H_{\sum_{i=2}^{m-1}(N_i-2\nu_i)+N_m}$. This yields $N_1=\sum_{i=2}^{m-1}(N_i-2\nu_i)+N_m$, fixing ν_{m-1}, and

$$\boxed{I_m=\sqrt{\pi}2^{N_1}N_1!\sum_{\nu_2,\ldots,\nu_{m-1}}\prod_{i=2}^{m-1}\binom{N_{i+1}}{\nu_i}\binom{\sum_{j=2}^{i-1}(N_j-2\nu_j)+N_i}{\nu_i}\nu_i!2^{\nu_i},} \tag{13.50}$$

where the limits on the summation indices are

$$0\le\nu_2\le\min(N_3,N_2),\ldots,\qquad 0\le\nu_{m-1}\le\min\left(N_m,\sum_{i=2}^{m-2}(N_i-2\nu_i)+N_{m-1}\right). \tag{13.51}$$

■

Example 13.1.4 APPLICATIONS OF THE PRODUCT FORMULAS

To check the expression I_m for $m = 3$, we note that the sum $\sum_{j=2}^{i-1}$ with $i - 1 = m - 2 = 1$ in the second binomial coefficient in I_m is empty (that is, zero), so only $N_i = N_{m-1} = N_2$ remains. Also, with $\nu_{m-2} = \nu_1$ the sum over the ν_i is that over ν_2, which is fixed by the constraint on the summation index ν_2: $N_1 = N_2 - 2\nu_2 + N_3$. Hence $\nu_2 = (N_2 + N_3 - N_1)/2 = N - N_1$, with $N = (N_1 + N_2 + N_3)/2$. That is, only the product remains in I_m. The general formula for I_m therefore gives

$$I_3 = \sqrt{\pi}2^{N_1}N_1!\binom{N_3}{\nu_2}\binom{N_2}{\nu_2}\nu_2!2^{\nu_2} = \frac{\sqrt{\pi}2^N N_1!N_2!N_3!}{(N-N_1)!(N-N_2)!(N-N_3)!},$$

which agrees with our earlier result of Example 13.1.2. The last expression is based on the following observations. The power of 2 has the exponent $N_1 + \nu_2 = N$. The factorials from the binomial coefficients are $N_3 - \nu_2 = (N_1 + N_3 - N_2)/2 = N - N_2$, $N_2 - \nu_2 = (N_1 + N_2 - N_3)/2 = N - N_3$.

Next let us consider $m = 4$, where we do not order the Hermite indices N_i as yet. The reason is that the general I_m expression was derived with a different grouping of the Hermite polynomials than the separate calculation of I_4 with which we compare. That is why we'll have to permute the indices to get the earlier result for I_4. That is a general conclusion: Different groupings of the Hermite polynomials just give different permutations of the Hermite indices in the general result.

We have two summations over ν_2 and $\nu_{m-1} = \nu_3$, which is fixed by the constraint $N_1 = N_2 - 2\nu_2 + N_3 - 2\nu_3 + N_4$. Hence

$$\nu_3 = \tfrac{1}{2}(N_2 + N_3 + N_4 - N_1) - \nu_2 = M - N_1 - \nu_2$$

with $M = \frac{1}{2}(N_1 + N_2 + N_3 + N_4)$. The exponent of 2 is $N_1 + \nu_2 + \nu_3 = M$. Therefore for $m = 4$ the I_m formula gives

$$\begin{aligned}
I_4 &= \sqrt{\pi}2^{N_1}N_1!\sum_{\nu_2\geq 0}\binom{N_3}{\nu_2}\binom{N_4}{\nu_3}\binom{N_2}{\nu_2}\binom{N_2-2\nu_2+N_3}{\nu_3}\nu_2!2^{\nu_2}\nu_3!2^{\nu_3} \\
&= \sum_{\nu_2\geq 0}\frac{\sqrt{\pi}2^M N_1!N_2!N_3!N_4!(N_2-2\nu_2+N_3)!}{\nu_2!\nu_3!(N_2-\nu_2)!(N_3-\nu_2)!(N_4-\nu_3)!(N_2+N_3-2\nu_2-\nu_3)!} \\
&= \sum_{\nu_2\geq 0}\frac{\sqrt{\pi}2^M N_1!N_2!N_3!N_4!(N_2+N_3-2\nu_2)!}{(N_2-\nu_2)!(N_3-\nu_2)!(N_4-\nu_3)!\nu_2!\nu_3!(N_2+N_3-2\nu_2-\nu_3)!} \\
&= \sum_{\nu_2\geq 0}\frac{\sqrt{\pi}2^M N_1!N_2!N_3!N_4!(N_2+N_3-2\nu_2)!}{\nu_2!(M-N_1-\nu_2)!(N_3-\nu_2)!(N_2-\nu_2)!(M-N_2-N_3+\nu_2)!(M-N_4-\nu_2)!}.
\end{aligned}$$

In the last expression we have substituted ν_3 and used

$$\begin{aligned}
N_4 - \nu_3 &= (N_1 - N_2 - N_3 + N_4) + \nu_2 = M - N_2 - N_3 + \nu_2, \\
N_2 + N_3 - 2\nu_2 - \nu_3 &= \frac{N_1 + N_2 + N_3 - N_4}{2} - \nu_2 = M - N_4 - \nu_2.
\end{aligned}$$

The upper limit is $\nu_2 \le \min(N_2, N_3, M - N_1, M - N_4)$, and the lower limit is $\nu_2 \ge \max(0, N_2 + N_3 - M)$. If we make the permutation $N_2 \leftrightarrow N_4$, $\nu_2 \to \nu$, then our previous I_4 result is obtained with upper limit $\nu \le \min(N_4, M - N_1) = \frac{1}{2}(N_2 + N_3 + N_4 - N_1)$ and lower limit $\nu \ge \max(0, N_3 + N_4 - M) = 0$ because $N_3 + N_4 - N_1 - N_2 \le 0$ for $N_1 \ge N_2 \ge N_3 \ge N_4 \ge 0$. ■

The Hermite polynomial product formula also applies to products of simple harmonic oscillator wave functions, $\int_{-\infty}^{\infty} e^{-mx^2/2} H_{N_1}(x) \cdots H_{N_m}(x)\, dx$, with a different exponential weight function. To evaluate such integrals we use the generalized Feldheim identity for $H_{N_2} \cdots H_{N_m}$ in conjunction with the integral (see Gradshteyn and Ryzhik, p. 845, in the Additional Readings),

$$\int_{-\infty}^{\infty} e^{-a^2x^2} H_m(x) H_n(x)\, dx = \frac{1}{2}\left(\frac{2}{a}\right)^{m+n+1} \left(1 - a^2\right)^{(m+n)/2} \Gamma\left(\frac{m+n+1}{2}\right) \cdot {}_2F_1\left(-m, -n; \frac{1-m-n}{2}; \frac{a^2}{2(a^2-1)}\right),$$

instead of the standard orthogonality integral for the remaining product of two Hermite polynomials. Here the hypergeometric function is the finite sum

$${}_2F_1\left(-m, -n; \frac{1-m-n}{2}; \frac{a^2}{2(a^2-1)}\right) = \sum_{\nu=0}^{\min(m,n)-1} \frac{(-m)_\nu (-n)_\nu}{\nu! (\frac{1-m-n}{2})_\nu} \left(\frac{a^2}{2(a^2-1)}\right)^\nu$$

with $(-m)_\nu = (-m)(1-m)\cdots(\nu - 1 - m)$ and $(-m)_0 \equiv 1$. This yields a result similar to I_m but somewhat more complicated.

The oscillator potential has also been employed extensively in calculations of nuclear structure (nuclear shell model) quark models of hadrons and the nuclear force.

There is a second independent solution of Eq. (13.13). This Hermite function of the second kind is an infinite series (Sections 9.5 and 9.6) and is of no physical interest, at least not yet.

Exercises

13.1.1 Assume the Hermite polynomials are known to be solutions of the differential equation (13.13). From this the recurrence relation, Eq. (13.3), and the values of $H_n(0)$ are also known.

(a) Assume the existence of a generating function

$$g(x,t) = \sum_{n=0}^{\infty} \frac{H_n(x) t^n}{n!}.$$

(b) Differentiate $g(x,t)$ with respect to x and using the recurrence relation develop a first-order PDE for $g(x,t)$.

(c) Integrate with respect to x, holding t fixed.

(d) Evaluate $g(0,t)$ using Eq. (13.5). Finally, show that

$$g(x,t) = \exp\left(-t^2 + 2tx\right).$$

13.1.2 In developing the properties of the Hermite polynomials, start at a number of different points, such as:

1. Hermite's ODE, Eq. (13.13),
2. Rodrigues' formula, Eq. (13.7),
3. Integral representation, Eq. (13.8),
4. Generating function, Eq. (13.1),
5. Gram–Schmidt construction of a complete set of orthogonal polynomials over $(-\infty, \infty)$ with a weighting factor of $\exp(-x^2)$, Section 10.3.

Outline how you can go from any one of these starting points to all the other points.

13.1.3 Prove that

$$\left(2x - \frac{d}{dx}\right)^n 1 = H_n(x).$$

Hint. Check out the first couple of examples and then use mathematical induction.

13.1.4 Prove that

$$|H_n(ix)| \le |H_n(x)|.$$

13.1.5 Rewrite the series form of $H_n(x)$, Eq. (13.9), as an **ascending** power series.

ANS. $H_{2n}(x) = (-1)^n \sum_{s=0}^{n} (-1)^{2s} (2x)^{2s} \dfrac{(2n)!}{\overline{(2s)!(n-s)!}}$,

$$H_{2n+1}(x) = (-1)^n \sum_{s=0}^{n} (-1)^s (2x)^{2s+1} \frac{(2n+1)!}{(2s+1)!(n-s)!}.$$

13.1.6 (a) Expand x^{2r} in a series of even-order Hermite polynomials.
(b) Expand x^{2r+1} in a series of odd-order Hermite polynomials.

ANS. (a) $x^{2r} = \dfrac{(2r)!}{2^{2r}} \sum_{n=0}^{r} \dfrac{H_{2n}(x)}{(2n)!(r-n)!}$

(b) $x^{2r+1} = \dfrac{(2r+1)!}{2^{2r+1}} \sum_{n=0}^{r} \dfrac{H_{2n+1}(x)}{(2n+1)!(r-n)!}, \qquad r = 0, 1, 2, \ldots.$

Hint. Use a Rodrigues representation and integrate by parts.

13.1.7 Show that

(a) $$\int_{-\infty}^{\infty} H_n(x) \exp\left[-\frac{x^2}{2}\right] dx = \begin{cases} 2\pi n!/(n/2)!, & n \text{ even} \\ 0, & n \text{ odd.} \end{cases}$$

(b) $$\int_{-\infty}^{\infty} x H_n(x) \exp\left[-\frac{x^2}{2}\right] dx = \begin{cases} 0, & n \text{ even} \\ 2\pi \dfrac{(n+1)!}{((n+1)/2)!}, & n \text{ odd.} \end{cases}$$

13.1.8 Show that

$$\int_{-\infty}^{\infty} x^m e^{-x^2} H_n(x)\,dx = 0 \qquad \text{for } m \text{ an integer}, \qquad 0 \le m \le n-1.$$

13.1.9 The transition probability between two oscillator states m and n depends on

$$\int_{-\infty}^{\infty} x e^{-x^2} H_n(x) H_m(x)\,dx.$$

Show that this integral equals $\pi^{1/2}2^{n-1}n!\delta_{m,n-1} + \pi^{1/2}2^n(n+1)!\delta_{m,n+1}$. This result shows that such transitions can occur only between states of adjacent energy levels, $m = n \pm 1$.
Hint. Multiply the generating function (Eq. (13.1)) by itself using two different sets of variables (x, s) and (x, t). Alternatively, the factor x may be eliminated by the recurrence relation, Eq. (13.2).

13.1.10 Show that

$$\int_{-\infty}^{\infty} x^2 e^{-x^2} H_n(x) H_n(x)\,dx = \pi^{1/2}2^n n!\left(n + \frac{1}{2}\right).$$

This integral occurs in the calculation of the mean-square displacement of our quantum oscillator.
Hint. Use the recurrence relation, Eq. (13.2), and the orthogonality integral.

13.1.11 Evaluate

$$\int_{-\infty}^{\infty} x^2 e^{-x^2} H_n(x) H_m(x)\,dx$$

in terms of n and m and appropriate Kronecker delta functions.

ANS. $2^{n-1}\pi^{1/2}(2n+1)n!\delta_{nm} + 2^n\pi^{1/2}(n+2)!\delta_{n+2,m} + 2^{n-2}\pi^{1/2}n!\delta_{n-2,m}$.

13.1.12 Show that

$$\int_{-\infty}^{\infty} x^r e^{-x^2} H_n(x) H_{n+p}(x)\,dx = \begin{cases} 0, & p > r \\ 2^n\pi^{1/2}(n+r)!, & p = r, \end{cases}$$

with n, p, and r nonnegative integers.
Hint. Use the recurrence relation, Eq. (13.2), p times.

13.1.13 (a) Using the Cauchy integral formula, develop an integral representation of $H_n(x)$ based on Eq. (13.1) with the contour enclosing the point $z = -x$.

ANS. $H_n(x) = \dfrac{n!}{2\pi i} e^{x^2} \oint \dfrac{e^{-z^2}}{(z+x)^{n+1}}\,dz.$

(b) Show by direct substitution that this result satisfies the Hermite equation.

13.1.14 With

$$\psi_n(x) = e^{-x^2/2}\frac{H_n(x)}{(2^n n!\pi^{1/2})^{1/2}},$$

verify that

$$\hat{a}_n\psi_n(x) = \frac{1}{\sqrt{2}}\left(x + \frac{d}{dx}\right)\psi_n(x) = n^{1/2}\psi_{n-1}(x),$$

$$\hat{a}_n^\dagger\psi_n(x) = \frac{1}{\sqrt{2}}\left(x - \frac{d}{dx}\right)\psi_n(x) = (n+1)^{1/2}\psi_{n+1}(x).$$

Note. The usual quantum mechanical operator approach establishes these raising and lowering properties before the form of $\psi_n(x)$ is known.

13.1.15 (a) Verify the operator identity

$$x - \frac{d}{dx} = -\exp\left[\frac{x^2}{2}\right]\frac{d}{dx}\exp\left[-\frac{x^2}{2}\right].$$

(b) The normalized simple harmonic oscillator wave function is

$$\psi_n(x) = \left(\pi^{1/2}2^n n!\right)^{-1/2}\exp\left[-\frac{x^2}{2}\right]H_n(x).$$

Show that this may be written as

$$\psi_n(x) = \left(\pi^{1/2}2^n n!\right)^{-1/2}\left(x - \frac{d}{dx}\right)^n\exp\left[-\frac{x^2}{2}\right].$$

Note. This corresponds to an n-fold application of the raising operator of Exercise 13.1.14.

13.1.16 (a) Show that the simple oscillator Hamiltonian (from Eq. (13.38)) may be written as

$$\mathcal{H} = -\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2}x^2 = \frac{1}{2}\left(\hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a}\right).$$

Hint. Express E in units of $\hbar\omega$.

(b) Using the creation–annihilation operator formulation of part (a), show that

$$\mathcal{H}\psi(x) = \left(n + \frac{1}{2}\right)\psi(x).$$

This means the energy eigenvalues are $E = (n + \frac{1}{2})(\hbar\omega)$, in agreement with Eq. (13.40).

13.1.17 Write a program that will generate the coefficients a_s, in the polynomial form of the Hermite polynomial $H_n(x) = \sum_{s=0}^{n} a_s x^s$.

13.1.18 A function $f(x)$ is expanded in a Hermite series:

$$f(x) = \sum_{n=0}^{\infty} a_n H_n(x).$$

From the orthogonality and normalization of the Hermite polynomials the coefficient a_n is given by

$$a_n = \frac{1}{2^n \pi^{1/2} n!} \int_{-\infty}^{\infty} f(x) H_n(x) e^{-x^2}\, dx.$$

For $f(x) = x^8$ determine the Hermite coefficients a_n by the Gauss–Hermite quadrature. Check your coefficients against AMS-55, Table 22.12 (for the reference see footnote 4 in Chapter 5 or the General References at book's end).

13.1.19 (a) In analogy with Exercise 12.2.13, set up the matrix of even Hermite polynomial coefficients that will transform an even Hermite series into an even power series:

$$\mathsf{B} = \begin{pmatrix} 1 & -2 & 12 & \cdots \\ 0 & 4 & -48 & \cdots \\ 0 & 0 & 16 & \cdots \\ \vdots & \vdots & \vdots & \cdots \end{pmatrix}.$$

Extend B to handle an even polynomial series through $H_8(x)$.

(b) Invert your matrix to obtain matrix A, which will transform an even power series (through x^8) into a series of even Hermite polynomials. Check the elements of A against those listed in AMS-55 (Table 22.12, in the General References at book's end).

(c) Finally, using matrix multiplication, determine the Hermite series equivalent to $f(x) = x^8$.

13.1.20 Write a subroutine that will transform a finite power series, $\sum_{n=0}^{N} a_n x^n$, into a Hermite series, $\sum_{n=0}^{N} b_n H_n(x)$. Use the recurrence relation, Eq. (13.2).
Note. Both Exercises 13.1.19 and 13.1.20 are faster and more accurate than the Gaussian quadrature, Exercise 13.1.18, if $f(x)$ is available as a power series.

13.1.21 Write a subroutine for evaluating Hermite polynomial matrix elements of the form

$$M_{pqr} = \int_{-\infty}^{\infty} H_p(x) H_q(x) x^r e^{-x^2}\, dx,$$

using the 10-point Gauss–Hermite quadrature (for $p + q + r \leq 19$). Include a parity check and set equal to zero the integrals with odd-parity integrand. Also, check to see if r is in the range $|p - q| \leq r$. Otherwise $M_{pqr} = 0$. Check your results against the specific cases listed in Exercises 13.1.9, 13.1.10, 13.1.11, and 13.1.12.

13.1.22 Calculate and tabulate the normalized linear oscillator wave functions

$$\psi_n(x) = 2^{-n/2} \pi^{-1/4} (n!)^{-1/2} H_n(x) \exp\left(-\frac{x^2}{2}\right) \qquad \text{for } x = 0.0(0.1)5.0$$

and $n = 0(1)5$. If a plotting routine is available, plot your results.

13.1.23 Evaluate $\int_{-\infty}^{\infty} e^{-2x^2} H_{N_1}(x) \cdots H_{N_4}(x)\, dx$ in closed form.
Hint. $\int_{-\infty}^{\infty} e^{-2x^2} H_{N_1}(x) H_{N_2}(x) H_{N_3}(x)\, dx = \frac{1}{\pi} 2^{(N_1+N_2+N_3-1)/2} \cdot \Gamma(s - N_1)\Gamma(s - N_2) \cdot \Gamma(s - N_3), s = (N_1 + N_2 + N_3 + 1)/2$ or $\int_{-\infty}^{\infty} e^{-2x^2} H_{N_1}(x) H_{N_2}(x)\, dx = (-1)^{(N_1+N_2-1)/2} 2^{(N_1+N_2-1)/2} \cdot \Gamma((N_1 + N_2 + 1)/2)$ may be helpful. Prove these formulas (see Gradshteyn and Ryzhik, no. 7.375 on p. 844, in the Additional Readings).

13.2 LAGUERRE FUNCTIONS

Differential Equation — Laguerre Polynomials

If we start with the appropriate generating function, it is possible to develop the Laguerre polynomials in analogy with the Hermite polynomials. Alternatively, a series solution may be developed by the methods of Section 9.5. Instead, to illustrate a different technique, let us start with Laguerre's ODE and obtain a solution in the form of a contour integral, as we did with the integral representation for the modified Bessel function $K_\nu(x)$ (Section 11.6). From this integral representation a generating function will be derived.

Laguerre's ODE (which derives from the radial ODE of Schrödinger's PDE for the hydrogen atom) is

$$\boxed{xy''(x) + (1-x)y'(x) + ny(x) = 0.} \tag{13.52}$$

We shall attempt to represent y, or rather y_n, since y will depend on the parameter n, a nonnegative integer, by the contour integral

$$y_n(x) = \frac{1}{2\pi i}\oint \frac{e^{-xz/(1-z)}}{(1-z)z^{n+1}}dz \tag{13.53a}$$

and demonstrate that it satisfies Laguerre's ODE. The contour includes the origin but does not enclose the point $z = 1$. By differentiating the exponential in Eq. (13.53a) we obtain

$$y_n'(x) = -\frac{1}{2\pi i}\oint \frac{e^{-xz/(1-z)}}{(1-z)^2 z^n}dz, \tag{13.53b}$$

$$y_n''(x) = \frac{1}{2\pi i}\oint \frac{e^{-xz/(1-z)}}{(1-z)^3 z^{n-1}}dz. \tag{13.53c}$$

Substituting into the left-hand side of Eq. (13.52), we obtain

$$\frac{1}{2\pi i}\oint \left[\frac{x}{(1-z)^3 z^{n-1}} - \frac{1-x}{(1-z)^2 z^n} + \frac{n}{(1-z)z^{n+1}}\right] e^{-xz/(1-z)}\,dz,$$

which is equal to

$$-\frac{1}{2\pi i}\oint \frac{d}{dz}\left[\frac{e^{-xz/(1-z)}}{(1-z)z^n}\right] dz. \tag{13.54}$$

If we integrate our perfect differential around a closed contour (Fig. 13.3), the integral will vanish, thus verifying that $y_n(x)$ (Eq. (13.53a)) is a solution of Laguerre's equation.

It has become customary to define $L_n(x)$, the Laguerre polynomial (Fig. 13.4), by[5]

$$L_n(x) = \frac{1}{2\pi i}\oint \frac{e^{-xz/(1-z)}}{(1-z)z^{n+1}}dz. \tag{13.55}$$

[5]Other definitions of $L_n(x)$ are in use. The definitions here of the Laguerre polynomial $L_n(x)$ and the associated Laguerre polynomial $L_n^k(x)$ agree with AMS-55, Chapter 22. (For the full ref. see footnote 4 in Chapter 5 or the General References at book's end.)

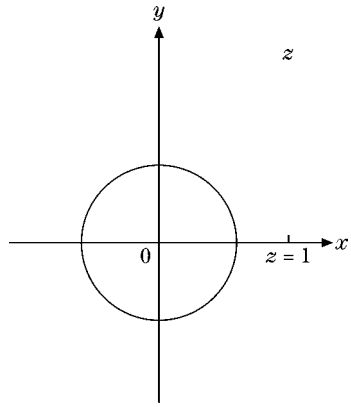

FIGURE 13.3 Laguerre polynomial contour.

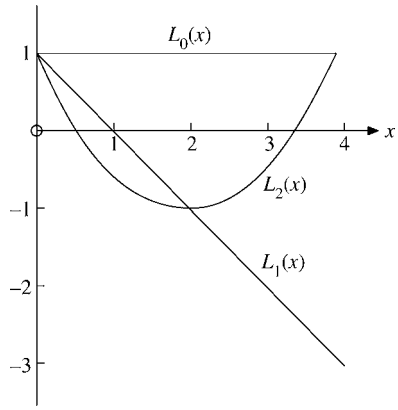

FIGURE 13.4 Laguerre polynomials.

This is exactly what we would obtain from the series

$$\boxed{g(x,z)=\frac{e^{-xz/(1-z)}}{1-z}=\sum_{n=0}^{\infty}L_n(x)z^n,} \qquad |z|<1, \tag{13.56}$$

if we multiplied $g(x,z)$ by z^{-n-1} and integrated around the origin. As in the development of the calculus of residues (Section 7.1), only the z^{-1} term in the series survives. On this basis we identify $g(x,z)$ as the generating function for the Laguerre polynomials.

With the transformation

$$\frac{xz}{1-z} = s - x, \qquad \text{or} \qquad z = \frac{s-x}{s}, \tag{13.57}$$

$$L_n(x) = \frac{e^x}{2\pi i} \oint \frac{s^n e^{-s}}{(s-x)^{n+1}}\, ds, \tag{13.58}$$

the new contour enclosing the point $s = x$ in the s-plane. By Cauchy's integral formula (for derivatives),

$$\boxed{L_n(x) = \frac{e^x}{n!}\frac{d^n}{dx^n}\left(x^n e^{-x}\right)} \qquad \text{(integral } n\text{)}, \tag{13.59}$$

giving Rodrigues' formula for Laguerre polynomials. From these representations of $L_n(x)$ we find the series form (for integral n),

$$\begin{aligned} L_n(x) &= \frac{(-1)^n}{n!}\left[x^n - \frac{n^2}{1!}x^{n-1} + \frac{n^2(n-1)^2}{2!}x^{n-2} - \cdots + (-1)^n n!\right] \\ &= \sum_{m=0}^{n} \frac{(-1)^m n! x^m}{(n-m)!m!m!} = \sum_{s=0}^{n} \frac{(-1)^{n-s} n! x^{n-s}}{(n-s)!(n-s)!s!} \end{aligned} \tag{13.60}$$

and the specific polynomials listed in Table 13.2 (Exercise 13.2.1). Clearly, the definition of Laguerre polynomials in Eqs. (13.55), (13.56), (13.59), and (13.60) are equivalent. Practical applications will decide which approach is used as one's starting point. Equation (13.59) is most convenient for generating Table 13.2, Eq. (13.56) for deriving recursion relations from which the ODE (13.52) is recovered.

By differentiating the generating function in Eq. (13.56) with respect to x and z, we obtain recurrence relations for the Laguerre polynomials as follows. Using the product rule for differentiation we verify the identities

$$(1-z)^2\frac{\partial g}{\partial z} = (1-x-z)g(x,z), \qquad (z-1)\frac{\partial g}{\partial x} = zg(x,z). \tag{13.61}$$

Table 13.2 Laguerre Polynomials

$L_0(x) = 1$
$L_1(x) = -x + 1$
$2!L_2(x) = x^2 - 4x + 2$
$3!L_3(x) = -x^3 + 9x^2 - 18x + 6$
$4!L_4(x) = x^4 - 16x^3 + 72x^2 - 96x + 24$
$5!L_5(x) = -x^5 + 25x^4 - 200x^3 + 600x^2 - 600x + 120$
$6!L_6(x) = x^6 - 36x^5 + 450x^4 - 2400x^3 + 5400x^2 - 4320x + 720$

Writing the left-hand and right-hand sides of the first identity in terms of Laguerre polynomials using Eq. (13.56) we obtain

$$\sum_n \big[(n+1)L_{n+1}(x) - 2nL_n(x) + (n-1)L_{n-1}(x)\big]z^n$$
$$= \sum_n \big[(1-x)L_n(x) - L_{n-1}(x)\big]z^n.$$

Equating coefficients of z^n yields

$$(n+1)L_{n+1}(x) = (2n+1-x)L_n(x) - nL_{n-1}(x). \tag{13.62}$$

To get the second recursion relation we use both identities of Eqs. (13.61) to verify the third identity,

$$x\frac{\partial g}{\partial x} = z\frac{\partial g}{\partial z} - z\frac{\partial (zg)}{\partial z}, \tag{13.63}$$

which, when written similarly in terms of Laguerre polynomials, is seen to be equivalent to

$$xL_n'(x) = nL_n(x) - nL_{n-1}(x). \tag{13.64}$$

Equation (13.61), modified to read

$$L_{n+1}(x) = 2L_n(x) - L_{n-1}(x) - \frac{1}{n+1}\big[(1+x)L_n(x) - L_{n-1}(x)\big], \tag{13.65}$$

for reasons of economy and numerical stability, is used for computation of numerical values of $L_n(x)$. The computer starts with known numerical values of $L_0(x)$ and $L_1(x)$, Table 13.2, and works up step by step. This is the same technique discussed for computing Legendre polynomials, Section 12.2.

Also, from Eq. (13.56) we find

$$g(0,z) = \frac{1}{1-z} = \sum_{n=0}^{\infty} z^n = \sum_{n=0}^{\infty} L_n(0)z^n,$$

which yields the special values of Laguerre polynomials

$$L_n(0) = 1. \tag{13.66}$$

As is seen from the form of the generating function, from the form of Laguerre's ODE, or from Table 13.2, the Laguerre polynomials have neither odd nor even symmetry under the parity transformation $x \to -x$.

The Laguerre ODE is not self-adjoint, and the Laguerre polynomials $L_n(x)$ do not by themselves form an orthogonal set. However, following the method of Section 10.1, if we multiply Eq. (13.52) by e^{-x} (Exercise 10.1.1) we obtain

$$\int_0^{\infty} e^{-x} L_m(x) L_n(x)\, dx = \delta_{mn}. \tag{13.67}$$

This orthogonality is a consequence of the Sturm–Liouville theory, Section 10.1. The normalization follows from the generating function. It is sometimes convenient to define orthogonalized Laguerre functions (with unit weighting function) by

$$\varphi_n(x) = e^{-x/2} L_n(x). \tag{13.68}$$

Our new orthonormal function, $\varphi_n(x)$, satisfies the ODE

$$x\varphi_n''(x) + \varphi_n'(x) + \left(n + \frac{1}{2} - \frac{x}{4}\right)\varphi_n(x) = 0, \tag{13.69}$$

which is seen to have the (self-adjoint) Sturm–Liouville form. Note that the interval $(0 \le x < \infty)$ was used because Sturm–Liouville boundary conditions are satisfied at its endpoints.

Associated Laguerre Polynomials

In many applications, particularly in quantum mechanics, we need the associated Laguerre polynomials defined by[6]

$$\boxed{L_n^k(x) = (-1)^k \frac{d^k}{dx^k} L_{n+k}(x).} \tag{13.70}$$

From the series form of $L_n(x)$ we verify that the lowest associated Laguerre polynomials are given by

$$\begin{aligned} L_0^k(x) &= 1, \\ L_1^k(x) &= -x + k + 1, \\ L_2^k(x) &= \frac{x^2}{2} - (k+2)x + \frac{(k+2)(k+1)}{2}. \end{aligned} \tag{13.71}$$

In general,

$$L_n^k(x) = \sum_{m=0}^{n} (-1)^m \frac{(n+k)!}{(n-m)!(k+m)!m!} x^m, \qquad k > -1. \tag{13.72}$$

A generating function may be developed by differentiating the Laguerre generating function k times to yield

$$\begin{aligned} (-1)^k \frac{d^k}{dx^k} \frac{e^{-xz/(1-z)}}{1-z} &= (-1)^k \sum_{n=0}^{\infty} \frac{d^k}{dx^k} L_{n+k}(x) z^{n+k} = \sum_{n=0}^{\infty} L_n^k(x) z^{n+k} \\ &= \left(\frac{z}{1-z}\right)^k \frac{e^{xz/(1-z)}}{1-z}. \end{aligned}$$

[6]Some authors use $\mathfrak{L}_{n+k}^k(x) = (d^k/dx^k)[L_{n+k}(x)]$. Hence our $L_n^k(x) = (-1)^k \mathfrak{L}_{n+k}^k(x)$.

From the last two members of this equation, canceling the common factor z^k, we obtain

$$\boxed{\frac{e^{-xz/(1-z)}}{(1-z)^{k+1}} = \sum_{n=0}^{\infty} L_n^k(x) z^n,} \qquad |z| < 1. \tag{13.73}$$

From this, for $x = 0$, the binomial expansion

$$\frac{1}{(1-z)^{k+1}} = \sum_{n=0}^{\infty} \binom{-k-1}{n} (-z)^n = \sum_{n=0}^{\infty} L_n^k(0) z^n$$

yields

$$L_n^k(0) = \frac{(n+k)!}{n!k!}. \tag{13.74}$$

Recurrence relations can be derived from the generating function or by differentiating the Laguerre polynomial recurrence relations. Among the numerous possibilities are

$$(n+1)L_{n+1}^k(x) = (2n+k+1-x)L_n^k(x) - (n+k)L_{n-1}^k(x), \tag{13.75}$$

$$x\frac{dL_n^k(x)}{dx} = nL_n^k(x) - (n+k)L_{n-1}^k(x). \tag{13.76}$$

Thus, we obtain from differentiating Laguerre's ODE once

$$x\frac{dL_n''}{dx} + L_n'' - L_n' + (1-x)\frac{dL_n'}{dx} + n\frac{dL_n}{dx} = 0,$$

and eventually from differentiating Laguerre's ODE k times

$$x\frac{d^k}{dx^k}L_n'' + k\frac{d^{k-1}}{dx^{k-1}}L_n'' - k\frac{d^{k-1}}{dx^{k-1}}L_n' + (1-x)\frac{d^k}{dx^k}L_n' + n\frac{d^k}{dx^k}L_n = 0.$$

Adjusting the index $n \to n+k$, we have the associated Laguerre ODE

$$\boxed{x\frac{d^2L_n^k(x)}{dx^2} + (k+1-x)\frac{dL_n^k(x)}{dx} + nL_n^k(x) = 0.} \tag{13.77}$$

When associated Laguerre polynomials appear in a physical problem it is usually because that physical problem involves Eq. (13.77). The most important application is the bound states of the hydrogen atom, which are derived in upcoming Example 13.2.1.

A Rodrigues representation of the associated Laguerre polynomial

$$L_n^k(x) = \frac{e^x x^{-k}}{n!}\frac{d^n}{dx^n}\left(e^{-x}x^{n+k}\right) \tag{13.78}$$

may be obtained from substituting Eq. (13.59) into Eq. (13.70). Note that all these formulas for associated Legendre polynomials $L_n^k(x)$ reduce to the corresponding expressions for $L_n(x)$ when $k = 0$.

The associated Laguerre equation (13.77) is not self-adjoint, but it can be put in self-adjoint form by multiplying by $e^{-x}x^k$, which becomes the weighting function (Section 10.1). We obtain

$$\int_0^\infty e^{-x}x^k L_n^k(x)L_m^k(x)\,dx = \frac{(n+k)!}{n!}\delta_{mn}. \tag{13.79}$$

Equation (13.79) shows the same orthogonality interval $(0,\infty)$ as that for the Laguerre polynomials, but with a new weighting function we have a new set of orthogonal polynomials, the associated Laguerre polynomials.

By letting $\psi_n^k(x) = e^{-x/2}x^{k/2}L_n^k(x)$, $\psi_n^k(x)$ satisfies the self-adjoint ODE

$$x\frac{d^2\psi_n^k(x)}{dx^2} + \frac{d\psi_n^k(x)}{dx} + \left(-\frac{x}{4} + \frac{2n+k+1}{2} - \frac{k^2}{4x}\right)\psi_n^k(x) = 0. \tag{13.80}$$

The $\psi_n^k(x)$ are sometimes called **Laguerre functions**. Equation (13.67) is the special case $k = 0$ of Eq. (13.79).

A further useful form is given by defining[7]

$$\Phi_n^k(x) = e^{-x/2}x^{(k+1)/2}L_n^k(x). \tag{13.81}$$

Substitution into the associated Laguerre equation yields

$$\frac{d^2\Phi_n^k(x)}{dx^2} + \left(-\frac{1}{4} + \frac{2n+k+1}{2x} - \frac{k^2-1}{4x^2}\right)\Phi_n^k(x) = 0. \tag{13.82}$$

The corresponding normalization integral $\int_0^\infty |\Phi_n^k(x)|^2\,dx$ is

$$\int_0^\infty e^{-x}x^{k+1}\left[L_n^k(x)\right]^2 dx = \frac{(n+k)!}{n!}(2n+k+1). \tag{13.83}$$

Notice that the $\Phi_n^k(x)$ do **not** form an orthogonal set (except with x^{-1} as a weighting function) because of the x^{-1} in the term $(2n+k+1)/2x$. (The Laguerre functions $L_\nu^\mu(x)$ in which the indices ν and μ are **not** integers may be defined using the confluent hypergeometric functions of Section 13.5.)

Example 13.2.1 THE HYDROGEN ATOM

The most important application of the Laguerre polynomials is in the solution of the Schrödinger equation for the hydrogen atom. This equation is

$$-\frac{\hbar^2}{2m}\nabla^2\psi - \frac{Ze^2}{4\pi\epsilon_0 r}\psi = E\psi, \tag{13.84}$$

in which $Z = 1$ for hydrogen, 2 for ionized helium, and so on. Separating variables, we find that the angular dependence of ψ is the spherical harmonics $Y_L^M(\theta,\varphi)$. The radial part, $R(r)$, satisfies the equation

$$-\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) - \frac{Ze^2}{4\pi\epsilon_0 r}R + \frac{\hbar^2}{2m}\frac{L(L+1)}{r^2}R = ER. \tag{13.85}$$

[7] This corresponds to modifying the function ψ in Eq. (13.80) to eliminate the first derivative (compare Exercise 9.6.11).

For bound states, $R \to 0$ as $r \to \infty$, and R is finite at the origin, $r = 0$. We do not consider continuum states with positive energy. Only when the latter are included do hydrogen wave functions form a complete set.

By use of the abbreviations (resulting from rescaling r to the dimensionless radial variable ρ)

$$\rho = \alpha r \qquad \text{with} \quad \alpha^2 = -\frac{8mE}{\hbar^2}, \qquad E < 0, \qquad \lambda = \frac{mZe^2}{2\pi\epsilon_0 \alpha \hbar^2}, \tag{13.86}$$

Eq. (13.85) becomes

$$\frac{1}{\rho^2}\frac{d}{d\rho}\left(\rho^2 \frac{d\chi(\rho)}{d\rho}\right) + \left(\frac{\lambda}{\rho} - \frac{1}{4} - \frac{L(L+1)}{\rho^2}\right)\chi(\rho) = 0, \tag{13.87}$$

where $\chi(\rho) = R(\rho/\alpha)$. A comparison with Eq. (13.82) for $\Phi_n^k(x)$ shows that Eq. (13.87) is satisfied by

$$\rho\chi(\rho) = e^{-\rho/2}\rho^{L+1} L_{\lambda-L-1}^{2L+1}(\rho), \tag{13.88}$$

in which k is replaced by $2L+1$ and n by $\lambda - L - 1$, upon using

$$\frac{1}{\rho^2}\frac{d}{d\rho}\rho^2\frac{d\chi}{d\rho} = \frac{1}{\rho}\frac{d^2}{d\rho^2}(\rho\chi).$$

We must restrict the parameter λ by requiring it to be an integer n, $n = 1, 2, 3, \ldots$.[8] This is necessary because the Laguerre function of nonintegral n would diverge[9] as $\rho^n e^\rho$, which is unacceptable for our physical problem, in which

$$\lim_{r\to\infty} R(r) = 0.$$

This restriction on λ, imposed by our boundary condition, has the effect of quantizing the energy,

$$E_n = -\frac{Z^2 m}{2n^2\hbar^2}\left(\frac{e^2}{4\pi\epsilon_0}\right)^2. \tag{13.89}$$

The negative sign reflects the fact that we are dealing here with bound states ($E < 0$), corresponding to an electron that is unable to escape to infinity, where the Coulomb potential goes to zero. Using this result for E_n, we have

$$\alpha = \frac{me^2}{2\pi\epsilon_0\hbar^2}\cdot\frac{Z}{n} = \frac{2Z}{na_0}, \qquad \rho = \frac{2Z}{na_0}r, \tag{13.90}$$

with

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{me^2}, \qquad \text{the Bohr radius.}$$

[8] This is the conventional notation for λ. It is not the same n as the index n in $\Phi_n^k(x)$.

[9] This may be shown, as in Exercise 9.5.5.

Thus, the final normalized hydrogen wave function is written as

$$\boxed{\psi_{nLM}(r,\theta,\varphi)=\left[\left(\frac{2Z}{na_0}\right)^3\frac{(n-L-1)!}{2n(n+L)!}\right]^{1/2}e^{-\alpha r/2}(\alpha r)^L L_{n-L-1}^{2L+1}(\alpha r)Y_L^M(\theta,\varphi).} \tag{13.91}$$

Regular solutions exist for $n \geq L+1$, so the lowest state with $L=1$ (called a 2P state) occurs only with $n=2$. ■

Exercises

13.2.1 Show with the aid of the Leibniz formula that the series expansion of $L_n(x)$ (Eq. (13.60)) follows from the Rodrigues representation (Eq. (13.59)).

13.2.2 (a) Using the explicit series form (Eq. (13.60)) show that

$$L_n'(0) \equiv -n,$$
$$L_n''(0) \equiv \tfrac{1}{2}n(n-1).$$

(b) Repeat without using the explicit series form of $L_n(x)$.

13.2.3 From the generating function derive the Rodrigues representation

$$L_n^k(x)=\frac{e^x x^{-k}}{n!}\frac{d^n}{dx^n}\left(e^{-x}x^{n+k}\right).$$

13.2.4 Derive the normalization relation (Eq. (13.79)) for the associated Laguerre polynomials.

13.2.5 Expand x^r in a series of associated Laguerre polynomials $L_n^k(x)$, k fixed and n ranging from 0 to r (or to ∞ if r is not an integer).
Hint. The Rodrigues form of $L_n^k(x)$ will be useful.

$$\text{ANS. } x^r=(r+k)!r!\sum_{n=0}^{r}\frac{(-1)^n L_n^k(x)}{(n+k)!(r-n)!}, \qquad 0\leq x<\infty.$$

13.2.6 Expand e^{-ax} in a series of associated Laguerre polynomials $L_n^k(x)$, k fixed and n ranging from 0 to ∞.

(a) Evaluate directly the coefficients in your assumed expansion.
(b) Develop the desired expansion from the generating function.

$$\text{ANS. } e^{-ax}=\frac{1}{(1+a)^{1+k}}\sum_{n=0}^{\infty}\left(\frac{a}{1+a}\right)^n\frac{L_n^k(x)}{(n+k)!}, \qquad 0\leq x<\infty.$$

13.2.7 Show that

$$\int_0^\infty e^{-x}x^{k+1}L_n^k(x)L_m^k(x)\,dx=\frac{(n+k)!}{n!}(2n+k+1)\delta_{nm}.$$

Hint. Note that

$$xL_n^k=-(n+k)L_{n-1}^k+(2n+k+1)L_n^k-(n+1)L_{n+1}^k.$$

13.2.8 Assume that a particular problem in quantum mechanics has led to the ODE

$$\frac{d^2y}{dx^2} - \left[\frac{k^2-1}{4x^2} - \frac{2n+k+1}{2x} + \frac{1}{4}\right]y = 0$$

for nonnegative integers n, k. Write $y(x)$ as

$$y(x) = A(x)B(x)C(x),$$

with the requirement that

(a) $A(x)$ be a **negative** exponential giving the required asymptotic behavior of $y(x)$ and
(b) $B(x)$ be a **positive** power of x giving the behavior of $y(x)$ for $0 \le x \ll 1$.

Determine $A(x)$ and $B(x)$. Find the relation between $C(x)$ and the associated Laguerre polynomial.

ANS. $A(x) = e^{-x/2}, \qquad B(x) = x^{(k+1)/2}, \qquad C(x) = L_n^k(x).$

13.2.9 From Eq. (13.91) the normalized radial part of the hydrogenic wave function is

$$R_{nL}(r) = \left[\alpha^3 \frac{(n-L-1)!}{2n(n+L)!}\right]^{1/2} e^{-\alpha r}(\alpha r)^L L_{n-L-1}^{2L+1}(\alpha r),$$

in which $\alpha = 2Z/na_0 = 2Zme^2/4\pi\varepsilon_0\hbar^2$. Evaluate

(a) $\langle r\rangle = \displaystyle\int_0^\infty r R_{nL}(\alpha r) R_{nL}(\alpha r) r^2\, dr,$

(b) $\langle r^{-1}\rangle = \displaystyle\int_0^\infty r^{-1} R_{nL}(\alpha r) R_{nL}(\alpha r) r^2\, dr.$

The quantity $\langle r\rangle$ is the average displacement of the electron from the nucleus, whereas $\langle r^{-1}\rangle$ is the average of the reciprocal displacement.

ANS. $\langle r\rangle = \dfrac{a_0}{2}\big[3n^2 - L(L+1)\big], \qquad \langle r^{-1}\rangle = \dfrac{1}{n^2 a_0}.$

13.2.10 Derive the recurrence relation for the hydrogen wave function expectation values:

$$\frac{s+2}{n^2}\langle r^{s+1}\rangle - (2s+3)a_0\langle r^s\rangle + \frac{s+1}{4}\big[(2L+1)^2 - (s+1)^2\big]a_0^2\langle r^{s-1}\rangle = 0,$$

with $s \ge -2L-1$, $\langle r^s\rangle \equiv \overline{r^s}$.
Hint. Transform Eq. (13.87) into a form analogous to Eq. (13.80). Multiply by $\rho^{s+2}u' - c\rho^{s+1}u$. Here $u = \rho\Phi$. Adjust c to cancel terms that do not yield expectation values.

13.2.11 The hydrogen wave functions, Eq. (13.91), are mutually orthogonal, as they should be, since they are eigenfunctions of the self-adjoint Schrödinger equation

$$\int \psi^*_{n_1L_1M_1}\psi_{n_2L_2M_2} r^2\, dr\, d\Omega = \delta_{n_1n_2}\delta_{L_1L_2}\delta_{M_1M_2}.$$

Yet the radial integral has the (misleading) form

$$\int_0^\infty e^{-\alpha r/2}(\alpha r)^L L_{n_1-L-1}^{2L+1}(\alpha r)e^{-\alpha r/2}(\alpha r)^L L_{n_2-L-1}^{2L+1}(\alpha r)r^2\,dr,$$

which **appears** to match Eq. (13.83) and not the associated Laguerre orthogonality relation, Eq. (13.79). How do you resolve this paradox?

ANS. The parameter α is dependent on n. The first three α, previously shown, are $2Z/n_1a_0$. The last three are $2Z/n_2a_0$. For $n_1 = n_2$ Eq. (13.83) applies. For $n_1 \neq n_2$ neither Eq. (13.79) nor Eq. (13.83) is applicable.

13.2.12 A quantum mechanical analysis of the Stark effect (parabolic coordinate) leads to the ODE

$$\frac{d}{d\xi}\left(\xi\frac{du}{d\xi}\right) + \left(\frac{1}{2}E\xi + L - \frac{m^2}{4\xi} - \frac{1}{4}F\xi^2\right)u = 0.$$

Here F is a measure of the perturbation energy introduced by an external electric field. Find the unperturbed wave functions ($F = 0$) in terms of associated Laguerre polynomials.

ANS. $u(\xi) = e^{-\varepsilon\xi/2}\xi^{m/2}L_p^m(\varepsilon\xi)$, with $\varepsilon = \sqrt{-2E} > 0$, $p = L/\varepsilon - (m+1)/2$, a nonnegative integer.

13.2.13 The wave equation for the three-dimensional harmonic oscillator is

$$-\frac{\hbar^2}{2M}\nabla^2\psi + \frac{1}{2}M\omega^2r^2\psi = E\psi.$$

Here ω is the angular frequency of the corresponding classical oscillator. Show that the radial part of ψ (in spherical polar coordinates) may be written in terms of associated Laguerre functions of argument (βr^2), where $\beta = M\omega/\hbar$.

Hint. As in Exercise 13.2.8, split off radial factors of r^l and $e^{-\beta r^2/2}$. The associated Laguerre function will have the form $L_{1/2(n-l-1)}^{l+1/2}(\beta r^2)$.

13.2.14 Write a computer program that will generate the coefficients a_s in the polynomial form of the Laguerre polynomial $L_n(x) = \sum_{s=0}^n a_s x^s$.

13.2.15 Write a computer program that will transform a finite power series $\sum_{n=0}^N a_n x^n$ into a Laguerre series $\sum_{n=0}^N b_n L_n(x)$. Use the recurrence relation, Eq. (13.62).

13.2.16 Tabulate $L_{10}(x)$ for $x = 0.0(0.1)30.0$. This will include the 10 roots of L_{10}. Beyond $x = 30.0$, $L_{10}(x)$ is monotonic increasing. If graphic software is available, plot your results.

Check value. Eighth root = 16.279.

13.2.17 Determine the 10 roots of $L_{10}(x)$ using root-finding software. You may use your knowledge of the approximate location of the roots or develop a search routine to look for the roots. The 10 roots of $L_{10}(x)$ are the evaluation points for the 10-point Gauss–Laguerre quadrature. Check your values by comparing with AMS-55, Table 25.9. (For the reference see footnote 4 in Chapter 5 or the General References at book's end.)

13.2.18 Calculate the coefficients of a Laguerre series expansion $(L_n(x), k = 0)$ of the exponential e^{-x}. Evaluate the coefficients by the Gauss–Laguerre quadrature (compare Eq. (10.64)). Check your results against the values given in Exercise 13.2.6.
Note. Direct application of the Gauss–Laguerre quadrature with $f(x)L_n(x)e^{-x}$ gives poor accuracy because of the extra e^{-x}. Try a change of variable, $y = 2x$, so that the function appearing in the integrand will be simply $L_n(y/2)$.

13.2.19 (a) Write a subroutine to calculate the Laguerre matrix elements

$$M_{mnp} = \int_0^\infty L_m(x)L_n(x)x^p e^{-x}\,dx.$$

Include a check of the condition $|m - n| \le p \le m + n$. (If p is outside this range, $M_{mnp} = 0$. Why?)
Note. A 10-point Gauss–Laguerre quadrature will give accurate results for $m + n + p \le 19$.

(b) Call your subroutine to calculate a variety of Laguerre matrix elements. Check M_{mn1} against Exercise 13.2.7.

13.2.20 Write a subroutine to calculate the numerical value of $L_n^k(x)$ for specified values of n, k, and x. Require that n and k be nonnegative integers and $x \ge 0$. (a) Use the recurrence relation, Eq. (13.75), to generate $L_n^k(x)$, $n = 2, 3, 4, \ldots$.
Hint. Starting with known values of L_0^k and L_1^k,

13.2.21 Show that $\int_{-\infty}^{\infty} x^n e^{-x^2} H_n(xy)\,dx = \sqrt{\pi}\,n!P_n(y)$, where P_n is a Legendre polynomial.

13.2.22 Write a program to calculate the normalized hydrogen radial wave function $\psi_{nL}(r)$. This is ψ_{nLM} of Eq. (13.91), omitting the spherical harmonic $Y_L^M(\theta, \varphi)$. Take $Z = 1$ and $a_0 = 1$ (which means that r is being expressed in units of Bohr radii). Accept n and L as input data. Tabulate $\psi_{nL}(r)$ for $r = 0.0(0.2)R$ with R taken large enough to exhibit the significant features of ψ. This means roughly $R = 5$ for $n = 1$, $R = 10$ for $n = 2$, and $R = 30$ for $n = 3$.

13.3 CHEBYSHEV POLYNOMIALS

In this section two types of Chebyshev polynomials are developed as special cases of ultraspherical polynomials. Their properties follow from the ultraspherical polynomial generating function. The primary importance of the Chebyshev polynomials is in numerical analysis.

Generating Functions

In Section 12.1 the generating function for the ultraspherical, or Gegenbauer, polynomials

$$\frac{1}{(1 - 2xt + t^2)^\alpha} = \sum_{n=0}^{\infty} C_n^{(\alpha)}(x)t^n, \qquad |x| < 1, \qquad |t| < 1 \tag{13.92}$$

was mentioned, with $\alpha = \frac{1}{2}$ giving rise to the Legendre polynomials. In this section we first take $\alpha = 1$ and then $\alpha = 0$ to generate two sets of polynomials known as the Chebyshev polynomials.

Type II

With $\alpha = 1$ and $C_n^{(1)}(x) = U_n(x)$, Eq. (13.92) gives

$$\frac{1}{1-2xt+t^2} = \sum_{n=0}^{\infty} U_n(x)t^n, \qquad |x| < 1, \qquad |t| < 1. \tag{13.93}$$

These functions $U_n(x)$ generated by $(1-2xt+t^2)^{-1}$ are labeled Chebyshev polynomials type II. Although these polynomials have few applications in mathematical physics, one unusual application is in the development of four-dimensional spherical harmonics used in angular momentum theory.

Type I

With $\alpha = 0$ there is a difficulty. Indeed, our generating function reduces to the constant 1. We may avoid this problem by first differentiating Eq. (13.92) with respect to t. This yields

$$\frac{-\alpha(-2x+2t)}{(1-2xt+t^2)^{\alpha+1}} = \sum_{n=1}^{\infty} nC_n^{(\alpha)}(x)t^{n-1}, \tag{13.94}$$

or

$$\frac{x-t}{(1-2xt+t^2)^{\alpha+1}} = \sum_{n=1}^{\infty} \frac{n}{2}\left[\frac{C_n^{(\alpha)}(x)}{\alpha}\right]t^{n-1}. \tag{13.95}$$

We define $C_n^{(0)}(x)$ by

$$C_n^{(0)}(x) = \lim_{\alpha\to 0} \frac{C_n^{(\alpha)}(x)}{\alpha}. \tag{13.96}$$

The purpose of differentiating with respect to t was to get α in the denominator and to create an indeterminate form. Now multiplying Eq. (13.95) by $2t$ and adding $1 = (1-2xt+t^2)/(1-2xt+t^2)$, we obtain

$$\frac{1-t^2}{1-2xt+t^2} = 1 + 2\sum_{n=1}^{\infty} \frac{n}{2}C_n^{(0)}(x)t^n. \tag{13.97}$$

We define $T_n(x)$ by

$$T_n(x) = \begin{cases} 1, & n = 0 \\ \dfrac{n}{2}C_n^{(0)}(x), & n > 0. \end{cases} \tag{13.98}$$

Notice the special treatment for $n = 0$. This is similar to the treatment of the $n = 0$ term in the Fourier series. Also, note that $C_n^{(0)}$ is the limit indicated in Eq. (13.96) and not a literal substitution of $\alpha = 0$ into the generating function series. With these new labels,

$$\frac{1-t^2}{1-2xt+t^2} = T_0(x) + 2\sum_{n=1}^{\infty} T_n(x)t^n, \qquad |x| \le 1, \qquad |t| < 1. \tag{13.99}$$

We call $T_n(x)$ the type I Chebyshev polynomials. Note that the notation and spelling of the name for these functions differs from reference to reference. Here we follow the usage of AMS-55 (for the full reference see footnote 4 in Chapter 5).

Differentiating the generating function (Eqs. (13.99)) with respect to t and multiplying by the denominator, $1-2xt+t^2$, we obtain

$$-t-(t-x)\left[T_0(x)+2\sum_{n=1}^{\infty}T_n(x)t^n\right]=\left(1-2xt+t^2\right)\sum_{n=1}^{\infty}nT_n(x)t^{n-1}$$

$$=\sum_{n=1}^{\infty}\left[nT_nt^{n-1}-2xnT_nt^n+nT_nt^{n+1}\right],$$

from which the recurrence relation

$$T_{n+1}(x)-2xT_n(x)+T_{n-1}(x)=0 \tag{13.100}$$

follows by shifting the summation index so as to get the same power, t^n, in each term and then comparing coefficients of t^n. Similarly treating Eq. (13.93) we find

$$-\frac{2(t-x)}{1-2xt+t^2}=\left(1-2xt+t^2\right)\sum_{n=1}^{\infty}nU_n(x)t^{n-1}$$

from which the recursion relation

$$U_{n+1}(x)-2xU_n(x)+U_{n-1}(x)=0 \tag{13.101}$$

follows upon comparing coefficients of like powers of t (see Table 13.3).

Then, using the generating functions for the first few values of n and these recurrence relations for the higher-order polynomials, we get Tables 13.4 and 13.5 (see also Figs. 13.5 and 13.6).

As with the Hermite polynomials, Section 13.1, the recurrence relations, Eqs. (13.100) and (13.101), together with the known values of $T_0(x)$, $T_1(x)$, $U_0(x)$, and $U_1(x)$, provide a convenient — that is, for a computer — means of getting the numerical value of any $T_n(x_0)$ or $U_n(x_0)$, with x_0 a given number.

Table 13.3 Recursion relation[a] $P_{n+1}(x)=(A_nx+B_n)P_n(x)-C_nP_{n-1}(x)$

	$P_n(x)$	A_n	B_n	C_n
Legendre	$P_n(x)$	$\frac{2n+1}{n+1}$	0	$\frac{1}{n+1}$
Chebyshev I	$T_n(x)$	2	0	1
Shifted Chebyshev I	$T_n^*(x)$	4	-2	1
Chebyshev II	$U_n(x)$	2	0	1
Shifted Chebyshev II	$U_n^*(x)$	4	-2	1
Associated Laguerre	$L_n^{(k)}(x)$	$-\frac{1}{n+1}$	$\frac{2n+k+1}{n+1}$	$\frac{n+k}{n+1}$
Hermite	$H_n(x)$	2	0	$2n$

[a] P_n denotes any of the orthogonal polynomials.

Table 13.4 Chebyshev polynomials, type I

$T_0 = 1$
$T_1 = x$
$T_2 = 2x^2 - 1$
$T_3 = 4x^3 - 3x$
$T_4 = 8x^4 - 8x^2 + 1$
$T_5 = 16x^5 - 20x^3 + 5x$
$T_6 = 32x^6 - 48x^4 + 18x^2 - 1$

Table 13.5 Chebyshev polynomials, type II

$U_0 = 1$
$U_1 = 2x$
$U_2 = 4x^2 - 1$
$U_3 = 8x^3 - 4x$
$U_4 = 16x^4 - 12x^2 + 1$
$U_5 = 32x^5 - 32x^3 + 6x$
$U_6 = 64x^6 - 80x^4 + 24x^2 - 1$

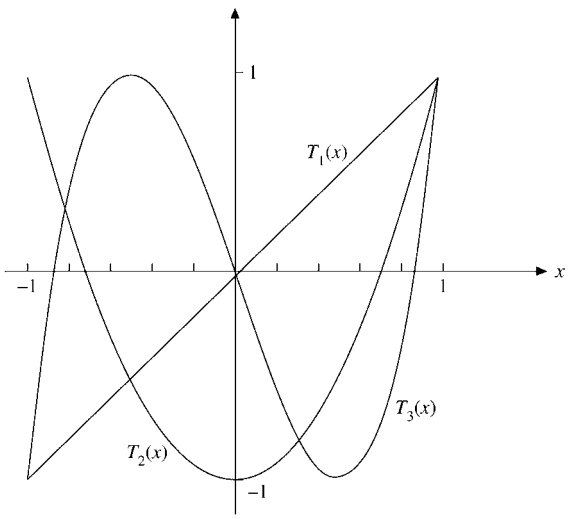

FIGURE 13.5 Chebyshev polynomials $T_n(x)$.

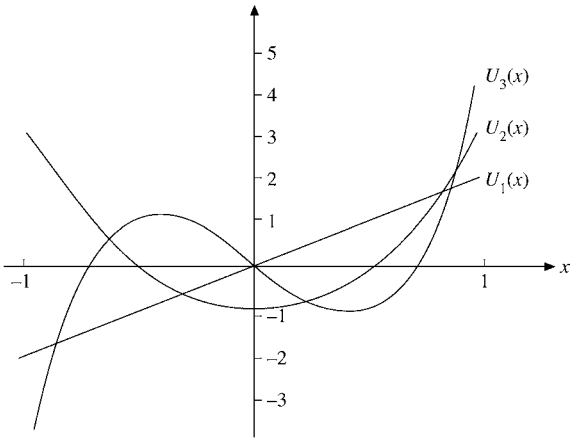

FIGURE 13.6 Chebyshev polynomials $U_n(x)$.

Again, from the generating functions, we can obtain the special values of various polynomials:

$$\begin{aligned} T_n(1) &= 1, \qquad T_n(-1) = (-1)^n, \\ T_{2n}(0) &= (-1)^n, \qquad T_{2n+1}(0) = 0; \end{aligned} \tag{13.102}$$

$$\begin{aligned} U_n(1) &= n+1, \qquad U_n(-1) = (-1)^n(n+1), \\ U_{2n}(0) &= (-1)^n, \qquad U_{2n+1}(0) = 0. \end{aligned} \tag{13.103}$$

For example, comparing the power series

$$\frac{1-t^2}{(1-t)^2} = \frac{1+t}{1-t} = \sum_{n=0}^{\infty} \left(t^n + t^{n+1}\right)$$

with Eq. (13.99) for $x = 1$ gives $T_n(1)$, and for $x = -1$ a similar expansion of $(1-t)/(1+t)$ gives $T_n(-1)$, while replacing $t \to -t^2$ in the first power series yields $T_n(0)$. The power series for $(1 \pm t)^{-2}$ and $(1+t^2)^{-1}$ generate the corresponding $U_n(\pm 1)$, $U_n(0)$.

The parity relations for T_n and U_n follow from their generating functions, with the substitutions $t \to -t, x \to -x$, which leave them invariant; these are

$$T_n(x) = (-1)^n T_n(-x), \qquad U_n(x) = (-1)^n U_n(-x). \tag{13.104}$$

Rodrigues' representations of $T_n(x)$ and $U_n(x)$ are

$$T_n(x) = \frac{(-1)^n \pi^{1/2} (1-x^2)^{1/2}}{2^n (n-\frac{1}{2})!} \frac{d^n}{dx^n} \left[\left(1-x^2\right)^{n-1/2}\right] \tag{13.105}$$

and

$$U_n(x) = \frac{(-1)^n (n+1) \pi^{1/2}}{2^{n+1} (n+\frac{1}{2})! (1-x^2)^{1/2}} \frac{d^n}{dx^n} \left[\left(1-x^2\right)^{n+1/2}\right]. \tag{13.106}$$

Recurrence Relations — Derivatives

Differentiation of the generating functions for $T_n(x)$ and $U_n(x)$ with respect to the variable x leads to a variety of recurrence relations involving derivatives. For example, from Eq. (13.99) we thus obtain

$$\left(1-2xt+t^2\right) 2 \sum_{n=1}^{\infty} T_n'(x) t^n = 2t \left[T_0(x) + 2 \sum_{n=1}^{\infty} T_n(x) t^n \right],$$

from which we extract the recursion

$$2T_{n-1}(x) = T_n'(x) - 2x T_{n-1}'(x) + T_{n-2}'(x), \tag{13.107}$$

which is the derivative of Eq. (13.100) for $n \to n-1$. Among the more useful recursions we thus find are

$$\left(1-x^2\right) T_n'(x) = -nx T_n(x) + n T_{n-1}(x) \tag{13.108}$$

and

$$\left(1-x^2\right)U_n'(x) = -nxU_n(x) + (n+1)U_{n-1}(x). \tag{13.109}$$

Manipulating a variety of these recursions as in Section 12.2 for Legendre polynomials one can eliminate the index $n-1$ also in favor of T_n'' and establish that $T_n(x)$, the Chebyshev polynomial type I, satisfies the ODE

$$\left(1-x^2\right)T_n''(x) - xT_n'(x) + n^2T_n(x) = 0. \tag{13.110}$$

The Chebyshev polynomial of type II, $U_n(x)$, satisfies

$$\left(1-x^2\right)U_n''(x) - 3xU_n'(x) + n(n+2)U_n(x) = 0. \tag{13.111}$$

Chebyshev polynomials may be defined starting from these ODEs, but our emphasis has been on generating functions.

The ultraspherical equation

$$\left(1-x^2\right)\frac{d^2}{dx^2}C_n^{(\alpha)}(x) - (2\alpha+1)x\frac{d}{dx}C_n^{(\alpha)}(x) + n(n+2\alpha)C_n^{(\alpha)}(x) = 0 \tag{13.112}$$

is a generalization of these differential equations, reducing to Eq. (13.110) for $\alpha = 0$ and Eq. (13.111) for $\alpha = 1$ (and to Legendre's equation for $\alpha = \frac{1}{2}$).

Trigonometric Form

At this point in the development of the properties of the Chebyshev solutions it is beneficial to change variables, replacing x by $\cos\theta$. With $x = \cos\theta$ and $d/dx = (-1/\sin\theta)(d/d\theta)$, we verify that

$$\left(1-x^2\right)\frac{d^2T_n}{dx^2} = \frac{d^2T_n}{d\theta^2} - \cot\theta\frac{dT_n}{d\theta}, \qquad xT_n' = -\cot\theta\frac{dT_n}{d\theta}.$$

Adding these terms, Eq. (13.110) becomes

$$\frac{d^2T_n}{d\theta^2} + n^2T_n = 0, \tag{13.113}$$

the simple harmonic oscillator equation with solutions $\cos n\theta$ and $\sin n\theta$. The special values (boundary conditions at $x = 0, 1$) identify

$$T_n = \cos n\theta = \cos n(\arccos x). \tag{13.114a}$$

A second linearly independent solution of Eqs. (13.110) and (13.113) is labeled

$$V_n = \sin n\theta = \sin n(\arccos x). \tag{13.114b}$$

The corresponding solutions of the type II Chebyshev equation, Eq. (13.111), become

$$U_n = \frac{\sin(n+1)\theta}{\sin\theta}, \tag{13.115a}$$

$$W_n = \frac{\cos(n+1)\theta}{\sin\theta}. \tag{13.115b}$$

The two sets of solutions, type I and type II, are related by

$$V_n(x) = \left(1-x^2\right)^{1/2} U_{n-1}(x), \tag{13.116a}$$

$$W_n(x) = \left(1-x^2\right)^{-1/2} T_{n+1}(x). \tag{13.116b}$$

As already seen from generating functions, $T_n(x)$ and $U_n(x)$ are polynomials. Clearly, $V_n(x)$ and $W_n(x)$ are **not** polynomials. From

$$\begin{aligned} T_n(x) + i\,V_n(x) &= \cos n\theta + i \sin n\theta \\ &= (\cos\theta + i\sin\theta)^n = \left[x + i\left(1-x^2\right)^{1/2}\right]^n, \qquad |x| \le 1 \end{aligned} \tag{13.117}$$

we obtain expansions

$$T_n(x) = x^n - \binom{n}{2} x^{n-2}\left(1-x^2\right) + \binom{n}{4} x^{n-4}\left(1-x^2\right)^2 - \cdots \tag{13.118a}$$

and

$$V_n(x) = \sqrt{1-x^2}\left[\binom{n}{1} x^{n-1} - \binom{n}{3} x^{n-3}\left(1-x^2\right) + \cdots\right]. \tag{13.118b}$$

From the generating functions, or from the ODEs, power-series representations are

$$T_n(x) = \frac{n}{2} \sum_{m=0}^{[n/2]} (-1)^m \frac{(n-m-1)!}{m!(n-2m)!} (2x)^{n-2m}, \tag{13.119a}$$

for $n \ge 1$, with $[n/2]$ the largest integer below $n/2$ and

$$U_n(x) = \sum_{m=0}^{[n/2]} (-1)^m \frac{(n-m)!}{m!(n-2m)!} (2x)^{n-2m}. \tag{13.119b}$$

Orthogonality

If Eq. (13.110) is put into self-adjoint form (Section 10.1), we obtain $w(x) = (1-x^2)^{-1/2}$ as a weighting factor. For Eq. (13.111) the corresponding weighting factor is $(1-x^2)^{+1/2}$. The resulting orthogonality integrals,

$$\int_{-1}^{1} T_m(x) T_n(x) \left(1-x^2\right)^{-1/2} dx = \begin{cases} 0, & m \ne n, \\ \dfrac{\pi}{2}, & m = n \ne 0, \\ \pi, & m = n = 0, \end{cases} \tag{13.120}$$

$$\int_{-1}^{1} V_m(x) V_n(x) \left(1-x^2\right)^{-1/2} dx = \begin{cases} 0, & m \ne n, \\ \dfrac{\pi}{2}, & m = n \ne 0, \\ 0, & m = n = 0, \end{cases} \tag{13.121}$$

$$\int_{-1}^{1} U_m(x) U_n(x) \left(1-x^2\right)^{1/2} dx = \frac{\pi}{2} \delta_{m,n}, \tag{13.122}$$

and

$$\int_{-1}^{1} W_m(x)W_n(x)\left(1-x^2\right)^{1/2}dx = \frac{\pi}{2}\delta_{m,n}, \tag{13.123}$$

are a direct consequence of the Sturm–Liouville theory, Chapter 10. The normalization values may best be obtained by using $x = \cos\theta$ and converting these four integrals into Fourier normalization integrals (for the half-period interval $[0, \pi]$).

Exercises

13.3.1 Another Chebyshev generating function is

$$\frac{1-xt}{1-2xt+t^2} = \sum_{n=0}^{\infty} X_n(x)t^n, \qquad |t| < 1.$$

How is $X_n(x)$ related to $T_n(x)$ and $U_n(x)$?

13.3.2 Given

$$\left(1-x^2\right)U_n''(x) - 3xU_n'(x) + n(n+2)U_n(x) = 0,$$

show that $V_n(x)$ (Eq. (13.116a)) satisfies

$$\left(1-x^2\right)V_n''(x) - xV_n'(x) + n^2V_n(x) = 0,$$

which is Chebyshev's equation.

13.3.3 Show that the Wronskian of $T_n(x)$ and $V_n(x)$ is given by

$$T_n(x)V_n'(x) - T_n'(x)V_n(x) = -\frac{n}{(1-x^2)^{1/2}}.$$

This verifies that T_n and $V_n (n \neq 0)$ are independent solutions of Eq. (13.110). Conversely, for $n = 0$, we do not have linear independence. What happens at $n = 0$? Where is the "second" solution?

13.3.4 Show that $W_n(x) = (1-x^2)^{-1/2}T_{n+1}(x)$ is a solution of

$$\left(1-x^2\right)W_n''(x) - 3xW_n'(x) + n(n+2)W_n(x) = 0.$$

13.3.5 Evaluate the Wronskian of $U_n(x)$ and $W_n(x) = (1-x^2)^{-1/2}T_{n+1}(x)$.

13.3.6 $V_n(x) = (1-x^2)^{1/2}U_{n-1}(x)$ is not defined for $n = 0$. Show that a second and independent solution of the Chebyshev differential equation for $T_n(x)$ $(n = 0)$ is $V_0(x) = \arccos x$ (or $\arcsin x$).

13.3.7 Show that $V_n(x)$ satisfies the same three-term recurrence relation as $T_n(x)$ (Eq. (13.100)).

13.3.8 Verify the series solutions for $T_n(x)$ and $U_n(x)$ (Eqs. (13.109a) and (13.119b)).

13.3.9 Transform the series form of $T_n(x)$, Eq. (13.119a), into an **ascending** power series.

$$\text{ANS. } T_{2n}(x) = (-1)^n n \sum_{m=0}^{n} (-1)^m \frac{(n+m-1)!}{(n-m)!(2m)!}(2x)^{2m}, \qquad n \geq 1,$$

$$T_{2n+1}(x) = \frac{2n+1}{2} \sum_{m=0}^{n} \frac{(-1)^{m+n}(n+m)!}{(n-m)!(2m+1)!}(2x)^{2m+1}.$$

13.3.10 Rewrite the series form of $U_n(x)$, Eq. (13.119b), as an ascending power series.

$$\text{ANS. } U_{2n}(x) = (-1)^n \sum_{m=0}^{n} (-1)^m \frac{(n+m)!}{(n-m)!(2m)!}(2x)^{2m},$$

$$U_{2n+1}(x) = (-1)^n \sum_{m=0}^{n} (-1)^m \frac{(n+m+1)!}{(n-m)!(2m+1)!}(2x)^{2m+1}.$$

13.3.11 Derive the Rodrigues representation of $T_n(x)$,

$$T_n(x) = \frac{(-1)^n \pi^{1/2}(1-x^2)^{1/2}}{2^n (n-\frac{1}{2})!} \frac{d^n}{dx^n}\left[\left(1-x^2\right)^{n-1/2}\right].$$

Hint. One possibility is to use the hypergeometric function relation

$${}_2F_1(a,b;c;z) = (1-z)^{-a}\, {}_2F_1\left(a, c-b; c; \frac{-z}{1-z}\right),$$

with $z = (1-x)/2$. An alternate approach is to develop a first-order differential equation for $y = (1-x^2)^{n-1/2}$. Repeated differentiation of this equation leads to the Chebyshev equation.

13.3.12 (a) From the differential equation for T_n (in self-adjoint form) show that

$$\int_{-1}^{1} \frac{dT_m(x)}{dx}\frac{dT_n(x)}{dx}\left(1-x^2\right)^{1/2} dx = 0, \qquad m \neq n.$$

(b) Confirm the preceding result by showing that

$$\frac{dT_n(x)}{dx} = nU_{n-1}(x).$$

13.3.13 The expansion of a power of x in a Chebyshev series leads to the integral

$$I_{mn} = \int_{-1}^{1} x^m T_n(x) \frac{dx}{\sqrt{1-x^2}}.$$

(a) Show that this integral vanishes for $m < n$.
(b) Show that this integral vanishes for $m + n$ odd.

13.3.14 Evaluate the integral

$$I_{mn} = \int_{-1}^{1} x^m T_n(x) \frac{dx}{\sqrt{1-x^2}}$$

for $m \geq n$ and $m + n$ even by each of two methods:

(a) Operate with x as the variable replacing T_n by its Rodrigues representation.
(b) Using $x = \cos\theta$ transform the integral to a form with θ as the variable.

$$\text{ANS. } I_{mn} = \pi \frac{m!}{(m-n)!} \frac{(m-n-1)!!}{(m+n)!!}, \qquad m \geq n, \qquad m+n \text{ even.}$$

13.3.15 Establish the following bounds, $-1 \leq x \leq 1$:

$$\text{(a) } |U_n(x)| \leq n+1, \qquad \text{(b) } \left|\frac{d}{dx} T_n(x)\right| \leq n^2.$$

13.3.16 (a) Establish the following bound, $-1 \leq x \leq 1$: $|V_n(x)| \leq 1$.
(b) Show that $W_n(x)$ is unbounded in $-1 \leq x \leq 1$.

13.3.17 Verify the orthogonality-normalization integrals for

(a) $T_m(x)$, $T_n(x)$, (b) $V_m(x)$, $V_n(x)$,
(c) $U_m(x)$, $U_n(x)$, (d) $W_m(x)$, $W_n(x)$.

Hint. All these can be converted to Fourier orthogonality-normalization integrals.

13.3.18 Show whether

(a) $T_m(x)$ and $V_n(x)$ are or are not orthogonal over the interval $[-1, 1]$ with respect to the weighting factor $(1-x^2)^{-1/2}$.
(b) $U_m(x)$ and $W_n(x)$ are or are not orthogonal over the interval $[-1, 1]$ with respect to the weighting factor $(1-x^2)^{1/2}$.

13.3.19 Derive (a) $T_{n+1}(x) + T_{n-1}(x) = 2xT_n(x)$,
(b) $T_{m+n}(x) + T_{m-n}(x) = 2T_m(x)T_n(x)$,

from the "corresponding" cosine identities.

13.3.20 A number of equations relate the two types of Chebyshev polynomials. As examples show that

$$T_n(x) = U_n(x) - xU_{n-1}(x)$$

and

$$\left(1-x^2\right)U_n(x) = xT_{n+1}(x) - T_{n+2}(x).$$

13.3.21 Show that

$$\frac{dV_n(x)}{dx} = -n\frac{T_n(x)}{\sqrt{1-x^2}}$$

(a) using the trigonometric forms of V_n and T_n,
(b) using the Rodrigues representation.

13.3.22 Starting with $x = \cos\theta$ and $T_n(\cos\theta) = \cos n\theta$, expand

$$x^k = \left(\frac{e^{i\theta} + e^{-i\theta}}{2}\right)^k$$

and show that

$$x^k = \frac{1}{2^{k-1}}\left[T_k(x) + \binom{k}{1} T_{k-2}(x) + \binom{k}{2} T_{k-4} + \cdots\right],$$

the series in brackets terminating with $\binom{k}{m}T_1(x)$ for $k = 2m+1$ or $\frac{1}{2}\binom{k}{m}T_0$ for $k = 2m$.

13.3.23 (a) Calculate and tabulate the Chebyshev functions $V_1(x)$, $V_2(x)$, and $V_3(x)$ for $x = -1.0(0.1)1.0$.

(b) A second solution of the Chebyshev differential equation, Eq. (13.100), for $n = 0$ is $y(x) = \sin^{-1} x$. Tabulate and plot this function over the same range: $-1.0(0.1)1.0$.

13.3.24 Write a computer program that will generate the coefficients a_s in the polynomial form of the Chebyshev polynomial $T_n(x) = \sum_{s=0}^{n} a_s x^s$.

13.3.25 Tabulate $T_{10}(x)$ for 0.00(0.01)1.00. This will include the five positive roots of T_{10}. If graphics software is available, plot your results.

13.3.26 Determine the five positive roots of $T_{10}(x)$ by calling a root-finding subroutine. Use your knowledge of the approximate location of these roots from Exercise 13.3.25 or write a search routine to look for the roots. These five positive roots (and their negatives) are the evaluation points of the 10-point Gauss–Chebyshev quadrature method.

Check values. $x_k = \cos\left[(2k-1)\pi/20\right], \qquad k = 1, 2, 3, 4, 5.$

13.3.27 Develop the following Chebyshev expansions (for $[-1, 1]$):

(a) $$\left(1 - x^2\right)^{1/2} = \frac{2}{\pi}\left[1 - 2\sum_{s=1}^{\infty}\left(4s^2 - 1\right)^{-1} T_{2s}(x)\right].$$

(b) $$\left.\begin{array}{l} +1, \quad 0 < x \le 1 \\ -1, \; -1 \le x < 0 \end{array}\right\} = \frac{4}{\pi}\sum_{s=0}^{\infty}(-1)^s(2s+1)^{-1}T_{2s+1}(x).$$

13.3.28 (a) For the interval $[-1, 1]$ show that

$$|x| = \frac{1}{2} + \sum_{s=1}^{\infty}(-1)^{s+1}\frac{(2s-3)!!}{(2s+2)!!}(4s+1)P_{2s}(x)$$

$$= \frac{2}{\pi} + \frac{4}{\pi}\sum_{s=1}^{\infty}(-1)^{s+1}\frac{1}{4s^2-1}T_{2s}(x).$$

(b) Show that the ratio of the coefficient of $T_{2s}(x)$ to that of $P_{2s}(x)$ approaches $(\pi s)^{-1}$ as $s \to \infty$. This illustrates the relatively rapid convergence of the Chebyshev series.

Hint. With the Legendre recurrence relations, rewrite $xP_n(x)$ as a linear combination of derivatives. The trigonometric substitution $x = \cos\theta$, $T_n(x) = \cos n\theta$ is most helpful for the Chebyshev part.

13.3.29 Show that

$$\frac{\pi^2}{8} = 1 + 2\sum_{s=1}^{\infty}\left(4s^2 - 1\right)^{-2}.$$

Hint. Apply Parseval's identity (or the completeness relation) to the results of Exercise 13.3.28.

13.3.30 Show that

(a) $$\cos^{-1} x = \frac{\pi}{2} - \frac{4}{\pi}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}T_{2n+1}(x).$$

(b) $$\sin^{-1} x = \frac{4}{\pi}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}T_{2n+1}(x).$$

13.4 HYPERGEOMETRIC FUNCTIONS

In Chapter 9 the hypergeometric equation[10]

$$\boxed{x(1-x)y''(x) + \left[c - (a+b+1)x\right]y'(x) - ab\,y(x) = 0} \tag{13.124}$$

was introduced as a canonical form of a linear second-order ODE with regular singularities at $x = 0$, 1, and ∞. One solution is

$$\begin{aligned} y(x) &= {}_2F_1(a, b; c; x) \\ &= 1 + \frac{a\cdot b}{c}\frac{x}{1!} + \frac{a(a+1)b(b+1)}{c(c+1)}\frac{x^2}{2!} + \cdots, \qquad c \neq 0, -1, -2, -3, \ldots, \end{aligned} \tag{13.125}$$

which is known as the **hypergeometric function** or **hypergeometric series**. The range of convergence for $c > a + b$ is $|x| < 1$ and $x = 1$, and is $x = -1$ for $c > a + b - 1$. In terms of the often-used Pochhammer symbol,

$$\begin{aligned} (a)_n &= a(a+1)(a+2)\cdots(a+n-1) = \frac{(a+n-1)!}{(a-1)!}, \\ (a)_0 &= 1, \end{aligned} \tag{13.126}$$

the hypergeometric function becomes

$$\boxed{{}_2F_1(a, b; c; x) = \sum_{n=0}^{\infty}\frac{(a)_n(b)_n}{(c)_n}\frac{x^n}{n!}.} \tag{13.127}$$

[10] This is sometimes called Gauss' ODE. The solutions then become Gauss functions.

In this form the subscripts 2 and 1 become clear. The leading subscript 2 indicates that two Pochhammer symbols appear in the numerator and the final subscript 1 indicates one Pochhammer symbol in the denominator.[11] (The confluent hypergeometric function ${}_1F_1$ with one Pochhammer symbol in the numerator and one in the denominator appears in Section 13.5.)

From the form of Eq. (13.125) we see that the parameter c may not be zero or a negative integer. On the other hand, if a or b equals 0 or a negative integer, the series terminates and the hypergeometric function becomes a polynomial. Many more or less elementary functions can be represented by the hypergeometric function.[12] Comparing the power series we verify that

$$\ln(1+x) = x\ {}_2F_1(1,1;2;-x). \tag{13.128}$$

For the complete elliptic integrals K and E,

$$K\left(k^2\right) = \int_0^{\pi/2} \left(1-k^2\sin^2\theta\right)^{-1/2} d\theta = \frac{\pi}{2}\ {}_2F_1\left(\frac{1}{2},\frac{1}{2};1;k^2\right), \tag{13.129}$$

$$E\left(k^2\right) = \int_0^{\pi/2} \left(1-k^2\sin\theta\right)^{1/2} d\theta = \frac{\pi}{2}\ {}_2F_1\left(\frac{1}{2},-\frac{1}{2};1;k^2\right). \tag{13.130}$$

The explicit series forms and other properties of the elliptic integrals are developed in Section 5.8.

The hypergeometric equation as a second-order linear ODE has a second independent solution. The usual form is

$$y(x) = x^{1-c}\ {}_2F_1(a+1-c, b+1-c; 2-c; x), \qquad c \neq 2, 3, 4, \ldots. \tag{13.131}$$

If c is an integer either the two solutions coincide or (barring a rescue by integral a or integral b) one of the solutions will blow up (see Exercise 13.4.1). In such a case the second solution is expected to include a logarithmic term.

Alternate forms of the hypergeometric ODE include

$$\left(1-z^2\right)\frac{d^2}{dz^2}y\left(\frac{1-z}{2}\right) - \left[(a+b+1)z - (a+b+1-2c)\right]\frac{d}{dz}y\left(\frac{1-z}{2}\right) - ab\ y\left(\frac{1-z}{2}\right) = 0, \tag{13.132}$$

$$\left(1-z^2\right)\frac{d^2}{dz^2}y(z^2) - \left[(2a+2b+1)z + \frac{1-2c}{z}\right]\frac{d}{dz}y\left(z^2\right) - 4ab\ y\left(z^2\right) = 0. \tag{13.133}$$

[11] The Pochhammer symbol is often useful in other expressions involving factorials, for instance,

$$(1-z)^{-a} = \sum_{n=0}^{\infty} (a)_n z^n/n!, \qquad |z| < 1.$$

[12] With three parameters, a, b, and c, we can represent almost anything.

Contiguous Function Relations

The parameters a, b, and c enter in the same way as the parameter n of Bessel, Legendre, and other special functions. As we found with these functions, we expect recurrence relations involving unit changes in the parameters a, b, and c. The usual nomenclature for the hypergeometric functions, in which one parameter changes by $+$ or -1, is a "contiguous function." Generalizing this term to include simultaneous unit changes in more than one parameter, we find 26 functions contiguous to ${}_2F_1(a, b; c; x)$. Taking them two at a time, we can develop the formidable total of 325 equations among the contiguous functions. One typical example is

$$\begin{aligned}(a-b)\big\{c(a+b-1)+1-a^2-b^2+\big[(a-b)^2-1\big](1-x)\big\}\,{}_2F_1(a,b;c;x)\\ =(c-a)(a-b+1)b\,{}_2F_1(a-1,b+1;c;x)\\ +(c-b)(a-b-1)a\,{}_2F_1(a+1,b-1;c;x).\end{aligned}\tag{13.134}$$

Another contiguous function relation appears in Exercise 13.4.10.

Hypergeometric Representations

Since the ultraspherical equation (13.112) in Section 13.3 is a special case of Eq. (13.124), we see that ultraspherical functions (and Legendre and Chebyshev functions) may be expressed as hypergeometric functions. For the ultraspherical function we obtain

$$C_n^\beta(x)=\frac{(n+2\beta)!}{2^\beta n!\beta!}\,{}_2F_1\left(-n,n+2\beta+1;1+\beta;\frac{1-x}{2}\right)\tag{13.135}$$

upon comparing its ODE with Eq. (13.124) and the power-series solutions. For Legendre and associated Legendre functions we find similarly

$$P_n(x)={}_2F_1\left(-n,n+1;1;\frac{1-x}{2}\right),\tag{13.136}$$

$$P_n^m(x)=\frac{(n+m)!}{(n-m)!}\frac{(1-x^2)^{m/2}}{2^m m!}\,{}_2F_1\left(m-n,m+n+1;m+1;\frac{1-x}{2}\right).\tag{13.137}$$

Alternate forms are

$$\begin{aligned}P_{2n}(x)&=(-1)^n\frac{(2n)!}{2^{2n}n!n!}\,{}_2F_1\left(-n,n+\frac{1}{2};\frac{1}{2};x^2\right)\\ &=(-1)^n\frac{(2n-1)!!}{(2n)!!}\,{}_2F_1\left(-n,n+\frac{1}{2};\frac{1}{2};x^2\right),\end{aligned}\tag{13.138}$$

$$\begin{aligned}P_{2n+1}(x)&=(-1)^n\frac{(2n+1)!}{2^{2n}n!n!}\,{}_2F_1\left(-n,n+\frac{3}{2};\frac{3}{2};x^2\right)x\\ &=(-1)^n\frac{(2n+1)!!}{(2n)!!}\,{}_2F_1\left(-n,n+\frac{3}{2};\frac{3}{2};x^2\right)x.\end{aligned}\tag{13.139}$$

In terms of hypergeometric functions, the Chebyshev functions become

$$T_n(x) = {}_2F_1\left(-n, n; \frac{1}{2}; \frac{1-x}{2}\right), \tag{13.140}$$

$$U_n(x) = (n+1)\,{}_2F_1\left(-n, n+2; \frac{3}{2}; \frac{1-x}{2}\right), \tag{13.141}$$

$$V_n(x) = n\sqrt{1-x^2}\,{}_2F_1\left(-n+1, n+1; \frac{3}{2}; \frac{1-x}{2}\right). \tag{13.142}$$

The leading factors are determined by direct comparison of complete power series, comparison of coefficients of particular powers of the variable, or evaluation at $x = 0$ or 1, and so on.

The hypergeometric series may be used to define functions with nonintegral indices. The physical applications are minimal.

Exercises

13.4.1 (a) For c, an integer, and a and b nonintegral, show that

$${}_2F_1(a, b; c; x) \qquad \text{and} \qquad x^{1-c}\,{}_2F_1(a+1-c, b+1-c; 2-c; x)$$

yield only one solution to the hypergeometric equation.

(b) What happens if a is an integer, say, $a = -1$, and $c = -2$?

13.4.2 Find the Legendre, Chebyshev I, and Chebyshev II recurrence relations corresponding to the contiguous hypergeometric function Eq. (13.134).

13.4.3 Transform the following polynomials into hypergeometric functions of argument x^2. (a) $T_{2n}(x)$; (b) $x^{-1}T_{2n+1}(x)$; (c) $U_{2n}(x)$; (d) $x^{-1}U_{2n+1}(x)$.

ANS. (a) $T_{2n}(x) = (-1)^n\,{}_2F_1(-n, n; \frac{1}{2}; x^2)$.
(b) $x^{-1}T_{2n+1}(x) = (-1)^n(2n+1)\,{}_2F_1(-n, n+1; \frac{3}{2}; x^2)$.
(c) $U_{2n}(x) = (-1)^n\,{}_2F_1(-n, n+1; \frac{1}{2}; x^2)$.
(d) $x^{-1}U_{2n+1}(x) = (-1)^n(2n+2)\,{}_2F_1(-n, n+2; \frac{3}{2}; x^2)$.

13.4.4 Derive or verify the leading factor in the hypergeometric representations of the Chebyshev functions.

13.4.5 Verify that the Legendre function of the second kind, $Q_\nu(z)$, is given by

$$Q_\nu(z) = \frac{\pi^{1/2}\nu!}{(\nu+\frac{1}{2})!(2z)^{\nu+1}}\,{}_2F_1\left(\frac{\nu}{2}+\frac{1}{2}, \frac{\nu}{2}+1; \frac{\nu}{2}+\frac{3}{2}; z^{-2}\right),$$

$$|z| > 1, \qquad |\arg z| < \pi, \qquad \nu \neq -1, -2, -3, \ldots.$$

13.4.6 Analogous to the incomplete gamma function, we may define an incomplete beta function by

$$B_x(a, b) = \int_0^x t^{a-1}(1-t)^{b-1}\,dt.$$

Show that

$$B_x(a,b) = a^{-1}x^a\, {}_2F_1(a, 1-b; a+1; x).$$

13.4.7 Verify the integral representation

$${}_2F_1(a,b;c;z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}\,dt.$$

What restrictions must be placed on the parameters b and c and on the variable z?
Note. The restriction on $|z|$ can be dropped — analytic continuation. For nonintegral a the real axis in the z-plane from 1 to ∞ is a cut line.
Hint. The integral is suspiciously like a beta function and can be expanded into a series of beta functions.

ANS. $\Re(c) > \Re(b) > 0$, and $|z| < 1$.

13.4.8 Prove that

$${}_2F_1(a,b;c;1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}, \qquad c \neq 0, -1, -2, \ldots \qquad c > a+b.$$

Hint. Here is a chance to use the integral representation, Exercise 13.4.7.

13.4.9 Prove that

$${}_2F_1(a,b;c;x) = (1-x)^{-a}\, {}_2F_1\left(a, c-b; c; \frac{-x}{1-x}\right).$$

Hint. Try an integral representation.
Note. This relation is useful in developing a Rodrigues representation of $T_n(x)$ (compare Exercise 13.3.11).

13.4.10 Verify that

$${}_2F_1(-n,b;c;1) = \frac{(c-b)_n}{(c)_n}.$$

Hint. Here is a chance to use the contiguous function relation $[2a - c + (b-a)x] \cdot {}_2F_1(a,b;c;x) = a(1-x)\,{}_2F_1(a+1,b;c;x) - (c-a)\,{}_2F_1(a-1,b;c;x)$ and mathematical induction. Alternatively, use the integral representation and the beta function.

13.5 CONFLUENT HYPERGEOMETRIC FUNCTIONS

The confluent hypergeometric equation[13]

$$\boxed{xy''(x) + (c-x)y'(x) - ay(x) = 0} \tag{13.143}$$

[13] This is often called **Kummer's equation**. The solutions, then, are **Kummer functions**.

has a regular singularity at $x = 0$ and an irregular one at $x = \infty$. It is obtained from the hypergeometric equation of Section 13.4 by merging (by hand: $x(1-x) \to x$ in Eq. (13.124)) two of the latter's three singularities. One solution of the confluent hypergeometric equation is

$$y(x) = {}_1F_1(a; c; x) = M(a, c, x)$$
$$= 1 + \frac{a}{c}\frac{x}{1!} + \frac{a(a+1)}{c(c+1)}\frac{x^2}{2!} + \cdots, \qquad c \neq 0, -1, -2, \ldots. \tag{13.144}$$

This solution is convergent for all finite x (or complex z). In terms of the Pochhammer symbols, we have

$$\boxed{M(a, c, x) = \sum_{n=0}^{\infty} \frac{(a)_n}{(c)_n}\frac{x^n}{n!}.} \tag{13.145}$$

Clearly, $M(a, c, x)$ becomes a polynomial if the parameter a is 0 or a negative integer. Numerous more or less elementary functions may be represented by the confluent hypergeometric function. Examples are the error function and the incomplete gamma function (from Eq. (8.69)):

$$\operatorname{erf}(x) = \frac{2}{\pi^{1/2}}\int_0^x e^{-t^2}\,dt = \frac{2}{\pi^{1/2}} x M\left(\frac{1}{2}, \frac{3}{2}, -x^2\right), \tag{13.146}$$

$$\gamma(a, x) = \int_0^x e^{-t} t^{a-1}\,dt = a^{-1}x^a M(a, a+1, -x), \qquad \Re(a) > 0. \tag{13.147}$$

Clearly, this coincides with the first solution for $c = a$. The error function and the incomplete gamma function are discussed further in Section 8.5.

A second solution of Eq. (13.143) is given by

$$y(x) = x^{1-c} M(a + 1 - c, 2 - c, x), \qquad c \neq 2, 3, 4, \ldots. \tag{13.148}$$

The standard form of the second solution of Eq. (13.143) is a linear combination of Eqs. (13.144) and (13.148):

$$U(a, c, x) = \frac{\pi}{\sin \pi c}\left[\frac{M(a, c, x)}{(a-c)!(c-1)!} - \frac{x^{1-c} M(a+1-c, 2-c, x)}{(a-1)!(1-c)!}\right]. \tag{13.149}$$

Note the resemblance to our definition of the Neumann function, Eq. (11.60). As with our Neumann function, Eq. (11.60), this definition of $U(a, c, x)$ becomes indeterminate in this case for c an integer.

An alternate form of the confluent hypergeometric equation that will be useful later is obtained by changing the independent variable from x to x^2:

$$\frac{d^2}{dx^2}y\left(x^2\right) + \left[\frac{2c-1}{x} - 2x\right]\frac{d}{dx}y\left(x^2\right) - 4ay\left(x^2\right) = 0. \tag{13.150}$$

As with the hypergeometric functions, contiguous functions exist in which the parameters a and c are changed by ± 1. Including the cases of simultaneous changes in the

two parameters,[14] we have eight possibilities. Taking the original function and pairs of the contiguous functions, we can develop a total of 28 equations.[15]

Integral Representations

It is frequently convenient to have the confluent hypergeometric functions in integral form. We find (Exercise 13.5.10)

$$M(a,c,x)=\frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)}\int_0^1 e^{xt}t^{a-1}(1-t)^{c-a-1}\,dt,\qquad \Re(c)>\Re(a)>0, \tag{13.151}$$

$$U(a,c,x)=\frac{1}{\Gamma(a)}\int_0^\infty e^{-xt}t^{a-1}(1+t)^{c-a-1}\,dt,\qquad \Re(x)>0,\qquad \Re(a)>0. \tag{13.152}$$

Three important techniques for deriving or verifying integral representations are as follows:

1. Transformation of generating function expansions and Rodrigues representations: The Bessel and Legendre functions provide examples of this approach.
2. Direct integration to yield a series: This direct technique is useful for a Bessel function representation (Exercise 11.1.18) and a hypergeometric integral (Exercise 13.4.7).
3. (a) Verification that the integral representation satisfies the ODE. (b) Exclusion of the other solution. (c) Verification of normalization. This is the method used in Section 11.5 to establish an integral representation of the modified Bessel function $K_\nu(z)$. It will work here to establish Eqs. (13.151) and (13.152).

Bessel and Modified Bessel Functions

Kummer's first formula,

$$M(a,c,x)=e^x M(c-a,c,-x), \tag{13.153}$$

is useful in representing the Bessel and modified Bessel functions. The formula may be verified by series expansion or by use of an integral representation (compare Exercise 13.5.10).

As expected from the form of the confluent hypergeometric equation and the character of its singularities, the confluent hypergeometric functions are useful in representing a number of the special functions of mathematical physics. For the Bessel functions,

$$J_\nu(x)=\frac{e^{-ix}}{\nu!}\left(\frac{x}{2}\right)^\nu M\left(\nu+\frac{1}{2},2\nu+1,2ix\right), \tag{13.154}$$

whereas for the modified Bessel functions of the first kind,

$$I_\nu(x)=\frac{e^{-x}}{\nu!}\left(\frac{x}{2}\right)^\nu M\left(\nu+\frac{1}{2},2\nu+1,2x\right). \tag{13.155}$$

[14] Slater refers to these as **associated functions**.

[15] The recurrence relations for Bessel, Hermite, and Laguerre functions are special cases of these equations.

Hermite Functions

The Hermite functions are given by

$$H_{2n}(x) = (-1)^n \frac{(2n)!}{n!} M\left(-n, \frac{1}{2}, x^2\right), \tag{13.156}$$

$$H_{2n+1}(x) = (-1)^n \frac{2(2n+1)!}{n!} x M\left(-n, \frac{3}{2}, x^2\right), \tag{13.157}$$

using Eq. (13.150).

Comparing the Laguerre ODE with the confluent hypergeometric equation (13.143), we have

$$L_n(x) = M(-n, 1, x). \tag{13.158}$$

The constant is fixed as unity by noting Eq. (13.66) for $x = 0$. For the associated Laguerre functions,

$$L_n^m(x) = (-1)^m \frac{d^m}{dx^m} L_{n+m}(x) = \frac{(n+m)!}{n!m!} M(-n, m+1, x). \tag{13.159}$$

Alternate verification is obtained by comparing Eq. (13.159) with the power-series solution (Eq. (13.72) of Section 13.2). Note that in the hypergeometric form, as distinct from a Rodrigues representation, the indices n and m need not be integers, and, if they are not integers, $L_n^m(x)$ will not be a polynomial.

Miscellaneous Cases

There are certain advantages in expressing our special functions in terms of hypergeometric and confluent hypergeometric functions. If the general behavior of the latter functions is known, the behavior of the special functions we have investigated follows as a series of special cases. This may be useful in determining asymptotic behavior or evaluating normalization integrals. The asymptotic behavior of $M(a, c, x)$ and $U(a, c, x)$ may be conveniently obtained from integral representations of these functions, Eqs. (13.151) and (13.152). The further advantage is that the relations between the special functions are clarified. For instance, an examination of Eqs. (13.156), (13.157), and (13.159) suggests that the Laguerre and Hermite functions are related.

The confluent hypergeometric equation (13.143) is clearly not self-adjoint. For this and other reasons it is convenient to define

$$M_{k\mu}(x) = e^{-x/2} x^{\mu+1/2} M\left(\mu - k + \frac{1}{2}, 2\mu + 1, x\right). \tag{13.160}$$

This new function, $M_{k\mu}(x)$, is a Whittaker function that satisfies the self-adjoint equation

$$M''_{k\mu}(x) + \left(-\frac{1}{4} + \frac{k}{x} + \frac{\frac{1}{4} - \mu^2}{x^2}\right) M_{k\mu}(x) = 0. \tag{13.161}$$

The corresponding second solution is

$$W_{k\mu}(x) = e^{-x/2} x^{\mu+1/2} U\left(\mu - k + \frac{1}{2}, 2\mu + 1, x\right). \tag{13.162}$$

Exercises

13.5.1 Verify the confluent hypergeometric representation of the error function

$$\operatorname{erf}(x) = \frac{2x}{\pi^{1/2}} M\left(\frac{1}{2}, \frac{3}{2}, -x^2\right).$$

13.5.2 Show that the Fresnel integrals $C(x)$ and $S(x)$ of Exercise 5.10.2 may be expressed in terms of the confluent hypergeometric function as

$$C(x) + iS(x) = xM\left(\frac{1}{2}, \frac{3}{2}, \frac{i\pi x^2}{2}\right).$$

13.5.3 By direct differentiation and substitution verify that

$$y = ax^{-a} \int_0^x e^{-t} t^{a-1}\, dt = ax^{-a}\gamma(a, x)$$

satisfies

$$xy'' + (a + 1 + x)y' + ay = 0.$$

13.5.4 Show that the modified Bessel function of the second kind, $K_\nu(x)$, is given by

$$K_\nu(x) = \pi^{1/2} e^{-x} (2x)^\nu U\left(\nu + \frac{1}{2}, 2\nu + 1, 2x\right).$$

13.5.5 Show that the cosine and sine integrals of Section 8.5 may be expressed in terms of confluent hypergeometric functions as

$$\operatorname{Ci}(x) + i \operatorname{si}(x) = -e^{ix} U(1, 1, -ix).$$

This relation is useful in numerical computation of $\operatorname{Ci}(x)$ and $\operatorname{si}(x)$ for large values of x.

13.5.6 Verify the confluent hypergeometric form of the Hermite polynomial $H_{2n+1}(x)$ (Eq. (13.157)) by showing that

(a) $H_{2n+1}(x)/x$ satisfies the confluent hypergeometric equation with $a = -n, c = \frac{3}{2}$ and argument x^2,

(b) $\displaystyle\lim_{x\to 0} \frac{H_{2n+1}(x)}{x} = (-1)^n \frac{2(2n+1)!}{n!}$.

13.5.7 Show that the contiguous confluent hypergeometric function equation

$$(c - a)M(a - 1, c, x) + (2a - c + x)M(a, c, x) - aM(a + 1, c, x) = 0$$

leads to the associated Laguerre function recurrence relation (Eq. (13.75)).

13.5.8 Verify the Kummer transformations:

(a) $M(a, c, x) = e^x M(c - a, c, -x)$

(b) $U(a, c, x) = x^{1-c} U(a - c + 1, 2 - c, x)$.

13.5.9 Prove that

(a) $$\frac{d^n}{dx^n}M(a,c,x)=\frac{(a)_n}{(b)_n}M(a+n,b+n,x),$$

(b) $$\frac{d^n}{dx^n}U(a,c,x)=(-1)^n(a)_nU(a+n,c+n,x).$$

13.5.10 Verify the following integral representations:

(a) $$M(a,c,x)=\frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)}\int_0^1 e^{xt}t^{a-1}(1-t)^{c-a-1}\,dt, \qquad \Re(c)>\Re(a)>0.$$

(b) $$U(a,c,x)=\frac{1}{\Gamma(a)}\int_0^\infty e^{-xt}t^{a-1}(1+t)^{c-a-1}\,dt, \qquad \Re(x)>0, \qquad \Re(a)>0.$$

Under what conditions can you accept $\Re(x)=0$ in part (b)?

13.5.11 From the integral representation of $M(a,c,x)$, Exercise 13.5.10(a), show that

$$M(a,c,x)=e^xM(c-a,c,-x).$$

Hint. Replace the variable of integration t by $1-s$ to release a factor e^x from the integral.

13.5.12 From the integral representation of $U(a,c,x)$, Exercise 13.5.10(b), show that the exponential integral is given by

$$E_1(x)=e^{-x}U(1,1,x).$$

Hint. Replace the variable of integration t in $E_1(x)$ by $x(1+s)$.

13.5.13 From the integral representations of $M(a,c,x)$ and $U(a,c,x)$ in Exercise 13.5.10 develop asymptotic expansions of
(a) $M(a,c,x)$, (b) $U(a,c,x)$.
Hint. You can use the technique that was employed with $K_\nu(z)$, Section 11.6.

ANS. (a) $$\frac{\Gamma(c)}{\Gamma(a)}\frac{e^x}{x^{c-a}}\left\{1+\frac{(1-a)(c-a)}{1!x}+\frac{(1-a)(2-a)(c-a)(c-a+1)}{2!x^2}+\cdots\right\}$$

(b) $$\frac{1}{x^a}\left\{1+\frac{a(1+a-c)}{1!(-x)}+\frac{a(a+1)(1+a-c)(2+a-c)}{2!(-x)^2}+\cdots\right\}.$$

13.5.14 Show that the Wronskian of the two confluent hypergeometric functions $M(a,c,x)$ and $U(a,c,x)$ is given by

$$MU'-M'U=-\frac{(c-1)!}{(a-1)!}\frac{e^x}{x^c}.$$

What happens if a is 0 or a negative integer?

13.5.15 The Coulomb wave equation (radial part of the Schrödinger equation with Coulomb potential) is

$$\frac{d^2y}{d\rho^2} + \left[1 - \frac{2\eta}{\rho} - \frac{L(L+1)}{\rho^2}\right] y = 0.$$

Show that a regular solution $y = F_L(\eta, \rho)$ is given by

$$F_L(\eta, \rho) = C_L(\eta)\rho^{L+1}e^{-i\rho}M(L+1-i\eta, 2L+2, 2i\rho).$$

13.5.16 (a) Show that the radial part of the hydrogen wave function, Eq. (13.81), may be written as

$$e^{-\alpha r/2}(\alpha r)^L L_{n-L-1}^{2L+1}(\alpha r)$$
$$= \frac{(n+L)!}{(n-L-1)!(2L+1)!}e^{-\alpha r/2}(\alpha r)^L M(L+1-n, 2L+2, \alpha r).$$

(b) It was assumed previously that the total (kinetic + potential) energy E of the electron was negative. Rewrite the (unnormalized) radial wave function for the free electron, $E > 0$.

ANS. $e^{i\alpha r/2}(\alpha r)^L M(L+1-in, 2L+2, -i\alpha r)$, outgoing wave. This representation provides a powerful alternative technique for the calculation of photoionization and recombination coefficients.

13.5.17 Evaluate

(a) $\displaystyle\int_0^\infty \left[M_{k\mu}(x)\right]^2 dx$, (b) $\displaystyle\int_0^\infty \left[M_{k\mu}(x)\right]^2 \frac{dx}{x}$,

(c) $\displaystyle\int_0^\infty \left[M_{k\mu}(x)\right]^2 \frac{dx}{x^{1-a}}$,

where $2\mu = 0, 1, 2, \ldots, k - \mu - \frac{1}{2} = 0, 1, 2, \ldots, a > -2\mu - 1$.

ANS. (a) $(2\mu)!2k$. (b) $(2\mu)!$. (c) $(2\mu)!(2k)^a$.

13.6 MATHIEU FUNCTIONS

When PDEs such as Laplace's, Poisson's, and the wave equation are solved with cylindrical or spherical boundary conditions by separating variables in polar coordinates, we find radial solutions, which are the Bessel functions of Chapter 11, and angular solutions, which are $\sin m\varphi, \cos m\varphi$ in cylindrical cases and spherical harmonics in spherical cases. Examples are electromagnetic waves in resonant cavities, vibrating circular drumheads, and coaxial wave guides.

When in such cylindrical problems the circular boundary condition becomes elliptical we are led to the angular and radial Mathieu functions, which, therefore, might be called elliptic cylinder functions. In fact, in 1868 Mathieu developed the leading terms of series solutions of the vibrating elliptical drumhead, and Whittaker and others in the early 1900s derived higher-order terms as well.

Here our goal is to give an introduction to the rich and complex properties of Mathieu functions.

Separation of Variables in Elliptical Coordinates

Elliptical cylinder coordinates ξ, η, z, which are appropriate for elliptical boundary conditions, are expressed in rectangular coordinates as

$$x = c\cosh\xi\cos\eta, \qquad y = c\sinh\xi\sin\eta, \qquad z = z, \tag{13.163}$$
$$0 \le \xi < \infty, \qquad 0 \le \eta \le 2\pi,$$

where the parameter $2c > 0$ is the distance between the foci of the confocal ellipses described by these coordinates (Fig. 13.7). We want to show that in the limit $c \to 0$ the foci of the ellipses coalesce to the center of circles. We work at constant z-coordinate mostly, $z = 0$, say. Indeed for fixed radial variable ξ =const. we can eliminate the angular variable η to obtain from Eq. (13.163)

$$\frac{x^2}{c^2\cosh^2\xi} + \frac{y^2}{c^2\sinh^2\xi} = 1, \tag{13.164}$$

describing confocal ellipses centered at the origin of the x, y-plane with major and minor half-axes

$$a = c\cosh\xi, \qquad b = c\sinh\xi, \tag{13.165}$$

respectively. Since

$$\frac{b}{a} = \tanh\xi = \sqrt{1 - \frac{1}{\cosh^2\xi}} \equiv \sqrt{1 - e^2}, \tag{13.166}$$

the eccentricity $e = 1/\cosh\xi$ of the ellipse with $0 \le e \le 1$, and the distance between the foci $2ae = 2c$, providing a geometrical interpretation of the radial coordinate ξ and the

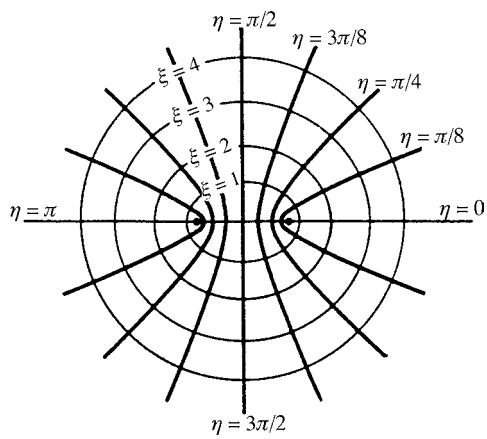

FIGURE 13.7 Elliptical coordinates ξ, η.

parameter c. As $\xi \to \infty$, $e \to 0$ and the ellipses become circles, which is indicated in Fig. 13.7. As $\xi \to 0$, the ellipse becomes more elongated until, at $\xi = 0$, it has shrunk to the line segment between the foci.

When $\eta =$const. we eliminate ξ to find confocal hyperbolas

$$\frac{x^2}{c^2 \cos^2 \eta} - \frac{y^2}{c^2 \sin^2 \eta} = 1, \tag{13.167}$$

which are also plotted in Fig. 13.7. Differentiating the ellipse, we obtain

$$\frac{x\,dx}{\cosh^2 \xi} + \frac{y\,dy}{\sinh^2 \xi} = 0, \tag{13.168}$$

which means that the tangent vector (dx, dy) of the ellipse is perpendicular to the vector $(\frac{x}{\cosh^2 \xi}, \frac{y}{\sinh^2 \xi})$. For the hyperbola the orthogonality condition is

$$\frac{x\,dx}{\cos^2 \eta} - \frac{y\,dy}{\sin^2 \eta} = 0, \tag{13.169}$$

so the scalar product of the ellipse and hyperbola tangent vectors at each of their intersection points (x, y) of Eq. (13.163) obey

$$\frac{x^2}{\cosh^2 \xi \cos^2 \eta} - \frac{y^2}{\sinh^2 \xi \sin^2 \eta} = c^2 - c^2 = 0. \tag{13.170}$$

This means that these confocal ellipses and hyperbolas form an orthogonal coordinate system, in the sense of Section 2.1. To extract the scale factors h_ξ, h_η from the differentials of the elliptical coordinates

$$\begin{aligned} dx &= c \sinh \xi \cos \eta\, d\xi - c \cosh \xi \sin \eta\, d\eta, \\ dy &= c \cosh \xi \sin \eta\, d\xi + c \sinh \xi \cos \eta\, d\eta, \end{aligned} \tag{13.171}$$

we sum their squares, finding

$$\begin{aligned} dx^2 + dy^2 &= c^2 \big(\sinh^2 \xi \cos^2 \eta + \cosh^2 \xi \sin^2 \eta\big)\big(d\xi^2 + d\eta^2\big) \\ &= c^2 \big(\cosh^2 \xi - \cos^2 \eta\big)\big(d\xi^2 + d\eta^2\big) \equiv h_\xi^2\, d\xi^2 + h_\eta^2\, d\eta^2 \end{aligned} \tag{13.172}$$

and yielding

$$h_\xi = h_\eta = c\big(\cosh^2 \xi - \cos^2 \eta\big)^{1/2}. \tag{13.173}$$

Note that there is no cross term involving $d\xi\, d\eta$, showing again that we are dealing with orthogonal coordinates.

Now we are ready to derive Mathieu's differential equations.

Example 13.6.1 ELLIPTICAL DRUM

We consider vibrations of an elliptical drumhead with vertical displacement $z = z(x, y, t)$ governed by the wave equation

$$\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = \frac{1}{v^2}\frac{\partial^2 z}{\partial t^2}, \tag{13.174}$$

where the velocity squared $v^2 = T/\rho$ with tension T and mass density ρ is a constant. We first separate the harmonic time dependence, writing

$$z(x, y, t) = u(x, y)w(t), \tag{13.175}$$

where $w(t) = \cos(\omega t + \delta)$, with ω the frequency and δ a constant phase. Substituting this function z into Eq. (13.174) yields

$$\frac{1}{u}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) = \frac{1}{v^2 w}\frac{\partial^2 w}{\partial t^2} = -\frac{\omega^2}{v^2} = -k^2 = \text{const.}, \tag{13.176}$$

that is, the two-dimensional Helmholtz equation for the displacement u. We now use Eq. (2.22) to convert the Laplacian ∇^2 to the elliptical coordinates, where we drop the z-coordinate. This gives

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + k^2 u = \frac{1}{h_\xi^2}\left(\frac{\partial^2 u}{\partial \xi^2} + \frac{\partial^2 u}{\partial \eta^2}\right) + k^2 u = 0, \tag{13.177}$$

that is, the Helmholtz equation in elliptical ξ, η coordinates,

$$\frac{\partial^2 u}{\partial \xi^2} + \frac{\partial^2 u}{\partial \eta^2} + c^2 k^2 \left(\cosh^2 \xi - \cos^2 \eta\right) u = 0. \tag{13.178}$$

Lastly, we separate ξ and η, writing $u(\xi, \eta) = R(\xi)\Phi(\eta)$, which yields

$$\frac{1}{R}\frac{d^2 R}{d\xi^2} + c^2 k^2 \cosh^2 \xi = c^2 k^2 \cos^2 \eta - \frac{1}{\Phi}\frac{d^2 \Phi}{d\eta^2} = \lambda + \frac{1}{2}c^2 k^2, \tag{13.179}$$

where $\lambda + c^2 k^2/2$ is the separation constant. Writing $\cosh 2\xi, \cos 2\eta$ instead of $\cosh^2 \xi$, $\cos^2 \eta$ (which motivates the special form of the separation constant in Eq. (13.179)) we find the linear, second-order ODE

$$\frac{d^2 R}{d\xi^2} - (\lambda - 2q \cosh 2\xi) R(\xi) = 0, \qquad q = \frac{1}{4}c^2 k^2, \tag{13.180}$$

which is also called the **radial Mathieu equation**, and

$$\frac{d^2 \Phi}{d\eta^2} + (\lambda - 2q \cos 2\eta)\Phi(\eta) = 0, \tag{13.181}$$

the **angular**, or **modified**, **Mathieu equation**. Note that the eigenvalue $\lambda(q)$ is a function of the continuous parameter q in the Mathieu ODEs. It is this parameter dependence that complicates the analysis of Mathieu functions and makes them among the most difficult special functions used in physics. ■

Clearly, all finite points are regular points of both ODEs, while infinity is an essential singularity for both ODEs, which are of the Sturm–Liouville type (Chapter 10) with coefficient functions $p \equiv 1$ and

$$q(\xi) = -\lambda + 2q\cosh 2\xi, \qquad q(\eta) = \lambda - 2q\cos 2\eta. \tag{13.182}$$

(These functions q must not be confused with the parameter q.) As a consequence, their solutions form orthogonal sets of functions. The substitution $\eta \to i\xi$ transforms the angular to the radial Mathieu ODE, so their solutions are closely related.

Using the Lindemann–Stieltjes substitution $z = \cos^2\eta$, $dz/d\eta = -\sin 2\eta$, the angular Mathieu ODE is transformed into an ODE with coefficients that are algebraic in the variable z (using $\frac{d}{d\eta} = \frac{dz}{d\eta}\frac{d}{dz} = -\sin 2\eta \frac{d}{dz}$ and $\frac{d^2}{d\eta^2} = -2\cos 2\eta \frac{d}{dz} + \sin^2 2\eta \frac{d^2}{dz^2}$):

$$4z(1-z)\frac{d^2\Phi}{dz^2} + 2(1-2z)\frac{d\Phi}{dz} + \big[\lambda + 2q(1-2z)\big]\Phi = 0. \tag{13.183}$$

This ODE has regular singularities at $z = 0$ and $z = 1$, whereas the point at infinity is an essential singularity (Chapter 9). By comparison, the hypergeometric ODE has three regular singularities. But not all ODEs with two regular singularities and one essential singularity can be transformed into an ODE of the Mathieu type.

Example 13.6.2 THE QUANTUM PENDULUM

A plane pendulum of length l and mass m with gravitational potential $V(\theta) = -mgl\cos\theta$ is called a **quantum pendulum** if its wave function Ψ obeys the Schrödinger equation

$$-\frac{\hbar^2}{2ml^2}\frac{d^2\Psi}{d\theta^2} + \big[V(\theta) - E\big]\Psi = 0, \tag{13.184}$$

where the variable θ is the angular displacement from the vertical direction. (For further details and illustrations we refer to Gutiérrez-Vega *et al*. in the Additional Readings.) A boundary condition applies to Ψ so as to be single-valued; that is, $\Psi(\theta + 2\pi) = \Psi(\theta)$. Substituting

$$\theta = 2\eta, \qquad \lambda = \frac{8Eml^2}{\hbar^2}, \qquad q = -\frac{4m^2gl^3}{\hbar^2} \tag{13.185}$$

into the Schrödinger equation yields the angular Mathieu ODE for $\Psi(2(\eta + \pi)) = \Psi(2\eta)$. ■

For many other applications involving Mathieu functions we refer to Ruby in the Additional Readings.

Our main focus will be on the solutions of the angular Mathieu ODE, which has the important property that its coefficient function is periodic with period π.

General Properties of Mathieu Functions

In physics applications the angular Mathieu functions are required to be single-valued, that is, periodic with period 2π. Let us start with some nomenclature. Since Mathieu's ODEs are invariant under parity ($\eta \to -\eta$), Mathieu functions have definite parity. Those of odd parity that have period 2π and, for small q, start with $\sin(2n+1)\eta$ are called $\mathrm{se}_{2n+1}(\eta, q)$, with n an integer, $n = 0, 1, 2, \ldots$ (se is short for sine-elliptic). Mathieu functions of odd parity and period π that start with $\sin 2n\eta$ for small q are called $\mathrm{se}_{2n}(\eta, q)$, with $n = 1, 2, \ldots$. Mathieu functions of even parity, period π that start with $\cos 2n\eta$ for small q are called $\mathrm{ce}_{2n}(\eta, q)$ (ce is short for cosine-elliptic), while those with period 2π that start with $\cos(2n+1)\eta$, $n = 0, 1, \ldots$, for small q are called $\mathrm{ce}_{2n+1}(\eta, q)$. In the limit where the parameter $q \to 0$ (and the Mathieu ODE becomes the classical harmonic oscillator ODE), Mathieu functions reduce to these trigonometric functions.

The periodicity condition $\Phi(\eta + 2\pi) = \Phi(\eta)$ is sufficient to determine a set of eigenvalues λ in terms of q. An elementary analog of this result is the fact that a solution of the classical harmonic oscillator ODE $u''(\eta) + \lambda u(\eta) = 0$ has period 2π if, and only if, $\lambda = n^2$ is the square of an integer. Such problems will be pursued in Section 14.7 as applications of Fourier series.

Example 13.6.3 RADIAL MATHIEU FUNCTIONS

Upon replacing the angular elliptic variable $\eta \to i\xi$, the angular Mathieu ODE, Eq. (13.181), becomes the radial ODE, Eq. (13.180). This motivates the definitions of radial Mathieu functions as

$$\mathrm{Ce}_{2n+p}(\xi, q) = \mathrm{ce}_{2n+p}(i\xi, q), \qquad p = 0, 1; \qquad n = 0, 1, \ldots,$$
$$\mathrm{Se}_{2n+p}(\xi, q) = -i\,\mathrm{se}_{2n+p}(i\xi, q), \qquad p = 0, 1; \qquad n = 1, 2, \ldots.$$

Because these functions are differentiable, they correspond to the regular solutions of the radial Mathieu ODE. Of course, they are no longer periodic but are oscillatory (Fig. 13.8).

In physical problems involving elliptical coordinates, the radial Mathieu ODE, Eq. (13.180), plays a role corresponding to Bessel's ODE in cylindrical geometry. Because there are four families of independent Bessel functions — the regular solutions J_n and irregular Neumann functions N_n, along with the modified Bessel functions I_n and K_n — we expect four kinds of radial Mathieu functions. Because of parity, the solutions split into even and odd Mathieu functions and so there are eight kinds. For $q > 0$,

$$\mathrm{Je}_{2n}(\xi, q) = \mathrm{Ce}_{2n}(\xi, q), \qquad \mathrm{Je}_{2n+1}(\xi, q) = \mathrm{Ce}_{2n+1}(\xi, q),$$
$$\mathrm{Jo}_{2n}(\xi, q) = \mathrm{Se}_{2n}(\xi, q), \qquad \mathrm{Jo}_{2n+1}(\xi, q) = \mathrm{Se}_{2n+1}(\xi, q), \qquad \text{regular or first kind;}$$
$$\mathrm{Ne}_n(\xi, q), \mathrm{No}_n(\xi, q), \qquad \text{irregular or second kind;}$$

for $q < 0$, the solutions of the radial Mathieu ODE are denoted by

$$\mathrm{Ie}_n(\xi, q), \mathrm{Io}_n(\xi, q), \qquad \text{regular or first kind,}$$
$$\mathrm{Ke}_n(\xi, q), \mathrm{Ko}_n(\xi, q), \qquad \text{irregular or second kind}$$

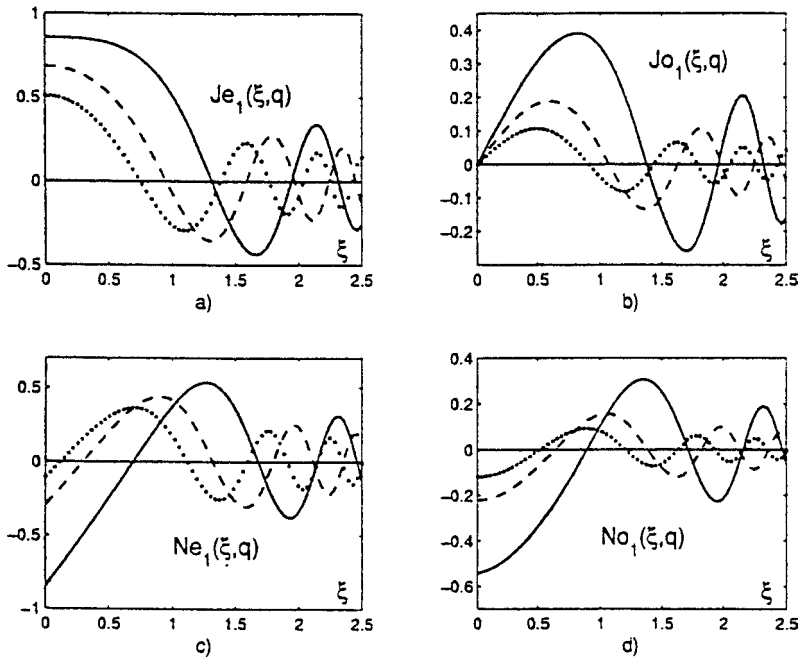

FIGURE 13.8 Radial Mathieu functions: $q = 1$ (solid line), $q = 2$ (dashed line), $q = 3$ (dotted line). (From Gutiérrez-Vega *et al., Am. J. Phys.* **71**: 233 (2003).)

and are known as the **evanescent radial Mathieu functions**. Mathieu functions corresponding to the Hankel functions can be similarly defined. In Fig. 13.8 some of them are plotted.

In applications such as a vibrating drumhead with elliptical boundary conditions (see Example 13.6.1), the solution can be expanded in even and odd Mathieu functions:

$$ze_n \equiv \mathrm{Je}_n(\xi, q)\mathrm{ce}_n(\eta, q)\cos(\omega_n t), \qquad m \geq 0,$$
$$zo_n \equiv \mathrm{Jo}_n(\xi, q)\mathrm{se}_n(\eta, q)\cos(\omega_n t), \qquad m \geq 1.$$

They obey Dirichlet boundary conditions, $ze_n(\xi_0, \eta, t) = 0 = zo_n(\xi_0, \eta, t)$, which hold provided the radial functions satisfy $\mathrm{Je}_n(\xi_0, q) = 0 = \mathrm{Jo}_n(\xi_0, q)$ at the elliptical boundary, where $\xi = \xi_0$.

When the focal distance $c \to 0$, the angular Mathieu functions become the conventional trigonometric functions, while the radial Mathieu functions become Bessel functions.

In the case of oscillations of a confocal annular elliptic lake, the modes have to include the Mathieu functions of the second kind and are thus given by

$$ze_n \equiv \big[A\mathrm{Je}_n(\xi, q) + B\mathrm{Ne}_n(\xi, q)\big]\mathrm{ce}_n(\eta, q)\cos(\omega_n t), \qquad m \geq 0,$$
$$zo_n \equiv \big[A\mathrm{Jo}_n(\xi, q) + B\mathrm{No}_n(\xi, q)\big]\mathrm{se}_n(\eta, q)\cos(\omega_n t), \qquad m \geq 1,$$

with A, B constants. These standing wave solutions must obey Neumann boundary conditions at the inner ($\xi = \xi_0$) and outer ($\xi = \xi_1$) elliptical boundaries; that is, the normal

derivatives (a prime denotes $d/d\xi$) of ze_n and zo_n vanish at each point of the boundaries. For even modes, we have $\mathrm{ze}_n'(\xi_1, m, q) = 0 = \mathrm{ze}_n'(\xi_2, m, q)$. The implied radial constraints are similar to Eqs. (11.81) and (11.82) of Example 11.3.1. Numerical examples and plots, also for traveling waves, are given in Gutiérrez-Vega *et al*. in the Additional Readings. ■

For zeros of Mathieu functions, their asymptotic expansions, and a more complete listing of formulas we refer to Abramowitz and Stegun (AMS-55) in the Additional Readings, *Am. J. Phys.* **71**, Jahnke and Emde and Gradshteyn and Ryzhik in the Additional Readings.

To illustrate and support the nomenclature, we want to show[16] that there is an angular Mathieu function that is

- even in η and of period π if and only if $\Phi_1'(\pi/2) = 0$;
- odd and of period π if and only if $\Phi_2(\pi/2) = 0$;
- even and of period 2π if and only if $\Phi_1(\pi/2) = 0$;
- odd and of period 2π if and only if $\Phi_2'(\pi/2) = 0$,

where $\Phi_1(\eta), \Phi_2(\eta)$ are two linearly independent solutions of the angular Mathieu ODE so that

$$\Phi_1(0) = 1, \quad \Phi_1'(0) = 0; \qquad \Phi_2(0) = 0, \quad \Phi_2'(0) = 1. \tag{13.186}$$

Since the Mathieu ODE is a linear second-order ODE, we know (Chapter 9) that these initial conditions are realistic. The first case just given corresponds to $\mathrm{ce}_{2n}(\eta, q)$, with $\Phi_1'(\pi/2) = -2n \sin 2n\eta|_{\eta=\pi/2} + \cdots = 0$ for $n = 1, 2, \ldots$. The second is the $\mathrm{se}_{2n}(\eta, q)$, with $\Phi_2(\pi/2) = \sin 2n\eta|_{\pi/2} + \cdots = 0$. The third case is the $\mathrm{ce}_{2n+1}(\eta, q)$, with $\Phi_1(\pi/2) = \cos(2n+1)\pi/2 + \cdots = 0$. The fourth case is the $\mathrm{se}_{2n+1}(\eta, q)$.

The key to the proof is Floquet's approach to linear second-order ODEs with periodic coefficient functions, such as Mathieu's angular ODE or the simple pendulum (Exercise 13.6.1). If $\Phi_1(\eta), \Phi_2(\eta)$ are two linearly independent solutions of the ODE, any other solution Φ can be expressed as

$$\Phi(\eta) = c_1\Phi_1(\eta) + c_2\Phi_2(\eta), \tag{13.187}$$

with constants c_1, c_2. Now, $\Phi_k(\eta + 2\pi)$ are also solutions because such an ODE is invariant under the translation $\eta \to \eta + 2\pi$, and in particular

$$\begin{aligned} \Phi_1(\eta + 2\pi) &= a_1\Phi_1(\eta) + a_2\Phi_2(\eta), \\ \Phi_2(\eta + 2\pi) &= b_1\Phi_1(\eta) + b_2\Phi_2(\eta), \end{aligned} \tag{13.188}$$

with constants a_i, b_j. Substituting Eq. (13.188) into Eq. (13.187) we get

$$\Phi(\eta + 2\pi) = (c_1a_1 + c_2b_1)\Phi_1(\eta) + (c_2b_2 + c_1a_2)\Phi_2(\eta), \tag{13.189}$$

where the constants c_i can be chosen as solutions of the eigenvalue equations

$$\begin{aligned} a_1c_1 + b_1c_2 &= \lambda c_1, \\ a_2c_1 + b_2c_2 &= \lambda c_2. \end{aligned} \tag{13.190}$$

[16] See Hochstadt in the Additional Readings.

$$= 0. \tag{13.191}$$

A useful **corollary** is obtained if we define μ and y by $\lambda = \exp(2\pi\mu)$ and $y(\eta) \equiv \exp(-\mu\eta)\Phi(\eta)$, so

Then **Floquet's theorem** states that $\Phi(\eta + 2\pi) = \lambda\Phi(\eta)$, where $y(\eta)$ is a root of (13.192)

Thus, $\Phi(\eta) = e^{\mu\eta}y(\eta)$, with y a periodic function of η with period 2π.

Let us apply Floquet's argument to the $\Phi_k(\eta + \pi)$, which are also solutions of Mathieu's ODE because the latter is invariant under the translation $\eta \to \eta + \pi$. Using the special values in Eq. (13.186) we know that

$$\begin{aligned} \Phi_1(\eta + \pi) &= \Phi_1(\pi)\Phi_1(\eta) + \Phi_1'(\pi)\Phi_2(\eta), \\ \Phi_2(\eta + \pi) &= \Phi_2(\pi)\Phi_1(\eta) + \Phi_2'(\pi)\Phi_2(\eta), \end{aligned} \tag{13.193}$$

because these linear combinations of $\Phi_k(\eta)$ are solutions of Mathieu's ODE with the correct values $\Phi_i(\eta + \pi)$, $\Phi_i'(\eta + \pi)$ for $\eta = 0$. Therefore,

$$\Phi_i(\eta + \pi) = \lambda_i \Phi_i(\eta), \tag{13.194}$$

where the λ_i are the roots of

$$\begin{vmatrix} \Phi_1(\pi) - \lambda & \Phi_2(\pi) \\ \Phi_1'(\pi) & \Phi_2'(\pi) - \lambda \end{vmatrix} = 0. \tag{13.195}$$

The constant term in the characteristic polynomial is given by the Wronskian

$$W\big(\Phi_1(\eta), \Phi_2(\eta)\big) = C, \tag{13.196}$$

a constant because the coefficient of $d\Phi/d\eta$ in the angular Mathieu ODE vanishes, implying $dW/d\eta = 0$. In fact, using Eq. (13.186),

$$\begin{aligned} W\big(\Phi_1(0), \Phi_2(0)\big) &= \Phi_1(0)\Phi_2'(0) - \Phi_1'(0)\Phi_2(0) = 1 \\ &= W\big(\Phi_1(\pi), \Phi_2(\pi)\big), \end{aligned} \tag{13.197}$$

so the eigenvalue Eq. (13.195) for λ becomes

$$\lambda^2 - \big[\Phi_1(\pi) + \Phi_2'(\pi)\big]\lambda + 1 = 0 \tag{13.198}$$

with $\lambda_1\lambda_2 = 1$ and $\lambda_1 + \lambda_2 = \Phi_1(\pi) + \Phi_2'(\pi)$. If $|\lambda_1| = |\lambda_2| = 1$, then $\lambda_2 = \exp(i\phi)$ and $\lambda_2 = \exp(-i\phi)$, so $\lambda_1 + \lambda_2 = 2\cos\phi$. For $\phi \neq 0, \pm 2\pi$, this case corresponds to $|\Phi_i(\eta)|$ bounded as $\eta \to \infty$ in steps of π using Eq. (13.194). These cases do not yield periodic Mathieu functions, and this is also the case when $|\Phi_i(\pi)| > 1$ and the Φ_i have period 2π with $\Phi_i(\eta + 2\pi) = \Phi_i(\eta)$. Thus, then $|\Phi_1(\pi) + \Phi_2'(\pi)| \geq 2$, where both solutions remain

Because the angular Mathieu ODE is invariant under a parity transformation $\eta \to -\eta$, it is convenient to consider solutions

$$\Phi_e(\eta) = \frac{1}{2}\big[\Phi(\eta) + \Phi(-\eta)\big], \qquad \Phi_o(\eta) = \frac{1}{2}\big[\Phi(\eta) - \Phi(-\eta)\big] \tag{13.199}$$

of definite parity, which obey the same initial conditions as Φ_i. We now relabel $\Phi_e \to \Phi_1$, $\Phi_o \to \Phi_2$, taking Φ_1 to be even and Φ_2 to be odd under parity. These solutions of definite parity of Mathieu's ODE are called **Mathieu functions** and are labeled according to our nomenclature discussed earlier.

If $\Phi_1(\eta)$ has period π, then $\Phi_1'(\eta + \pi)$ also has period π but is odd under parity. Substituting $\eta = -\pi/2$ we obtain

$$\Phi_1'\left(\frac{\pi}{2}\right) = \Phi_1'\left(-\frac{\pi}{2}\right) = -\Phi_1'\left(\frac{\pi}{2}\right), \qquad \text{so} \qquad \Phi_1'\left(\frac{\pi}{2}\right) = 0. \tag{13.200}$$

Conversely, if $\Phi_1'(\pi/2) = 0$, then $\Phi_1(\eta)$ has period π. To see this, we use

$$\Phi_1(\eta + \pi) = c_1\Phi_1(\eta) + c_2\Phi_2(\eta). \tag{13.201}$$

This expansion is valid because $\Phi_1(\eta + \pi)$ is a solution of the angular Mathieu ODE. We now determine the coefficients c_i, setting $\eta = -\pi/2$, and recall that Φ_1 and Φ_2 are even and are odd under parity, whereas Φ_2 and Φ_1 are odd. This yields

$$\begin{aligned} \Phi_1\left(\frac{\pi}{2}\right) &= c_1\Phi_1\left(\frac{\pi}{2}\right) - c_2\Phi_2\left(\frac{\pi}{2}\right), \\ \Phi_1'\left(\frac{\pi}{2}\right) &= -c_1\Phi_1'\left(\frac{\pi}{2}\right) + c_2\Phi_2'\left(\frac{\pi}{2}\right). \end{aligned} \tag{13.202}$$

Since $\Phi_1'(\pi/2) = 0$, $\Phi_2(\pi/2) \neq 0$, or the Wronskian would vanish and $\Phi_2 \sim \Phi_1$ would follow. Hence $c_2 = 1$ follows from the second equation and $c_1 = 1$ from the first. Thus, $\Phi_1(\eta + \pi) = \Phi_1(\eta)$. The other bulleted cases listed earlier can be proved similarly.

Because the Mathieu ODEs are of the Sturm–Liouville type, Mathieu functions represent orthogonal systems of functions. So, for m, n nonnegative integers, the orthogonality relations and normalizations are

$$\begin{aligned} &\int_{-\pi}^{\pi} \mathrm{ce}_m \mathrm{ce}_n \, d\eta = \int_{-\pi}^{\pi} \mathrm{se}_m \mathrm{se}_n \, d\eta = 0, \qquad \text{if } m \neq n; \\ &\int_{-\pi}^{\pi} \mathrm{ce}_m \mathrm{se}_n \, d\eta = 0; \\ &\int_{-\pi}^{\pi} [\mathrm{ce}_{2n}]^2 \, d\eta = \int_{-\pi}^{\pi} [\mathrm{se}_{2n}]^2 \, d\eta = \pi, \qquad \text{if } n \geq 1; \qquad \int_0^{\pi} \big[\mathrm{ce}_0(\eta, q)\big]^2 \, d\eta = \pi. \end{aligned} \tag{13.203}$$

If a function $f(\eta)$ is periodic with period π, then it can be expanded in a series of orthogonal Mathieu functions as

$$f(\eta) = \frac{1}{2}a_0 \, \mathrm{ce}_0(\eta, q) + \sum_{n=1}^{\infty}\big[a_n \, \mathrm{ce}_{2n}(\eta, q) + b_n \, \mathrm{se}_{2n}(\eta, q)\big] \tag{13.204}$$

with

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(\eta)\mathrm{ce}_{2n}(\eta, q)\,d\eta, \qquad n \geq 0;$$
$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(\eta)\mathrm{se}_{2n}(\eta, q)\,d\eta, \qquad n \geq 1. \tag{13.205}$$

Similar expansions exist for functions of period 2π in terms of ce_{2n+1} and se_{2n+1}.

Series expansions of Mathieu functions will be derived in Section 14.7.

Exercises

13.6.1 For the simple pendulum ODE of Section 5.8, apply Floquet's method and derive the properties of its solutions similar to those marked by bullets before Eq. (13.186).

13.6.2 Derive a Mathieu function analog for the Rayleigh expansion of a plane wave forcos $(k\cos\eta\cos\theta)$ and $\sin(k\cos\eta\cos\theta)$.

Additional Readings

Abramowitz, M., and I. A. Stegun, eds., *Handbook of Mathematical Functions*, Applied Mathematics Series-55 (AMS-55). Washington, DC: National Bureau of Standards (1964). Paperback edition, New York: Dover (1974). Chapter 22 is a detailed summary of the properties and representations of orthogonal polynomials. Other chapters summarize properties of Bessel, Legendre, hypergeometric, and confluent hypergeometric functions and much more.

Buchholz, H., *The Confluent Hypergeometric Function*. New York: Springer-Verlag (1953); translated (1969). Buchholz strongly emphasizes the Whittaker rather than the Kummer forms. Applications to a variety of other transcendental functions.

Erdelyi, A., W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions*, 3 vols. New York: McGraw-Hill (1953). Reprinted Krieger (1981). A detailed, almost exhaustive listing of the properties of the special functions of mathematical physics.

Fox, L. and I. B. Parker, *Chebyshev Polynomials in Numerical Analysis*. Oxford: Oxford University Press (1968). A detailed, thorough, but very readable account of Chebyshev polynomials and their applications in numerical analysis.

Gradshteyn, I. S., and I. M. Ryzhik, *Table of Integrals, Series and Products*, New York: Academic Press (1980).

Gutiérrez-Vega, J. C., R. M. Rodríguez-Dagnino, M. A. Meneses-Nava and S. Chávez-Cerda, *Am. J. Phys.* **71**: 233 (2003).

Hochstadt, H., *Special Functions of Mathematical Physics*. New York: Holt, Rinehart and Winston (1961), reprinted Dover (1986).

Jahnke, E., and F. Emde, *Table of Functions*. Leipzig: Teubner (1933); New York: Dover (1943).

Lebedev, N. N., *Special Functions and their Applications* (translated by R. A. Silverman). Englewood Cliffs, NJ: Prentice-Hall (1965). Paperback, New York: Dover (1972).

Luke, Y. L., *The Special Functions and Their Approximations*. New York: Academic Press (1969). Two volumes: Volume 1 is a thorough theoretical treatment of gamma functions, hypergeometric functions, confluent hypergeometric functions, and related functions. Volume 2 develops approximations and other techniques for numerical work.

Luke, Y. L., *Mathematical Functions and Their Approximations*. New York: Academic Press (1975). This is an updated supplement to *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables* (AMS-55).

Mathieu, E., *J. de Math. Pures et Appl.* **13**: 137–203 (1868).

McLachlan, N. W., *Theory and Applications of Mathieu Functions*. Oxford, UK: Clarendon Press (1947).

Magnus, W., F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics*. New York: Springer (1966). An excellent summary of just what the title says, including the topics of Chapters 10 to 13.

Rainville, E. D., *Special Functions*. New York: Macmillan (1960), reprinted Chelsea (1971). This book is a coherent, comprehensive account of almost all the special functions of mathematical physics that the reader is likely to encounter.

Rowland, D. R., *Am. J. Phys.* **72**: 758–766 (2004).

Ruby, L., *Am. J. Phys.* **64**: 39–44 (1996).

Sansone, G., *Orthogonal Functions* (translated by A. H. Diamond). New York: Interscience (1959). Reprinted Dover (1991).

Slater, L. J., *Confluent Hypergeometric Functions*. Cambridge, UK: Cambridge University Press (1960). This is a clear and detailed development of the properties of the confluent hypergeometric functions and of relations of the confluent hypergeometric equation to other ODEs of mathematical physics.

Sneddon, I. N., *Special Functions of Mathematical Physics and Chemistry*, 3rd ed. New York: Longman (1980).

Whittaker, E. T., and G. N. Watson, *A Course of Modern Analysis*. Cambridge, UK: Cambridge University Press, reprinted (1997). The classic text on special functions and real and complex analysis.

CHAPTER 14

FOURIER SERIES

14.1 GENERAL PROPERTIES

Periodic phenomena involving waves, rotating machines (harmonic motion), or other repetitive driving forces are described by periodic functions. Fourier series are a basic tool for solving ordinary differential equations (ODEs) and partial differential equations (PDEs) with periodic boundary conditions. Fourier integrals for nonperiodic phenomena are developed in Chapter 15. The common name for the field is **Fourier analysis**.

A Fourier series is defined as an expansion of a function or representation of a function in a series of sines and cosines, such as

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + \sum_{n=1}^{\infty} b_n \sin nx. \tag{14.1}$$

The coefficients a_0, a_n, and b_n are related to the periodic function $f(x)$ by definite integrals:

$$a_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos nx \, dx, \tag{14.2}$$

$$b_n = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin nx \, dx, \qquad n = 0, 1, 2, \ldots, \tag{14.3}$$

which are subject to the requirement that the integrals exist. Notice that a_0 is singled out for special treatment by the inclusion of the factor $\frac{1}{2}$. This is done so that Eq. (14.2) will apply to all $a_n, n = 0$ as well as $n > 0$.

The conditions imposed on $f(x)$ to make Eq. (14.1) valid are that $f(x)$ have only a finite number of finite discontinuities and only a finite number of extreme values, maxima, and minima in the interval $[0, 2\pi]$.[1] Functions satisfying these conditions may be called

[1] These conditions are **sufficient** but not **necessary**.

piecewise regular. The conditions themselves are known as the **Dirichlet conditions.** Although there are some functions that do not obey these Dirichlet conditions, they may well be labeled pathological for purposes of Fourier expansions. In the vast majority of physical problems involving a Fourier series these conditions will be satisfied. In most physical problems we shall be interested in functions that are square integrable (in the Hilbert space L^2 of Section 10.4). In this space the sines and cosines form a complete orthogonal set. And this in turn means that Eq. (14.1) is valid, in the sense of convergence in the mean.

Expressing $\cos nx$ and $\sin nx$ in exponential form, we may rewrite Eq. (14.1) as

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx}, \tag{14.4}$$

in which

$$c_n = \frac{1}{2}(a_n - ib_n), \qquad c_{-n} = \frac{1}{2}(a_n + ib_n), \qquad n > 0, \tag{14.5a}$$

and

$$c_0 = \frac{1}{2}a_0. \tag{14.5b}$$

Complex Variables — Abel's Theorem

Consider a function $f(z)$ represented by a convergent power series

$$f(z) = \sum_{n=0}^{\infty} C_n z^n = \sum_{n=0}^{\infty} C_n r^n e^{in\theta}. \tag{14.6}$$

This is our Fourier exponential series, Eq. (14.4). Separating real and imaginary parts we get

$$u(r,\theta) = \sum_{n=0}^{\infty} C_n r^n \cos n\theta, \qquad v(r,\theta) = \sum_{n=1}^{\infty} C_n r^n \sin n\theta, \tag{14.7a}$$

the Fourier cosine and sine series. Abel's theorem asserts that if $u(1,\theta)$ and $v(1,\theta)$ are convergent for a given θ, then

$$u(1,\theta) + iv(1,\theta) = \lim_{r\to 1} f\left(re^{i\theta}\right). \tag{14.7b}$$

An application of this appears as Exercise 14.1.9 and in Example 14.1.1.

Example 14.1.1 SUMMATION OF A FOURIER SERIES

Usually in this chapter we shall be concerned with finding the coefficients of the Fourier expansion of a known function. Occasionally, we may wish to reverse this process and determine the function represented by a given Fourier series.

Consider the series $\sum_{n=1}^{\infty}(1/n)\cos nx, x \in (0, 2\pi)$. Since this series is only conditionally convergent (and diverges at $x = 0$), we take

$$\sum_{n=1}^{\infty}\frac{\cos nx}{n} = \lim_{r\to 1}\sum_{n=1}^{\infty}\frac{r^n\cos nx}{n}, \tag{14.8}$$

absolutely convergent for $|r| < 1$. Our procedure is to try forming power series by transforming the trigonometric functions into exponential form:

$$\sum_{n=1}^{\infty}\frac{r^n\cos nx}{n} = \frac{1}{2}\sum_{n=1}^{\infty}\frac{r^n e^{inx}}{n} + \frac{1}{2}\sum_{n=1}^{\infty}\frac{r^n e^{-inx}}{n}. \tag{14.9}$$

Now, these power series may be identified as Maclaurin expansions of $-\ln(1-z), z = re^{ix}, re^{-ix}$ (Eq. (5.95)), and

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{r^n\cos nx}{n} &= -\frac{1}{2}\left[\ln\left(1-re^{ix}\right)+\ln\left(1-re^{-ix}\right)\right]\\ &= -\ln\left[\left(1+r^2\right)-2r\cos x\right]^{1/2}.\end{aligned} \tag{14.10}$$

Letting $r = 1$ and using Abel's theorem, we see that

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{\cos nx}{n} &= -\ln(2-2\cos x)^{1/2}\\ &= -\ln\left(2\sin\frac{x}{2}\right), \qquad x \in (0, 2\pi).^{2}\end{aligned} \tag{14.11}$$

Both sides of this expression diverge as $x \to 0$ and 2π. ■

Completeness

The problem of establishing completeness may be approached in a number of different ways. One way is to transform the trigonometric Fourier series into exponential form and to compare it with a Laurent series. If we expand $f(z)$ in a Laurent series[3] (assuming $f(z)$ is analytic),

$$f(z) = \sum_{n=-\infty}^{\infty} d_n z^n. \tag{14.12}$$

On the unit circle $z = e^{i\theta}$ and

$$f(z) = f(e^{i\theta}) = \sum_{n=-\infty}^{\infty} d_n e^{in\theta}. \tag{14.13}$$

[2]The limits may be shifted to $[-\pi, \pi]$ (and $x \neq 0$) using $|x|$ on the right-hand side.
[3]Section 6.5.

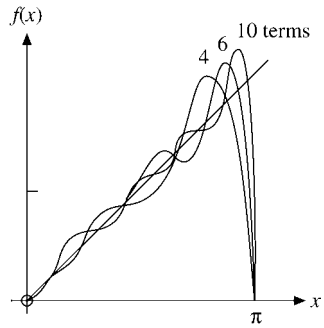

FIGURE 14.1 Fourier representation of sawtooth wave.

The Laurent expansion on the unit circle (Eq. (14.13)) has the same form as the complex Fourier series (Eq. (14.12)), which shows the equivalence between the two expansions. Since the Laurent series as a power series has the property of completeness, we see that the Fourier functions e^{inx} form a complete set. There is a significant limitation here. Laurent series and complex power series cannot handle discontinuities such as a square wave or the sawtooth wave of Fig. 14.1, except on the circle of convergence.

The theory of vector spaces provides a second approach to the completeness of the sines and cosines. Here completeness is established by the Weierstrass theorem for two variables.

The Fourier expansion and the completeness property may be expected, for the functions $\sin nx, \cos nx, e^{inx}$ are all eigenfunctions of a self-adjoint linear ODE,

$$y'' + n^2 y = 0. \tag{14.14}$$

We obtain orthogonal eigenfunctions for different values of the eigenvalue n for the interval $[0, 2\pi]$ that satisfy the boundary conditions in the Sturm–Liouville theory (Chapter 10). Different eigenfunctions for the same eigenvalue n are orthogonal. We have

$$\int_0^{2\pi} \sin mx \sin nx\, dx = \begin{cases} \pi\delta_{mn}, & m \neq 0, \\ 0, & m = 0, \end{cases} \tag{14.15}$$

$$\int_0^{2\pi} \cos mx \cos nx\, dx = \begin{cases} \pi\delta_{mn}, & m \neq 0, \\ 2\pi, & m = n = 0, \end{cases} \tag{14.16}$$

$$\int_0^{2\pi} \sin mx \cos nx\, dx = 0 \qquad \text{for all integral } m \text{ and } n. \tag{14.17}$$

Note that any interval $x_0 \leq x \leq x_0 + 2\pi$ will be equally satisfactory. Frequently, we shall use $x_0 = -\pi$ to obtain the interval $-\pi \leq x \leq \pi$. For the complex eigenfunctions $e^{\pm inx}$ orthogonality is usually **defined** in terms of the complex conjugate of one of the two factors,

$$\int_0^{2\pi} \left(e^{imx}\right)^* e^{inx}\, dx = 2\pi\delta_{mn}. \tag{14.18}$$

This agrees with the treatment of the spherical harmonics (Section 12.6).

Sturm–Liouville Theory

The Sturm–Liouville theory guarantees the validity of Eq. (14.1) (for functions satisfying the Dirichlet conditions) and, by use of the orthogonality relations, Eqs. (14.15), (14.16), and (14.17), allows us to compute the expansion coefficients a_n, b_n, as shown in Eqs. (14.2), and (14.3). Substituting Eqs. (14.2) and (14.3) into Eq. (14.1), we write our Fourier expansion as

$$\begin{aligned} f(x) &= \frac{1}{2\pi}\int_0^{2\pi} f(t)\,dt \\ &\quad + \frac{1}{\pi}\sum_{n=1}^{\infty}\left(\cos nx\int_0^{2\pi} f(t)\cos nt\,dt + \sin nx\int_0^{2\pi} f(t)\sin nt\,dt\right) \\ &= \frac{1}{2\pi}\int_0^{2\pi} f(t)\,dt + \frac{1}{\pi}\sum_{n=1}^{\infty}\int_0^{2\pi} f(t)\cos n(t-x)\,dt, \end{aligned} \tag{14.19}$$

the first (constant) term being the average value of $f(x)$ over the interval $[0, 2\pi]$. Equation (14.19) offers one approach to the development of the Fourier integral and Fourier transforms, Section 15.1.

Another way of describing what we are doing here is to say that $f(x)$ is part of an infinite-dimensional Hilbert space, with the orthogonal $\cos nx$ and $\sin nx$ as the basis. (They can always be renormalized to unity if desired.) The statement that $\cos nx$ and $\sin nx$ $(n = 0, 1, 2, \ldots)$ span this Hilbert space is equivalent to saying that they form a complete set. Finally, the expansion coefficients a_n and b_n correspond to the projections of $f(x)$, with the integral inner products (Eqs. (14.2) and (14.3)) playing the role of the dot product of Section 1.3. These points are outlined in Section 10.4.

Example 14.1.2 SAWTOOTH WAVE

An idea of the convergence of a Fourier series and the error in using only a finite number of terms in the series may be obtained by considering the expansion of

$$f(x) = \begin{cases} x, & 0 \le x < \pi, \\ x - 2\pi, & \pi < x \le 2\pi. \end{cases} \tag{14.20}$$

This is a sawtooth wave, and for convenience we shall shift our interval from $[0, 2\pi]$ to $[-\pi, \pi]$. In this interval we have $f(x) = x$. Using Eqs. (14.2) and (14.3), we show the expansion to be

$$f(x) = x = 2\left[\sin x - \frac{\sin 2x}{2} + \frac{\sin 3x}{3} - \cdots + (-1)^{n+1}\frac{\sin nx}{n} + \cdots\right]. \tag{14.21}$$

Figure 14.1 shows $f(x)$ for $0 \le x < \pi$ for the sum of 4, 6, and 10 terms of the series. Three features deserve comment.

1. There is a steady increase in the accuracy of the representation as the number of terms included is increased.

2. All the curves pass through the midpoint, $f(x) = 0$, at $x = \pi$.
3. In the vicinity of $x = \pi$ there is an overshoot that persists and shows no sign of diminishing.

As a matter of incidental interest, setting $x = \pi/2$ in Eq. (14.21) provides an alternate derivation of Leibniz' formula, Exercise 5.7.6. ■

Behavior of Discontinuities

The behavior of the sawtooth wave $f(x)$ at $x = \pi$ is an example of a general rule that at a finite discontinuity the series converges to the arithmetic mean. For a discontinuity at $x = x_0$ the series yields

$$f(x_0) = \tfrac{1}{2}\big[f(x_0 + 0) + f(x_0 - 0)\big], \tag{14.22}$$

the arithmetic mean of the right and left approaches to $x = x_0$. A general proof using partial sums, as in Section 14.5, is given by Jeffreys and Jeffreys and by Carslaw (see the Additional Readings). The proof may be simplified by the use of Dirac delta functions — Exercise 14.5.1.

The overshoot of the sawtooth wave just before $x = \pi$ in Fig. 14.1 is an example of the Gibbs phenomenon, discussed in Section 14.5.

Exercises

14.1.1 A function $f(x)$ (quadratically integrable) is to be represented by a **finite** Fourier series. A convenient measure of the accuracy of the series is given by the integrated square of the deviation,

$$\Delta_p = \int_0^{2\pi} \left[f(x) - \frac{a_0}{2} - \sum_{n=1}^{p} (a_n \cos nx + b_n \sin nx) \right]^2 dx.$$

Show that the requirement that Δ_p be minimized, that is,

$$\frac{\partial \Delta_p}{\partial a_n} = 0, \qquad \frac{\partial \Delta_p}{\partial b_n} = 0,$$

for all n, leads to choosing a_n and b_n as given in Eqs. (14.2) and (14.3).
Note. Your coefficients a_n and b_n are independent of p. This independence is a consequence of orthogonality and would not hold for powers of x, fitting a curve with polynomials.

14.1.2 In the analysis of a complex waveform (ocean tides, earthquakes, musical tones, etc.) it might be more convenient to have the Fourier series written as

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \alpha_n \cos(nx - \theta_n).$$

Show that this is equivalent to Eq. (14.1) with

$$a_n = \alpha_n \cos\theta_n, \qquad \alpha_n^2 = a_n^2 + b_n^2,$$
$$b_n = \alpha_n \sin\theta_n, \qquad \tan\theta_n = b_n/a_n.$$

Note. The coefficients α_n^2 as a function of n define what is called the **power spectrum**. The importance of α_n^2 lies in their invariance under a shift in the phase θ_n.

14.1.3 A function $f(x)$ is expanded in an exponential Fourier series

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx}.$$

If $f(x)$ is real, $f(x) = f^*(x)$, what restriction is imposed on the coefficients c_n?

14.1.4 Assuming that $\int_{-\pi}^{\pi} [f(x)]^2\, dx$ is finite, show that

$$\lim_{m\to\infty} a_m = 0, \qquad \lim_{m\to\infty} b_m = 0.$$

Hint. Integrate $[f(x) - s_n(x)]^2$, where $s_n(x)$ is the nth partial sum, and use Bessel's inequality, Section 10.4. For our finite interval the assumption that $f(x)$ is square integrable ($\int_{-\pi}^{\pi} |f(x)|^2\, dx$ is finite) implies that $\int_{-\pi}^{\pi} |f(x)|\, dx$ is also finite. The converse does not hold.

14.1.5 Apply the summation technique of this section to show that

$$\sum_{n=1}^{\infty} \frac{\sin nx}{n} = \begin{cases} \frac{1}{2}(\pi - x), & 0 < x \le \pi \\ -\frac{1}{2}(\pi + x), & -\pi \le x < 0 \end{cases}$$

(Fig. 14.2).

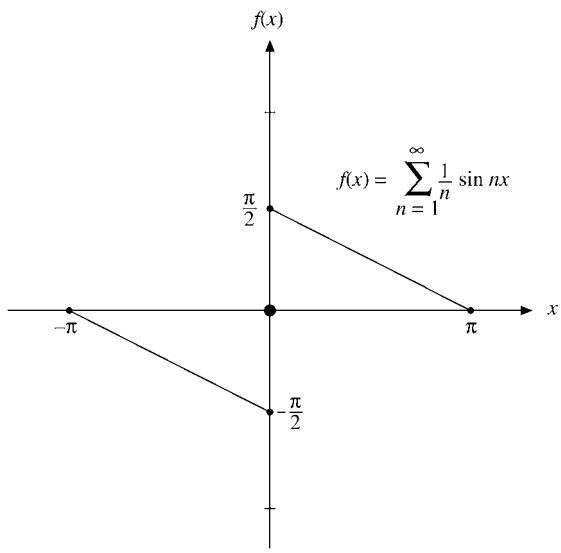

FIGURE 14.2 Reverse sawtooth wave.

14.1.6 Sum the trigonometric series

$$\sum_{n=1}^{\infty}(-1)^{n+1}\frac{\sin nx}{n}$$

and show that it equals $x/2$.

14.1.7 Sum the trigonometric series

$$\sum_{n=0}^{\infty}\frac{\sin(2n+1)x}{2n+1}$$

and show that it equals

$$\begin{cases}\pi/4, & 0<x<\pi\\ -\pi/4, & -\pi<x<0.\end{cases}$$

14.1.8 Calculate the sum of the finite Fourier sine series for the sawtooth wave, $f(x)=x,(-\pi,\pi)$, Eq. (14.21). Use 4-, 6-, 8-, and 10-term series and $x/\pi = 0.00(0.02)1.00$. If a plotting routine is available, plot your results and compare with Fig. 14.1.

14.1.9 Let $f(z)=\ln(1+z)=\sum_{n=1}^{\infty}(-1)^{n+1}z^n/n$. (This series converges to $\ln(1+z)$ for $|z|\le 1$, except at the point $z=-1$.)

(a) From the real parts show that

$$\ln\left(2\cos\frac{\theta}{2}\right)=\sum_{n=1}^{\infty}(-1)^{n+1}\frac{\cos n\theta}{n},\qquad -\pi<\theta<\pi.$$

(b) Using a change of variable, transform part (a) into

$$-\ln\left(2\sin\frac{\theta}{2}\right)=\sum_{n=1}^{\infty}\frac{\cos n\theta}{n},\qquad 0<\theta<2\pi.$$

14.2 ADVANTAGES, USES OF FOURIER SERIES

Discontinuous Functions

One of the advantages of a Fourier representation over some other representation, such as a Taylor series, is that it can represent a discontinuous function. An example is the sawtooth wave in the preceding section. Other examples are considered in Section 14.3 and in the exercises.

Periodic Functions

Related to this advantage is the usefulness of a Fourier series in representing a periodic function. If $f(x)$ has a period of 2π, perhaps it is only natural that we expand it in a series of functions with period $2\pi, 2\pi/2, 2\pi/3, \ldots$. This guarantees that if our periodic $f(x)$ is represented over one interval $[0, 2\pi]$ or $[-\pi,\pi]$, the representation holds for all finite x.

At this point we may conveniently consider the properties of symmetry. Using the interval $[-\pi, \pi]$, $\sin x$ is odd and $\cos x$ is an even function of x. Hence, by Eqs. (14.2) and (14.3),[4] if $f(x)$ is odd, all $a_n = 0$ and if $f(x)$ is even, all $b_n = 0$. In other words,

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx, \qquad f(x) \text{ even}, \tag{14.23}$$

$$f(x) = \sum_{n=1}^{\infty} b_n \sin nx, \qquad f(x) \text{ odd}, \tag{14.24}$$

Frequently these properties are helpful in expanding a given function.

We have noted that the Fourier series is periodic. This is important in considering whether Eq. (14.1) holds outside the initial interval. Suppose we are given only that

$$f(x) = x, \qquad 0 \le x < \pi \tag{14.25}$$

and are asked to represent $f(x)$ by a series expansion. Let us take three of the infinite number of possible expansions.

1. If we assume a Taylor expansion, we have

$$f(x) = x, \tag{14.26}$$

   a one-term series. This (one-term) series is defined for all finite x.
2. Using the Fourier cosine series (Eq. (14.23)), thereby assuming the function is represented faithfully in the interval $[0, \pi)$ and extended to neighboring intervals using the known symmetry properties, we predict that

$$\begin{aligned} f(x) &= -x, & -\pi < x \le 0, \\ f(x) &= 2\pi - x, & \pi < x < 2\pi. \end{aligned} \tag{14.27}$$

3. Finally, from the Fourier sine series (Eq. (14.24)), we have

$$\begin{aligned} f(x) &= x, & -\pi < x \le 0, \\ f(x) &= x - 2\pi, & \pi < x < 2\pi. \end{aligned} \tag{14.28}$$

These three possibilities — Taylor series, Fourier cosine series, and Fourier sine series — are each perfectly valid in the original interval, $[0, \pi]$. Outside, however, their behavior is strikingly different (compare Fig. 14.3). Which of the three, then, is correct? This question has no answer, unless we are given more information about $f(x)$. It may be any of the three or none of them. Our Fourier expansions are valid over the basic interval. Unless the function $f(x)$ is known to be periodic with a period equal to our basic interval or to $(1/n)$th of our basic interval, there is no assurance whatever that the representation (Eq. (14.1)) will have any meaning outside the basic interval.

In addition to the advantages of representing discontinuous and periodic functions, there is a third very real advantage in using a Fourier series. Suppose that we are solving the equation of motion of an oscillating particle subject to a periodic driving force. The Fourier

[4] With the range of integration $-\pi \le x \le \pi$.

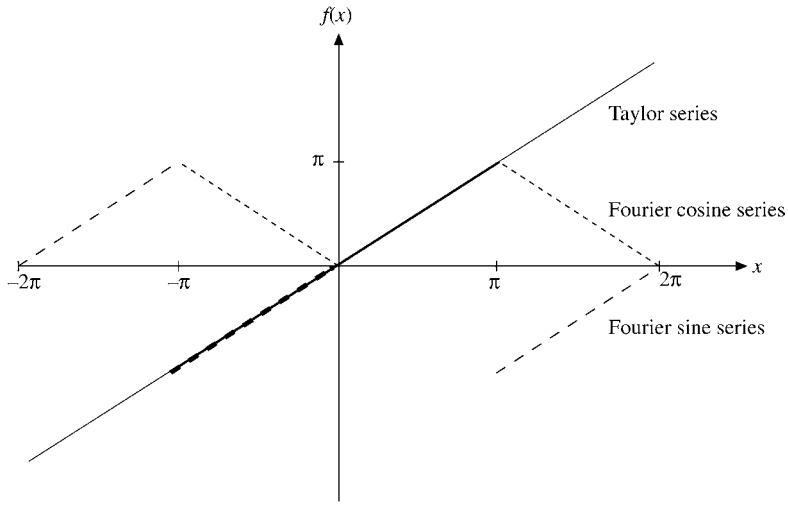

FIGURE 14.3 Comparison of Fourier cosine series, Fourier sine series, and Taylor series.

expansion of the driving force then gives us the fundamental term and a series of harmonics. The (linear) ODE may be solved for each of these harmonics individually, a process that may be much easier than dealing with the original driving force. Then, as long as the ODE is linear, all the solutions may be added together to obtain the final solution.[5] This is more than just a clever mathematical trick.

- It corresponds to finding the response of the system to the fundamental frequency and to each of the harmonic frequencies.

One question that is sometimes raised is: "Were the harmonics there all along, or were they created by our Fourier analysis?" One answer compares the functional resolution into harmonics with the resolution of a vector into rectangular components. The components may have been present, in the sense that they may be isolated and observed, but the resolution is certainly not unique. Hence many authors prefer to say that the harmonics were created by our choice of expansion. Other expansions in other sets of orthogonal functions would give different results. For further discussion we refer to a series of notes and letters in the *American Journal of Physics*.[6]

Change of Interval

So far attention has been restricted to an interval of length 2π. This restriction may easily be relaxed. If $f(x)$ is periodic with a period $2L$, we may write

[5] One of the nastier features of nonlinear differential equations is that this principle of superposition is not valid.

[6] B. L. Robinson, Concerning frequencies resulting from distortion. *Am. J. Phys.* **21**: 391 (1953); F. W. Van Name, Jr., Concerning frequencies resulting from distortion. *ibid.* **22**: 94 (1954).

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left[a_n \cos\frac{n\pi x}{L} + b_n \sin\frac{n\pi x}{L}\right], \tag{14.29}$$

with

$$a_n = \frac{1}{L}\int_{-L}^{L} f(t)\cos\frac{n\pi t}{L}dt, \qquad n = 0, 1, 2, 3, \ldots, \tag{14.30}$$

$$b_n = \frac{1}{L}\int_{-L}^{L} f(t)\sin\frac{n\pi t}{L}dt, \qquad n = 1, 2, 3, \ldots, \tag{14.31}$$

replacing x in Eq. (14.1) with $\pi x/L$ and t in Eqs. (14.2) and (14.3) with $\pi t/L$. (For convenience the interval in Eqs. (14.2) and (14.3) is shifted to $-\pi \le t \le \pi$.) The choice of the symmetric interval $(-L, L)$ is **not** essential. For $f(x)$ periodic with a period of $2L$, any interval $(x_0, x_0 + 2L)$ will do. The choice is a matter of convenience or literally personal preference.

Exercises

14.2.1 The boundary conditions (such as $\psi(0) = \psi(l) = 0$) may suggest solutions of the form $\sin(n\pi x/l)$ and eliminate the corresponding cosines.

(a) Verify that the boundary conditions used in the Sturm–Liouville theory are satisfied for the interval $(0, l)$. Note that this is only half the usual Fourier interval.

(b) Show that the set of functions $\varphi_n(x) = \sin(n\pi x/l),\ n = 1, 2, 3, \ldots,$ satisfies an orthogonality relation

$$\int_0^l \varphi_m(x)\varphi_n(x)\,dx = \frac{l}{2}\delta_{mn}, \qquad n > 0.$$

14.2.2 (a) Expand $f(x) = x$ in the interval $(0, 2L)$. Sketch the series you have found (right-hand side of Ans.) over $(-2L, 2L)$.

$$\text{ANS. } x = L - \frac{2L}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin\left(\frac{n\pi x}{L}\right).$$

(b) Expand $f(x) = x$ as a sine series in the **half** interval $(0, L)$. Sketch the series you have found (right-hand side of Ans.) over $(-2L, 2L)$.

$$\text{ANS. } x = \frac{4L}{\pi}\sum_{n=0}^{\infty}\frac{1}{2n+1}\sin\left(\frac{(2n+1)\pi x}{L}\right).$$

14.2.3 In some problems it is convenient to approximate $\sin\pi x$ over the interval $[0, 1]$ by a parabola $ax(1 - x)$, where a is a constant. To get a feeling for the accuracy of this approximation, expand $4x(1 - x)$ in a Fourier sine series $(-1 \le x \le 1)$:

$$f(x) = \left\{\begin{array}{ll} 4x(1-x), & 0 \le x \le 1 \\ 4x(1+x), & -1 \le x \le 0 \end{array}\right\} = \sum_{n=1}^{\infty} b_n \sin n\pi x.$$

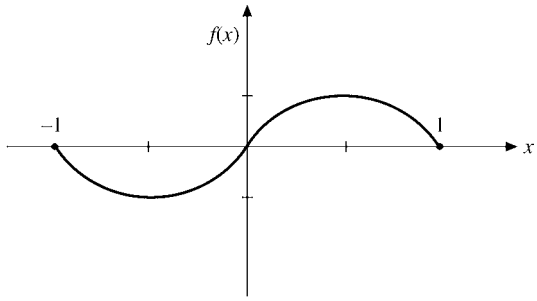

FIGURE 14.4 Parabolic sine wave.

ANS. $b_n = \dfrac{32}{\pi^3} \cdot \dfrac{1}{n^3}$, n odd
$b_n = 0$, n even.

(Fig. 14.4.)

14.3 APPLICATIONS OF FOURIER SERIES

Example 14.3.1 SQUARE WAVE — HIGH FREQUENCIES

One application of Fourier series, the analysis of a "square" wave (Fig. 14.5) in terms of its Fourier components, occurs in electronic circuits designed to handle sharply rising pulses. Suppose that our wave is defined by

$$f(x) = 0, \qquad -\pi < x < 0,$$
$$f(x) = h, \qquad 0 < x < \pi. \tag{14.32}$$

From Eqs. (14.2) and (14.3) we find

$$a_0 = \frac{1}{\pi}\int_0^\pi h\,dt = h, \tag{14.33}$$

$$a_n = \frac{1}{\pi}\int_0^\pi h\cos nt\,dt = 0, \qquad n = 1, 2, 3, \ldots, \tag{14.34}$$

$$b_n = \frac{1}{\pi}\int_0^\pi h\sin nt\,dt = \frac{h}{n\pi}(1 - \cos n\pi); \tag{14.35}$$

$$b_n = \frac{2h}{n\pi}, \qquad n \text{ odd}, \tag{14.36}$$

$$b_n = 0, \qquad n \text{ even}. \tag{14.37}$$

The resulting series is

$$f(x) = \frac{h}{2} + \frac{2h}{\pi}\left(\frac{\sin x}{1} + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + \cdots\right). \tag{14.38}$$

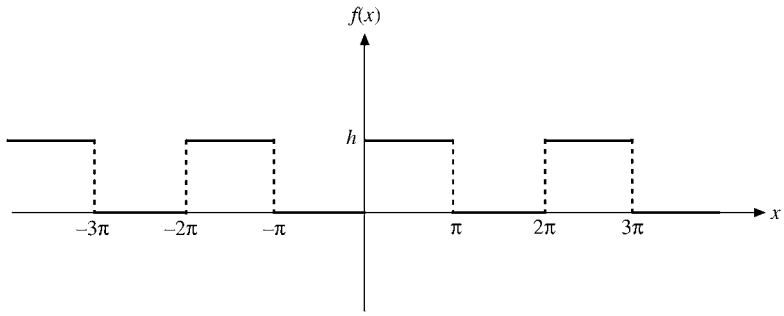

FIGURE 14.5 Square wave.

Except for the first term, which represents an average of $f(x)$ over the interval $[-\pi, \pi]$, all the cosine terms have vanished. Since $f(x) - h/2$ is odd, we have a Fourier sine series. Although only the odd terms in the sine series occur, they fall only as n^{-1}. This **conditional convergence** is like that of the alternating harmonic series. Physically this means that our square wave contains a lot of **high-frequency components**. If the electronic apparatus will not pass these components, our square-wave input will emerge more or less rounded off, perhaps as an amorphous blob. ■

Example 14.3.2 FULL-WAVE RECTIFIER

As a second example, let us ask how well the output of a full-wave rectifier approaches pure direct current (Fig. 14.6). Our rectifier may be thought of as having passed the positive peaks of an incoming sine wave and inverting the negative peaks. This yields

$$\begin{aligned} f(t) &= \sin\omega t, \qquad 0 < \omega t < \pi, \\ f(t) &= -\sin\omega t, \qquad -\pi < \omega t < 0. \end{aligned} \tag{14.39}$$

Since $f(t)$ defined here is even, no terms of the form $\sin n\omega t$ will appear. Again, from Eqs. (14.2) and (14.3), we have

$$\begin{aligned} a_0 &= -\frac{1}{\pi}\int_{-\pi}^{0} \sin\omega t\, d(\omega t) + \frac{1}{\pi}\int_{0}^{\pi} \sin\omega t\, d(\omega t) \\ &= \frac{2}{\pi}\int_{0}^{\pi} \sin\omega t\, d(\omega t) = \frac{4}{\pi}, \end{aligned} \tag{14.40}$$

$$\begin{aligned} a_n &= \frac{2}{\pi}\int_{0}^{\pi} \sin\omega t \cos n\omega t\, d(\omega t) \\ &= -\frac{2}{\pi}\frac{2}{n^2-1}, \qquad n \text{ even}, \\ &= 0, \qquad n \text{ odd}. \end{aligned} \tag{14.41}$$

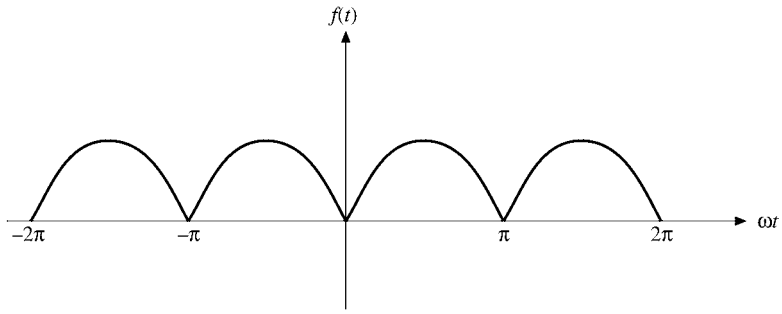

FIGURE 14.6 Full-wave rectifier.

Note that $[0, \pi]$ is not an orthogonality interval for both sines and cosines together and we do not get zero for even n. The resulting series is

$$f(t) = \frac{2}{\pi} - \frac{4}{\pi} \sum_{n=2,4,6,\ldots}^{\infty} \frac{\cos n\omega t}{n^2 - 1}. \tag{14.42}$$

The original frequency, ω, has been eliminated. The lowest-frequency oscillation is 2ω. The high-frequency components fall off as n^{-2}, showing that the full-wave rectifier does a fairly good job of approximating direct current. Whether this good approximation is adequate depends on the particular application. If the remaining ac components are objectionable, they may be further suppressed by appropriate filter circuits. These two examples bring out two features characteristic of Fourier expansions.[7]

- If $f(x)$ has discontinuities (as in the square wave in Example 14.3.1), we can expect the nth coefficient to be decreasing as $\mathcal{O}(1/n)$. Convergence is conditional only.
- If $f(x)$ is continuous (although possibly with discontinuous derivatives, as in the full-wave rectifier of Example 14.3.2), we can expect the nth coefficient to be decreasing as $1/n^2$, that is, absolute convergence. ■

Example 14.3.3 INFINITE SERIES, RIEMANN ZETA FUNCTION

As a final example, we consider the problem of expanding x^2. Let

$$f(x) = x^2, \qquad -\pi < x < \pi. \tag{14.43}$$

Since $f(x)$ is even, all $b_n = 0$. For the a_n we have

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} x^2\, dx = \frac{2\pi^2}{3}, \tag{14.44}$$

[7] G. Raisbeek, Order of magnitude of Fourier coefficients. *Am. Math. Mon.* **62**: 149–155 (1955).

$$a_n = \frac{2}{\pi}\int_0^{\pi} x^2 \cos nx\, dx$$
$$= \frac{2}{\pi}\cdot(-1)^n \frac{2\pi}{n^2}$$
$$= (-1)^n \frac{4}{n^2}. \tag{14.45}$$

From this we obtain

$$x^2 = \frac{\pi^2}{3} + 4\sum_{n=1}^{\infty}(-1)^n \frac{\cos nx}{n^2}. \tag{14.46}$$

As it stands, Eq. (14.46) is of no particular importance. But if we set $x = \pi$,

$$\cos n\pi = (-1)^n \tag{14.47}$$

and Eq. (14.46) becomes[8]

$$\pi^2 = \frac{\pi^2}{3} + 4\sum_{n=1}^{\infty}\frac{1}{n^2}, \tag{14.48}$$

or

$$\frac{\pi^2}{6} = \sum_{n=1}^{\infty}\frac{1}{n^2} \equiv \zeta(2), \tag{14.49}$$

thus yielding the Riemann zeta function, $\zeta(2)$, in closed form (in agreement with the Bernoulli number result of Section 5.9). From our expansion of x^2 and expansions of other powers of x, numerous other infinite series can be evaluated. A few are included in this list of exercises:

	Fourier series	Reference
1.	$\displaystyle\sum_{n=1}^{\infty}\frac{1}{n}\sin nx = \begin{cases} -\frac{1}{2}(\pi + x), & -\pi \le x < 0 \\ \frac{1}{2}(\pi - x), & 0 \le x < \pi \end{cases}$	Exercise 14.1.5 Exercise 14.3.3
2.	$\displaystyle\sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n}\sin nx = \frac{1}{2}x, \qquad -\pi < x < \pi$	Exercise 14.1.6 Exercise 14.3.2
3.	$\displaystyle\sum_{n=0}^{\infty}\frac{1}{2n+1}\sin(2n+1)x = \begin{cases} -\pi/4, & -\pi < x < 0 \\ +\pi/4, & 0 < x < \pi \end{cases}$	Exercise 14.1.7 Eq. (14.38)
4.	$\displaystyle\sum_{n=1}^{\infty}\frac{\cos nx}{n} = -\ln\left[2\sin\left(\frac{\lvert x\rvert}{2}\right)\right], \qquad -\pi < x < \pi$	Eq. (14.11) Exercise 14.1.9(b)
5.	$\displaystyle\sum_{n=1}^{\infty}(-1)^n\frac{1}{n}\cos nx = -\ln\left[2\cos\left(\frac{x}{2}\right)\right], \qquad -\pi < x < \pi$	Exercise 14.1.9(a)
6.	$\displaystyle\sum_{n=0}^{\infty}\frac{1}{2n+1}\cos(2n+1)x = \frac{1}{2}\ln\left[\cot\frac{\lvert x\rvert}{2}\right], \qquad -\pi < x < \pi$	

[8]Note that the point $x = \pi$ is not a point of discontinuity.

The square-wave Fourier series from Eq. (14.38) and item (3) in the table,

$$g(x) = \sum_{n=0}^{\infty} \frac{\sin(2n+1)x}{2n+1} = (-1)^m \frac{\pi}{4}, \qquad m\pi < x < (m+1)\pi, \tag{14.50}$$

can be used to derive Riemann's **functional equation for the zeta function**. Its defining Dirichlet series can be written in various forms:

$$\begin{aligned}\zeta(s) = \sum_{n=1}^{\infty} n^{-s} &= 1 + \sum_{n=1}^{\infty} (2n)^{-s} + \sum_{n=1}^{\infty} (2n+1)^{-s} \\ &= 2^{-s}\zeta(s) + \sum_{n=0}^{\infty} (2n+1)^{-s}\end{aligned}$$

implying that the function $\lambda(s)$ defined in Section 5.9 (along with $\eta(s)$) satisfies

$$\lambda(s) \equiv \sum_{n=0}^{\infty} (2n+1)^{-s} = \left(1 - 2^{-s}\right)\zeta(s). \tag{14.51}$$

Here s is a complex variable. Both Dirichlet series converge for $\sigma = \Re s > 1$. Alternatively, using Eq. (14.51), we have

$$\eta(s) \equiv \sum_{n=1}^{\infty} (-1)^{n-1} n^{-s} = \sum_{n=0}^{\infty} (2n+1)^{-s} - \sum_{n=1}^{\infty} (2n)^{-s} = \left(1 - 2^{1-s}\right)\zeta(s), \tag{14.52}$$

which converges already for $\Re s > 0$ using the Leibniz convergence criterion (see Section 5.3).

Another approach to Dirichlet series starts from Euler's integral for the gamma function,

$$\int_0^{\infty} y^{s-1} e^{-ny}\, dy = n^{-s} \int_0^{\infty} e^{-y} y^{s-1}\, dy = n^{-s}\Gamma(s), \tag{14.53}$$

which may be summed using the geometric series

$$\sum_{n=1}^{\infty} e^{-ny} = \frac{e^{-y}}{1 - e^{-y}} = \frac{1}{e^y - 1}$$

to yield the integral representation for the zeta function:

$$\int_0^{\infty} \frac{y^{s-1}}{e^y - 1} dy = \zeta(s)\Gamma(s). \tag{14.54}$$

If we combine the alternative forms of Eq. (14.53),

$$\int_0^{\infty} y^{s-1} e^{-iny}\, dy = n^{-s}\Gamma(s) e^{-i\pi s/2},$$

$$\int_0^{\infty} y^{s-1} e^{iny}\, dy = n^{-s}\Gamma(s) e^{i\pi s/2},$$

we obtain

$$\int_0^\infty y^{s-1}\sin(ny)\,dy = n^{-s}\Gamma(s)\sin\frac{\pi s}{2}. \tag{14.55}$$

Dividing both sides of Eq. (14.55) by n and summing over all odd n yields, for $\sigma = \Re(s) > 0$,

$$\int_0^\infty g(y)y^{s-1}\,dy = \left(1-2^{-s-1}\right)\zeta(s+1)\Gamma(s)\sin\frac{\pi s}{2}, \tag{14.56}$$

using Eqs. (14.50) and (14.51). Here, the interchange of summation and integration is justified by uniform convergence. This relation is at the heart of the functional equation. If we divide the integration range into intervals $m\pi < y < (m+1)\pi$ and substitute Eq. (14.50) into Eq. (14.56) we find

$$\begin{aligned}\int_0^\infty g(y)y^{s-1}\,dy &= \frac{\pi}{4}\sum_{m=0}^{\infty}(-1)^m\int_{m\pi}^{(m+1)\pi} y^{s-1}\,dy \\ &= \frac{\pi^{s+1}}{4s}\left\{\sum_{m=1}^{\infty}(-1)^m\left[(m+1)^s - m^s\right]+1\right\} \\ &= \frac{\pi^{s+1}}{2s}\left(1-2^{s+1}\right)\zeta(-s),\end{aligned} \tag{14.57}$$

using Eq. (14.52). The series in Eq. (14.57) converges for $\Re s < 1$ to an analytic function. Comparing Eqs. (14.56) and (14.57) for the common area of convergence to analytic functions, $0 < \sigma = \Re s < 1$, we get the **functional equation**

$$\frac{\pi^{s+1}}{2s}\left(1-2^{s+1}\right)\zeta(-s) = \left(1-2^{-s-1}\right)\zeta(s+1)\Gamma(s)\sin\frac{\pi s}{2},$$

which can be rewritten as

$$\zeta(1-s) = 2(2\pi)^{-s}\zeta(s)\Gamma(s)\cos\frac{\pi s}{2}. \tag{14.58}$$

This functional equation provides an analytic continuation of $\zeta(s)$ into the negative half-plane of s. For $s \to 1$ the pole of $\zeta(s)$ and the zero of $\cos(\pi s/2)$ cancel in Eq. (14.58), so $\zeta(0) = -1/2$ results. Since $\cos(\pi s/2) = 0$ for $s = 2m+1 =$ odd integer, Eq. (14.58) gives $\zeta(-2m) = 0$, the trivial zeros of the zeta function for $m = 1, 2, \ldots$. All other zeros must lie in the "critical strip" $0 < \sigma = \Re s < 1$. They are closely related to the distribution of prime numbers because the prime number product for $\zeta(s)$ (see Section 5.9) can be converted into a Dirichlet series over prime powers for $\zeta'/\zeta = d\ln\zeta(s)/ds$. From here on we sketch ideas only, without proofs. Using the inverse Mellin transform (see Section 16.2) yields the relation

$$\sum_{\substack{p^m<x,\,p=\text{prime}\\ m=1,2,\ldots}} \ln p = -\frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty}\frac{\zeta'(s)}{\zeta(s)s}x^s\,ds \tag{14.59}$$

for $\sigma > 1$, which is a cornerstone of analytic number theory. Since zeros of $\zeta(s)$ become simple poles of ζ'/ζ, the asymptotic distribution of prime numbers is directly related by

Eq. (14.59) to the zeros of the Riemann zeta function. Riemann conjectured that all zeros lie on the line $\sigma = 1/2$, that is, have the form $1/2 + it$ with real t. If so, one could shift the line of integration to the left to $\sigma = 1/2 + \varepsilon$, the simple pole of $\zeta(s)$ at $s = 1$ giving rise to the residue x, while the integral along the line $\sigma = 1/2 + \varepsilon$ is of order $\mathcal{O}(x^{1/2+\varepsilon})$. Hence, the remarkably small remainder in the asymptotic estimate

$$\sum_{p<x} \ln p \sim x + \mathcal{O}\big(x^{1/2+\varepsilon}\big), \qquad x \to \infty$$

would result for arbitrarily small ε. This is equivalent to the estimate for the number of primes below x,

$$\pi(x) = \sum_{p<x} 1 = \int_2^x (\ln t)^{-1}\, dt + \mathcal{O}\big(x^{1/2+\varepsilon}\big), \qquad x \to \infty.$$

In fact, numerical studies have shown that the first 300×10^9 zeros are simple and lie all on the critical line $\sigma = 1/2$. For more details the reader is referred to the classic text by E. C. Titchmarsh and D. R. Heath-Brown, *The Theory of the Riemann Zeta Function*, Oxford, UK: Clarendon Press (1986); H. M. Edwards, *Riemann's Zeta Function*, New York: Academic Press (1974) and Dover (2003); J. Van de Lune, H. J. J. Te Riele, and D. T. Winter, On the zeros of the Riemann zeta function in the critical strip. IV. *Math. Comput.* **47**: 667 (1986). Popular accounts can be found in M. du Sautoy, *The Music of the Primes: Searching to Solve the Greatest Mystery in Mathematics*, New York: HarperCollins (2003); J. Derbyshire, *Prime Obsession: Bernhard Riemann and the Greatest Unsolved Problem in Mathematics*. Washington, DC: Joseph Henry Press (2003); K. Sabbagh, *The Riemann Hypothesis: The Greatest Unsolved Problem in Mathematics*, New York: Farrar, Straus and Giroux (2003).

More recently the statistics of the zeros ρ of the Riemann zeta function on the critical line played a prominent role in the development of theories of chaos (see Chapter 18 for an introduction). Assuming that there is a quantum mechanical system whose energies are the imaginary parts of the ρ, then primes determine the primitive periodic orbits of the associated classically chaotic system. For this case Gutzwiller's trace formula, which relates quantum energy levels and classical periodic orbits, plays a central role and can be better understood using properties of the zeta function and primes. For more details see Sections 12.6 and 12.7 by J. Keating, in *The Nature of Chaos* (T. Mullin, ed.), Oxford, UK: Clarendon Press (1993), and references therein. ■

Exercises

14.3.1 Develop the Fourier series representation of

$$f(t) = \begin{cases} 0, & -\pi \le \omega t \le 0, \\ \sin \omega t, & 0 \le \omega t \le \pi. \end{cases}$$

This is the output of a simple half-wave rectifier. It is also an approximation of the solar thermal effect that produces "tides" in the atmosphere.

$$\text{ANS. } f(t) = \frac{1}{\pi} + \frac{1}{2}\sin \omega t - \frac{2}{\pi} \sum_{n=2,4,6,\ldots}^{\infty} \frac{\cos n\omega t}{n^2 - 1}.$$

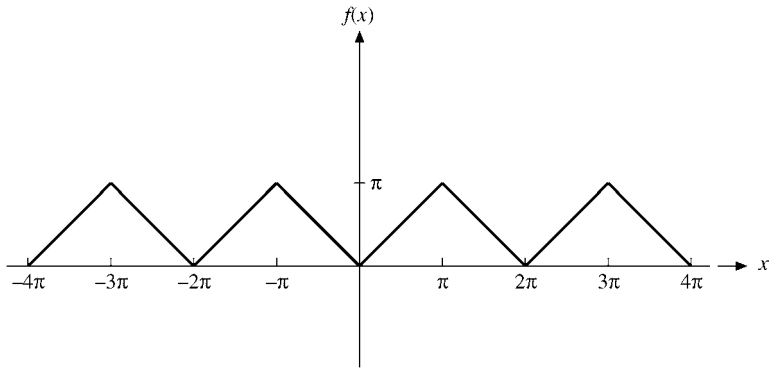

FIGURE 14.7 Triangular wave.

14.3.2 A sawtooth wave is given by

$$f(x) = x, \qquad -\pi < x < \pi.$$

Show that

$$f(x) = 2\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin nx.$$

14.3.3 A different sawtooth wave is described by

$$f(x) = \begin{cases} -\frac{1}{2}(\pi + x), & -\pi \le x < 0 \\ +\frac{1}{2}(\pi - x), & 0 < x \le \pi. \end{cases}$$

Show that $f(x) = \sum_{n=1}^{\infty} (\sin nx/n)$.

14.3.4 A triangular wave (Fig. 14.7) is represented by

$$f(x) = \begin{cases} x, & 0 < x < \pi \\ -x, & -\pi < x < 0. \end{cases}$$

Represent $f(x)$ by a Fourier series.

$$\text{ANS. } f(x) = \frac{\pi}{2} - \frac{4}{\pi} \sum_{n=1,3,5,\ldots} \frac{\cos nx}{n^2}.$$

14.3.5 Expand

$$f(x) = \begin{cases} 1, & x^2 < x_0^2 \\ 0, & x^2 > x_0^2 \end{cases}$$

in the interval $[-\pi, \pi]$.
Note. This variable-width square wave is of some importance in electronic music.

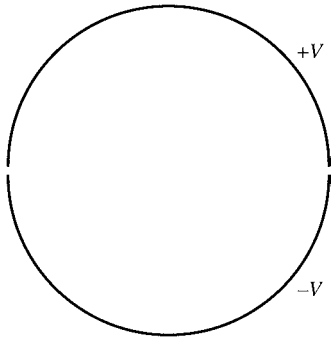

FIGURE 14.8 Cross section of split tube.

14.3.6 A metal cylindrical tube of radius a is split lengthwise into two nontouching halves. The top half is maintained at a potential $+V$, the bottom half at a potential $-V$ (Fig. 14.8). Separate the variables in Laplace's equation and solve for the electrostatic potential for $r \le a$. Observe the resemblance between your solution for $r = a$ and the Fourier series for a square wave.

14.3.7 A metal cylinder is placed in a (previously) uniform electric field, E_0, with the axis of the cylinder perpendicular to that of the original field.

(a) Find the perturbed electrostatic potential.
(b) Find the induced surface charge on the cylinder as a function of angular position.

14.3.8 Transform the Fourier expansion of a square wave, Eq. (14.38), into a power series. Show that the coefficients of x^1 form a **divergent** series. Repeat for the coefficients of x^3.
A power series cannot handle a discontinuity. These infinite coefficients are the result of attempting to beat this basic limitation on power series.

14.3.9 (a) Show that the Fourier expansion of $\cos ax$ is

$$\cos ax = \frac{2a \sin a\pi}{\pi}\left\{\frac{1}{2a^2} - \frac{\cos x}{a^2 - 1^2} + \frac{\cos 2x}{a^2 - 2^2} - \cdots\right\},$$

$$a_n = (-1)^n \frac{2a \sin a\pi}{\pi(a^2 - n^2)}.$$

(b) From the preceding result show that

$$a\pi \cot a\pi = 1 - 2\sum_{p=1}^{\infty} \zeta(2p) a^{2p}.$$

This provides an alternate derivation of the relation between the Riemann zeta function and the Bernoulli numbers, Eq. (5.152).

14.3.10 Derive the Fourier series expansion of the Dirac delta function $\delta(x)$ in the interval $-\pi < x < \pi$.

(a) What significance can be attached to the constant term?
(b) In what region is this representation valid?
(c) With the identity

$$\sum_{n=1}^{N} \cos nx = \frac{\sin(Nx/2)}{\sin(x/2)} \cos\left[\left(N + \frac{1}{2}\right)\frac{x}{2}\right],$$

show that your Fourier representation of $\delta(x)$ is consistent with Eq. (1.190).

14.3.11 Expand $\delta(x - t)$ in a Fourier series. Compare your result with the bilinear form of Eq. (1.190).

$$\text{ANS. } \delta(x-t) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{n=1}^{\infty}(\cos nx \cos nt + \sin nx \sin nt)$$
$$= \frac{1}{2\pi} + \frac{1}{\pi}\sum_{n=1}^{\infty}\cos n(x-t).$$

14.3.12 Verify that

$$\delta(\varphi_1 - \varphi_2) = \frac{1}{2\pi}\sum_{m=-\infty}^{\infty} e^{im(\varphi_1-\varphi_2)}$$

is a Dirac delta function by showing that it satisfies the definition of a Dirac delta function:

$$\int_{-\pi}^{\pi} f(\varphi_1)\frac{1}{2\pi}\sum_{m=-\infty}^{\infty} e^{im(\varphi_1-\varphi_2)}\,d\varphi_1 = f(\varphi_2).$$

Hint. Represent $f(\varphi_1)$ by an exponential Fourier series.
Note. The continuum analog of this expression is developed in Section 15.2. The most important application of this expression is in the determination of Green's functions, Section 9.7.

14.3.13 (a) Using

$$f(x) = x^2, \qquad -\pi < x < \pi,$$

show that

$$\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^2} = \frac{\pi^2}{12} = \eta(2).$$

(b) Using the Fourier series for a triangular wave developed in Exercise 14.3.4, show that

$$\sum_{n=1}^{\infty}\frac{1}{(2n-1)^2} = \frac{\pi^2}{8} = \lambda(2).$$

(c) Using

$$f(x) = x^4, \qquad -\pi < x < \pi,$$

show that

$$\sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90} = \zeta(4), \qquad \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^4} = \frac{7\pi^4}{720} = \eta(4).$$

(d) Using

$$f(x) = \begin{cases} x(\pi - x), & 0 < x < \pi, \\ x(\pi + x), & \pi < x < 0, \end{cases}$$

derive

$$f(x) = \frac{8}{\pi} \sum_{n=1,3,5,\ldots}^{\infty} \frac{\sin nx}{n^3}$$

and show that

$$\sum_{n=1,3,5,\ldots}^{\infty} (-1)^{(n-1)/2} \frac{1}{n^3} = 1 - \frac{1}{3^3} + \frac{1}{5^3} - \frac{1}{7^3} + \cdots = \frac{\pi^3}{32} = \beta(3).$$

(e) Using the Fourier series for a square wave, show that

$$\sum_{n=1,3,5,\ldots}^{\infty} (-1)^{(n-1)/2} \frac{1}{n} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots = \frac{\pi}{4} = \beta(1).$$

This is Leibniz' formula for π, obtained by a different technique in Exercise 5.7.6. *Note.* The $\eta(2), \eta(4), \lambda(2), \beta(1)$, and $\beta(3)$ functions are defined by the indicated series. General definitions appear in Section 5.9.

14.3.14 (a) Find the Fourier series representation of

$$f(x) = \begin{cases} 0, & -\pi < x \le 0 \\ x, & 0 \le x < \pi. \end{cases}$$

(b) From the Fourier expansion show that

$$\frac{\pi^2}{8} = 1 + \frac{1}{3^2} + \frac{1}{5^2} + \cdots.$$

14.3.15 A symmetric triangular pulse of adjustable height and width is described by

$$f(x) = \begin{cases} a(1 - x/b), & 0 \le |x| \le b \\ 0, & b \le |x| \le \pi. \end{cases}$$

(a) Show that the Fourier coefficients are

$$a_0 = \frac{ab}{\pi}, \qquad a_n = \frac{2ab}{\pi(nb)^2}(1 - \cos nb).$$

Sum the finite Fourier series through $n = 10$ and through $n = 100$ for $x/\pi = 0(1/9)1$. Take $a = 1$ and $b = \pi/2$.

(b) Call a Fourier analysis subroutine (if available) to calculate the Fourier coefficients of $f(x), a_0$ through a_{10}.

14.3.16 (a) Using a Fourier analysis subroutine, calculate the Fourier cosine coefficients a_0 through a_{10} of

$$f(x) = \left[1 - \left(\frac{x}{\pi}\right)^2\right]^{1/2}, \qquad x \in [-\pi, \pi].$$

(b) Spot-check by calculating some of the preceding coefficients by direct numerical quadrature.

Check values. $a_0 = 0.785, a_2 = 0.284.$

14.3.17 Using a Fourier analysis subroutine, calculate the Fourier coefficients through a_{10} and b_{10} for

(a) a full-wave rectifier, Example 14.3.2,
(b) a half-wave rectifier, Exercise 14.3.1. Check your results against the analytic forms given (Eq. (14.41) and Exercise 14.3.1).

14.4 PROPERTIES OF FOURIER SERIES

Convergence

It might be noted, first, that our Fourier series should not be expected to be uniformly convergent if it represents a discontinuous function. A uniformly convergent series of continuous functions $(\sin nx, \cos nx)$ always yields a continuous function (compare Section 5.5). If, however,

(a) $f(x)$ is continuous, $-\pi \le x \le \pi$,
(b) $f(-\pi) = f(+\pi)$, and
(c) $f'(x)$ is sectionally continuous,

the Fourier series for $f(x)$ will converge uniformly. These restrictions do not demand that $f(x)$ be periodic, but they will be satisfied by continuous, differentiable, periodic functions (period of 2π). For a proof of uniform convergence we refer to the literature.[9] With or without a discontinuity in $f(x)$, the Fourier series will yield convergence in the mean, Section 10.4.

[9] See, for instance, R. V. Churchill, *Fourier Series and Boundary Value Problems*, 5th ed., New York: McGraw-Hill (1993), Section 38.

Integration

Term-by-term integration of the series

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + \sum_{n=1}^{\infty} b_n \sin nx \tag{14.60}$$

yields

$$\int_{x_0}^{x} f(x)\,dx = \frac{a_0 x}{2}\bigg|_{x_0}^{x} + \sum_{n=1}^{\infty} \frac{a_n}{n} \sin nx\bigg|_{x_0}^{x} - \sum_{n=1}^{\infty} \frac{b_n}{n} \cos nx\bigg|_{x_0}^{x}. \tag{14.61}$$

Clearly, the effect of integration is to place an additional power of n in the denominator of each coefficient. This results in more rapid convergence than before. Consequently, a convergent Fourier series may always be integrated term by term, the resulting series converging uniformly to the integral of the original function. Indeed, term-by-term integration may be valid even if the original series (Eq. (14.60)) is not itself convergent. The function $f(x)$ need only be integrable. A discussion will be found in Jeffreys and Jeffreys, Section 14.06 (see the Additional Readings).

Strictly speaking, Eq. (14.61) may not be a Fourier series; that is, if $a_0 \neq 0$, there will be a term $\frac{1}{2}a_0 x$. However,

$$\int_{x_0}^{x} f(x)\,dx - \frac{1}{2}a_0 x \tag{14.62}$$

will still be a Fourier series.

Differentiation

The situation regarding differentiation is quite different from that of integration. Here the word is caution. Consider the series for

$$f(x) = x, \qquad -\pi < x < \pi. \tag{14.63}$$

We readily find (compare Exercise 14.3.2) that the Fourier series is

$$x = 2\sum_{n=1}^{\infty} (-1)^{n+1} \frac{\sin nx}{n}, \qquad -\pi < x < \pi. \tag{14.64}$$

Differentiating term by term, we obtain

$$1 = 2\sum_{n=1}^{\infty} (-1)^{n+1} \cos nx, \tag{14.65}$$

which is not convergent. **Warning**: Check your derivative for convergence.

For a triangular wave (Exercise 14.3.4), in which the convergence is more rapid (and uniform),

$$f(x) = \frac{\pi}{2} - \frac{4}{\pi} \sum_{n=1,\text{odd}}^{\infty} \frac{\cos nx}{n^2}. \tag{14.66}$$

Differentiating term by term we get

$$f'(x) = \frac{4}{\pi} \sum_{n=1,\text{odd}}^{\infty} \frac{\sin nx}{n}, \tag{14.67}$$

which is the Fourier expansion of a square wave,

$$f'(x) = \begin{cases} 1, & 0 < x < \pi, \\ -1, & -\pi < x < 0. \end{cases} \tag{14.68}$$

Inspection of Fig. 14.7 verifies that this is indeed the derivative of our triangular wave.

- As the inverse of integration, the operation of differentiation has placed an additional factor n in the numerator of each term. This reduces the rate of convergence and may, as in the first case mentioned, render the differentiated series divergent.
- In general, term-by-term differentiation is permissible under the same conditions listed for uniform convergence.

Exercises

14.4.1 Show that integration of the Fourier expansion of $f(x) = x, -\pi < x < \pi$, leads to

$$\frac{\pi^2}{12} = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2} = 1 - \frac{1}{4} + \frac{1}{9} - \frac{1}{16} + \cdots.$$

14.4.2 Parseval's identity.

(a) Assuming that the Fourier expansion of $f(x)$ is uniformly convergent, show that

$$\frac{1}{\pi} \int_{-\pi}^{\pi} \big[f(x)\big]^2 \, dx = \frac{a_0^2}{2} + \sum_{n=1}^{\infty} \big(a_n^2 + b_n^2\big).$$

This is Parseval's identity. It is actually a special case of the completeness relation, Eq. (10.73).

(b) Given

$$x^2 = \frac{\pi^2}{3} + 4 \sum_{n=1}^{\infty} \frac{(-1)^n \cos nx}{n^2}, \qquad -\pi \le x \le \pi,$$

apply Parseval's identity to obtain $\zeta(4)$ in closed form.

(c) The condition of uniform convergence is not necessary. Show this by applying the Parseval identity to the square wave

$$f(x) = \begin{cases} -1, & -\pi < x < 0 \\ 1, & 0 < x < \pi \end{cases}$$

$$= \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{\sin(2n-1)x}{2n-1}.$$

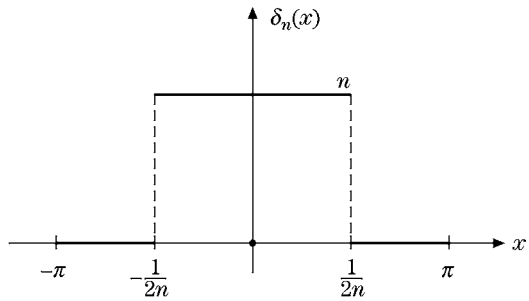

FIGURE 14.9 Rectangular pulse.

14.4.3 Show that integrating the Fourier expansion of the Dirac delta function (Exercise 14.3.10) leads to the Fourier representation of the square wave, Eq. (14.38), with $h = 1$.
Note. Integrating the constant term $(1/2\pi)$ leads to a term $x/2\pi$. What are you going to do with this?

14.4.4 Integrate the Fourier expansion of the unit step function

$$f(x) = \begin{cases} 0, & -\pi < x < 0 \\ x, & 0 < x < \pi. \end{cases}$$

Show that your integrated series agrees with Exercise 14.3.14.

14.4.5 In the interval $(-\pi, \pi)$,

$$\delta_n(x) = \begin{cases} n, & \text{for } |x| < \frac{1}{2n}, \\ 0, & \text{for } |x| > \frac{1}{2n} \end{cases}$$

(Fig. 14.9).

(a) Expand $\delta_n(x)$ as a Fourier cosine series.
(b) Show that your Fourier series agrees with a Fourier expansion of $\delta(x)$ in the limit as $n \to \infty$.

14.4.6 Confirm the delta function nature of your Fourier series of Exercise 14.4.4 by showing that for any $f(x)$ that is finite in the interval $[-\pi, \pi]$ and continuous at $x = 0$,

$$\int_{-\pi}^{\pi} f(x)\big[\text{Fourier expansion of } \delta_\infty(x)\big]\,dx = f(0).$$

14.4.7 (a) Show that the Dirac delta function $\delta(x - a)$, expanded in a Fourier sine series in the half-interval $(0, L)(0 < a < L)$, is given by

$$\delta(x - a) = \frac{2}{L}\sum_{n=1}^{\infty} \sin\left(\frac{n\pi a}{L}\right)\sin\left(\frac{n\pi x}{L}\right).$$

Note that this series actually describes

$$-\delta(x + a) + \delta(x - a) \qquad \text{in the interval} \qquad (-L, L).$$

(b) By integrating both sides of the preceding equation from 0 to x, show that the cosine expansion of the square wave

$$f(x) = \begin{cases} 0, & 0 \le x < a \\ 1, & a < x < L, \end{cases}$$

is

$$f(x) = \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin\left(\frac{n\pi a}{L}\right) - \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin\left(\frac{n\pi a}{L}\right)\cos\left(\frac{n\pi x}{L}\right),$$

for $0 \le x < L$.

(c) Verify that the term

$$\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin\left(\frac{n\pi a}{L}\right)$$

is $\langle f(x)\rangle$.

14.4.8 Verify the Fourier cosine expansion of the square wave, Exercise 14.4.7(b), by direct calculation of the Fourier coefficients.

14.4.9 (a) A string is clamped at both ends $x = 0$ and $x = L$. Assuming small-amplitude vibrations, we find that the amplitude $y(x, t)$ satisfies the wave equation

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 y}{\partial t^2}.$$

Here v is the wave velocity. The string is set in vibration by a sharp blow at $x = a$. Hence we have

$$y(x, 0) = 0, \qquad \frac{\partial y(x, t)}{\partial t} = Lv_0\delta(x - a) \qquad \text{at } t = 0.$$

The constant L is included to compensate for the dimensions (inverse length) of $\delta(x - a)$. With $\delta(x - a)$ given by Exercise 14.4.7(a), solve the wave equation subject to these initial conditions.

$$\text{ANS. } y(x, t) = \frac{2v_0 L}{\pi v}\sum_{n=1}^{\infty}\frac{1}{n}\sin\frac{n\pi a}{L}\sin\frac{n\pi x}{L}\sin\frac{n\pi vt}{L}.$$

(b) Show that the transverse velocity of the string $\partial y(x, t)/\partial t$ is given by

$$\frac{\partial y(x, t)}{\partial t} = 2v_0\sum_{n=1}^{\infty}\sin\frac{n\pi a}{L}\sin\frac{n\pi x}{L}\cos\frac{n\pi vt}{L}.$$

14.4.10 A string, clamped at $x = 0$ and at $x = 1$, is vibrating freely. Its motion is described by the wave equation

$$\frac{\partial^2 u(x, t)}{\partial t^2} = v^2\frac{\partial^2 u(x, t)}{\partial x^2}.$$

Assume a Fourier expansion of the form

$$u(x,t) = \sum_{n=1}^{\infty} b_n(t) \sin \frac{n\pi x}{l}$$

and determine the coefficients $b_n(t)$. The initial conditions are

$$u(x,0) = f(x) \qquad \text{and} \qquad \frac{\partial}{\partial t} u(x,0) = g(x).$$

Note. This is only half the conventional Fourier orthogonality integral interval. However, as long as only the sines are included here, the Sturm–Liouville boundary conditions are still satisfied and the functions are orthogonal.

$$\text{ANS. } b_n(t) = A_n \cos \frac{n\pi vt}{l} + B_n \sin \frac{n\pi vt}{l},$$

$$A_n = \frac{2}{l} \int_0^l f(x) \sin \frac{n\pi x}{l} dx, \; B_n = \frac{2}{n\pi v} \int_0^l g(x) \sin \frac{n\pi x}{l} dx.$$

14.4.11 (a) Let us continue the vibrating string problem, Exercise 14.4.10. The presence of a resisting medium will damp the vibrations according to the equation

$$\frac{\partial^2 u(x,t)}{\partial t^2} = v^2 \frac{\partial^2 u(x,t)}{\partial x^2} - k \frac{\partial u(x,t)}{\partial t}.$$

Assume a Fourier expansion

$$u(x,t) = \sum_{n=1}^{\infty} b_n(t) \sin \frac{n\pi x}{l}$$

and again determine the coefficients $b_n(t)$. Take the initial and boundary conditions to be the same as in Exercise 14.4.10. Assume the damping to be small.

(b) Repeat, but assume the damping to be large.

$$\text{ANS. (a) } b_n(t) = e^{-kt/2}\{A_n \cos \omega_n t + B_n \sin \omega_n t\},$$

$$A_n = \frac{2}{l} \int_0^l f(x) \sin \frac{n\pi x}{l} dx,$$

$$B_n = \frac{2}{\omega_n l} \int_0^l g(x) \sin \frac{n\pi x}{l} dx + \frac{k}{2\omega_n} A_n,$$

$$\omega_n^2 = \left(\frac{n\pi v}{l}\right) - \left(\frac{k}{2}\right)^2 > 0.$$

(b) $b_n(t) = e^{-kt/2}\{A_n \cosh \sigma_n t + B_n \sinh \sigma_n t\}$,

$$A_n = \frac{2}{l}\int_0^l f(x) \sin \frac{n\pi x}{l} dx,$$

$$B_n = \frac{2}{\sigma_n l}\int_0^l g(x) \sin \frac{n\pi x}{l} dx + \frac{k}{2\sigma_n} A_n,$$

$$\sigma_n^2 = \left(\frac{k}{2}\right)^2 - \left(\frac{n\pi v}{l}\right)^2 > 0.$$

14.4.12 Find the charge distribution over the interior surfaces of the semicircles of Exercise 14.3.6.
Note. You obtain a divergent series and this Fourier approach fails. Using conformal mapping techniques, we may show the charge density to be proportional to $\csc\theta$. Does $\csc\theta$ have a Fourier expansion?

14.4.13 Given

$$\varphi_1(x) = \sum_{n=1}^{\infty} \frac{\sin nx}{n} = \begin{cases} -\frac{1}{2}(\pi + x), & -\pi \le x < 0, \\ \frac{1}{2}(\pi - x), & 0 < x \le \pi, \end{cases}$$

show by integrating that

$$\varphi_2(x) \equiv \sum_{n=1}^{\infty} \frac{\cos nx}{n^2} = \begin{cases} \frac{1}{4}(\pi + x)^2 - \frac{\pi^2}{12}, & -\pi \le x \le 0 \\ \frac{1}{4}(\pi - x)^2 - \frac{\pi^2}{12}, & 0 \le x \le \pi. \end{cases}$$

14.4.14 Given

$$\psi_{2s}(x) = \sum_{n=1}^{\infty} \frac{\sin nx}{n^{2s}}, \qquad \psi_{2s+1}(x) = \sum_{n=1}^{\infty} \frac{\cos nx}{n^{2s+1}},$$

develop the following recurrence relations:

(a) $\psi_{2s}(x) = \displaystyle\int_0^x \psi_{2s-1}(x)\, dx$

(b) $\psi_{2s+1}(x) = \zeta(2s+1) - \displaystyle\int_0^x \psi_{2s}(x)\, dx.$

Note. The functions $\psi_s(x)$ and the $\varphi_s(x)$ of the preceding two exercises are known as **Clausen functions**. In theory they may be used to improve the rate of convergence of a Fourier series. As with the series of Chapter 5, there is always the question of how much analytical work we do and how much arithmetic work we demand that the computer do. As computers become steadily more powerful, the balance progressively shifts so that we are doing less and demanding that they do more.

14.4.15 Show that

$$f(x) = \sum_{n=1}^{\infty} \frac{\cos nx}{n+1}$$

may be written as

$$f(x) = \psi_1(x) - \varphi_2(x) + \sum_{n=1}^{\infty} \frac{\cos nx}{n^2(n+1)}.$$

Note. $\psi_1(x)$ and $\varphi_2(x)$ are defined in the preceding exercises.

14.5 GIBBS PHENOMENON

The Gibbs phenomenon is an overshoot, a peculiarity of the Fourier series and other eigenfunction series at a simple discontinuity. An example is seen in Fig. 14.1.

Summation of Series

In Section 14.1 the sum of the first several terms of the Fourier series for a sawtooth wave was plotted (Fig. 14.10). Now we develop analytic methods of summing the first r terms of our Fourier series.

From Eq. (14.19),

$$a_n \cos nx + b_n \sin nx = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos n(t-x)\, dt. \tag{14.69}$$

Then the rth partial sum becomes[10]

$$\begin{aligned} s_r(x) &= \frac{a_0}{2} + \sum_{n=1}^{r} (a_n \cos nx + b_n \sin nx) \\ &= \Re \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \left[\frac{1}{2} + \sum_{n=1}^{r} e^{-i(t-x)n} \right] dt. \end{aligned} \tag{14.70}$$

Summing the finite series of exponentials (geometric progression),[11] we obtain

$$s_r(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \frac{\sin[(r+\frac{1}{2})(t-x)]}{\sin \frac{1}{2}(t-x)} dt. \tag{14.71}$$

This is convergent at all points, including $t = x$. The factor

$$\frac{\sin[(r+\frac{1}{2})(t-x)]}{2\pi \sin \frac{1}{2}(t-x)}$$

is the Dirichlet kernel mentioned in Section 1.15 as a Dirac delta distribution.

[10] It is of some interest to note that this series also occurs in the analysis of the diffraction grating (r slits).

[11] Compare Exercise 6.1.7 with initial value $n = 1$.

Square Wave

For convenience of numerical calculation we consider the behavior of the Fourier series that represents the periodic square wave

$$f(x) = \begin{cases} \frac{h}{2}, & 0 < x < \pi, \\ -\frac{h}{2}, & -\pi < x < 0. \end{cases} \tag{14.72}$$

This is essentially the square wave used in Section 14.3, and we immediately see that the solution is

$$f(x) = \frac{2h}{\pi}\left(\frac{\sin x}{1} + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + \cdots\right). \tag{14.73}$$

Applying Eq. (14.71) to our square wave (Eq. (14.72)), we have the sum of the first r terms (plus $\frac{1}{2}a_0$, which is zero here):

$$\begin{aligned} s_r(x) &= \frac{h}{4\pi}\int_0^{\pi} \frac{\sin[(r+\frac{1}{2})(t-x)]}{\sin\frac{1}{2}(t-x)}dt - \frac{h}{4\pi}\int_{-\pi}^{0} \frac{\sin[(r+\frac{1}{2})(t-x)]}{\sin\frac{1}{2}(t-x)}dt \\ &= \frac{h}{4\pi}\int_0^{\pi} \frac{\sin[(r+\frac{1}{2})(t-x)]}{\sin\frac{1}{2}(t-x)}dt - \frac{h}{4\pi}\int_0^{\pi} \frac{\sin[(r+\frac{1}{2})(t+x)]}{\sin\frac{1}{2}(t+x)}dt. \end{aligned} \tag{14.74}$$

This last result follows from the transformation

$$\vec{t} - t$$

in the second integral. Replacing $t - x$ in the first term with s and $t + x$ in the second term with s, we obtain

$$s_r(x) = \frac{h}{4\pi}\int_{-x}^{\pi-x} \frac{\sin(r+\frac{1}{2})s}{\sin\frac{1}{2}s}ds - \frac{h}{4\pi}\int_{x}^{\pi+x} \frac{\sin(r+\frac{1}{2})s}{\sin\frac{1}{2}s}ds. \tag{14.75}$$

The intervals of integration are shown in Fig. 14.10(top). Because the integrands have the same mathematical form, the integrals from x to $\pi - x$ cancel, leaving the integral ranges shown in the bottom portion of Fig. 14.10:

$$s_r(x) = \frac{h}{4\pi}\int_{-x}^{x} \frac{\sin(r+\frac{1}{2})s}{\sin\frac{1}{2}s}ds - \frac{h}{4\pi}\int_{\pi-x}^{\pi+x} \frac{\sin(r+\frac{1}{2})s}{\sin\frac{1}{2}s}ds. \tag{14.76}$$

Consider the partial sum in the vicinity of the discontinuity at $x = 0$. As $x \to 0$, the second integral becomes negligible, and we associate the first integral with the discontinuity at $x = 0$. Using $r + \frac{1}{2} = p$ and $ps = \xi$ we obtain

$$s_r(x) = \frac{h}{2\pi}\int_0^{px} \frac{\sin\xi}{\sin(\xi/2p)}\frac{d\xi}{p}. \tag{14.77}$$

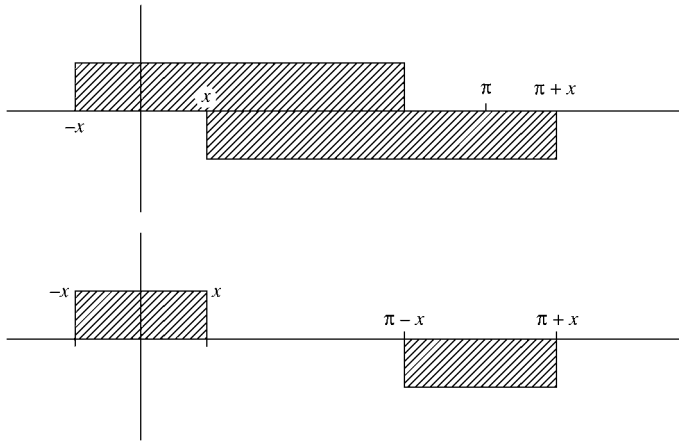

FIGURE 14.10 Intervals of integration — Eq. (14.75).

Calculation of Overshoot

Our partial sum $s_r(x)$ starts at zero when $x = 0$ (in agreement with Eq. (14.22)) and increases until $\xi = ps = \pi$, at which point the numerator, $\sin\xi$, goes negative. For large r, and therefore for large p, our denominator remains positive. We get the maximum value of the partial sum by taking the upper limit $px = \pi$. Right here we see that x, the location of the overshoot maximum, is inversely proportional to the number of terms taken:

$$x = \frac{\pi}{p} \approx \frac{\pi}{r}.$$

The maximum value of the partial sum is then

$$s_r(x)_{\max} = \frac{h}{2} \cdot \frac{1}{\pi} \int_0^{\pi} \frac{\sin\xi \, d\xi}{\sin(\xi/2p)p} \approx \frac{h}{2} \cdot \frac{2}{\pi} \int_0^{\pi} \frac{\sin\xi}{\xi} d\xi. \tag{14.78}$$

In terms of the sine integral, $\mathrm{si}(x)$ of Section 8.5,

$$\int_0^{\pi} \frac{\sin\xi}{\xi} d\xi = \frac{\pi}{2} + \mathrm{si}(\pi). \tag{14.79}$$

The integral is clearly greater than $\pi/2$, since it can be written as

$$\left(\int_0^{\infty} - \int_{\pi}^{3\pi} - \int_{3\pi}^{5\pi} - \cdots \right) \frac{\sin\xi}{\xi} d\xi = \int_0^{\pi} \frac{\sin\xi}{\xi} d\xi. \tag{14.80}$$

We saw in Example 7.1.4 that the integral from 0 to ∞ is $\pi/2$. From this integral we are subtracting a series of negative terms. A Gaussian quadrature or a power-series expansion and term-by-term integration yields

$$\frac{2}{\pi} \int_0^{\pi} \frac{\sin\xi}{\xi} d\xi = 1.1789797\ldots, \tag{14.81}$$

which means that the Fourier series tends to overshoot the positive corner by some 18 percent and to undershoot the negative corner by the same amount, as suggested in Fig. 14.11.

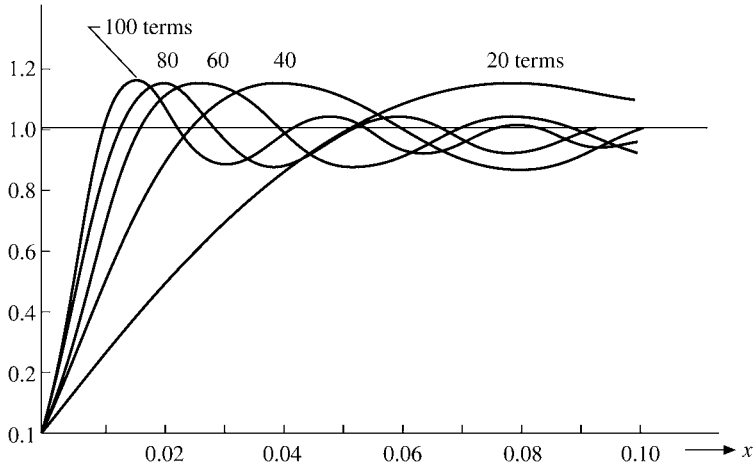

FIGURE 14.11 Square wave — Gibbs phenomenon.

The inclusion of more terms (increasing r) does nothing to remove this overshoot but merely moves it closer to the point of discontinuity. The overshoot is the Gibbs phenomenon, and because of it the Fourier series representation may be highly unreliable for precise numerical work, especially in the vicinity of a discontinuity.

The Gibbs phenomenon is not limited to the Fourier series. It occurs with other eigenfunction expansions. Exercise 12.3.27 is an example of the Gibbs phenomenon for a Legendre series. For more details, see W. J. Thompson, Fourier series and the Gibbs phenomenon, *Am. J. Phys.* **60**: 425 (1992).

Exercises

14.5.1 With the Fourier series for $f(x)$ takes on the isotropic shown that at a discontinuity $f(x)$ the Fourier series for $f(x)$ takes on the isotropic shown that at a discontinuity in hand limits:

$$f(x_0) = \frac{1}{2}\big[f(x_0+0) + f(x_0-0)\big].$$

In evaluating $\lim_{r\to\infty} s_r(x_0)$ you may find it convenient to identify part of the integrand as a Dirac delta function.

14.5.2 Determine the partial sum, s_n, of the series in Eq. (14.73) by using

$$\text{(a)}\ \frac{\sin mx}{m} = \int_0^x \cos my\,dy, \qquad \text{(b)}\ \sum_{p=1}^{n} \cos(2p-1)y = \frac{\sin 2ny}{2\sin y}.$$

Do you agree with the result given in Eq. (14.79)?

14.5.3 Evaluate the finite step function series, Eq. (14.73), $h = 2$, using 100, 200, 300, 400, and 500 terms for $x = 0.0000(0.0005)0.0200$. Sketch your results (five curves) or, if a plotting routine is available, plot your results.

14.5.4 (a) Calculate the value of the Gibbs phenomenon integral

$$I = \frac{2}{\pi}\int_0^{\pi} \frac{\sin t}{t}\,dt$$

by numerical quadrature accurate to 12 significant figures.

(b) Check your result by (1) expanding the integrand as a series, (2) integrating term by term, and (3) evaluating the integrated series. This calls for double precision calculation.

ANS. $I = 1.178979744472$.

14.6 DISCRETE FOURIER TRANSFORM

For many physicists the Fourier transform is automatically the continuous Fourier transform of Chapter 15. The use of digital computers, however, necessarily replaces a continuum of values by a discrete set; an integration is replaced by a summation. The continuous Fourier transform becomes the discrete Fourier transform and an appropriate topic for this chapter.

Orthogonality over Discrete Points

The orthogonality of the trigonometric functions and the imaginary exponentials is expressed in Eqs. (14.15) to (14.18). This is the usual orthogonality for functions: **integration** of a product of functions over the orthogonality interval. The sines, cosines, and imaginary exponentials have the remarkable property that they are also orthogonal over a series of discrete, equally spaced points over the period (the orthogonality interval).

Consider a set of $2N$ time values

$$t_k = 0, \frac{T}{2N}, \frac{2T}{2N}, \ldots, \frac{(2N-1)T}{2N} \tag{14.82}$$

for the time interval $(0, T)$. Then

$$t_k = \frac{kT}{2N}, \qquad k = 0, 1, 2, \ldots, 2N-1. \tag{14.83}$$

We shall prove that the exponential functions $\exp(2\pi i p t_k/T)$ and $\exp(2\pi i q t_k/T)$ satisfy an orthogonality relation over the discrete points t_k:

$$\sum_{k=0}^{2N-1}\left[\exp\left(\frac{2\pi i p t_k}{T}\right)\right]^* \exp\left(\frac{2\pi i q t_k}{T}\right) = 2N\delta_{p,q\pm 2nN}. \tag{14.84}$$

Here n, p, and q are all integers.

Replacing $q - p$ by s, we find that the left-hand side of Eq. (14.84) becomes

$$\sum_{k=0}^{2N-1}\exp\left(\frac{2\pi i s t_k}{T}\right) = \sum_{k=0}^{2N-1}\exp\left(\frac{2\pi i s k}{2N}\right).$$

This right-hand side is obtained by using Eq. (14.83) to replace T. This is a finite geometric series with an initial term 1 and a ratio

$$r = \exp\left(\frac{\pi i s}{N}\right).$$

From Eq. (5.3),

$$\sum_{k=0}^{2N-1} \exp\left(\frac{2\pi i s t_k}{T}\right) = \begin{cases} \dfrac{1-r^{2N}}{1-r} = 0, & r \neq 1 \\ 2N, & r = 1, \end{cases} \tag{14.85}$$

establishing Eq. (14.84), our basic orthogonality relation. The upper value, zero, is a consequence of

$$r^{2N} = \exp(2\pi i s) = 1$$

for s an integer. The lower value, $2N$, for $r = 1$ corresponds to $p = q$. The orthogonality of the corresponding trigonometric functions is left as Exercise 14.6.1.

Discrete Fourier Transform

To simplify the notation and to make more direct contact with physics, we introduce the (reciprocal) ω-space, or angular frequency, with

$$\omega_p = \frac{2\pi p}{T}, \qquad p = 0, 1, 2, \ldots, 2N-1. \tag{14.86}$$

We make p range over the same integers as k. The exponential $\exp(\pm 2\pi i p t_k/T)$ of Eq. (14.84) becomes $\exp(\pm i\omega_p t_k)$. The choice of whether to use the $+$ or the $-$ sign is a matter of convenience or convention. In quantum mechanics the negative sign is selected when expressing the time dependence.

Consider a function of time defined (measured) at the discrete time values t_k. Then we construct

$$F(\omega_p) = \frac{1}{2N} \sum_{k=0}^{2N-1} f(t_k) e^{i\omega_p t_k}. \tag{14.87}$$

Employing the orthogonality relation, we obtain

$$\frac{1}{2N} \sum_{p=0}^{2N-1} \left(e^{i\omega_p t_m}\right)^* e^{i\omega_p t_k} = \delta_{mk}, \tag{14.88}$$

and then replacing the subscript m by k, we find that the amplitudes $f(t_k)$ become

$$f(t_k) = \sum_{p=0}^{2N-1} F(\omega_p) e^{-i\omega_p t_k}. \tag{14.89}$$

The time function $f(t_k), k = 0, 1, 2, \ldots, 2N-1$, and the frequency function $F(\omega_p)$, $p = 0, 1, 2, \ldots, 2N-1$, are discrete Fourier transforms of each other.[12] Compare Eqs. (14.87)

[12] The two transform equations may be symmetrized with a resulting $(2N)^{-1/2}$ in each equation if desired.

and (14.89) with the corresponding continuous Fourier transforms, Eqs. (15.22) and (15.23) of Chapter 15.

Limitations

Taken as a pair of mathematical relations, the discrete Fourier transforms are exact. We can say that the $2N$ $2N$-component vectors $\exp(-i\omega_p t_k), k = 0, 1, 2, \ldots, 2N-1$, form a complete set[13] spanning the t_k-space. Then $f(t_k)$ in Eq. (14.89) is simply a particular linear combination of these vectors. Alternatively, we may take the $2N$ measured components $f(t_k)$ as defining a $2N$-component vector in t_k-space. Then, Eq. (14.87) yields the $2N$-component vector $F(\omega_p)$ in the **reciprocal** ω_p-space. Equations (14.87) and (14.89) become matrix equations, with $\exp(i\omega_p t_k)/(2N)^{1/2}$ the elements of a unitary matrix.

The limitations of the discrete Fourier transform arise when we apply Eqs. (14.87) and (14.89) to physical systems and attempt physical interpretation and the limit $F(\omega_p) \to F(\omega)$. Example 14.6.1 illustrates the problems that can occur. The most important precaution to be taken to avoid trouble is to take N sufficiently large so that there is no angular frequency component of a higher angular frequency than $\omega_N = 2\pi N/T$. For details on errors and limitations in the use of the discrete Fourier transform we refer to Hamming in the Additional Readings.

Example 14.6.1 DISCRETE FOURIER TRANSFORM — ALIASING

Consider the simple case of $T = 2\pi$, $N = 2$, and $f(t_k) = \cos t_k$. From

$$t_k = \frac{kT}{4} = \frac{k\pi}{2}, \qquad k = 0, 1, 2, 3, \tag{14.90}$$

$f(t_k) = \cos(t_k)$ is represented by the four-component vector

$$f(t_k) = (1, 0, -1, 0). \tag{14.91}$$

The frequencies, ω_p, are given by Eq. (14.86):

$$\omega_p = \frac{2\pi p}{T} = p. \tag{14.92}$$

Clearly, $\cos t_k$ implies a $p = 1$ component and no other frequency components.

The transformation matrix

$$\frac{\exp(i\omega_p t_k)}{2N} = \frac{\exp(ipk\pi/2)}{2N}$$

becomes

$$\frac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix}. \tag{14.93}$$

[13] By Eq. (14.85) these vectors are orthogonal and are therefore linearly independent.

Note that the $2N \times 2N$ matrix has only $2N$ independent components. It is the **repetition of values that makes the fast Fourier transform technique possible**.

Operating on column vector $f(t_k)$, we find that this matrix yields a column vector

$$F(\omega_p) = \left(0, \tfrac{1}{2}, 0, \tfrac{1}{2}\right). \tag{14.94}$$

Apparently, there is a $p = 3$ frequency component present. We reconstruct $f(t_k)$ by Eq. (14.89), obtaining

$$f(t_k) = \tfrac{1}{2}e^{-it_k} + \tfrac{1}{2}e^{-3it_k}. \tag{14.95}$$

Taking real parts, we can rewrite the equation as

$$\Re f(t_k) = \tfrac{1}{2}\cos t_k + \tfrac{1}{2}\cos 3t_k. \tag{14.96}$$

Obviously, this result, Eq. (14.96), is not identical with our original $f(t_k)\cos t_k$. But $\cos t_k = \frac{1}{2}\cos t_k + \frac{1}{2}\cos 3t_k$ at $t_k = 0, \pi/2, \pi$; and $3\pi/2$. The $\cos t_k$ and $\cos 3t_k$ mimic each other because of the limited number of data points (and the particular choice of data points). This error of one frequency mimicking another is known as **aliasing**. The problem can be minimized by taking more data points. ■

Fast Fourier Transform

The fast Fourier transform is a particular way of factoring and rearranging the terms in the sums of the discrete Fourier transform. Brought to the attention of the scientific community by Cooley and Tukey,[14] its importance lies in the drastic reduction in the number of numerical operations required. Because of the tremendous increase in speed achieved (and reduction in cost), the fast Fourier transform has been hailed as one of the few really significant advances in numerical analysis in the past few decades.

For N time values (measurements), a direct calculation of a discrete Fourier transform would mean about N^2 multiplications. For N a power of 2, the fast Fourier transform technique of Cooley and Tukey cuts the number of multiplications required to $(N/2)\log_2 N$. If $N = 1024$ $(= 2^{10})$, the fast Fourier transform achieves a computational reduction by a factor of over 200. This is why the fast Fourier transform is called fast and why it has revolutionized the digital processing of waveforms. Details on the internal operation will be found in the paper by Cooley and Tukey and in the paper by Bergland.[15]

[14] J. W. Cooley and J. W. Tukey, *Math. Comput.* **19**: 297 (1965).

[15] G. D. Bergland, A guided tour of the fast Fourier transform, *IEEE Spectrum*, July, pp. 41–52 (1969); see also, W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*, 2nd ed., Cambridge, UK: Cambridge University Press (1996), Section 12.3.

Exercises

14.6.1 Derive the trigonometric forms of discrete orthogonality corresponding to Eq. (14.84):

$$\sum_{k=0}^{2N-1} \cos\left(\frac{2\pi p t_k}{T}\right) \sin\left(\frac{2\pi q t_k}{T}\right) = 0$$

$$\sum_{k=0}^{2N-1} \cos\left(\frac{2\pi p t_k}{T}\right) \cos\left(\frac{2\pi q t_k}{T}\right) = \begin{cases} 0, & p \neq q \\ N, & p = q \neq 0, N \\ 2N, & p = q = 0, N \end{cases}$$

$$\sum_{k=0}^{2N-1} \sin\left(\frac{2\pi p t_k}{T}\right) \sin\left(\frac{2\pi q t_k}{T}\right) = \begin{cases} 0, & p \neq q \\ N, & p = q \neq 0, N \\ 0, & p = q = 0, N. \end{cases}$$

Hint. Trigonometric identities such as

$$\sin A \cos B = \tfrac{1}{2}\big[\sin(A + B) + \sin(A - B)\big]$$

are useful.

14.6.2 Equation (14.84) exhibits orthogonality summing over time points. Show that we have the same orthogonality summing over frequency points

$$\frac{1}{2N} \sum_{p=0}^{2N-1} \left(e^{i\omega_p t_m}\right)^* e^{i\omega_p t_k} = \delta_{mk}.$$

14.6.3 Show in detail how to go from

$$F(\omega_p) = \frac{1}{2N} \sum_{k=0}^{2N-1} f(t_k) e^{i\omega_p t_k}$$

to

$$f(t_k) = \sum_{p=0}^{2N-1} F(\omega_p) e^{-i\omega_p t_k}.$$

14.6.4 The functions $f(t_k)$ and $F(\omega_p)$ are discrete Fourier transforms of each other. Derive the following symmetry relations:

(a) If $f(t_k)$ is real, $F(\omega_p)$ is Hermitian symmetric; that is,

$$F(\omega_p) = F^*\left(\frac{4\pi N}{T} - \omega_p\right).$$

(b) If $f(t_k)$ is pure imaginary,

$$F(\omega_p) = -F^*\left(\frac{4\pi N}{T} - \omega_p\right).$$

Note. The symmetry of part (a) is an illustration of aliasing. The frequency $4\pi N/T - \omega_p$ masquerades as the frequency ω_p.

14.6.5 Given $N = 2$, $T = 2\pi$, and $f(t_k) = \sin t_k$,

(a) find $F(\omega_p)$, $p = 0, 1, 2, 3$, and
(b) reconstruct $f(t_k)$ from $F(\omega_p)$ and exhibit the aliasing of $\omega_1 = 1$ and $\omega_3 = 3$.

ANS. (a) $F(\omega_p) = (0, i/2, 0, -i/2)$

(b) $f(t_k) = \frac{1}{2}\sin t_k - \frac{1}{2}\sin 3t_k$.

14.6.6 Show that the Chebyshev polynomials $T_m(x)$ satisfy a discrete orthogonality relation

$$\frac{1}{2}T_m(-1)T_n(-1) + \sum_{s=1}^{N-1} T_m(x_s)T_n(x_s) + \frac{1}{2}T_m(1)T_n(1) = \begin{cases} 0, & m \neq n \\ N/2, & m = n \neq 0 \\ N, & m = n = 0. \end{cases}$$

Here, $x_s = \cos\theta_s$, where the $(N+1)\theta_s$ are equally spaced along the θ-axis:

$$\theta_s = \frac{s\pi}{N}, \qquad s = 0, 1, 2, \ldots, N.$$

14.7 Fourier Expansions of Mathieu Functions

As a realistic application of Fourier series we now derive first integral equations satisfied by Mathieu functions, from which subsequently their Fourier series are obtained.

Integral Equations and Fourier Series for Mathieu Functions

Our first goal is to establish Whittaker's integral equations that Mathieu functions satisfy, from which we then obtain their Fourier series representations.

We start from an integral representation

$$V(\mathbf{r}) = \int_{-\pi}^{\pi} f(z + ix\cos\theta + iy\sin\theta, \theta)\, d\theta \tag{14.97}$$

of a solution V of Laplace's equation with a twice differentiable function $f(v, \theta)$. Applying ∇^2 to V we verify that it obeys Laplace's PDE. Separating variables in Laplace's PDE suggests choosing the product form $f(v, \theta) = e^{kv}\phi(\theta)$. Substituting the elliptical variables of Eq. (13.163) we rewrite V as

$$R(\xi)\Phi(\eta)e^{kz} = \int_{-\pi}^{\pi} \phi(\theta) e^{k(z + ic\cosh\xi\cos\eta\cos\theta + ic\sinh\xi\sin\eta\sin\theta)}\, d\theta \tag{14.98}$$

with normalization $R(0) = 1$. Since ξ and η are independent variables we may set $\xi = 0$, which leads to Whittaker's integral representation

$$\Phi(\eta) = \int_{-\pi}^{\pi} \phi(\theta) \exp(ick\cos\theta\cos\eta)\, d\theta, \tag{14.99}$$

where $ck = 2\sqrt{q}$ from Eq. (13.180). Clearly, Φ is even in the variable η and periodic with period π. In order to prove that $\phi \sim \Phi$ we check how $\phi(\theta)$ is constrained when $\Phi(\eta)$ is taken to obey the angular Mathieu ODE

$$\begin{aligned}&\frac{d^2\Phi}{d\eta^2} + (\lambda - 2q\cos 2\eta)\Phi(\eta)\\ &\quad = \int_{-\pi}^{\pi} \phi(\theta)\exp(ick\cos\theta\cos\eta)\\ &\qquad \cdot\left[\lambda - 2q\cos 2\eta + (ick\cos\theta\sin\eta)^2 - ick\cos\theta\cos\eta\right]d\theta. \end{aligned} \tag{14.100}$$

Here we integrate the last term on the right-hand side by parts, obtaining

$$\begin{aligned}&\frac{d^2\Phi}{d\eta^2} + (\lambda - 2q\cos 2\eta)\Phi(\eta)\\ &\quad = \phi(\theta)(-ick\cos\eta\sin\theta)\exp(ick\cos\theta\cos\eta)\Big|_{\theta=-\pi}^{\pi}\\ &\qquad + \int_{-\pi}^{\pi} \phi(\theta)\exp(ick\cos\theta\cos\eta)[\lambda - 2q\cos 2\eta - ick\cos\theta\cos\eta]\,d\theta\\ &\qquad + \int_{-\pi}^{\pi}\left[-\phi'(\theta)(-ick\cos\eta\sin\theta) + \phi(\theta)ick\cos\eta\cos\theta\right]\exp(ick\cos\theta\cos\eta)\,d\theta\\ &\quad = \int_{-\pi}^{\pi} \exp ick\cos\theta\cos\eta\left[\phi(\theta)(\lambda - 2q\cos 2\eta) + \phi'(\theta)ick\cos\eta\sin\theta\right]d\theta, \end{aligned} \tag{14.101}$$

where the integrated term vanishes if $\phi(-\pi) = \phi(\pi)$, which we assume to be the case. Integrating once more by parts yields

$$\begin{aligned}&\frac{d^2\Phi}{d\eta^2} + (\lambda - 2q\cos 2\eta)\Phi(\eta) = -\phi'(\theta)\exp ick\cos\theta\cos\eta\Big|_{\theta=-\pi}^{\pi}\\ &\qquad + \int_{-\pi}^{\pi} \exp(ick\cos\theta\cos\eta)\left[\phi(\theta)(\lambda - 2q\cos 2\eta) + \phi''(\theta)\right]d\theta, \end{aligned} \tag{14.102}$$

where the integrated term vanishes if ϕ' is periodic with period π, which we assume is the case. Therefore, if $\phi(\theta)$ obeys the angular Mathieu ODE, so does the integral $\Phi(\eta)$, in Eq. (14.99). As a consequence, $\phi(\theta) \sim \Phi(\theta)$, where the constant may be a function of the parameter q.

Thus, we have the main result that a solution $\Phi(\eta)$ of Mathieu's ODE that is even in the variable η satisfies the integral equation

$$\boxed{\Phi(\eta) = \Lambda_n(q)\int_{-\pi}^{\pi} e^{2i\sqrt{q}\cos\theta\cos\eta}\Phi(\theta)\,d\theta.} \tag{14.103}$$

When these Mathieu functions are expanded in a Fourier cosine series and normalized so that the leading term is $\cos n\eta$, they are denoted by $\mathrm{ce}_n(\eta, q)$.

Similarly, solutions of Mathieu's ODE that are odd in η with leading term $\sin n\eta$ in a Fourier series are denoted by $\mathrm{se}_n(\eta, q)$, and they can similarly be shown to obey the integral equation

$$\boxed{\mathrm{se}_n(\eta, q) = s_n(q) \int_{-\pi}^{\pi} \sin\bigl(2i\sqrt{q}\sin\eta\sin\theta\bigr)\mathrm{se}_n(\theta, q)\, d\theta.} \tag{14.104}$$

We now come to the **Fourier expansion for the angular Mathieu functions** and start with

$$\mathrm{se}_1(\eta, q) = \sin\eta + \sum_{\nu=1}^{\infty} \beta_\nu(q)\sin(2\nu+1)\eta,\ \ \beta_\nu(q) = \sum_{\mu=\nu}^{\infty} \beta_\mu^{(\nu)} q^\mu \tag{14.105}$$

as a paradigm for the systematic construction of Mathieu functions of odd parity. Notice the key point that the coefficient β_ν of $\sin(2\nu+1)\eta$ in the Fourier series depends on the parameter q and is expanded in a power series. Moreover, se_1 is normalized so that the coefficient of the leading term, $\sin\eta$, is unity, that is, independent of q. This feature will become important when se_1 is substituted into the angular Mathieu ODE to determine the eigenvalue $\lambda(q)$.

The fact that the β_ν power series in q starts with exponent ν can be proved by a simpler but similar series for $\mathrm{se}_1(\eta, q)$:

$$\mathrm{se}_1(\eta, q) = \sum_{\nu=0}^{\infty} \gamma_\nu(q)\sin^{2\nu+1}\eta, \qquad \gamma_\nu(q) = \sum_{\mu=0}^{\infty} \gamma_\mu^{(\nu)} q^\mu, \tag{14.106}$$

which is useful for this demonstration alone. However, since we need to expand

$$\sin^{2\nu+1}\eta = \sum_{m=0}^{\nu} B_{\nu m}\sin(2m+1)\eta \tag{14.107}$$

with Fourier coefficients

$$B_{\nu m} = \frac{1}{\pi}\int_{-\pi}^{\pi} \sin^{2\nu+1}\eta\sin(2m+1)\eta\, d\eta = \frac{(-1)^m}{2^{2\nu}}\binom{2\nu+1}{\nu-m} \tag{14.108}$$

that we can look up in a table of integrals (see Gradshteyn and Ryzhik in the Additional Readings of Chapter 13), this proof gives us an opportunity to introduce the $B_{\nu m}$ that are nonzero only if $m \le \nu$ and are important ingredients of the recursion relations for the leading terms of se_1 (and all other Mathieu functions of odd parity). Substituting Eq. (14.107) into Eq. (14.106) we obtain

$$\mathrm{se}_1(\eta, q) = \sum_{\nu=0}^{\infty}\sum_{m=0}^{\nu} B_{\nu m}\gamma_\nu(q)\sin(2m+1)\eta. \tag{14.109}$$

Comparing this expression for se_1 with Eq. (14.105) we find

$$\beta_\nu(q) = \sum_{m=\nu}^{\infty} B_{m\nu}\gamma_m(q). \tag{14.110}$$

Here, the sum starts with $m = \nu$ because $B_{m\nu} = 0$ for $m < \nu$.

Next we substitute Eq. (14.106) into the integral Eq. (14.104) for $n = 1$, where we insert the power series for $\sin(2i\sqrt{q}\sin\eta\sin\theta)$. This yields

$$\begin{aligned}\frac{\mathrm{se}_1(\eta,q)}{2\pi s_1(q)} &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sin(2i\sqrt{q}\sin\eta\sin\theta)\mathrm{se}_1(\theta,q)\,d\theta \\ &= \frac{1}{2\pi s_1(q)}\sum_{m=0}^{\infty}\gamma_m(q)\sin^{2m+1}\eta \\ &= i\sqrt{q}\sum_{m,\nu=0}^{\infty}q^m\gamma_\nu(q)\sin^{2m+1}\eta\frac{2^{2m+1}}{(2m+1)!}\frac{1}{2\pi}\int_{-\pi}^{\pi}\sin^{2\nu+2m+2}\theta\,d\theta,\end{aligned} \tag{14.111}$$

from which we obtain the recursion relations

$$\gamma_m(q) = \frac{2^{2m+1}}{(2m+1)!}q^m i\sqrt{q}s_1(q)\sum_{\nu=0}^{\infty}\gamma_\nu(q)\int_{-\pi}^{\pi}\sin^{2\nu+2m+2}\theta\,d\theta, \tag{14.112}$$

upon comparing coefficients of $\sin^{2m+1}\eta$. This shows that the power series for $\gamma_m(q)$ starts with q^m. Using Eq. (14.110) proves that the power series for $\beta_m(q)$ also starts with q^m, and this confirms Eq. (14.105). The integral in Eq. (14.112) can be evaluated analytically and expressed via the beta function (Chapter 8) in terms of ratios of factorials, but we do not need this formula here.

Our next goal is to establish a recursion relation for the leading term $\beta_\nu^{(\nu)}$ of se_1, in Eq. (14.105). We substitute Eq. (14.105) into the integral Eq. (14.104) for $n = 1$, where we insert the power series for $\sin(2i\sqrt{q}\sin\eta\sin\theta)$ again, along with the expansion

$$\frac{1}{2\pi s_1(q)} = i\sum_{m=0}^{\infty}\alpha_m q^{m+1/2}. \tag{14.113}$$

Here, the extra factor, $i\sqrt{q}$, cancels the corresponding factor from the sine in the integral equation. This yields

$$\begin{aligned}&\sum_{\nu=0}^{\infty}\beta_\mu^{(\lambda)}q^{\mu+\nu}\sin^{2\nu+1}\eta\frac{2^{2\nu+1}}{(2\nu+1)!}\frac{1}{2\pi}\int_{-\pi}^{\pi}\sin^{2\nu+2\lambda+2}\theta\,d\theta \\ &\quad= \sum_{m,\mu=0}^{\infty}\alpha_m q^{m+\mu}\beta_\mu^{(\nu)}\sin(2\nu+1)\eta.\end{aligned} \tag{14.114}$$

Here, we replace $\sin^{2\nu+1}\eta$ by $\sin(2m+1)\eta$ using Eq. (14.107). Upon comparing the coefficients of $q^N\sin(2\nu+1)\eta$ for $N = \mu+\nu$ we obtain the recursion relation

$$\sum_{\nu=n}^{N}\sum_{\lambda=0}^{N-\nu}\beta_{N-\nu}^{(\lambda)}\frac{2^{2\nu}}{(2\nu+1)!}B_{\nu n}B_{\nu\lambda} = \sum_{m=0}^{N-n}\alpha_m\beta_{N-m}^{(n)}. \tag{14.115}$$

Now we substitute Eq. (14.108) to obtain the **main recursion relation** for the leading coefficients $\beta_\nu^{(\nu)}$ of se_1:

$$\sum_{\nu=n}^{N}\sum_{\lambda=0}^{N-\nu}\frac{\beta_{N-\nu}^{(\lambda)}}{2^{2\nu}(2\nu+1)!}\binom{2\nu+1}{\nu-n}\binom{2\nu+1}{\nu-\lambda}=\sum_{m=0}^{N-n}\alpha_m\beta_{N-m}^{(n)}. \tag{14.116}$$

Example 14.7.1 LEADING COEFFICIENTS OF se_1

We evaluate Eq. (14.116) starting with $N=0, n=0$. For this case we find $\beta_0^{(0)}=\alpha_0\beta_0^{(0)}$, or $\alpha_0=1$ because the coefficient of $\sin\eta$ in se_1, $\beta_0^{(0)}=1$, by normalization. For $N=1$, $n=0$ Eq. (14.116) yields

$$\alpha_0\beta_1^{(0)}+\alpha_1\beta_0^{(0)}=\beta_1^{(0)}+\frac{1}{4\cdot 3!}\binom{3}{1}\left[\beta_0^{(0)}\binom{3}{1}+\beta_0^{(1)}\binom{3}{0}\right], \tag{14.117}$$

where $\beta_0^{(1)}=0$ and $\beta_1^{(0)}$ drops out, a general feature. Of course, $\beta_1^{(0)}=0$ because $\sin\eta$ in se_1 has coefficient unity. This yields $\alpha_1=3/8$.

The case $N=1, n=1$ yields

$$\frac{-1}{4\cdot 3!}\binom{3}{0}\beta_0^{(0)}\binom{3}{1}=\alpha_0\beta_1^{(1)}, \tag{14.118}$$

or $\beta_1^{(1)}=-1/8$. The leading term is obtained from the general case $n=N$,

$$\frac{(-1)^N}{2^{2N}(2N+1)!}\binom{2N+1}{0}\beta_0^{(0)}\binom{2N+1}{N}=\alpha_0\beta_N^{(N)}, \tag{14.119}$$

as

$$\boxed{\beta_N^{(N)}=\frac{(-1)^N}{2^{2N}(2N+1)!}\binom{2N+1}{N},} \tag{14.120}$$

which was first derived by Mathieu. For $N=1$ this formula reproduces our earlier result, $\beta_1^{(1)}=-1/8$. ■

In order to determine the first nonleading term $\beta_{N+1}^{(N)}$ of se_1, Eq. (14.105), and the eigenvalue $\lambda_1(q)$ we substitute se_1 into the angular Mathieu ODE, Eq. (13.181), using the trigonometric identities

$$2\cos 2\eta\sin(2\nu+1)\eta=\sin(2\nu+3)\eta+\sin(2\nu-1)\eta$$

and

$$\frac{d^2\sin(2\nu+1)\eta}{d\eta^2}=-(2\nu+1)^2\sin(2\nu+1)\eta.$$

This yields

$$
\begin{aligned}
0 &= \frac{d^2\mathrm{se}_1}{d\eta^2} + (\lambda_1 - 2q\cos 2\eta)\mathrm{se}_1 = q(\sin\eta - \sin 3\eta) + \lambda_1 \sin\eta - \sin\eta \\
&\quad + \sum_{\nu=1}^{\infty}\big[\lambda_1 - (2\nu+1)^2\big]\left[\frac{(-1)^\nu q^\nu}{2^{2\nu}\nu!(\nu+1)!} + \beta_{\nu+1}^{(\nu)} q^{\nu+1} + \cdots\right]\sin(2\nu+1)\eta \\
&\quad - q\sum_{\nu=1}^{\infty}\left[\frac{(-q)^\nu}{2^{2\nu}\nu!(\nu+1)!} + \beta_{\nu+1}^{(\nu)} q^{\nu+1} + \cdots\right]\big(\sin(2\nu+3)\eta + \sin(2\nu-1)\eta\big) \\
&= \left(\lambda_1 - 1 + q - q\left[-\frac{q}{2^2 2!} + \beta_2^{(1)} q^2 + \cdots\right]\right)\sin\eta \\
&\quad + \sin 3\eta\left[-q - q\left(\frac{q^2}{2^4 2!3!} + \beta_3^{(2)} q^3\right) + \big(\lambda_1 - 3^2\big)\left(-\frac{q}{2^2 2!} + \beta_2^{(1)} q^2\right)\right] \\
&\quad + \sin(2\nu+1)\eta\big[\lambda_1 - (2\nu+1)^2\big]\left(\frac{(-q)^\nu}{2^{2\nu}\nu!(\nu+1)!} + \beta_{\nu+1}^{(\nu)} q^{\nu+1}\right) \\
&\quad - q\sin(2\nu+1)\eta\left(\frac{(-q)^{\nu+1}}{2^{2(\nu+1)}(\nu+1)!(\nu+2)!} + \beta_{\nu+2}^{(\nu+1)} q^{\nu+2}\right) \\
&\quad - q\sin(2\nu+1)\eta\left(\frac{(-q)^{\nu-1}}{2^{2(\nu-1)}(\nu-1)!\nu!} + \beta_{\nu}^{(\nu-1)} q^{\nu}\right) + \cdots . \qquad (14.121)
\end{aligned}
$$

In this series the coefficient of each power of q within different sine terms must vanish; that of $\sin\eta$ being zero yields the eigenvalue

$$\lambda_1(q) = 1 - q - \frac{1}{8}q^2 + \beta_2^{(1)} q^3 + \cdots , \qquad (14.122)$$

with $\beta_2^{(1)} = 1/2^6$ coming from the vanishing coefficient of q^2 in $\sin 3\eta$. Setting the coefficient of $(-q)^\nu$ in $\sin(2\nu+1)\eta$ equal to zero yields the identity

$$\big[1 - (2\nu+1)^2\big]\frac{1}{2^{2\nu}\nu!(\nu+1)!} + \frac{1}{2^{2(\nu-1)}(\nu-1)!\nu!} = 0, \qquad (14.123)$$

which verifies the correct determination of the leading terms $\beta_\nu^{(\nu)}$ in Eq. (14.120). The vanishing coefficient of $q^{\nu+1}$ in $\sin(2\nu+1)\eta$ yields

$$\frac{(-1)^{\nu+1}}{2^{2\nu}\nu!(\nu+1)!} + \big[1 - (2\nu+1)^2\big]\beta_{\nu+1}^{(\nu)} - \beta_\nu^{(\nu-1)} = 0, \qquad (14.124)$$

which implies the main **recursion relation for nonleading coefficients**,

$$4\nu(\nu+1)\beta_{\nu+1}^{(\nu)} = -\beta_\nu^{(\nu-1)} + \frac{(-1)^{\nu+1}}{2^{2\nu}\nu!(\nu+1)!}, \qquad (14.125)$$

for the first nonleading terms. We verify that

$$\boxed{\beta_{\nu+1}^{(\nu)} = \frac{(-1)^{\nu+1}\nu}{2^{2\nu+2}(\nu+1)!^2}} \qquad (14.126)$$

satisfies this recursion relation. Higher nonleading terms may be obtained by setting to zero the coefficient of $q^{\nu+2}$, etc. Altogether we have derived the Fourier series for

$$\boxed{\mathrm{se}_1(\eta,q)=\sin\eta+\sum_{\nu=1}^{\infty}\left[\frac{(-q)^{\nu}}{2^{2\nu}\nu!(\nu+1)!}+\frac{(-q)^{\nu+1}\nu}{2^{2\nu+2}(\nu+1)!^2}+\cdots\right]\cdot\sin(2\nu+1)\eta.} \tag{14.127}$$

A similar treatment yields the Fourier series for $\mathrm{se}_{2n+1}(\eta,q)$ and $\mathrm{se}_{2n}(\eta,q)$. An invariance of Mathieu's ODE leads to the symmetry relation

$$\mathrm{ce}_{2n+1}(\eta,q)=(-1)^n\mathrm{se}_{2n+1}(\eta+\pi/2,-q), \tag{14.128}$$

which allows us to determine the ce_{2n+1} of period 2π from se_{2n+1}. Similarly, $\mathrm{ce}_{2n}(\eta+\pi/2,-q)=\mathrm{se}_{2n}(\eta,q)$ relates these Mathieu functions of period π to each other.

Finally, we briefly outline a derivation of the Fourier series for

$$\mathrm{ce}_0(\eta,q)=1+\sum_{n=1}^{\infty}\beta_n(q)\cos 2n\eta,\qquad \beta_n(q)=\sum_{m=n}^{\infty}\beta_m^{(n)}q^m, \tag{14.129}$$

as a paradigm for the Mathieu functions of period π. Note that this normalization agrees with Whittaker and Watson and with Hochstadt in the Additional Readings of Chapter 13, whereas in AMS-55 (for the full reference see footnote 4 in Chapter 5) ce_0 differs by a factor of $1/\sqrt{2}$. The symmetry relation from the Mathieu ODE,

$$\mathrm{ce}_0\left(\frac{\pi}{2}-\eta,-q\right)=\mathrm{ce}_0(\eta,q), \tag{14.130}$$

implies

$$\beta_n(-q)=(-1)^n\beta_n(q); \tag{14.131}$$

that is, β_{2n} contains only even powers of q and β_{2n+1} only odd powers.

The fact that the power series for $\beta_n(q)$ in Eq. (14.129) starts with q^n can be proved by the similar expansion

$$\mathrm{ce}_0(\eta,q)=\sum_{n=0}^{\infty}\gamma_n(q)\cos^{2n}\eta,\qquad \gamma_n(q)=\sum_{\mu=0}^{\infty}\gamma_{\mu}^{(n)}q^{\mu}, \tag{14.132}$$

as for se_1 in Eqs. (14.105) to (14.112). Substituting Eq. (14.132) into the integral equation

$$\mathrm{ce}_0(\eta,q)=c_0(q)\int_{-\pi}^{\pi}e^{2i\sqrt{q}\cos\theta\cos\eta}\mathrm{ce}(\theta,q)\,d\theta, \tag{14.133}$$

inserting the power series for the exponential function (odd powers $\cos^{2m+1}\theta$ drop out) and equating the coefficients of $\cos^{2m}\eta$ yields

$$\gamma_m(q)=c_1(q)(-q)^m\frac{2^{2m}}{(2m)!}\sum_{\mu=0}^{\infty}\gamma_{\mu}(q)\int_{-\pi}^{\pi}\cos^{2m+2\mu}\theta\,d\theta. \tag{14.134}$$

This recursion relation shows that the power series for $\gamma_m(q)$ starts with q^m. We expand

$$\cos^{2n}\eta = \sum_{m=0}^{n} A_{nm}\cos 2m\eta \tag{14.135}$$

with Fourier coefficients

$$A_{nm} = \frac{2}{\pi}\int_{-\pi/2}^{\pi/2}\cos^{2n}\eta\cos 2m\eta\,d\eta = \frac{1}{2^{2n-1}}\binom{2n}{n-m}, \tag{14.136}$$

which are nonzero only when $m \le n$. Using this result to replace the cosine powers in Eq. (14.132) by $\cos 2m\eta$ we obtain

$$\beta_n(q) = \sum_{m=n}^{\infty} A_{mn}\gamma_m(q), \tag{14.137}$$

confirming Eq. (14.129).

Proceeding as for se_1 in Eqs. (14.113) to (14.120) we substitute Eq. (14.129) into the integral Eq. (14.133) and obtain

$$\sum_{m,\mu,\nu,\lambda=0}^{\infty}(-1)^m\frac{2^{2m}}{(2m)!}q^{\mu+m}\sum_{\nu=0}^{m}A_{m\nu}\cos(2\nu\eta)A_{m\lambda}$$

$$= \sum_{m,\mu,\nu=0}^{\infty}\alpha_m\beta_\mu^{(\nu)}q^{m+\mu}\cos(2\nu\eta), \tag{14.138}$$

with

$$\frac{1}{2\pi c_1(q)} = \sum_{m=0}^{\infty}\alpha_m q^m. \tag{14.139}$$

Upon comparing the coefficient of $q^N\cos(2\nu\eta)$ with $N = m+\mu$, we extract the **recursion relation for leading coefficients** $\beta_n^{(n)}$ of ce_0

$$\sum_{m=\nu}^{N}(-1)^m\frac{2^{2m}}{(2m)!}A_{m\nu}\sum_{\lambda=0}^{m}\beta_{N-m}^{(\lambda)}A_{m\lambda} = \sum_{m=0}^{N}\alpha_m\beta_{N-m}^{(\nu)}, \tag{14.140}$$

with A_{nm} in Eq. (14.136).

Example 14.7.2 LEADING COEFFICIENTS FOR ce_0

The case $N = 0$, $\nu = 0$ of Eq. (14.140) yields

$$A_{00}^2\beta_0^{(0)} = \alpha_0\beta_0^{(0)}, \tag{14.141}$$

with $A_{00} = 1$ and $\beta_0^{(0)} = 1$ from normalizing the leading term of ce_0 to unity so that $\alpha_0 = 1$ results.

The case $N = 1$, $\nu = 0$ yields

$$A_{00}\beta_1^{(0)}A_{00} - 2A_{10}\left[\beta_0^{(0)}A_{10} + \beta_0^{(1)}A_{11}\right] = \alpha_0\beta_1^{(0)} + \alpha_1\beta_0^{(0)}, \tag{14.142}$$

with $\beta_0^{(1)} = 0$ by Eq. (14.129). This simplifies to

$$\beta_1^{(0)} - \frac{1}{2}\beta_0^{(0)} = \alpha_1\beta_0^{(0)} + \beta_1^{(0)}, \tag{14.143}$$

where $\beta_1^{(0)}$ drops out. We know already that $\beta_1^{(0)} = 0$ from the leading term unity of ce_0. Therefore, $\alpha_1 = -1/2$.

For the case $N = 1$, $\nu = 1$ we obtain

$$-2A_{11}\beta_0^{(0)}A_{10} = \alpha_0\beta_1^{(1)}, \tag{14.144}$$

with $A_{10} = 1/2 = A_{11}$, from which $\beta_1^{(1)} = -1/2$ follows. For the case $N = 2$, $\nu = 2$ we find

$$\frac{2^4}{4!2^3}\beta_0^{(0)}A_{20} = \alpha_0\beta_2^{(2)}, \tag{14.145}$$

with $A_{20} = 3/8$, from which $\beta_2^{(2)} = 2^{-4}$ follows. The general case N, $\nu = N$ yields

$$(-1)^N\frac{2^{2N}}{(2N)!}A_{NN}\beta_0^{(0)}A_{N0} = \alpha_0\beta_N^{(N)}, \tag{14.146}$$

with $A_{NN} = 2^{-2N+1}$, $A_{N0} = \frac{1}{2^{2N-1}}\binom{2N}{N}$, from which the leading term

$$\boxed{\beta_N^{(N)} = \frac{(-1)^N}{2^{2N-1}(2N)!}\binom{2N}{N} = \frac{(-1)^N}{2^{2N-1}N!^2}} \tag{14.147}$$

follows. ■

The nonleading terms $\beta_{N+1}^{(N)}$ of ce_0 are best determined from the angular Mathieu ODE by substitution of Eq. (14.129), in analogy with se_1, Eqs. (14.121) to (14.127). Using the identities

$$\begin{aligned} 2\cos(2n\eta)\cos 2\eta &= \cos(2n+2)\eta + \cos(2n-2)\eta, \\ \frac{d^2}{d\eta^2}\cos(2n\eta) &= -(2n)^2\cos(2n\eta), \end{aligned} \tag{14.148}$$

we obtain

$$\begin{aligned} \frac{d^2\mathrm{ce}_0}{d\eta^2} + \big(\lambda_0(q) - 2q\cos 2\eta\big)\mathrm{ce}_0 = 0 \\ = \lambda_0(q) - q\left(-\frac{q}{2} + \frac{7}{2^7}q^3\right) + \cdots \\ + \sum_{n=1}^{\infty}\big(\lambda_0 - 4n^2\big)\cos(2n\eta)\left[\frac{(-q)^n}{2^{2n-1}n!^2} + \beta_{n+2}^{(n)}q^{n+2}\right] \\ - q\sum_{n=1}^{\infty}\left[\frac{(-q)^n}{2^{2n-1}n!^2} + \beta_{n+2}^{(n)}q^{n+2}\right] \\ \times\big[\cos(2n+2)\eta + \cos(2n-2)\eta\big]. \end{aligned} \tag{14.149}$$

Setting the coefficient of $\cos(2n\eta)$ for $n = 0$ to zero yields the eigenvalue

$$\lambda_0 = -\frac{1}{2}q^2 + \frac{7}{2^7}q^4 + \cdots . \tag{14.150}$$

The coefficient of $\cos(2n\eta)q^n$ yields an identity,

$$(-1)^{n+1}\frac{4n^2}{2^{2n-1}n!^2} + \frac{(-1)^n}{2^{2n-3}(n-1)!^2} = 0, \tag{14.151}$$

which shows that the leading term in Eq. (14.147) was correctly determined. The coefficient of $q^{n+2}\cos(2n\eta)$ yields the recursion relation

$$-4n^2\beta_{n+2}^{(n)} - \beta_{n+1}^{(n-1)} + \frac{(-1)^{n+1}}{2^{2n}n!^2} + \frac{(-1)^n}{2^{2n+1}(n+1)!^2} = 0. \tag{14.152}$$

It is straightforward to check that

$$\beta_{n+2}^{(n)} = (-1)^{n+1}\frac{n(3n+4)}{2^{2n+3}(n+1)!^2} \tag{14.153}$$

satisfies this recursion relation. Altogether we have derived the formula

$$\begin{aligned}\mathrm{ce}_0(\eta, q) &= 1 + \cos 2\eta\left[-\frac{1}{2}q^2 + \frac{7}{2^7}q^3 + \cdots\right] \\ &\quad + \cos 4\eta\left[\frac{q^2}{2^5} + \cdots\right] + \cos 6\eta\left[-\frac{q^3}{2^7 3^2} + \cdots\right] \\ &= 1 + \sum_{n=1}^{\infty}\cos(2n\eta)\left[\frac{(-q)^n}{2^{2n-1}n!^2} + \frac{(-1)^{n+1}n(3n+4)q^{n+2}}{2^{2n+3}(n+1)!^2} + \cdots\right].\end{aligned} \tag{14.154}$$

Similarly one can derive

$$\mathrm{ce}_1(\eta, q) = \cos\eta + \sum_{n=1}^{\infty}\cos(2n+1)\eta\left[\frac{(-q)^n}{2^{2n}n!(n+1)!} - \frac{(-q)^{n+1}n}{2^{2n+2}(n+1)1^2} + \cdots\right], \tag{14.155}$$

whose eigenvalue is given by the power series

$$\lambda_1(q) = 1 + q - \frac{1}{8}q^2 - \frac{1}{2^6}q^3 + \cdots . \tag{14.156}$$

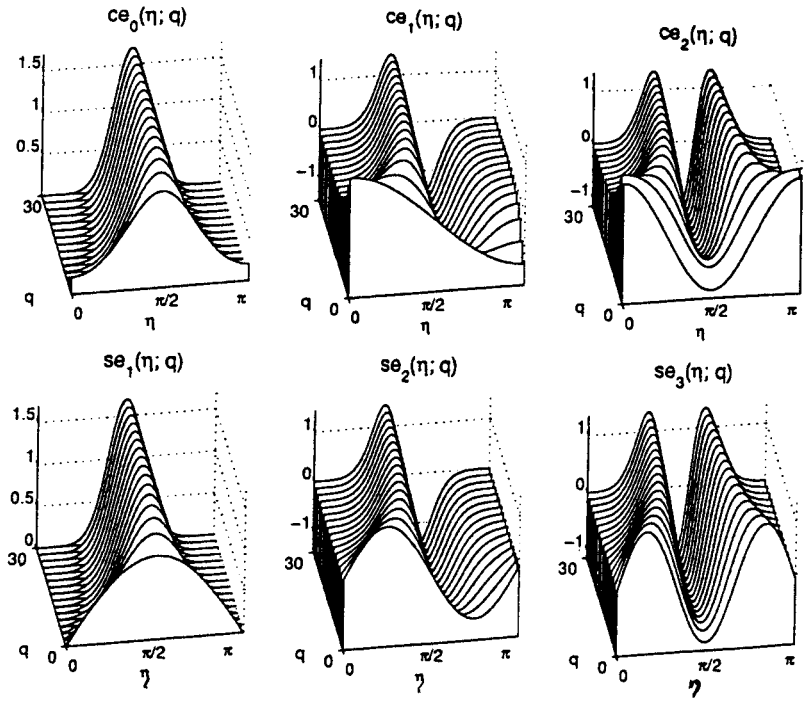

FIGURE 14.12 Angular Mathieu functions. (From Gutiérrez-Vega *et al.*, *Am. J. Phys.* **71**: 233 (2003).)

Exercises

14.7.1 Determine the nonleading coefficients $\beta_{n+2}^{(n)}$ for se_1. Derive a suitable recursion relation.

14.7.2 Determine the nonleading coefficients $\beta_{n+4}^{(n)}$ for ce_0. Derive the corresponding recursion relation.

14.7.3 Derive the formula for ce_1, Eq. (14.155), and its eigenvalue, Eq. (14.156).

Additional Readings

Carslaw, H. S., *Introduction to the Theory of Fourier's Series and Integrals*, 2nd ed. London: Macmillan (1921); 3rd ed., paperback, New York: Dover (1952). This is a detailed and classic work; includes a considerable discussion of Gibbs phenomenon in Chapter IX.

Hamming, R. W., *Numerical Methods for Scientists and Engineers*, 2nd ed. New York: McGraw-Hill (1973), reprinted Dover (1987). Chapter 33 provides an excellent description of the fast Fourier transform.

Jeffreys, H., and B. S. Jeffreys, *Methods of Mathematical Physics*, 3rd ed. Cambridge, UK: Cambridge University Press (1972).

Kufner, A., and J. Kadlec, *Fourier Series*. London: Iliffe (1971). This book is a clear account of Fourier series in the context of Hilbert space.

Lanczos, C., *Applied Analysis*, Englewood Cliffs, NJ: Prentice-Hall (1956), reprinted Dover (1988). The book gives a well-written presentation of the Lanczos convergence technique (which suppresses the Gibbs phenomenon oscillations). This and several other topics are presented from the point of view of a mathematician who wants useful numerical results and not just abstract existence theorems.

Oberhettinger, F., *Fourier Expansions, A Collection of Formulas*. New York, Academic Press (1973).

Zygmund, A., *Trigonometric Series*. Cambridge, UK: Cambridge University Press (1988). The volume contains an extremely complete exposition, including relatively recent results in the realm of pure mathematics.

CHAPTER 15

INTEGRAL TRANSFORMS

15.1 INTEGRAL TRANSFORMS

Frequently in mathematical physics we encounter pairs of functions related by an expression of the form

$$g(\alpha) = \int_a^b f(t)K(\alpha, t)\,dt. \tag{15.1}$$

The function $g(\alpha)$ is called the (integral) transform of $f(t)$ by the kernel $K(\alpha, t)$. The operation may also be described as mapping a function $f(t)$ in t-space into another function, $g(\alpha)$, in α-space. This interpretation takes on physical significance in the time-frequency relation of Fourier transforms, as in Example 15.3.1, and in the real space–momentum space relations in quantum physics of Section 15.6.

Fourier Transform

One of the most useful of the infinite number of possible transforms is the Fourier transform, given by

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)e^{i\omega t}\,dt. \tag{15.2}$$

Two modifications of this form, developed in Section 15.3, are the Fourier cosine and Fourier sine transforms:

$$g_c(\omega) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} f(t)\cos\omega t\,dt, \tag{15.3}$$

$$g_s(\omega) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} f(t)\sin\omega t\,dt. \tag{15.4}$$

The Fourier transform is based on the kernel $e^{i\omega t}$ and its real and imaginary parts taken separately, $\cos\omega t$ and $\sin\omega t$. Because these kernels are the functions used to describe waves, Fourier transforms appear frequently in studies of waves and the extraction of information from waves, particularly when phase information is involved. The output of a stellar interferometer, for instance, involves a Fourier transform of the brightness across a stellar disk. The electron distribution in an atom may be obtained from a Fourier transform of the amplitude of scattered X-rays. In quantum mechanics the physical origin of the Fourier relations of Section 15.6 is the wave nature of matter and our description of matter in terms of waves.

Example 15.1.1 FOURIER TRANSFORM OF GAUSSIAN

The Fourier transform of a Gaussian function $e^{-a^2t^2}$,

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-a^2t^2} e^{i\omega t}\, dt,$$

can be done analytically by completing the square in the exponent,

$$-a^2t^2 + i\omega t = -a^2\left(t - \frac{i\omega}{2a^2}\right)^2 - \frac{\omega^2}{4a^2},$$

which we check by evaluating the square. Substituting this identity we obtain

$$g(\omega) = \frac{1}{\sqrt{2\pi}} e^{-\omega^2/4a^2} \int_{-\infty}^{\infty} e^{-a^2t^2}\, dt,$$

upon shifting the integration variable $t \to t + \frac{i\omega}{2a^2}$. This is justified by an application of Cauchy's theorem to the rectangle with vertices $-T$, T, $T + \frac{i\omega}{2a^2}$, $-T + \frac{i\omega}{2a^2}$ for $T \to \infty$, noting that the integrand has no singularities in this region and that the integrals over the sides from $\pm T$ to $\pm T + \frac{i\omega}{2a^2}$ become negligible for $T \to \infty$. Finally we rescale the integration variable as $\xi = at$ in the integral (see Eqs. (8.6) and (8.8)):

$$\int_{-\infty}^{\infty} e^{-a^2t^2}\, dt = \frac{1}{a} \int_{-\infty}^{\infty} e^{-\xi^2}\, d\xi = \frac{\sqrt{\pi}}{a}.$$

Substituting these results we find

$$g(\omega) = \frac{1}{a\sqrt{2}} \exp\left(-\frac{\omega^2}{4a^2}\right),$$

again a Gaussian, but in ω-space. The bigger a is, that is, the narrower the original Gaussian $e^{-a^2t^2}$ is, the wider is its Fourier transform $\sim e^{-\omega^2/4a^2}$. ■

Laplace, Mellin, and Hankel Transforms

Three other useful kernels are

$$e^{-\alpha t}, \qquad t J_n(\alpha t), \qquad t^{\alpha-1}.$$

These give rise to the following transforms

$$g(\alpha) = \int_0^\infty f(t) e^{-\alpha t} dt, \qquad \text{Laplace transform} \tag{15.5}$$

$$g(\alpha) = \int_0^\infty f(t) t J_n(\alpha t)\, dt, \qquad \text{Hankel transform (Fourier–Bessel)} \tag{15.6}$$

$$g(\alpha) = \int_0^\infty f(t) t^{\alpha-1} dt, \qquad \text{Mellin transform.} \tag{15.7}$$

Clearly, the possible types are unlimited. These transforms have been useful in mathematical analysis and in physical applications. We have actually used the Mellin transform without calling it by name; that is, $g(\alpha) = (\alpha - 1)!$ is the Mellin transform of $f(t) = e^{-t}$. See E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, 2nd ed., New York: Oxford University Press (1937), for more Mellin transforms. Of course, we could just as well say $g(\alpha) = n!/\alpha^{n+1}$ is the Laplace transform of $f(t) = t^n$. Of the three, the Laplace transform is by far the most used. It is discussed at length in Sections 15.8 to 15.12. The Hankel transform, a Fourier transform for a Bessel function expansion, represents a limiting case of a Fourier–Bessel series. It occurs in potential problems in cylindrical coordinates and has been applied extensively in acoustics.

Linearity

All these integral transforms are linear; that is,

$$\int_a^b \big[c_1 f_1(t) + c_2 f_2(t)\big] K(\alpha, t)\, dt$$

$$= c_1 \int_a^b f_1(t) K(\alpha, t)\, dt + c_2 \int_a^b f_2(t) K(\alpha, t)\, dt, \tag{15.8}$$

$$\int_a^b c f(t) K(\alpha, t)\, dt = c \int_a^b f(t) K(\alpha, t)\, dt, \tag{15.9}$$

where c_1 and c_2 are constants and $f_1(t)$ and $f_2(t)$ are functions for which the transform operation is defined.

Representing our linear integral transform by the operator $\mathcal{L}$, we obtain

$$g(\alpha) = \mathcal{L} f(t). \tag{15.10}$$

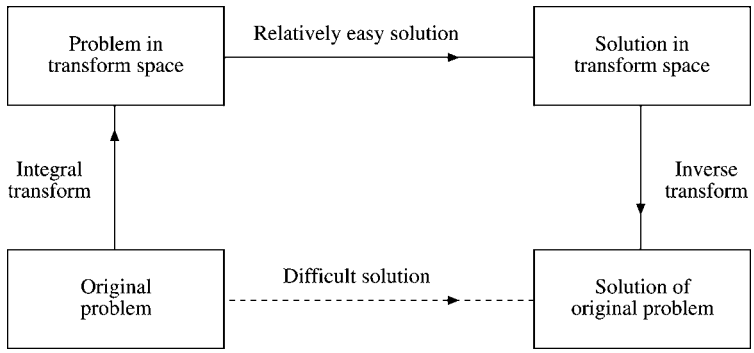

FIGURE 15.1 Schematic integral transforms.

We expect an inverse operator $\mathcal{L}^{-1}$ exists such that[1]

$$f(t) = \mathcal{L}^{-1} g(\alpha). \tag{15.11}$$

For our three Fourier transforms $\mathcal{L}^{-1}$ is given in Section 15.3. In general, the determination of the inverse transform is the main problem in using integral transforms. The inverse Laplace transform is discussed in Section 15.12. For details of the inverse Hankel and inverse Mellin transforms we refer to the Additional Readings at the end of the chapter.

Integral transforms have many special physical applications and interpretations that are noted in the remainder of this chapter. The most common application is outlined in Fig. 15.1. Perhaps an original problem can be solved only with difficulty, if at all, in the original coordinates (space). It often happens that the transform of the problem can be solved relatively easily. Then the inverse transform returns the solution from the transform coordinates to the original system. Example 15.4.1 and Exercise 15.4.1 illustrate this technique.

Exercises

15.1.1 The Fourier transforms for a function of two variables are

$$F(u, v) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int f(x, y) e^{i(ux+vy)}\, dx\, dy,$$

$$f(x, y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int F(u, v) e^{-i(ux+vy)}\, du\, dv.$$

Using $f(x, y) = f([x^2 + y^2]^{1/2})$, show that the zero-order Hankel transforms

$$F(\rho) = \int_0^{\infty} r f(r) J_0(\rho r)\, dr,$$

$$f(r) = \int_0^{\infty} \rho F(\rho) J_0(\rho r)\, d\rho,$$

are a special case of the Fourier transforms.

[1] Expectation is not proof, and here proof of existence is complicated because we are actually in an **infinite**-dimensional Hilbert space. We shall prove existence in the special cases of interest by actual construction.

This technique may be generalized to derive the Hankel transforms of order $\nu = 0, \frac{1}{2}, 1, \frac{1}{2}, \ldots$ (compare I. N. Sneddon, *Fourier Transforms*, New York: McGraw-Hill (1951)). A more general approach, valid for $\nu > -\frac{1}{2}$, is presented in Sneddon's *The Use of Integral Transforms* (New York: McGraw-Hill (1972)). It might also be noted that the Hankel transforms of nonintegral order $\nu = \pm\frac{1}{2}$ reduce to Fourier sine and cosine transforms.

15.1.2 Assuming the validity of the Hankel transform–inverse transform pair of equations

$$g(\alpha) = \int_0^\infty f(t) J_n(\alpha t) t\, dt,$$

$$f(t) = \int_0^\infty g(\alpha) J_n(\alpha t) \alpha\, d\alpha,$$

show that the Dirac delta function has a Bessel integral representation

$$\delta(t - t') = t \int_0^\infty J_n(\alpha t) J_n(\alpha t') \alpha\, d\alpha.$$

This expression is useful in developing Green's functions in cylindrical coordinates, where the eigenfunctions are Bessel functions.

15.1.3 From the Fourier transforms, Eqs. (15.22) and (15.23), show that the transformation

$$t \to \ln x$$

$$i\omega \to \alpha - \gamma$$

leads to

$$G(\alpha) = \int_0^\infty F(x) x^{\alpha-1} dx$$

and

$$F(x) = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} G(\alpha) x^{-\alpha} d\alpha.$$

These are the Mellin transforms. A similar change of variables is employed in Section 15.12 to derive the inverse Laplace transform.

15.1.4 Verify the following Mellin transforms:

(a) $$\int_0^\infty x^{\alpha-1} \sin(kx)\, dx = k^{-\alpha}(\alpha - 1)! \sin\frac{\pi\alpha}{2}, \qquad -1 < \alpha < 1.$$

(b) $$\int_0^\infty x^{\alpha-1} \cos(kx)\, dx = k^{-\alpha}(\alpha - 1)! \cos\frac{\pi\alpha}{2}, \qquad 0 < \alpha < 1.$$

Hint. You can force the integrals into a tractable form by inserting a convergence factor e^{-bx} and (after integrating) letting $b \to 0$. Also, $\cos kx + i \sin kx = \exp ikx$.

15.2 DEVELOPMENT OF THE FOURIER INTEGRAL

In Chapter 14 it was shown that Fourier series are useful in representing certain functions (1) over a limited range $[0, 2\pi]$, $[-L, L]$, and so on, or (2) for the infinite interval $(-\infty, \infty)$, **if the function is periodic**. We now turn our attention to the problem of representing a nonperiodic function over the infinite range. Physically this means resolving a single pulse or wave packet into sinusoidal waves.

We have seen (Section 14.2) that for the interval $[-L, L]$ the coefficients a_n and b_n could be written as

$$a_n = \frac{1}{L}\int_{-L}^{L} f(t)\cos\frac{n\pi t}{L}dt, \tag{15.12}$$

$$b_n = \frac{1}{L}\int_{-L}^{L} f(t)\sin\frac{n\pi t}{L}dt. \tag{15.13}$$

The resulting Fourier series is

$$f(x) = \frac{1}{2L}\int_{-L}^{L} f(t)\,dt + \frac{1}{L}\sum_{n=1}^{\infty}\cos\frac{n\pi x}{L}\int_{-L}^{L} f(t)\cos\frac{n\pi t}{L}dt + \frac{1}{L}\sum_{n=1}^{\infty}\sin\frac{n\pi x}{L}\int_{-L}^{L} f(t)\sin\frac{n\pi t}{L}dt, \tag{15.14}$$

or

$$f(x) = \frac{1}{2L}\int_{-L}^{L} f(t)\,dt + \frac{1}{L}\sum_{n=1}^{\infty}\int_{-L}^{L} f(t)\cos\frac{n\pi}{L}(t-x)\,dt. \tag{15.15}$$

We now let the parameter L approach infinity, transforming the finite interval $[-L, L]$ into the infinite interval $(-\infty, \infty)$. We set

$$\frac{n\pi}{L} = \omega, \qquad \frac{\pi}{L} = \Delta\omega, \qquad \text{with } L \to \infty.$$

Then we have

$$f(x) \to \frac{1}{\pi}\sum_{n=1}^{\infty}\Delta\omega\int_{-\infty}^{\infty} f(t)\cos\omega(t-x)\,dt, \tag{15.16}$$

or

$$f(x) = \frac{1}{\pi}\int_{0}^{\infty} d\omega\int_{-\infty}^{\infty} f(t)\cos\omega(t-x)\,dt, \tag{15.17}$$

replacing the infinite sum by the integral over ω. The first term (corresponding to a_0) has vanished, assuming that $\int_{-\infty}^{\infty} f(t)\,dt$ exists.

It must be emphasized that this result (Eq. (15.17)) is purely formal. It is not intended as a rigorous derivation, but it can be made rigorous (compare I. N. Sneddon, *Fourier Transforms*, Section 3.2). We take Eq. (15.17) as the Fourier integral. It is subject to the conditions that $f(x)$ is (1) piecewise continuous, (2) piecewise differentiable, and (3) absolutely integrable — that is, $\int_{-\infty}^{\infty} |f(x)|\,dx$ is finite.

Fourier Integral — Exponential Form

Our Fourier integral (Eq. (15.17)) may be put into exponential form by noting that

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} f(t) \cos \omega(t - x)\, dt, \tag{15.18}$$

whereas

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} f(t) \sin \omega(t - x)\, dt = 0; \tag{15.19}$$

$\cos \omega(t - x)$ is an even function of ω and $\sin \omega(t - x)$ is an odd function of ω. Adding Eqs. (15.18) and (15.19) (with a factor i), we obtain the **Fourier integral theorem**

$$\boxed{f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega x}\, d\omega \int_{-\infty}^{\infty} f(t) e^{i\omega t}\, dt.} \tag{15.20}$$

The variable ω introduced here is an arbitrary mathematical variable. In many physical problems, however, it corresponds to the angular frequency ω. We may then interpret Eq. (15.18) or (15.20) as a representation of $f(x)$ in terms of a distribution of infinitely long sinusoidal wave trains of angular frequency ω, in which this frequency is a **continuous** variable.

Dirac Delta Function Derivation

If the order of integration of Eq. (15.20) is reversed, we may rewrite it as

$$\boxed{f(x) = \int_{-\infty}^{\infty} f(t) \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega(t-x)} d\omega \right\} dt.} \tag{15.20a}$$

Apparently the quantity in curly brackets behaves as a delta function $\delta(t - x)$. We might take Eq. (15.20a) as presenting us with a representation of the Dirac delta function. Alternatively, we take it as a clue to a new derivation of the Fourier integral theorem.

From Eq. (1.171b) (shifting the singularity from $t = 0$ to $t = x$),

$$f(x) = \lim_{n\to\infty} \int_{-\infty}^{\infty} f(t) \delta_n(t - x) dt, \tag{15.21a}$$

where $\delta_n(t - x)$ is a sequence defining the distribution $\delta(t - x)$. Note that Eq. (15.21a) assumes that $f(t)$ is continuous at $t = x$. We take $\delta_n(t - x)$ to be

$$\delta_n(t - x) = \frac{\sin n(t - x)}{\pi(t - x)} = \frac{1}{2\pi} \int_{-n}^{n} e^{i\omega(t-x)}\, d\omega, \tag{15.21b}$$

using Eq. (1.174). Substituting into Eq. (15.21a), we have

$$f(x) = \lim_{n\to\infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) \int_{-n}^{n} e^{i\omega(t-x)} d\omega\, dt. \tag{15.21c}$$

Interchanging the order of integration and then taking the limit as $n \to \infty$, we have Eq. (15.20), the Fourier integral theorem.

With the understanding that it belongs under an integral sign, as in Eq. (15.21a), the identification

$$\boxed{\delta(t-x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega(t-x)}\,d\omega} \tag{15.21d}$$

provides a very useful representation of the delta function.

15.3 FOURIER TRANSFORMS — INVERSION THEOREM

Let us **define** $g(\omega)$, the Fourier transform of the function $f(t)$, by

$$g(\omega) \equiv \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(t)e^{i\omega t}\,dt. \tag{15.22}$$

Exponential Transform

Then, from Eq. (15.20), we have the inverse relation,

$$f(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g(\omega)e^{-i\omega t}\,d\omega. \tag{15.23}$$

Note that Eqs. (15.22) and (15.23) are almost but not quite symmetrical, differing in the sign of i.

Here two points deserve comment. First, the $1/\sqrt{2\pi}$ symmetry is a matter of choice, not of necessity. Many authors will attach the entire $1/2\pi$ factor of Eq. (15.20) to one of the two equations: Eq. (15.22) or Eq. (15.23). Second, although the Fourier integral, Eq. (15.20), has received much attention in the mathematics literature, we shall be primarily interested in the Fourier transform and its inverse. They are the equations with physical significance.

When we move the Fourier transform pair to three-dimensional space, it becomes

$$g(\mathbf{k}) = \frac{1}{(2\pi)^{3/2}}\int f(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}\,d^3r, \tag{15.23a}$$

$$f(\mathbf{r}) = \frac{1}{(2\pi)^{3/2}}\int g(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{r}}\,d^3k. \tag{15.23b}$$

The integrals are over all space. Verification, if desired, follows immediately by substituting the left-hand side of one equation into the integrand of the other equation and using the three-dimensional delta function.[2] Equation (15.23b) may be interpreted as an expansion of a function $f(\mathbf{r})$ in a continuum of plane wave eigenfunctions; $g(\mathbf{k})$ then becomes the amplitude of the wave, $\exp(-i\mathbf{k}\cdot\mathbf{r})$.

[2] $\delta(\mathbf{r}_1-\mathbf{r}_2) = \delta(x_1-x_2)\delta(y_1-y_2)\delta(z_1-z_2)$ with Fourier integral $\delta(x_1-x_2) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp[ik_1(x_1-x_2)]\,dk_1$, etc.

Cosine Transform

If $f(x)$ is odd or even, these transforms may be expressed in a somewhat different form. Consider first an even function f_c with $f_c(x) = f_c(-x)$. Writing the exponential of Eq. (15.22) in trigonometric form, we have

$$g_c(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f_c(t)(\cos\omega t + i \sin\omega t)\, dt$$

$$= \sqrt{\frac{2}{\pi}} \int_0^{\infty} f_c(t) \cos\omega t\, dt, \tag{15.24}$$

the $\sin\omega t$ dependence vanishing on integration over the symmetric interval $(-\infty, \infty)$. Similarly, since $\cos\omega t$ is even, Eqs. (15.23) transforms to

$$f_c(x) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} g_c(\omega) \cos\omega x\, d\omega. \tag{15.25}$$

Equations (15.24) and (15.25) are known as Fourier cosine transforms.

Sine Transform

The corresponding pair of Fourier sine transforms is obtained by assuming that $f_s(x) = -f_s(-x)$, odd, and applying the same symmetry arguments. The equations are

$$g_s(\omega) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} f_s(t) \sin\omega t\, dt,^{3} \tag{15.26}$$

$$f_s(x) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} g_s(\omega) \sin\omega x\, d\omega. \tag{15.27}$$

From the last equation we may develop the physical interpretation that $f(x)$ is being described by a continuum of sine waves. The amplitude of $\sin\omega x$ is given by $\sqrt{2/\pi}\, g_s(\omega)$, in which $g_s(\omega)$ is the Fourier sine transform of $f(x)$. It will be seen that Eq. (15.27) is the integral analog of the summation (Eq. (14.24)). Similar interpretations hold for the cosine and exponential cases.

If we take Eqs. (15.22), (15.24), and (15.26) as the direct integral transforms, described by $\mathcal{L}$ in Eq. (15.10) (Section 15.1), the corresponding inverse transforms, $\mathcal{L}^{-1}$ of Eq. (15.11), are given by Eqs. (15.23), (15.25), and (15.27).

Note that the Fourier cosine transforms and the Fourier sine transforms each involve only positive values (and zero) of the arguments. We use the parity of $f(x)$ to establish the transforms; but once the transforms are established, the behavior of the functions f and g for negative argument is irrelevant. In effect, the transform equations themselves impose a **definite parity**: **even for the Fourier cosine** transform and **odd for the Fourier sine** transform.

[3] Note that a factor $-i$ has been absorbed into this $g(\omega)$.

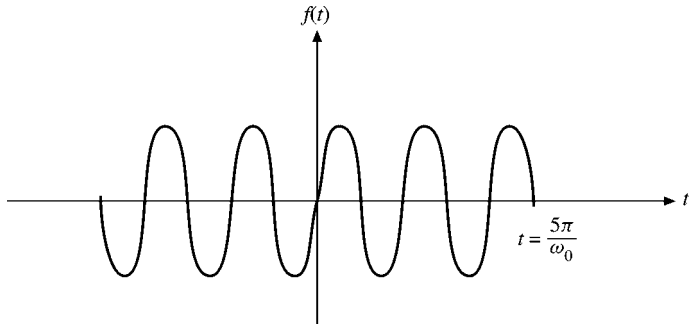

FIGURE 15.2 Finite wave train.

Example 15.3.1 FINITE WAVE TRAIN

An important application of the Fourier transform is the resolution of a finite pulse into sinusoidal waves. Imagine that an infinite wave train $\sin\omega_0 t$ is clipped by Kerr cell or saturable dye cell shutters so that we have

$$f(t) = \begin{cases} \sin\omega_0 t, & |t| < \dfrac{N\pi}{\omega_0}, \\ 0, & |t| > \dfrac{N\pi}{\omega_0}. \end{cases} \tag{15.28}$$

This corresponds to N cycles of our original wave train (Fig. 15.2). Since $f(t)$ is odd, we may use the Fourier sine transform (Eq. (15.26)) to obtain

$$g_s(\omega) = \sqrt{\frac{2}{\pi}} \int_0^{N\pi/\omega_0} \sin\omega_0 t \sin\omega t \, dt. \tag{15.29}$$

Integrating, we find our amplitude function:

$$g_s(\omega) = \sqrt{\frac{2}{\pi}} \left[\frac{\sin[(\omega_0 - \omega)(N\pi/\omega_0)]}{2(\omega_0 - \omega)} - \frac{\sin[(\omega_0 + \omega)(N\pi/\omega_0)]}{2(\omega_0 + \omega)} \right]. \tag{15.30}$$

It is of considerable interest to see how $g_s(\omega)$ depends on frequency. For large ω_0 and $\omega \approx \omega_0$, only the first term will be of any importance because of the denominators. It is plotted in Fig. 15.3. This is the amplitude curve for the single-slit diffraction pattern.

There are zeros at

$$\frac{\omega_0 - \omega}{\omega_0} = \frac{\Delta\omega}{\omega_0} = \pm\frac{1}{N}, \pm\frac{2}{N}, \qquad \text{and so on.} \tag{15.31}$$

For large N, $g_s(\omega)$ may also be interpreted as a Dirac delta distribution, as in Section 1.15. Since the contributions outside the central maximum are small in this case, we may take

$$\Delta\omega = \frac{\omega_0}{N} \tag{15.32}$$

as a good measure of the spread in frequency of our wave pulse. Clearly, if N is large (a long pulse), the frequency spread will be small. On the other hand, if our pulse is clipped

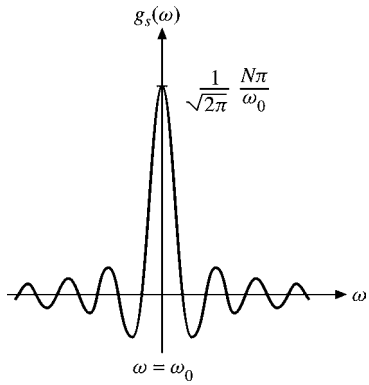

FIGURE 15.3 Fourier transform of finite wave train.

short, N small, the frequency distribution will be wider and the secondary maxima are more important. ■

Uncertainty Principle

Here is a classical analog of the famous uncertainty principle of quantum mechanics. If we are dealing with electromagnetic waves,

$$\frac{h\omega}{2\pi} = E, \qquad \text{energy (of our photon)}$$
$$\frac{h\Delta\omega}{2\pi} = \Delta E, \tag{15.33}$$

h being Planck's constant. Here ΔE represents an uncertainty in the energy of our pulse. There is also an uncertainty in the time, for our wave of N cycles requires $2N\pi/\omega_0$ seconds to pass. Taking

$$\Delta t = \frac{2N\pi}{\omega_0}, \tag{15.34}$$

we have the product of these two uncertainties:

$$\Delta E \cdot \Delta t = \frac{h\,\Delta\omega}{2\pi} \cdot \frac{2\pi N}{\omega_0} = h\frac{\omega_0}{2\pi N} \cdot \frac{2\pi N}{\omega_0} = h. \tag{15.35}$$

The Heisenberg uncertainty principle actually states

$$\Delta E \cdot \Delta t \geq \frac{h}{4\pi}, \tag{15.36}$$

and this is clearly satisfied in our example.

Exercises

15.3.1 (a) Show that $g(-\omega) = g^*(\omega)$ is a necessary and sufficient condition for $f(x)$ to be real.

(b) Show that $g(-\omega) = -g^*(\omega)$ is a necessary and sufficient condition for $f(x)$ to be pure imaginary.

Note. The condition of part (a) is used in the development of the dispersion relations of Section 7.2.

15.3.2 Let $F(\omega)$ be the Fourier (exponential) transform of $f(x)$ and $G(\omega)$ be the Fourier transform of $g(x) = f(x + a)$. Show that

$$G(\omega) = e^{-ia\omega} F(\omega).$$

15.3.3 The function

$$f(x) = \begin{cases} 1, & |x| < 1 \\ 0, & |x| > 1 \end{cases}$$

is a symmetrical finite step function.

(a) Find the $g_c(\omega)$, Fourier cosine transform of $f(x)$.

(b) Taking the inverse cosine transform, show that

$$f(x) = \frac{2}{\pi} \int_0^\infty \frac{\sin\omega \cos\omega x}{\omega} d\omega.$$

(c) From part (b) show that

$$\int_0^\infty \frac{\sin\omega \cos\omega x}{\omega} d\omega = \begin{cases} 0, & |x| > 1, \\ \frac{\pi}{4}, & |x| = 1, \\ \frac{\pi}{2}, & |x| < 1. \end{cases}$$

15.3.4 (a) Show that the Fourier sine and cosine transforms of e^{-at} are

$$g_s(\omega) = \sqrt{\frac{2}{\pi}} \frac{\omega}{\omega^2 + a^2}, \qquad g_c(\omega) = \sqrt{\frac{2}{\pi}} \frac{a}{\omega^2 + a^2}.$$

Hint. Each of the transforms can be related to the other by integration by parts.

(b) Show that

$$\int_0^\infty \frac{\omega \sin\omega x}{\omega^2 + a^2} d\omega = \frac{\pi}{2} e^{-ax}, \qquad x > 0,$$

$$\int_0^\infty \frac{\cos\omega x}{\omega^2 + a^2} d\omega = \frac{\pi}{2a} e^{-ax}, \qquad x > 0.$$

These results are also obtained by contour integration (Exercise 7.1.14).

15.3.5 Find the Fourier transform of the triangular pulse (Fig. 15.4).

$$f(x) = \begin{cases} h\big(1 - a|x|\big), & |x| < \frac{1}{a}, \\ 0, & |x| > \frac{1}{a}. \end{cases}$$

Note. This function provides another delta sequence with $h = a$ and $a \to \infty$.

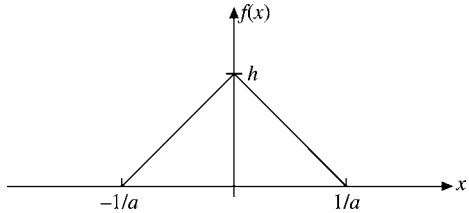

FIGURE 15.4 Triangular pulse.

15.3.6 Define a sequence

$$\delta_n(x) = \begin{cases} n, & |x| < \frac{1}{2n}, \\ 0, & |x| > \frac{1}{2n}. \end{cases}$$

(This is Eq. (1.172).) Express $\delta_n(x)$ as a Fourier integral (via the Fourier integral theorem, inverse transform, etc.). Finally, show that we may write

$$\delta(x) = \lim_{n\to\infty} \delta_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ikx}\, dk.$$

15.3.7 Using the sequence

$$\delta_n(x) = \frac{n}{\sqrt{\pi}} \exp\left(-n^2x^2\right),$$

show that

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ikx}\, dk.$$

Note. Remember that $\delta(x)$ is defined in terms of its behavior as part of an integrand (Section 1.15), especially Eqs. (1.178) and (1.179).

15.3.8 Derive sine and cosine representations of $\delta(t-x)$ that are comparable to the exponential representation, Eq. (15.21d).

$$\text{ANS. } \frac{2}{\pi} \int_0^{\infty} \sin\omega t \sin\omega x\, d\omega, \quad \frac{2}{\pi} \int_0^{\infty} \cos\omega t \cos\omega x\, d\omega.$$

15.3.9 In a resonant cavity an electromagnetic oscillation of frequency ω_0 dies out as

$$A(t) = A_0 e^{-\omega_0 t/2Q} e^{-i\omega_0 t}, \qquad t > 0.$$

(Take $A(t) = 0$ for $t < 0$.) The parameter Q is a measure of the ratio of stored energy to energy loss per cycle. Calculate the frequency distribution of the oscillation, $a^*(\omega)a(\omega)$, where $a(\omega)$ is the Fourier transform of $A(t)$.
Note. The larger Q is, the sharper your resonance line will be.

$$\text{ANS. } a^*(\omega)a(\omega) = \frac{A_0^2}{2\pi} \frac{1}{(\omega-\omega_0)^2 + (\omega_0/2Q)^2}.$$

15.3.10 Prove that

$$\frac{\hbar}{2\pi i}\int_{-\infty}^{\infty}\frac{e^{-i\omega t}\,d\omega}{E_0 - i\Gamma/2 - \hbar\omega} = \begin{cases} \exp\left(-\frac{\Gamma t}{2\hbar}\right)\exp\left(-\frac{iE_0 t}{\hbar}\right), & t > 0, \\ 0, & t < 0. \end{cases}$$

This Fourier integral appears in a variety of problems in quantum mechanics: WKB barrier penetration, scattering, time-dependent perturbation theory, and so on.
Hint. Try contour integration.

15.3.11 Verify that the following are Fourier integral transforms of one another:

(a) $\sqrt{\frac{2}{\pi}}\cdot\frac{1}{\sqrt{a^2-x^2}}, \quad |x| < a, \quad$ and $J_0(ay)$,
$0, \quad |x| > a,$
$0, \quad |x| < a,$

(b) $-\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{x^2+a^2}}, \quad |x| > a, \quad$ and $N_0(a|y|)$,

(c) $\sqrt{\frac{\pi}{2}}\cdot\frac{1}{\sqrt{x^2+a^2}}$ and $K_0\big(a|y|\big)$.

(d) Can you suggest why $I_0(ay)$ is not included in this list?

Hint. J_0, N_0, and K_0 may be transformed most easily by using an exponential representation, reversing the order of integration, and employing the Dirac delta function exponential representation (Section 15.2). These cases can be treated equally well as Fourier cosine transforms.
Note. The K_0 relation appears as a consequence of a Green's function equation in Exercise 9.7.14.

15.3.12 A calculation of the magnetic field of a circular current loop in circular cylindrical coordinates leads to the integral

$$\int_0^{\infty}\cos kz\, k\, K_1(ka)\,dk.$$

Show that this integral is equal to

$$\frac{\pi a}{2(z^2+a^2)^{3/2}}.$$

Hint. Try differentiating Exercise 15.3.11(c).

15.3.13 As an extension of Exercise 15.3.11, show that

(a) $\int_0^{\infty} J_0(y)\,dy = 1$, (b) $\int_0^{\infty} N_0(y)\,dy = 0$, (c) $\int_0^{\infty} K_0(y)\,dy = \frac{\pi}{2}$.

15.3.14 The Fourier integral, Eq. (15.18), has been held meaningless for $f(t) = \cos\alpha t$. Show that the Fourier integral can be extended to cover $f(t) = \cos\alpha t$ by use of the Dirac delta function.

15.3.15 Show that

$$\int_0^\infty \sin ka\, J_0(k\rho)\, dk = \begin{cases} \left(a^2-\rho^2\right)^{-1/2}, & \rho < a, \\ 0, & \rho > a. \end{cases}$$

Here a and ρ are positive. The equation comes from the determination of the distribution of charge on an isolated conducting disk, radius a. Note that the function on the right has an **infinite** discontinuity at $\rho = a$.
Note. A Laplace transform approach appears in Exercise 15.10.8.

15.3.16 The function $f(r)$ has a Fourier exponential transform,

$$g(\mathbf{k}) = \frac{1}{(2\pi)^{3/2}} \int f(\mathbf{r}) e^{i\mathbf{k}\cdot\mathbf{r}}\, d^3r = \frac{1}{(2\pi)^{3/2}k^2}.$$

Determine $f(\mathbf{r})$.
Hint. Use spherical polar coordinates in k-space.

ANS. $f(\mathbf{r}) = \dfrac{1}{4\pi r}$.

15.3.17 (a) Calculate the Fourier exponential transform of $f(x) = e^{-a|x|}$.
(b) Calculate the inverse transform by employing the calculus of residues (Section 7.1).

15.3.18 Show that the following are Fourier transforms of each other

$$i^n J_n(t) \qquad \text{and} \qquad \begin{cases} \sqrt{\dfrac{2}{\pi}} T_n(x)\left(1-x^2\right)^{-1/2}, & |x| < 1, \\ 0, & |x| > 1. \end{cases}$$

$T_n(x)$ is the nth-order Chebyshev polynomial.
Hint. With $T_n(\cos\theta) = \cos n\theta$, the transform of $T_n(x)(1-x^2)^{-1/2}$ leads to an integral representation of $J_n(t)$.

15.3.19 Show that the Fourier exponential transform of

$$f(\mu) = \begin{cases} P_n(\mu), & |\mu| \le 1, \\ 0, & |\mu| > 1 \end{cases}$$

is $(2i^n/2\pi)\, j_n(kr)$. Here $P_n(\mu)$ is a Legendre polynomial and $j_n(kr)$ is a spherical Bessel function.

15.3.20 Show that the three-dimensional Fourier exponential transform of a radially symmetric function may be rewritten as a Fourier sine transform:

$$\frac{1}{(2\pi)^{3/2}} \int_{-\infty}^{\infty} f(r) e^{i\mathbf{k}\cdot\mathbf{r}}\, d^3x = \frac{1}{k}\sqrt{\frac{2}{\pi}} \int_0^\infty \left[r f(r)\right] \sin kr\, dr.$$

15.3.21 (a) Show that $f(x) = x^{-1/2}$ is a **self-reciprocal** under both Fourier cosine and sine transforms; that is,

$$\sqrt{\frac{2}{\pi}} \int_0^\infty x^{-1/2} \cos xt \, dx = t^{-1/2},$$

$$\sqrt{\frac{2}{\pi}} \int_0^\infty x^{-1/2} \sin xt \, ds = t^{-1/2}.$$

(b) Use the preceding results to evaluate the Fresnel integrals $\int_0^\infty \cos(y^2)\, dy$ and $\int_0^\infty \sin(y^2)\, dy$.

15.4 FOURIER TRANSFORM OF DERIVATIVES

In Section 15.1, Fig. 15.1 outlines the overall technique of using Fourier transforms and inverse transforms to solve a problem. Here we take an initial step in solving a differential equation — obtaining the Fourier transform of a derivative.

Using the exponential form, we determine that the Fourier transform of $f(x)$ is

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{i\omega x} \, dx \tag{15.37}$$

and for $df(x)/dx$

$$g_1(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{df(x)}{dx} e^{i\omega x} \, dx. \tag{15.38}$$

Integrating Eq. (15.38) by parts, we obtain

$$g_1(\omega) = \frac{e^{i\omega x}}{\sqrt{2\pi}} f(x)\Big|_{-\infty}^{\infty} - \frac{i\omega}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{i\omega x} \, dx. \tag{15.39}$$

If $f(x)$ vanishes[4] as $x \to \pm\infty$, we have

$$g_1(\omega) = -i\omega\, g(\omega); \tag{15.40}$$

that is, the transform of the derivative is $(-i\omega)$ times the transform of the original function. This may readily be generalized to the nth derivative to yield

$$g_n(\omega) = (-i\omega)^n g(\omega), \tag{15.41}$$

provided all the integrated parts vanish as $x \to \pm\infty$. This is the power of the Fourier transform, the reason it is so useful in solving (partial) differential equations. The operation of differentiation has been replaced by a multiplication in ω-space.

[4] Apart from cases such as Exercise 15.3.6, $f(x)$ must vanish as $x \to \pm\infty$ in order for the Fourier transform of $f(x)$ to exist.

Example 15.4.1 WAVE EQUATION

This technique may be used to advantage in handling PDEs. To illustrate the technique, let us derive a familiar expression of elementary physics. An infinitely long string is vibrating freely. The amplitude y of the (small) vibrations satisfies the wave equation

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 y}{\partial t^2}. \tag{15.42}$$

We shall assume an initial condition

$$y(x, 0) = f(x), \tag{15.43}$$

where f is localized, that is, approaches zero at large x.

Applying our Fourier transform in x, which means multiplying by $e^{i\alpha x}$ and integrating over x, we obtain

$$\int_{-\infty}^{\infty} \frac{\partial^2 y(x,t)}{\partial x^2} e^{i\alpha x}\, dx = \frac{1}{v^2}\int_{-\infty}^{\infty} \frac{\partial^2 y(x,t)}{\partial t^2} e^{i\alpha x}\, dx \tag{15.44}$$

or

$$(-i\alpha)^2 Y(\alpha, t) = \frac{1}{v^2}\frac{\partial^2 Y(\alpha, t)}{\partial t^2}. \tag{15.45}$$

Here we have used

$$Y(\alpha, t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y(x, t) e^{i\alpha x}\, dx \tag{15.46}$$

and Eq. (15.41) for the second derivative. Note that the integrated part of Eq. (15.39) vanishes: The wave has not yet gone to $\pm\infty$ because it is propagating forward in time, and there is no source at infinity because $f(\pm\infty) = 0$. Since no derivatives with respect to α appear, Eq. (15.45) is actually an ODE — in fact, the linear oscillator equation. This transformation, from a PDE to an ODE, is a significant achievement. We solve Eq. (15.45) subject to the appropriate initial conditions. At $t = 0$, applying Eq. (15.43), Eq. (15.46) reduces to

$$Y(\alpha, 0) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(x) e^{i\alpha x}\, dx = F(\alpha). \tag{15.47}$$

The general solution of Eq. (15.45) in exponential form is

$$Y(\alpha, t) = F(\alpha) e^{\pm i v\alpha t}. \tag{15.48}$$

Using the inversion formula (Eq. (15.23)), we have

$$y(x, t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} Y(\alpha, t) e^{-i\alpha x}\, d\alpha, \tag{15.49}$$

and, by Eq. (15.48),

$$y(x, t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} F(\alpha) e^{-i\alpha(x \mp vt)}\, d\alpha. \tag{15.50}$$

Since $f(x)$ is the Fourier inverse transform of $F(\alpha)$,

$$y(x,t) = f(x \mp vt), \tag{15.51}$$

corresponding to waves advancing in the $+x$- and $-x$-directions, respectively.

The particular linear combinations of waves is given by the boundary condition of Eq. (15.43) and some other boundary condition, such as a restriction on $\partial y/\partial t$. ■

The accomplishment of the Fourier transform here deserves special emphasis.

- Our Fourier transform converted a PDE into an ODE, where the "degree of transcendence" of the problem was reduced.

In Section 15.9 Laplace transforms are used to convert ODEs (with constant coefficients) into algebraic equations. Again, the degree of transcendence is reduced. The problem is simplified — as outlined in Fig. 15.1.

Example 15.4.2 HEAT FLOW PDE

To illustrate another transformation of a PDE into an ODE, let us Fourier transform the heat flow partial differential equation

$$\frac{\partial \psi}{\partial t} = a^2 \frac{\partial^2 \psi}{\partial x^2},$$

where the solution $\psi(x,t)$ is the temperature in space as a function of time. By taking the Fourier transform of both sides of this equation (note that here only ω is the transform variable conjugate to x because t is the time in the heat flow PDE), where

$$\Psi(\omega,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi(x,t) e^{i\omega x}\, dx,$$

this yields an ODE for the Fourier transform Ψ of ψ in the time variable t,

$$\frac{\partial \Psi(\omega,t)}{\partial t} = -a^2\omega^2 \Psi(\omega,t).$$

Integrating we obtain

$$\ln \Psi = -a^2\omega^2 t + \ln C, \qquad \text{or} \qquad \Psi = Ce^{-a^2\omega^2 t},$$

where the integration constant C may still depend on ω and, in general, is determined by initial conditions. In fact, $C = \Psi(\omega, 0)$ is the initial spatial distribution of Ψ, so it is given by the transform (in x) of the initial distribution of ψ, namely, $\psi(x,0)$. Putting this solution back into our inverse Fourier transform, this yields

$$\psi(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} C(\omega) e^{-i\omega x} e^{-a^2\omega^2 t}\, d\omega.$$

For simplicity, we here take C ω-independent (assuming a delta-function initial temperature distribution) and integrate by completing the square in ω, as in Example 15.1.1,

making appropriate changes of variables and parameters ($a^2 \to a^2 t,\ \omega \to x, t \to -\omega$). This yields the particular solution of the heat flow PDE,

$$\psi(x,t) = \frac{C}{a\sqrt{2t}} \exp\left(-\frac{x^2}{4a^2 t}\right),$$

which appears as a clever guess in Chapter 8. In effect, we have shown that ψ is the inverse Fourier transform of $C\exp(-a^2\omega^2 t)$. ■

Example 15.4.3 INVERSION OF PDE

Derive a Fourier integral for the Green's function G_0 of Poisson's PDE, which is a solution of

$$\nabla^2 G_0(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}').$$

Once G_0 is known, the general solution of Poisson's PDE,

$$\nabla^2 \Phi = -4\pi\rho(\mathbf{r})$$

of electrostatics, is given as

$$\Phi(\mathbf{r}) = \int G_0(\mathbf{r}, \mathbf{r}') 4\pi\rho(\mathbf{r}')\, d^3 r'.$$

Applying ∇^2 to Φ and using the PDE the Green's function satisfies, we check that

$$\nabla^2\Phi(\mathbf{r}) = \int \nabla^2 G_0(\mathbf{r}, \mathbf{r}') 4\pi\rho(\mathbf{r}')\, d^3 r' = -\int \delta(\mathbf{r} - \mathbf{r}') 4\pi\rho(\mathbf{r}')\, d^3 r' = -4\pi\rho(\mathbf{r}).$$

Now we use the Fourier transform of G_0, which is g_0, and of that of the δ function, writing

$$\nabla^2 \int g_0(\mathbf{p}) e^{i\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}')} \frac{d^3 p}{(2\pi)^3} = -\int e^{i\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}')} \frac{d^3 p}{(2\pi)^3}.$$

Because the integrands of equal Fourier integrals must be the same (almost) everywhere, which follows from the inverse Fourier transform, and with

$$\nabla e^{i\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}')} = i\mathbf{p} e^{i\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}')},$$

this yields $-\mathbf{p}^2 g_0(\mathbf{p}) = -1$. Therefore, application of the Laplacian to a Fourier integral $f(\mathbf{r})$ corresponds to multiplying its Fourier transform $g(\mathbf{p})$ by $-\mathbf{p}^2$. Substituting this solution into the inverse Fourier transform for G_0 gives

$$G_0(\mathbf{r}, \mathbf{r}') = \int e^{i\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}')} \frac{d^3 p}{(2\pi)^3 \mathbf{p}^2} = \frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|}.$$

We can verify the last part of this result by applying ∇^2 to G_0 again and recalling from Chapter 1 that $\nabla^2 \frac{1}{|\mathbf{r}-\mathbf{r}'|} = -4\pi\delta(\mathbf{r} - \mathbf{r}')$.

The inverse Fourier transform can be evaluated using polar coordinates, exploiting the spherical symmetry of $\mathbf{p}^2$. For simplicity, we write $\mathbf{R} = \mathbf{r} - \mathbf{r}'$ and call θ the angle between $\mathbf{R}$ and $\mathbf{p}$,

$$\begin{aligned}\int e^{i\mathbf{p}\cdot\mathbf{R}} \frac{d^3p}{p^2} &= \int_0^\infty dp \int_{-1}^1 e^{ipR\cos\theta} d\cos\theta \int_0^{2\pi} d\varphi \\ &= \frac{2\pi}{iR} \int_0^\infty \frac{dp}{p} e^{ipR\cos\theta}\Big|_{\cos\theta=-1}^1 = \frac{4\pi}{R}\int_0^\infty \frac{\sin pR}{p} dp \\ &= \frac{4\pi}{R}\int_0^\infty \frac{\sin pR}{pR} d(pR) = \frac{2\pi^2}{R},\end{aligned}$$

where θ and φ are the angles of $\mathbf{p}$ and $\int_0^\infty \frac{\sin x}{x} dx = \frac{\pi}{2}$, from Example 7.1.4. Dividing by $(2\pi)^3$, we obtain $G_0(R) = 1/(4\pi R)$, as claimed. An evaluation of this Fourier transform by contour integration is given in Example 9.7.2. ■

Exercises

15.4.1 The one-dimensional Fermi age equation for the diffusion of neutrons slowing down in some medium (such as graphite) is

$$\frac{\partial^2 q(x,\tau)}{\partial x^2} = \frac{\partial q(x,\tau)}{\partial \tau}.$$

Here q is the number of neutrons that slow down, falling below some given energy per second per unit volume. The Fermi age, τ, is a measure of the energy loss.
If $q(x,0) = S\delta(x)$, corresponding to a plane source of neutrons at $x = 0$, emitting S neutrons per unit area per second, derive the solution

$$q = S\frac{e^{-x^2/4\tau}}{\sqrt{4\pi\tau}}.$$

Hint. Replace $q(x,\tau)$ with

$$p(k,\tau) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} q(x,\tau) e^{ikx}\, dx.$$

This is analogous to the diffusion of heat in an infinite medium.

15.4.2 Equation (15.41) yields

$$g_2(\omega) = -\omega^2 g(\omega)$$

for the Fourier transform of the second derivative of $f(x)$. The condition $f(x) \to 0$ for $x \to \pm\infty$ may be relaxed slightly. Find the least restrictive condition for the preceding equation for $g_2(\omega)$ to hold.

ANS. $\left[\frac{df(x)}{dx} - i\omega f(x)\right] e^{i\omega x}\Big|_{-\infty}^{\infty} = 0.$

15.4.3 The one-dimensional neutron diffusion equation with a (plane) source is

$$-D\frac{d^2\varphi(x)}{dx^2} + K^2 D\varphi(x) = Q\delta(x),$$

where $\varphi(x)$ is the neutron flux, $Q\delta(x)$ is the (plane) source at $x = 0$, and D and K^2 are constants. Apply a Fourier transform. Solve the equation in transform space. Transform your solution back into x-space.

ANS. $\varphi(x) = \dfrac{Q}{2KD}e^{-|Kx|}$.

15.4.4 For a point source at the origin, the three-dimensional neutron diffusion equation becomes

$$-D\nabla^2\varphi(\mathbf{r}) + K^2 D\varphi(\mathbf{r}) = Q\delta(\mathbf{r}).$$

Apply a three-dimensional Fourier transform. Solve the transformed equation. Transform the solution back into $\mathbf{r}$-space.

15.4.5 (a) Given that $F(\mathbf{k})$ is the three-dimensional Fourier transform of $f(\mathbf{r})$ and $F_1(\mathbf{k})$ is the three-dimensional Fourier transform of $\nabla f(\mathbf{r})$, show that

$$F_1(\mathbf{k}) = (-i\mathbf{k})F(\mathbf{k}).$$

This is a three-dimensional generalization of Eq. (15.40).

(b) Show that the three-dimensional Fourier transform of $\nabla \cdot \nabla f(\mathbf{r})$ is

$$F_2(\mathbf{k}) = (-i\mathbf{k})^2 F(\mathbf{k}).$$

Note. Vector $\mathbf{k}$ is a vector in the transform space. In Section 15.6 we shall have $\hbar\mathbf{k} = \mathbf{p}$, linear momentum.

15.5 CONVOLUTION THEOREM

We shall employ convolutions to solve differential equations, to normalize momentum wave functions (Section 15.6), and to investigate transfer functions (Section 15.7).

Let us consider two functions $f(x)$ and $g(x)$ with Fourier transforms $F(t)$ and $G(t)$, respectively. We define the operation

$$f * g \equiv \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g(y)f(x-y)\,dy \tag{15.52}$$

as the **convolution** of the two functions f and g over the interval $(-\infty, \infty)$. This form of an integral appears in probability theory in the determination of the probability density of two random, independent variables. Our solution of Poisson's equation, Eq. (9.148), may be interpreted as a convolution of a charge distribution, $\rho(\mathbf{r}_2)$, and a weighting function, $(4\pi\varepsilon_0|\mathbf{r}_1 - \mathbf{r}_2|)^{-1}$. In other works this is sometimes referred to as the **Faltung**, to use the

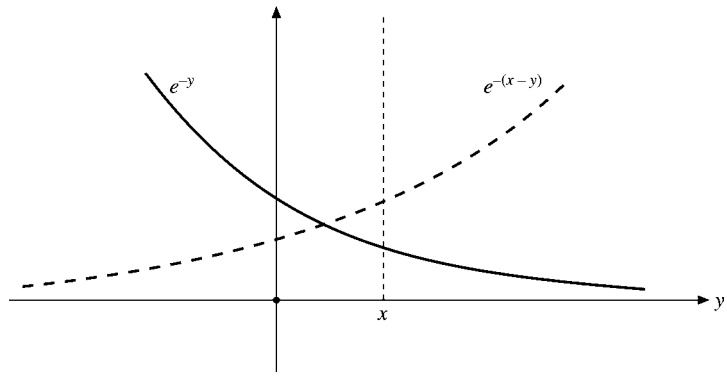

FIGURE 15.5

German term for "folding."[5] We now transform the integral in Eq. (15.52) by introducing the Fourier transforms:

$$\begin{aligned}\int_{-\infty}^{\infty} g(y)f(x-y)dy &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g(y)\int_{-\infty}^{\infty} F(t)e^{-it(x-y)}\,dt\,dy \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} F(t)\left[\int_{-\infty}^{\infty} g(y)e^{ity}dy\right]e^{-itx}\,dt \\ &= \int_{-\infty}^{\infty} F(t)G(t)e^{-itx}\,dt, \end{aligned} \tag{15.53}$$

interchanging the order of integration and transforming $g(y)$. This result may be interpreted as follows: The Fourier inverse transform of a **product** of Fourier transforms is the convolution of the original functions, $f * g$.

For the special case $x = 0$ we have

$$\int_{-\infty}^{\infty} F(t)G(t)\,dt = \int_{-\infty}^{\infty} f(-y)g(y)\,dy. \tag{15.54}$$

The minus sign in $-y$ suggests that modifications be tried. We now do this with g^* instead of g using a different technique.

Parseval's Relation

Results analogous to Eqs. (15.53) and (15.54) may be derived for the Fourier sine and cosine transforms (Exercises 15.5.1 and 15.5.3). Equation (15.54) and the corresponding sine and cosine convolutions are often labeled *Parseval's relations* by analogy with Parseval's theorem for Fourier series (Chapter 14, Exercise 14.4.2).

[5]For $f(y) = e^{-y}$, $f(y)$ and $f(x-y)$ are plotted in Fig. 15.5. Clearly, $f(y)$ and $f(x-y)$ are mirror images of each other in relation to the vertical line $y = x/2$, that is, we could generate $f(x-y)$ by folding over $f(y)$ on the line $y = x/2$.

The Parseval relation[6,7]

$$\int_{-\infty}^{\infty} F(\omega)G^*(\omega)\,d\omega = \int_{-\infty}^{\infty} f(t)g^*(t)\,dt \tag{15.55}$$

may be derived elegantly using the Dirac delta function representation, Eq. (15.21d). We have

$$\int_{-\infty}^{\infty} f(t)g^*(t)\,dt = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(\omega)e^{-i\omega t}d\omega \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} G^*(x)e^{ixt}\,dx\,dt, \tag{15.56}$$

with attention to the complex conjugation in the $G^*(x)$ to $g^*(t)$ transform. Integrating over t first, and using Eq. (15.21d), we obtain

$$\begin{aligned}\int_{-\infty}^{\infty} f(t)g^*(t)\,dt &= \int_{-\infty}^{\infty} F(\omega) \int_{-\infty}^{\infty} G^*(x)\delta(x-\omega)\,dx\,d\omega \\ &= \int_{-\infty}^{\infty} F(\omega)G^*(\omega)\,d\omega,\end{aligned} \tag{15.57}$$

our desired Parseval relation. If $f(t) = g(t)$, then the integrals in the Parseval relation are normalization integrals (Section 10.4). Equation (15.57) guarantees that if a function $f(t)$ is normalized to unity, its transform $F(\omega)$ is likewise normalized to unity. This is extremely important in quantum mechanics as developed in the next section.

It may be shown that the Fourier transform is a unitary operation (in the Hilbert space L^2, square integrable functions). The Parseval relation is a reflection of this unitary property — analogous to Exercise 3.4.26 for matrices.

In Fraunhofer diffraction optics the diffraction pattern (amplitude) appears as the transform of the function describing the aperture (compare Exercise 15.5.5). With intensity proportional to the square of the amplitude the Parseval relation implies that the energy passing through the aperture seems to be somewhere in the diffraction pattern — a statement of the conservation of energy. Parseval's relations may be developed independently of the inverse Fourier transform and then used rigorously to derive the inverse transform. Details are given by Morse and Feshbach,[8] Section 4.8 (see also Exercise 15.5.4).

Exercises

15.5.1 Work out the convolution equation corresponding to Eq. (15.53) for

(a) Fourier sine transforms

$$\frac{1}{2}\int_0^{\infty} g(y)\big[f(y+x) + f(y-x)\big]\,dy = \int_0^{\infty} F_s(s)G_s(s)\cos sx\,ds,$$

where f and g are odd functions.

[6] Note that all arguments are positive, in contrast to Eq. (15.54).

[7] Some authors prefer to restrict Parseval's name to series and refer to Eq. (15.55) as **Rayleigh's theorem**.

[8] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, New York: McGraw-Hill (1953).

(b) Fourier cosine transforms

$$\frac{1}{2}\int_0^\infty g(y)\big[f(y+x)+f(x-y)\big]\,dy=\int_0^\infty F_c(s)G_c(s)\cos sx\,ds,$$

where f and g are even functions.

15.5.2 $F(\rho)$ and $G(\rho)$ are the Hankel transforms of $f(r)$ and $g(r)$, respectively (Exercise 15.1.1). Derive the Hankel transform Parseval relation:

$$\int_0^\infty F^*(\rho)G(\rho)\rho\,d\rho=\int_0^\infty f^*(r)g(r)r\,dr.$$

15.5.3 Show that for both Fourier sine and Fourier cosine transforms Parseval's relation has the form

$$\int_0^\infty F(t)G(t)\,dt=\int_0^\infty f(y)g(y)\,dy.$$

15.5.4 Starting from Parseval's relation (Eq. (15.54)), let $g(y)=1$, $0\le y\le\alpha$, and zero elsewhere. From this derive the Fourier inverse transform (Eq. (15.23)).
Hint. Differentiate with respect to α.

15.5.5 (a) A rectangular pulse is described by

$$f(x)=\begin{cases}1, & |x|<a,\\ 0, & |x|>a.\end{cases}$$

Show that the Fourier exponential transform is

$$F(t)=\sqrt{\frac{2}{\pi}}\frac{\sin at}{t}.$$

This is the single-slit diffraction problem of physical optics. The slit is described by $f(x)$. The diffraction pattern **amplitude** is given by the Fourier transform $F(t)$.

(b) Use the Parseval relation to evaluate

$$\int_{-\infty}^\infty \frac{\sin^2 t}{t^2}dt.$$

This integral may also be evaluated by using the calculus of residues, Exercise 7.1.12.

ANS. (b) π.

15.5.6 Solve Poisson's equation, $\nabla^2\psi(\mathbf{r})=-\rho(\mathbf{r})/\varepsilon_0$, by the following sequence of operations:

(a) Take the Fourier transform of both sides of this equation. Solve for the Fourier transform of $\psi(\mathbf{r})$.

(b) Carry out the Fourier inverse transform by using a three-dimensional analog of the convolution theorem, Eq. (15.53).

15.5.7 (a) Given $f(x) = 1 - |x/2|$, $-2 \le x \le 2$, and zero elsewhere, show that the Fourier transform of $f(x)$ is

$$F(t) = \sqrt{\frac{2}{\pi}} \left(\frac{\sin t}{t} \right)^2.$$

(b) Using the Parseval relation, evaluate

$$\int_{-\infty}^{\infty} \left(\frac{\sin t}{t} \right)^4 dt.$$

ANS. (b) $\dfrac{2\pi}{3}$.

15.5.8 If $g(x)$ is an approximation to $f(x)$, the preceding relation indicates that the mean square deviation in t-space is equal to the mean square deviation in x-space.

With $F(t)$ and $G(t)$ the Fourier transforms of $f(x)$ and $g(x)$, respectively, show that

$$\int_{-\infty}^{\infty} |f(x) - g(x)|^2 dx = \int_{-\infty}^{\infty} |F(t) - G(t)|^2 dt.$$

15.5.9 Use the Parseval relation to evaluate

(a) $\displaystyle\int_{-\infty}^{\infty} \frac{d\omega}{(\omega^2 + a^2)^2}$, (b) $\displaystyle\int_{-\infty}^{\infty} \frac{\omega^2 d\omega}{(\omega^2 + a^2)^2}$.

Hint. Compare Exercise 15.3.4.

ANS. (a) $\dfrac{\pi}{2a^3}$, (b) $\dfrac{\pi}{2a}$.

15.6 MOMENTUM REPRESENTATION

In advanced dynamics and in quantum mechanics, linear momentum and spatial position occur on an equal footing. In this section we shall start with the usual space distribution and derive the corresponding momentum distribution. For the one-dimensional case our wave function $\psi(x)$ has the following properties:

1. $\psi^*(x)\psi(x)\,dx$ is the probability density of finding a quantum particle between x and $x + dx$, and
2. $$\int_{-\infty}^{\infty} \psi^*(x)\psi(x)\,dx = 1 \tag{15.58}$$
   corresponds to probability unity.
3. In addition, we have
   $$\langle x \rangle = \int_{-\infty}^{\infty} \psi^*(x) x \psi(x)\,dx \tag{15.59}$$
   for the **average** position of the particle along the x-axis. This is often called an *expectation value*.

We want a function $g(p)$ that will give the same information about the momentum:

1. $g^*(p)g(p)\,dp$ is the probability density that our quantum particle has a momentum between p and $p+dp$.

2. $$\int_{-\infty}^{\infty} g^*(p)g(p)\,dp = 1. \tag{15.60}$$

3. $$\langle p\rangle = \int_{-\infty}^{\infty} g^*(p)\,p\,g(p)\,dp. \tag{15.61}$$

As subsequently shown, such a function is given by the Fourier transform of our space function $\psi(x)$. Specifically,[9]

$$g(p) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \psi(x) e^{-ipx/\hbar}\,dx, \tag{15.62}$$

$$g^*(p) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \psi^*(x) e^{ipx/\hbar}\,dx. \tag{15.63}$$

The corresponding three-dimensional momentum function is

$$g(\mathbf{p}) = \frac{1}{(2\pi\hbar)^{3/2}} \iiint_{-\infty}^{\infty} \psi(\mathbf{r}) e^{-i\mathbf{r}\cdot\mathbf{p}/\hbar}\,d^3r.$$

To verify Eqs. (15.62) and (15.63), let us check on properties 2 and 3.

Property 2, the normalization, is automatically satisfied as a Parseval relation, Eq. (15.55). If the space function $\psi(x)$ is normalized to unity, the momentum function $g(p)$ is also normalized to unity.

To check on property 3, we must show that

$$\langle p\rangle = \int_{-\infty}^{\infty} g^*(p)\,p\,g(p)\,dp = \int_{-\infty}^{\infty} \psi^*(x)\frac{\hbar}{i}\frac{d}{dx}\psi(x)\,dx, \tag{15.64}$$

where $(\hbar/i)(d/dx)$ is the momentum operator in the space representation. We replace the momentum functions by Fourier-transformed space functions, and the first integral becomes

$$\frac{1}{2\pi\hbar} \iiint_{-\infty}^{\infty} p e^{-ip(x-x')/\hbar}\psi^*(x')\psi(x)\,dp\,dx'\,dx. \tag{15.65}$$

Now we use the plane-wave identity

$$p e^{-ip(x-x')/\hbar} = \frac{d}{dx}\left[-\frac{\hbar}{i}e^{-ip(x-x')/\hbar}\right], \tag{15.66}$$

[9]The $\hbar$ may be avoided by using the wave number k, $p = k\hbar$ (and $\mathbf{p} = \mathbf{k}\hbar$), so

$$\varphi(k) = \frac{1}{(2\pi)^{1/2}} \int \psi(x) e^{-ikx}\,dx.$$

An example of this notation appears in Section 16.1.

with p a constant, not an operator. Substituting into Eq. (15.65) and integrating by parts, holding x' and p constant, we obtain

$$\langle p\rangle = \iint_{-\infty}^{\infty}\left[\frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} e^{-ip(x-x')/\hbar}\,dp\right]\cdot\psi^*(x')\frac{\hbar}{i}\frac{d}{dx}\psi(x)\,dx'\,dx. \tag{15.67}$$

Here we assume $\psi(x)$ vanishes as $x\to\pm\infty$, eliminating the integrated part. Using the Dirac delta function, Eq. (15.21c), Eq. (15.67) reduces to Eq. (15.64) to verify our momentum representation.

Alternatively, if the integration over p is done first in Eq. (15.65), leading to

$$\int_{-\infty}^{\infty} pe^{-ip(x-x')/\hbar}\,dp = 2\pi i\hbar^2\delta'(x-x'),$$

and using Exercise 1.15.9, we can do the integration over x, which causes $\psi(x)$ to become $-d\psi(x')/dx'$. The remaining integral over x' is the right-hand side of Eq. (15.64).

Example 15.6.1 HYDROGEN ATOM

The hydrogen atom ground state[10] may be described by the spatial wave function

$$\psi(\mathbf{r}) = \left(\frac{1}{\pi a_0^3}\right)^{1/2} e^{-r/a_0}, \tag{15.68}$$

a_0 being the Bohr radius, $4\pi\varepsilon_0\hbar^2/me^2$. We now have a three-dimensional wave function. The transform corresponding to Eq. (15.62) is

$$g(\mathbf{p}) = \frac{1}{(2\pi\hbar)^{3/2}}\int \psi(\mathbf{r})e^{-i\mathbf{p}\cdot\mathbf{r}/\hbar}\,d^3r. \tag{15.69}$$

Substituting Eq. (15.68) into Eq. (15.69) and using

$$\int e^{-ar+i\mathbf{b}\cdot\mathbf{r}}\,d^3r = \frac{8\pi a}{(a^2+b^2)^2}, \tag{15.70}$$

we obtain the hydrogenic momentum wave function,

$$g(\mathbf{p}) = \frac{2^{3/2}}{\pi}\frac{a_0^{3/2}\hbar^{5/2}}{(a_0^2p^2+\hbar^2)^2}. \tag{15.71}$$

Such momentum functions have been found useful in problems like Compton scattering from atomic electrons, the wavelength distribution of the scattered radiation, depending on the momentum distribution of the target electrons.

The relation between the ordinary space representation and the momentum representation may be clarified by considering the basic commutation relations of quantum mechanics. We go from a classical Hamiltonian to the Schrödinger wave equation by requiring that momentum p and position x **not** commute. Instead, we require that

$$[p,x] \equiv px - xp = -i\hbar. \tag{15.72}$$

[10] See E. V. Ivash, A momentum representation treatment of the hydrogen atom problem. *Am. J. Phys.* **40**: 1095 (1972) for a momentum representation treatment of the hydrogen atom $l = 0$ states.

For the multidimensional case, Eq. (15.72) is replaced by

$$[p_i, x_j] = -i\hbar\delta_{ij}. \tag{15.73}$$

The Schrödinger (space) representation is obtained by using

$$x \to x: \quad p_i \to -i\hbar\frac{\partial}{\partial x_i},$$

replacing the momentum by a partial space derivative. We see that

$$[p, x]\psi(x) = -i\hbar\psi(x). \tag{15.74}$$

However, Eq. (15.72) can equally well be satisfied by using

$$p \to p: \quad x_j \to i\hbar\frac{\partial}{\partial p_j}.$$

This is the momentum representation. Then

$$[p, x]g(p) = -i\hbar g(p). \tag{15.75}$$

Hence the representation (x) is not unique; (p) is an alternate possibility.

In general, the Schrödinger representation (x) leading to the Schrödinger wave equation is more convenient because the potential energy V is generally given as a function of position $V(x, y, z)$. The momentum representation (p) usually leads to an integral equation (compare Chapter 16 for the pros and cons of the integral equations). For an exception, consider the harmonic oscillator. ■

Example 15.6.2 HARMONIC OSCILLATOR

The classical Hamiltonian (kinetic energy + potential energy = total energy) is

$$H(p, x) = \frac{p^2}{2m} + \frac{1}{2}kx^2 = E, \tag{15.76}$$

where k is the Hooke's law constant.

In the Schrödinger representation we obtain

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + \frac{1}{2}kx^2\psi(x) = E\psi(x). \tag{15.77}$$

For total energy E equal to $\sqrt{(k/m)}\hbar/2$ there is an unnormalized solution (Section 13.1),

$$\psi(x) = e^{-(\sqrt{mk}/2\hbar)x^2}. \tag{15.78}$$

The momentum representation leads to

$$\frac{p^2}{2m}g(p) - \frac{\hbar^2 k}{2}\frac{d^2g(p)}{dp^2} = Eg(p). \tag{15.79}$$

Again, for

$$E = \sqrt{\frac{k}{m}}\frac{\hbar}{2} \tag{15.80}$$

the momentum wave equation (15.79) is satisfied by the unnormalized

$$g(p) = e^{-p^2/(2\hbar\sqrt{mk})}. \tag{15.81}$$

Either representation, space or momentum (and an infinite number of other possibilities), may be used, depending on which is more convenient for the particular problem under attack.

The demonstration that $g(p)$ is the momentum wave function corresponding to Eq. (15.78)—that it is the Fourier inverse transform of Eq. (15.78)—is left as Exercise 15.6.3. ■

Exercises

15.6.1 The function $e^{i\mathbf{k}\cdot\mathbf{r}}$ describes a plane wave of momentum $\mathbf{p} = \hbar\mathbf{k}$ normalized to unit density. (Time dependence of $e^{-i\omega t}$ is assumed.) Show that these plane-wave functions satisfy an orthogonality relation

$$\int \left(e^{i\mathbf{k}\cdot\mathbf{r}}\right)^* e^{i\mathbf{k}'\cdot\mathbf{r}} dx\, dy\, dz = (2\pi)^3 \delta(\mathbf{k} - \mathbf{k}').$$

15.6.2 An infinite plane wave in quantum mechanics may be represented by the function

$$\psi(x) = e^{ip'x/\hbar}.$$

Find the corresponding momentum distribution function. Note that it has an infinity and that $\psi(x)$ is not normalized.

15.6.3 A linear quantum oscillator in its ground state has a wave function

$$\psi(x) = a^{-1/2}\pi^{-1/4}e^{-x^2/2a^2}.$$

Show that the corresponding momentum function is

$$g(p) = a^{1/2}\pi^{-1/4}\hbar^{-1/2}e^{-a^2p^2/2\hbar^2}.$$

15.6.4 The nth excited state of the linear quantum oscillator is described by

$$\psi_n(x) = a^{-1/2}2^{-n/2}\pi^{-1/4}(n!)^{-1/2}e^{-x^2/2a^2}H_n(x/a),$$

where $H_n(x/a)$ is the nth Hermite polynomial, Section 13.1. As an extension of Exercise 15.6.3, find the momentum function corresponding to $\psi_n(x)$.
Hint. $\psi_n(x)$ may be represented by $(\hat{a}^\dagger)^n\psi_0(x)$, where $\hat{a}^\dagger$ is the raising operator, Exercise 13.1.14 to 13.1.16.

15.6.5 A free particle in quantum mechanics is described by a plane wave

$$\psi_k(x,t) = e^{i[kx-(\hbar k^2/2m)t]}.$$

Combining waves of adjacent momentum with an amplitude weighting factor $\varphi(k)$, we form a wave packet

$$\Psi(x,t) = \int_{-\infty}^{\infty} \varphi(k)e^{i[kx-(\hbar k^2/2m)t]}\,dk.$$

(a) Solve for $\varphi(k)$ given that

$$\Psi(x,0) = e^{-x^2/2a^2}.$$

(b) Using the known value of $\varphi(k)$, integrate to get the explicit form of $\Psi(x,t)$. Note that this wave packet diffuses or spreads out with time.

$$\text{ANS. } \Psi(x,t) = \frac{e^{-\{x^2/2[a^2+(i\hbar/m)t]\}}}{[1+(i\hbar t/ma^2)]^{1/2}}.$$

Note. An interesting discussion of this problem from the evolution operator point of view is given by S. M. Blinder, Evolution of a Gaussian wave packet, *Am. J. Phys.* **36**: 525 (1968).

15.6.6 Find the time-dependent momentum wave function $g(k,t)$ corresponding to $\Psi(x,t)$ of Exercise 15.6.5. Show that the momentum wave packet $g^*(k,t)g(k,t)$ is **independent** of time.

15.6.7 The deuteron, Example 10.1.2, may be described reasonably well with a Hulthén wave function

$$\psi(\mathbf{r}) = A[e^{-\alpha r} - e^{-\beta r}]/r,$$

with A, α, and β constants. Find $g(\mathbf{p})$, the corresponding momentum function.
Note. The Fourier transform may be rewritten as Fourier sine and cosine transforms or as a Laplace transform, Section 15.8.

15.6.8 The nuclear form factor $F(k)$ and the charge distribution $\rho(r)$ are three-dimensional Fourier transforms of each other:

$$F(k) = \frac{1}{(2\pi)^{3/2}} \int \rho(r) e^{i\mathbf{k}\cdot\mathbf{r}}\, d^3r.$$

If the measured form factor is

$$F(k) = (2\pi)^{-3/2}\left(1 + \frac{k^2}{a^2}\right)^{-1},$$

find the corresponding charge distribution.

$$\text{ANS. } \rho(r) = \frac{a^2}{4\pi}\frac{e^{-ar}}{r}.$$

15.6.9 Check the normalization of the hydrogen momentum wave function

$$g(\mathbf{p}) = \frac{2^{3/2}}{\pi}\frac{a_0^{3/2}\hbar^{5/2}}{(a_0^2p^2+\hbar^2)^2}$$

by direct evaluation of the integral

$$\int g^*(\mathbf{p})g(\mathbf{p})\, d^3p.$$

15.6.10 With $\psi(\mathbf{r})$ a wave function in ordinary space and $\varphi(\mathbf{p})$ the corresponding momentum function, show that

(a) $$\frac{1}{(2\pi\hbar)^{3/2}}\int \mathbf{r}\psi(\mathbf{r})e^{-i\mathbf{r}\cdot\mathbf{p}/\hbar}\,d^3r = i\hbar\,\nabla_p\varphi(\mathbf{p}),$$

(b) $$\frac{1}{(2\pi\hbar)^{3/2}}\int \mathbf{r}^2\psi(\mathbf{r})e^{-\mathbf{r}\cdot\mathbf{p}/\hbar}\,d^3r = (i\hbar\,\nabla_p)^2\varphi(\mathbf{p}).$$

Note. ∇_p is the gradient in momentum space:

$$\hat{\mathbf{x}}\frac{\partial}{\partial p_x} + \hat{\mathbf{y}}\frac{\partial}{\partial p_y} + \hat{\mathbf{z}}\frac{\partial}{\partial p_z}.$$

These results may be extended to any positive integer power of r and therefore to any (analytic) function that may be expanded as a Maclaurin series in r.

15.6.11 The ordinary space wave function $\psi(\mathbf{r}, t)$ satisfies the time-dependent Schrödinger equation

$$i\hbar\frac{\partial\psi(\mathbf{r},t)}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi + V(\mathbf{r})\psi.$$

Show that the corresponding time-dependent momentum wave function satisfies the analogous equation,

$$i\hbar\frac{\partial\varphi(\mathbf{p},t)}{\partial t} = \frac{p^2}{2m}\varphi + V(i\hbar\,\nabla_p)\varphi.$$

Note. Assume that $V(\mathbf{r})$ may be expressed by a Maclaurin series and use Exercise 15.6.10. $V(i\hbar\,\nabla_p)$ is the same function of the variable $i\hbar\,\nabla_p$ that $V(\mathbf{r})$ is of the variable $\mathbf{r}$.

15.6.12 The one-dimensional time-independent Schrödinger wave equation is

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2} + V(x)\psi(x) = E\psi(x).$$

For the special case of $V(x)$ an analytic function of x, show that the corresponding momentum wave equation is

$$V\left(i\hbar\frac{d}{dp}\right)g(p) + \frac{p^2}{2m}g(p) = Eg(p).$$

Derive this momentum wave equation from the Fourier transform, Eq. (15.62), and its inverse. Do not use the substitution $x \to i\hbar(d/dp)$ directly.

15.7 TRANSFER FUNCTIONS

A time-dependent electrical pulse may be regarded as built-up as a superposition of plane waves of many frequencies. For angular frequency ω we have a contribution

$$F(\omega)e^{i\omega t}.$$

Then the complete pulse may be written as

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{i\omega t}\,d\omega. \tag{15.82}$$

FIGURE 15.6 Servomechanism or a stereo amplifier.

Because the angular frequency ω is related to the linear frequency ν by

$$\nu = \frac{\omega}{2\pi},$$

it is customary to associate the entire $1/2\pi$ factor with this integral.

But if ω is a frequency, what about the negative frequencies? The negative ω may be looked on as a mathematical device to avoid dealing with two functions ($\cos\omega t$ and $\sin\omega t$) separately (compare Section 14.1).

Because Eq. (15.82) has the form of a Fourier transform, we may solve for $F(\omega)$ by writing the inverse transform,

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}\,dt. \tag{15.83}$$

Equation (15.83) represents a **resolution of the pulse** $f(t)$ into its angular frequency components. Equation (15.82) is a **synthesis of the pulse** from its components.

Consider some device, such as a servomechanism or a stereo amplifier (Fig. 15.6), with an input $f(t)$ and an output $g(t)$. For an input of a single frequency ω, $f_\omega(t) = e^{i\omega t}$, the amplifier will alter the amplitude and may also change the phase. The changes will probably depend on the frequency. Hence

$$g_\omega(t) = \varphi(\omega) f_\omega(t). \tag{15.84}$$

This amplitudes- and phase-modifying function $\varphi(\omega)$ is called a **transfer** function. It usually will be complex:

$$\varphi(\omega) = u(\omega) + i\,v(\omega), \tag{15.85}$$

where the functions $u(\omega)$ and $v(\omega)$ are real.

In Eq. (15.84) we assume that the transfer function $\varphi(\omega)$ is independent of input amplitude and of the presence or absence of any other frequency components. That is, we are assuming a linear mapping of $f(t)$ onto $g(t)$. Then the total output may be obtained by integrating over the entire input, as modified by the amplifier

$$g(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi(\omega)F(\omega)e^{i\omega t}d\omega. \tag{15.86}$$

The transfer function is characteristic of the amplifier. Once the transfer function is known (measured or calculated), the output $g(t)$ can be calculated for any input $f(t)$. Let us consider $\varphi(\omega)$ as the Fourier (inverse) transform of some function $\Phi(t)$:

$$\varphi(\omega) = \int_{-\infty}^{\infty} \Phi(t)e^{-i\omega t}dt. \tag{15.87}$$

Then Eq. (15.86) is the Fourier transform of two inverse transforms. From Section 15.5 we obtain the convolution

$$g(t) = \int_{-\infty}^{\infty} f(\tau)\Phi(t-\tau)\,d\tau. \tag{15.88}$$

Interpreting Eq. (15.88), we have an input — a "cause" — $f(\tau)$, modified by $\Phi(t-\tau)$, producing an output — an "effect" — $g(t)$. Adopting the concept of **causality** — that the cause precedes the effect — we must require $\tau < t$. We do this by requiring

$$\Phi(t-\tau) = 0, \qquad \tau > t. \tag{15.89}$$

Then Eq. (15.88) becomes

$$g(t) = \int_{-\infty}^{t} f(\tau)\Phi(t-\tau)\,d\tau. \tag{15.90}$$

The adoption of Eq. (15.89) has profound consequences here and equivalently in dispersion theory, Section 7.2.

Significance of $\Phi(t)$

To see the significance of Φ, let $f(\tau)$ be a sudden impulse starting at $\tau = 0$,

$$f(\tau) = \delta(\tau),$$

where $\delta(\tau)$ is a Dirac delta distribution on the positive side of the origin. Then Eq. (15.90) becomes

$$\begin{aligned} g(t) &= \int_{-\infty}^{t} \delta(\tau)\Phi(t-\tau)\,d\tau, \\ g(t) &= \begin{cases} \Phi(t), & t > 0, \\ 0, & t < 0. \end{cases} \end{aligned} \tag{15.91}$$

This identifies $\Phi(t)$ as the output function corresponding to a unit impulse at $t = 0$. Equation (15.91) also serves to establish that $\Phi(t)$ is real. Our original transfer function gives the steady-state output corresponding to a unit-amplitude single-frequency input. $\Phi(t)$ and $\varphi(\omega)$ are Fourier transforms of each other.

From Eq. (15.87) we now have

$$\varphi(\omega) = \int_{0}^{\infty} \Phi(t)e^{-i\omega t}\,dt, \tag{15.92}$$

with the lower limit set equal to zero by causality (Eq. (15.89)). With $\Phi(t)$ real from Eq. (15.91) we separate real and imaginary parts and write

$$\begin{aligned} u(\omega) &= \int_{0}^{\infty} \Phi(t)\cos\omega t\,dt, \\ v(\omega) &= -\int_{0}^{\infty} \Phi(t)\sin\omega t\,dt, \qquad \omega > 0. \end{aligned} \tag{15.93}$$

From this we see that the real part of $\varphi(\omega)$, $u(\omega)$, is even, whereas the imaginary part of $\varphi(\omega)$, $v(\omega)$, is odd:

$$u(-\omega) = u(\omega), \qquad v(-\omega) = -v(\omega).$$

Compare this result with Exercise 15.3.1.

Interpreting Eq. (15.93) as Fourier cosine and sine transforms, we have

$$\begin{aligned}\Phi(t) &= \frac{2}{\pi}\int_0^\infty u(\omega)\cos\omega t\, d\omega \\ &= -\frac{2}{\pi}\int_0^\infty v(\omega)\sin\omega t\, d\omega, \qquad t > 0.\end{aligned} \tag{15.94}$$

Combining Eqs. (15.93) and (15.94), we obtain

$$v(\omega) = -\int_0^\infty \sin\omega t\left\{\frac{2}{\pi}\int_0^\infty u(\omega')\cos\omega' t\, d\omega'\right\} dt, \tag{15.95}$$

showing that if our transfer function has a real part, it will also have an imaginary part (and vice versa). Of course, this assumes that the Fourier transforms exist, thus excluding cases such as $\Phi(t) = 1$.

The imposition of causality has led to a mutual interdependence of the real and imaginary parts of the transfer function. The reader should compare this with the results of the dispersion theory of Section 7.2, also involving causality.

It may be helpful to show that the parity properties of $u(\omega)$ and $v(\omega)$ require $\Phi(t)$ to vanish for negative t. Inverting Eq. (15.87), we have

$$\Phi(t) = \frac{1}{2\pi}\int_{-\infty}^\infty \big[u(\omega) + i\,v(\omega)\big]\big[\cos\omega t + i\sin\omega t\big]\, d\omega. \tag{15.96}$$

With $u(\omega)$ even and $v(\omega)$ odd, Eq. (15.96) becomes

$$\Phi(t) = \frac{1}{\pi}\int_0^\infty u(\omega)\cos\omega t\, d\omega - \frac{1}{\pi}\int_0^\infty v(\omega)\sin\omega t\, d\omega. \tag{15.97}$$

From Eq. (15.94),

$$\int_0^\infty u(\omega)\cos\omega t\, d\omega = -\int_0^\infty v(\omega)\sin\omega t\, d\omega, \quad t > 0. \tag{15.98}$$

If we reverse the sign of t, $\sin\omega t$ reverses sign and, from Eq. (15.97),

$$\Phi(t) = 0, \qquad t < 0$$

(demonstrating the internal consistency of our analysis).

Exercise

15.7.1 Derive the convolution

$$g(t) = \int_{-\infty}^\infty f(\tau)\Phi(t-\tau)\, d\tau.$$

15.8 LAPLACE TRANSFORMS

Definition

The Laplace transform $f(s)$ or $\mathcal{L}$ of a function $F(t)$ is defined by[11]

$$f(s) = \mathcal{L}\{F(t)\} = \lim_{a\to\infty} \int_0^a e^{-st} F(t)\,dt = \int_0^\infty e^{-st} F(t)\,dt. \tag{15.99}$$

A few comments on the existence of the integral are in order. The infinite integral of $F(t)$,

$$\int_0^\infty F(t)\,dt,$$

need not exist. For instance, $F(t)$ may diverge exponentially for large t. However, if there is some constant s_0 such that

$$|e^{-s_0 t} F(t)| \le M, \tag{15.100}$$

a positive constant for sufficiently large $t, t > t_0$, the Laplace transform (Eq. (15.99)) will exist for $s > s_0$; $F(t)$ is said to be of *exponential order*. As a counterexample, $F(t) = e^{t^2}$ does not satisfy the condition given by Eq. (15.100) and is **not** of exponential order. $\mathcal{L}\{e^{t^2}\}$ does **not** exist.

The Laplace transform may also fail to exist because of a sufficiently strong singularity in the function $F(t)$ as $t \to 0$; that is,

$$\int_0^\infty e^{-st} t^n\,dt$$

diverges at the origin for $n \le -1$. The Laplace transform $\mathcal{L}\{t^n\}$ does not exist for $n \le -1$.

Since, for two functions $F(t)$ and $G(t)$, for which the integrals exist

$$\mathcal{L}\{aF(t) + bG(t)\} = a\mathcal{L}\{F(t)\} + b\mathcal{L}\{G(t)\}, \tag{15.101}$$

the operation denoted by $\mathcal{L}$ is **linear**.

Elementary Functions

To introduce the Laplace transform, let us apply the operation to some of the elementary functions. In all cases we assume that $F(t) = 0$ for $t < 0$. If

$$F(t) = 1, \qquad t > 0,$$

[11] This is sometimes called a **one-sided Laplace transform**; the integral from $-\infty$ to $+\infty$ is referred to as a **two-sided Laplace transform**. Some authors introduce an additional factor of s. This extra s appears to have little advantage and continually gets in the way (compare Jeffreys and Jeffreys, Section 14.13 — see the Additional Readings — for additional comments). Generally, we take s to be real and positive. It is possible to have s complex, provided $\Re(s) > 0$.

then

$$\mathcal{L}\{1\} = \int_0^\infty e^{-st}\,dt = \frac{1}{s}, \qquad \text{for } s > 0. \tag{15.102}$$

Again, let

$$F(t) = e^{kt}, \qquad t > 0.$$

The Laplace transform becomes

$$\mathcal{L}\{e^{kt}\} = \int_0^\infty e^{-st}e^{kt}\,dt = \frac{1}{s-k}, \qquad \text{for } s > k. \tag{15.103}$$

Using this relation, we obtain the Laplace transform of certain other functions. Since

$$\cosh kt = \frac{1}{2}\left(e^{kt} + e^{-kt}\right), \qquad \sinh kt = \frac{1}{2}\left(e^{kt} - e^{-kt}\right), \tag{15.104}$$

we have

$$\begin{aligned} \mathcal{L}\{\cosh kt\} &= \frac{1}{2}\left(\frac{1}{s-k} + \frac{1}{s+k}\right) = \frac{s}{s^2-k^2}, \\ \mathcal{L}\{\sinh kt\} &= \frac{1}{2}\left(\frac{1}{s-k} - \frac{1}{s+k}\right) = \frac{k}{s^2-k^2}, \end{aligned} \tag{15.105}$$

both valid for $s > k$. We have the relations

$$\cos kt = \cosh ikt, \qquad \sin kt = -i\sinh ikt. \tag{15.106}$$

Using Eqs. (15.105) with k replaced by ik, we find that the Laplace transforms are

$$\begin{aligned} \mathcal{L}\{\cos kt\} &= \frac{s}{s^2+k^2}, \\ \mathcal{L}\{\sin kt\} &= \frac{k}{s^2+k^2}, \end{aligned} \tag{15.107}$$

both valid for $s > 0$. Another derivation of this last transform is given in the next section. Note that $\lim_{s\to 0}\mathcal{L}\{\sin kt\} = 1/k$. The Laplace transform assigns a value of $1/k$ to $\int_0^\infty \sin kt\,dt$.

Finally, for $F(t) = t^n$, we have

$$\mathcal{L}\{t^n\} = \int_0^\infty e^{-st}t^n\,dt,$$

which is just the factorial function. Hence

$$\mathcal{L}\{t^n\} = \frac{n!}{s^{n+1}}, \qquad s > 0,\ n > -1. \tag{15.108}$$

Note that in all these transforms we have the variable s in the denominator — negative powers of s. In particular, $\lim_{s\to\infty} f(s) = 0$. The significance of this point is that if $f(s)$ involves positive powers of s ($\lim_{s\to\infty} f(s) \to \infty$), then no inverse transform exists.

Inverse Transform

There is little importance to these operations unless we can carry out the inverse transform, as in Fourier transforms. That is, with

$$\mathcal{L}\{F(t)\} = f(s),$$

then

$$\mathcal{L}^{-1}\{f(s)\} = F(t). \tag{15.109}$$

This inverse transform is **not** unique. Two functions $F_1(t)$ and $F_2(t)$ may have the same transform, $f(s)$. However, in this case

$$F_1(t) - F_2(t) = N(t),$$

where $N(t)$ is a null function (Fig. 15.7), indicating that

$$\int_0^{t_0} N(t)\,dt = 0,$$

for all positive t_0. This result is known as **Lerch's theorem**. Therefore to the physicist and engineer $N(t)$ may almost always be taken as zero and the inverse operation becomes unique.

The inverse transform can be determined in various ways. (1) A table of transforms can be built up and used to carry out the inverse transformation, exactly as a table of logarithms can be used to look up antilogarithms. The preceding transforms constitute the beginnings of such a table. For a more complete set of Laplace transforms see upcoming Table 15.2 or AMS-55, Chapter 29 (see footnote 4 in Chapter 5 for the reference). Employing partial fraction expansions and various operational theorems, which are considered in succeeding sections, facilitates use of the tables.

- There is some justification for suspecting that these tables are probably of more value in solving textbook exercises than in solving real-world problems.
- (2) A general technique for $\mathcal{L}^{-1}$ will be developed in Section 15.12 by using the calculus of residues.

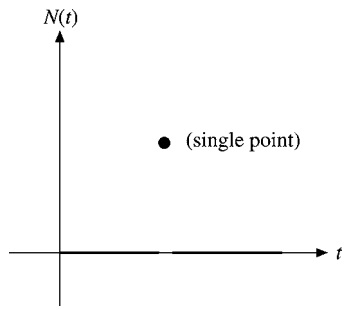

FIGURE 15.7 A possible null function.

- (3) For the difficulties and the possibilities of a numerical approach — numerical inversion — we refer to the Additional Readings.

Partial Fraction Expansion

Utilization of a table of transforms (or inverse transforms) is facilitated by expanding $f(s)$ in **partial fractions**.

Frequently $f(s)$, our transform, occurs in the form $g(s)/h(s)$, where $g(s)$ and $h(s)$ are polynomials with no common factors, $g(s)$ being of lower degree than $h(s)$. If the factors of $h(s)$ are all linear and distinct, then by the method of partial fractions we may write

$$f(s) = \frac{c_1}{s-a_1} + \frac{c_2}{s-a_2} + \cdots + \frac{c_n}{s-a_n}, \tag{15.110}$$

where the c_i are independent of s. The a_i are the roots of $h(s)$. If any one of the roots, say, a_1, is multiple (occurring m times), then $f(s)$ has the form

$$f(s) = \frac{c_{1,m}}{(s-a_1)^m} + \frac{c_{1,m-1}}{(s-a_1)^{m-1}} + \cdots + \frac{c_{1,1}}{s-a_1} + \sum_{i=2}^{n} \frac{c_i}{s-a_i}. \tag{15.111}$$

Finally, if one of the factors is quadratic, $(s^2 + ps + q)$, then the numerator, instead of being a simple constant, will have the form

$$\frac{as+b}{s^2+ps+q}.$$

There are various ways of determining the constants introduced. For instance, in Eq. (15.110) we may multiply through by $(s - a_i)$ and obtain

$$c_i = \lim_{s \to a_i} (s-a_i) f(s). \tag{15.112}$$

In elementary cases a direct solution is often the easiest.

Example 15.8.1 PARTIAL FRACTION EXPANSION

Let

$$f(s) = \frac{k^2}{s(s^2+k^2)} = \frac{c}{s} + \frac{as+b}{s^2+k^2}. \tag{15.113}$$

Putting the right side of the equation over a common denominator and equating like powers of s in the numerator, we obtain

$$\frac{k^2}{s(s^2+k^2)} = \frac{c(s^2+k^2)+s(as+b)}{s(s^2+k^2)}, \tag{15.114}$$

$$c + a = 0, \quad s^2; \qquad b = 0, \quad s^1; \qquad ck^2 = k^2, \quad s^0.$$

Solving these $(s \neq 0)$, we have

$$c = 1, \qquad b = 0, \qquad a = -1,$$

giving

$$f(s) = \frac{1}{s} - \frac{s}{s^2 + k^2}, \tag{15.115}$$

and

$$\mathcal{L}^{-1}\{f(s)\} = 1 - \cos kt \tag{15.116}$$

by Eqs. (15.102) and (15.106). ■

Example 15.8.2 A STEP FUNCTION

As one application of Laplace transforms, consider the evaluation of

$$F(t) = \int_0^\infty \frac{\sin tx}{x} dx. \tag{15.117}$$

Suppose we take the Laplace transform of this definite (and improper) integral:

$$\mathcal{L}\left\{\int_0^\infty \frac{\sin tx}{x} dx\right\} = \int_0^\infty e^{-st} \int_0^\infty \frac{\sin tx}{x} dx\, dt. \tag{15.118}$$

Now, interchanging the order of integration (which is justified),[12] we get

$$\int_0^\infty \frac{1}{x}\left[\int_0^\infty e^{-st} \sin tx\, dt\right] dx = \int_0^\infty \frac{dx}{s^2 + x^2}, \tag{15.119}$$

since the factor in square brackets is just the Laplace transform of $\sin tx$. From the integral tables,

$$\int_0^\infty \frac{dx}{s^2 + x^2} = \frac{1}{s} \tan^{-1}\left(\frac{x}{s}\right)\Big|_0^\infty = \frac{\pi}{2s} = f(s). \tag{15.120}$$

By Eq. (15.102) we carry out the inverse transformation to obtain

$$F(t) = \frac{\pi}{2}, \qquad t > 0, \tag{15.121}$$

in agreement with an evaluation by the calculus of residues (Section 7.1). It has been assumed that $t > 0$ in $F(t)$. For $F(-t)$ we need note only that $\sin(-tx) = -\sin tx$, giving $F(-t) = -F(t)$. Finally, if $t = 0$, $F(0)$ is clearly zero. Therefore

$$\int_0^\infty \frac{\sin tx}{x} dx = \frac{\pi}{2}\big[2u(t) - 1\big] = \begin{cases} \frac{\pi}{2}, & t > 0 \\ 0, & t = 0 \\ -\frac{\pi}{2}, & t < 0. \end{cases} \tag{15.122}$$

Note that $\int_0^\infty (\sin tx/x)\, dx$, taken as a function of t, describes a step function (Fig. 15.8), a step of height π at $t = 0$. This is consistent with Eq. (1.174). ■

The technique in the preceding example was to (1) introduce a second integration — the Laplace transform, (2) reverse the order of integration and integrate, and (3) take the

[12] See — in the Additional Readings — Jeffreys and Jeffreys (1966), Chapter 1 (uniform convergence of integrals).

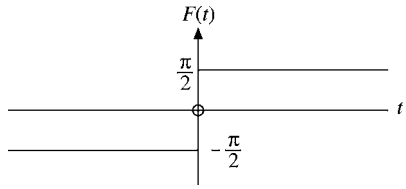

FIGURE 15.8 $F(t) = \int_0^\infty \frac{\sin tx}{x} dx$, a step function.

inverse Laplace transform. There are many opportunities where this technique of reversing the order of integration can be applied and proved useful. Exercise 15.8.6 is a variation of this.

Exercises

15.8.1 Prove that

$$\lim_{s\to\infty} sf(s) = \lim_{t\to+0} F(t).$$

Hint. Assume that $F(t)$ can be expressed as $F(t) = \sum_{n=0}^{\infty} a_n t^n$.

15.8.2 Show that

$$\frac{1}{\pi} \lim_{s\to 0} \mathcal{L}\{\cos xt\} = \delta(x).$$

15.8.3 Verify that

$$\mathcal{L}\left\{\frac{\cos at - \cos bt}{b^2 - a^2}\right\} = \frac{s}{(s^2 + a^2)(s^2 + b^2)}, \qquad a^2 \neq b^2.$$

15.8.4 Using partial fraction expansions, show that

(a) $$\mathcal{L}^{-1}\left\{\frac{1}{(s+a)(s+b)}\right\} = \frac{e^{-at} - e^{-bt}}{b-a}, \qquad a \neq b.$$

(b) $$\mathcal{L}^{-1}\left\{\frac{s}{(s+a)(s+b)}\right\} = \frac{ae^{-at} - be^{-bt}}{a-b}, \qquad a \neq b.$$

15.8.5 Using partial fraction expansions, show that for $a^2 \neq b^2$,

(a) $$\mathcal{L}^{-1}\left\{\frac{1}{(s^2 + a^2)(s^2 + b^2)}\right\} = -\frac{1}{a^2 - b^2}\left\{\frac{\sin at}{a} - \frac{\sin bt}{b}\right\},$$

(b) $$\mathcal{L}^{-1}\left\{\frac{s^2}{(s^2 + a^2)(s^2 + b^2)}\right\} = \frac{1}{a^2 - b^2}\{a \sin at - b \sin bt\}.$$

15.8.6 The electrostatic potential of a charged conducting disk is known to have the general form (circular cylindrical coordinates)

$$\Phi(\rho, z) = \int_0^\infty e^{-k|z|} J_0(k\rho) f(k)\, dk,$$

with $f(k)$ unknown. At large distances ($z \to \infty$) the potential must approach the Coulomb potential $Q/4\pi\varepsilon_0 z$. Show that

$$\lim_{k\to 0} f(k) = \frac{q}{4\pi\varepsilon_0}.$$

Hint. You may set $\rho = 0$ and assume a Maclaurin expansion of $f(k)$ or, using e^{-kz}, construct a delta sequence.

15.8.7 Show that

(a) $$\int_0^\infty \frac{\cos s}{s^\nu} ds = \frac{\pi}{2(\nu-1)!\cos(\nu\pi/2)}, \qquad 0 < \nu < 1,$$

(b) $$\int_0^\infty \frac{\sin s}{s^\nu} ds = \frac{\pi}{2(\nu-1)!\sin(\nu\pi/2)}, \qquad 0 < \nu < 2,$$

Why is ν restricted to (0, 1) for (a), to (0, 2) for (b)? These integrals may be interpreted as Fourier transforms of $s^{-\nu}$ and as Mellin transforms of $\sin s$ and $\cos s$.
Hint. Replace $s^{-\nu}$ by a Laplace transform integral: $\mathcal{L}\{t^{\nu-1}\}/(\nu-1)!$. Then integrate with respect to s. The resulting integral can be treated as a beta function (Section 8.4).

15.8.8 A function $F(t)$ can be expanded in a power series (Maclaurin); that is,

$$F(t) = \sum_{n=0}^\infty a_n t^n.$$

Then

$$\mathcal{L}\big\{F(t)\big\} = \int_0^\infty e^{-st} \sum_{n=0}^\infty a_n t^n dt = \sum_{n=0}^\infty a_n \int_0^\infty e^{-st} t^n dt.$$

Show that $f(s)$, the Laplace transform of $F(t)$, contains no powers of s greater than s^{-1}. Check your result by calculating $\mathcal{L}\{\delta(t)\}$, and comment on this fiasco.

15.8.9 Show that the Laplace transform of $M(a, c, x)$ is

$$\mathcal{L}\big\{M(a, c, x)\big\} = \frac{1}{s}\,{}_2F_1\left(a, 1; c, \frac{1}{s}\right).$$

15.9 LAPLACE TRANSFORM OF DERIVATIVES

Perhaps the main application of Laplace transforms is in converting differential equations into simpler forms that may be solved more easily. It will be seen, for instance, that coupled differential equations with constant coefficients transform to simultaneous linear algebraic equations.

Let us transform the first derivative of $F(t)$:

$$\mathcal{L}\{F'(t)\} = \int_0^\infty e^{-st} \frac{dF(t)}{dt} dt.$$

Integrating by parts, we obtain

$$\begin{aligned}\mathcal{L}\{F'(t)\} &= e^{-st}F(t)\Big|_0^\infty + s\int_0^\infty e^{-st}F(t)\,dt \\ &= s\mathcal{L}\{F(t)\} - F(0). \end{aligned} \tag{15.123}$$

Strictly speaking, $F(0) = F(+0)$[13]and dF/dt is required to be at least piecewise continuous for $0 \le t < \infty$. Naturally, both $F(t)$ and its derivative must be such that the integrals do not diverge. Incidentally, Eq. (15.123) provides another proof of Exercise 15.8.8. An extension gives

$$\mathcal{L}\{F^{(2)}(t)\} = s^2\mathcal{L}\{F(t)\} - sF(+0) - F'(+0), \tag{15.124}$$

$$\mathcal{L}\{F^{(n)}(t)\} = s^n\mathcal{L}\{F(t)\} - s^{n-1}F(+0) - \cdots - F^{(n-1)}(+0). \tag{15.125}$$

The Laplace transform, like the Fourier transform, replaces differentiation with multiplication. In the following examples ODEs become algebraic equations. Here is the power and the utility of the Laplace transform. But see Example 15.10.3 for what may happen if the coefficients are not constant.

Note how the initial conditions, $F(+0)$, $F'(+0)$, and so on, are incorporated into the transform. Equation (15.124) may be used to derive $\mathcal{L}\{\sin kt\}$. We use the identity

$$-k^2 \sin kt = \frac{d^2}{dt^2}\sin kt. \tag{15.126}$$

Then applying the Laplace transform operation, we have

$$\begin{aligned}-k^2\mathcal{L}\{\sin kt\} &= \mathcal{L}\left\{\frac{d^2}{dt^2}\sin kt\right\} \\ &= s^2\mathcal{L}\{\sin kt\} - s\sin(0) - \frac{d}{dt}\sin kt\Big|_{t=0}. \end{aligned} \tag{15.127}$$

Since $\sin(0) = 0$ and $d/dt \sin kt|_{t=0} = k$,

$$\mathcal{L}\{\sin kt\} = \frac{k}{s^2 + k^2}, \tag{15.128}$$

verifying Eq. (15.107).

[13]Zero is approached from the positive side.

Example 15.9.1 SIMPLE HARMONIC OSCILLATOR

As a physical example, consider a mass m oscillating under the influence of an ideal spring, spring constant k. As usual, friction is neglected. Then Newton's second law becomes

$$m\frac{d^2X(t)}{dt^2} + kX(t) = 0; \tag{15.129}$$

also, we take as initial conditions

$$X(0) = X_0, \qquad X'(0) = 0.$$

Applying the Laplace transform, we obtain

$$m\mathcal{L}\left\{\frac{d^2X}{dt^2}\right\} + k\mathcal{L}\{X(t)\} = 0, \tag{15.130}$$

and by use of Eq. (15.124) this becomes

$$ms^2x(s) - msX_0 + kx(s) = 0, \tag{15.131}$$

$$x(s) = X_0\frac{s}{s^2+\omega_0^2}, \qquad \text{with } \omega_0^2 \equiv \frac{k}{m}. \tag{15.132}$$

From Eq. (15.107) this is seen to be the transform of $\cos\omega_0 t$, which gives

$$X(t) = X_0\cos\omega_0 t, \tag{15.133}$$

as expected. ■

Example 15.9.2 EARTH'S NUTATION

A somewhat more involved example is the nutation of the earth's poles (force-free precession). If we treat the Earth as a rigid (oblate) spheroid, the Euler equations of motion reduce to

$$\frac{dX}{dt} = -aY, \qquad \frac{dY}{dt} = +aX, \tag{15.134}$$

where $a \equiv [(I_z - I_x)/I_z]\omega_z$, $X = \omega_x$, $Y = \omega_y$ with angular velocity vector $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)$ (Fig. 15.9), I_z = moment of inertia about the z-axis and $I_y = I_x$ moment of inertia about the x- (or y-)axis. The z-axis coincides with the axis of symmetry of the Earth. It differs from the axis for the Earth's daily rotation, $\boldsymbol{\omega}$, by some 15 meters, measured at the poles. Transformation of these coupled differential equations yields

$$sx(s) - X(0) = -ay(s), \qquad sy(s) - Y(0) = ax(s). \tag{15.135}$$

Combining to eliminate $y(s)$, we have

$$s^2x(s) - sX(0) + aY(0) = -a^2x(s),$$

or

$$x(s) = X(0)\frac{s}{s^2+a^2} - Y(0)\frac{a}{s^2+a^2}. \tag{15.136}$$

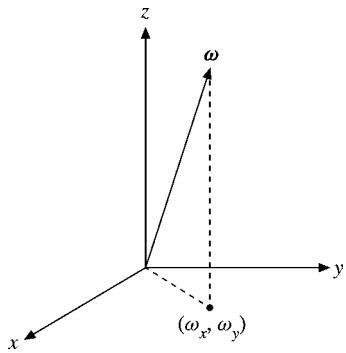

FIGURE 15.9

Hence

$$X(t) = X(0)\cos at - Y(0)\sin at. \tag{15.137}$$

Similarly,

$$Y(t) = X(0)\sin at + Y(0)\cos at. \tag{15.138}$$

This is seen to be a rotation of the vector (X, Y) counterclockwise (for $a > 0$) about the z-axis with angle $\theta = at$ and angular velocity a.

A direct interpretation may be found by choosing the time axis so that $Y(0) = 0$. Then

$$X(t) = X(0)\cos at, \qquad Y(t) = X(0)\sin at, \tag{15.139}$$

which are the parametric equations for rotation of (X, Y) in a circular orbit of radius $X(0)$, with angular velocity a in the counterclockwise sense.

In the case of the Earth's angular velocity, vector $X(0)$ is about 15 meters, whereas a, as defined here, corresponds to a period $(2\pi/a)$ of some 300 days. Actually because of departures from the idealized rigid body assumed in setting up Euler's equations, the period is about 427 days.[14] If in Eq. (15.134) we set

$$X(t) = L_x, \qquad Y(t) = L_y,$$

where L_x and L_y are the x- and y-components of the angular momentum $\mathbf{L}$, $a = -g_L B_z$, g_L is the gyromagnetic ratio, and B_z is the magnetic field (along the z-axis), then Eq. (15.134) describes the Larmor precession of charged bodies in a uniform magnetic field B_z. ■

[14]D. Menzel, ed., *Fundamental Formulas of Physics*, Englewood Cliffs, NJ: Prentice-Hall (1955), reprinted, 2nd ed., Dover (1960), p. 695.

Dirac Delta Function

For use with differential equations one further transform is helpful — the Dirac delta function:[15]

$$\mathcal{L}\{\delta(t-t_0)\} = \int_0^\infty e^{-st}\delta(t-t_0)\,dt = e^{-st_0}, \qquad \text{for } t_0 \geq 0, \tag{15.140}$$

and for $t_0 = 0$

$$\mathcal{L}\{\delta(t)\} = 1, \tag{15.141}$$

where it is assumed that we are using a representation of the delta function such that

$$\int_0^\infty \delta(t)\,dt = 1, \qquad \delta(t) = 0, \qquad \text{for } t > 0. \tag{15.142}$$

As an alternate method, $\delta(t)$ may be considered the limit as $\varepsilon \to 0$ of $F(t)$, where

$$F(t) = \begin{cases} 0, & t < 0, \\ \varepsilon^{-1}, & 0 < t < \varepsilon, \\ 0, & t > \varepsilon. \end{cases} \tag{15.143}$$

By direct calculation

$$\mathcal{L}\{F(t)\} = \frac{1 - e^{-\varepsilon s}}{\varepsilon s}. \tag{15.144}$$

Taking the limit of the integral (instead of the integral of the limit), we have

$$\lim_{\varepsilon \to 0} \mathcal{L}\{F(t)\} = 1,$$

or Eq. (15.141),

$$\mathcal{L}\{\delta(t)\} = 1.$$

This delta function is frequently called the **impulse** function because it is so useful in describing impulsive forces, that is, forces lasting only a short time.

Example 15.9.3 IMPULSIVE FORCE

Newton's second law for impulsive force acting on a particle of mass m becomes

$$m\frac{d^2X}{dt^2} = P\delta(t), \tag{15.145}$$

where P is a constant. Transforming, we obtain

$$ms^2x(s) - msX(0) - mX'(0) = P. \tag{15.146}$$

[15] Strictly speaking, the Dirac delta function is undefined. However, the integral over it is well defined. This approach is developed in Section 1.16 using delta sequences.

For a particle starting from rest, $X'(0) = 0$.[16] We shall also take $X(0) = 0$. Then

$$x(s) = \frac{P}{ms^2}, \tag{15.147}$$

and

$$X(t) = \frac{P}{m}t, \tag{15.148}$$

$$\frac{dX(t)}{dt} = \frac{P}{m}, \qquad \text{a constant.} \tag{15.149}$$

The effect of the impulse $P\delta(t)$ is to transfer (instantaneously) P units of linear momentum to the particle.

A similar analysis applies to the ballistic galvanometer. The torque on the galvanometer is given initially by $k\iota$, in which ι is a pulse of current and k is a proportionality constant. Since ι is of short duration, we set

$$k\iota = kq\ \delta(t), \tag{15.150}$$

where q is the total charge carried by the current ι. Then, with I the moment of inertia,

$$I\frac{d^2\theta}{dt^2} = kq\ \delta(t), \tag{15.151}$$

and, transforming as before, we find that the effect of the current pulse is a transfer of kq units of **angular** momentum to the galvanometer. ■

Exercises

15.9.1 Use the expression for the transform of a second derivative to obtain the transform of $\cos kt$.

15.9.2 A mass m is attached to one end of an unstretched spring, spring constant k (Fig. 15.10). At time $t = 0$ the free end of the spring experiences a constant acceleration a, away from the mass. Using Laplace transforms,

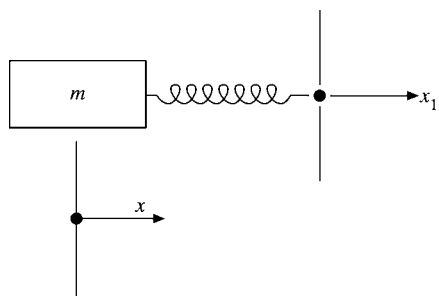

FIGURE 15.10 Spring.

[16] This should be $X'(+0)$. To include the effect of the impulse, consider that the impulse will occur at $t = \varepsilon$ and let $\varepsilon \to 0$.

(a) Find the position x of m as a function of time.
(b) Determine the limiting form of $x(t)$ for small t.

$$\text{ANS. (a) } x = \frac{1}{2}at^2 - \frac{a}{\omega^2}(1-\cos\omega t), \quad \omega^2 = \frac{k}{m},$$
$$\text{(b) } x = \frac{a\omega^2}{4!}t^4, \quad \omega t \ll 1.$$

15.9.3 Radioactive nuclei decay according to the law

$$\frac{dN}{dt} = -\lambda N,$$

N being the concentration of a given nuclide and λ being the particular decay constant. This equation may be interpreted as stating that the rate of decay is proportional to the number of these radioactive nuclei present. They all decay independently.
In a radioactive series of n different nuclides, starting with N_1,

$$\frac{dN_1}{dt} = -\lambda_1 N_1,$$
$$\frac{dN_2}{dt} = \lambda_1 N_1 - \lambda_2 N_2, \qquad \text{and so on.}$$
$$\frac{dN_n}{dt} = \lambda_{n-1} N_{n-1}, \qquad \text{stable.}$$

Find $N_1(t)$, $N_2(t)$, $N_3(t)$, $n = 3$, with $N_1(0) = N_0$, $N_2(0) = N_3(0) = 0$.

$$\text{ANS. } N_1(t) = N_0 e^{-\lambda_1 t}, \quad N_2(t) = N_0 \frac{\lambda_1}{\lambda_2 - \lambda_1}\left(e^{-\lambda_1 t} - e^{-\lambda_2 t}\right),$$
$$N_3(t) = N_0\left(1 - \frac{\lambda_2}{\lambda_2 - \lambda_1}e^{-\lambda_1 t} + \frac{\lambda_1}{\lambda_2 - \lambda_1}e^{-\lambda_2 t}\right).$$

Find an approximate expression for N_2 and N_3, valid for small t when $\lambda_1 \approx \lambda_2$.

$$\text{ANS. } N_2 \approx N_0\lambda_1 t, \quad N_3 \approx \frac{N_0}{2}\lambda_1\lambda_2 t^2.$$

Find approximate expressions for N_2 and N_3, valid for large t, when

(a) $\lambda_1 \gg \lambda_2$,
(b) $\lambda_1 \ll \lambda_2$.

$$\text{ANS. (a) } N_2 \approx N_0 e^{-\lambda_2 t},$$
$$N_3 \approx N_0\left(1 - e^{-\lambda_2 t}\right), \quad \lambda_1 t \gg 1.$$
$$\text{(b) } N_2 \approx N_0 \tfrac{\lambda_1}{\lambda_2} e^{-\lambda_1 t},$$
$$N_3 \approx N_0\left(1 - e^{-\lambda_1 t}\right), \quad \lambda_2 t \gg 1.$$

15.9.4 The formation of an isotope in a nuclear reactor is given by

$$\frac{dN_2}{dt} = nv\sigma_1 N_{10} - \lambda_2 N_2(t) - nv\sigma_2 N_2(t).$$

Here the product nv is the neutron flux, neutrons per cubic centimeter, times centimeters per second mean velocity; σ_1 and σ_2 (cm^2) are measures of the probability of neutron absorption by the original isotope, concentration N_{10}, which is assumed constant and the newly formed isotope, concentration N_2, respectively. The radioactive decay constant for the isotope is λ_2.

(a) Find the concentration N_2 of the new isotope as a function of time.
(b) If the original element is Eu^{153}, $\sigma_1 = 400$ barns $= 400 \times 10^{-24}$ cm^2, $\sigma_2 = 1000$ barns $= 1000 \times 10^{-24}$ cm^2, and $\lambda_2 = 1.4 \times 10^{-9}\,\text{s}^{-1}$. If $N_{10} = 10^{20}$ and $(nv) = 10^9\ \text{cm}^{-2}\,\text{s}^{-1}$, find N_2, the concentration of Eu^{154} after one year of continuous irradiation. Is the assumption that N_1 is constant justified?

15.9.5 In a nuclear reactor Xe^{135} is formed as both a direct fission product and a decay product of I^{135}, half-life, 6.7 hours. The half-life of Xe^{135} is 9.2 hours. Because Xe^{135} strongly absorbs thermal neutrons thereby "poisoning" the nuclear reactor, its concentration is a matter of great interest. The relevant equations are

$$\frac{dN_I}{dt} = \gamma_I \varphi \sigma_f N_U - \lambda_I N_I,$$
$$\frac{dN_X}{dt} = \lambda_I N_I + \gamma_X \varphi \sigma_f N_U - \lambda_X N_X - \varphi \sigma_X N_X.$$

Here $N_I =$ concentration of I^{135} (Xe^{135}, U^{235}). Assume

$N_U =$ constant,

$\gamma_I =$ yield of I^{135} per fission $= 0.060$,

$\gamma_X =$ yield of Xe^{135} direct from fission $= 0.003$,

$\lambda_I = I^{135}$ $\left(Xe^{135}\right)$ decay constant $= \dfrac{\ln 2}{t_{1/2}} = \dfrac{0.693}{t_{1/2}}$,

$\sigma_f =$ thermal neutron fission cross section for U^{235},

$\sigma_X =$ thermal neutron absorption cross section for Xe^{135}

$= 3.5 \times 10^6$ barns $= 3.5 \times 10^{-18}$ cm^2.

(σ_I the absorption cross section of I^{135}, is negligible.)

$\varphi =$ neutron flux $=$ neutrons/$\text{cm}^3 \times$ mean velocity (cm/s).

(a) Find $N_X(t)$ in terms of neutron flux φ and the product $\sigma_f N_U$.
(b) Find $N_X(t \to \infty)$.
(c) After N_X has reached equilibrium, the reactor is shut down, $\varphi = 0$. Find $N_X(t)$ following shutdown. Notice the increase in N_X, which may for a few hours interfere with starting the reactor up again.

15.10 OTHER PROPERTIES

Substitution

If we replace the parameter s by $s - a$ in the definition of the Laplace transform (Eq. (15.99)), we have

$$f(s-a) = \int_0^\infty e^{-(s-a)t} F(t)\,dt = \int_0^\infty e^{-st} e^{at} F(t)\,dt$$
$$= \mathcal{L}\{e^{at} F(t)\}. \tag{15.152}$$

Hence the replacement of s with $s - a$ corresponds to multiplying $F(t)$ by e^{at}, and conversely. This result can be used to good advantage in extending our table of transforms. From Eq. (15.107) we find immediately that

$$\mathcal{L}\{e^{at}\sin kt\} = \frac{k}{(s-a)^2 + k^2}; \tag{15.153}$$

also,

$$\mathcal{L}\{e^{at}\cos kt\} = \frac{s-a}{(s-a)^2 + k^2}, \qquad s > a.$$

Example 15.10.1 DAMPED OSCILLATOR

These expressions are useful when we consider an oscillating mass with damping proportional to the velocity. Equation (15.129), with such damping added, becomes

$$mX''(t) + bX'(t) + kX(t) = 0, \tag{15.154}$$

in which b is a proportionality constant. Let us assume that the particle starts from rest at $X(0) = X_0$, $X'(0) = 0$. The transformed equation is

$$m[s^2 x(s) - sX_0] + b[sx(s) - X_0] + kx(s) = 0, \tag{15.155}$$

and

$$x(s) = X_0 \frac{ms + b}{ms^2 + bs + k}. \tag{15.156}$$

This may be handled by completing the square of the denominator:

$$s^2 + \frac{b}{m}s + \frac{k}{m} = \left(s + \frac{b}{2m}\right)^2 + \left(\frac{k}{m} - \frac{b^2}{4m^2}\right). \tag{15.157}$$

If the damping is small, $b^2 < 4\,km$, the last term is positive and will be denoted by ω_1^2:

$$x(s) = X_0 \frac{s + b/m}{(s + b/2m)^2 + \omega_1^2}$$
$$= X_0 \frac{s + b/2m}{(s + b/2m)^2 + \omega_1^2} + X_0 \frac{(b/2m\omega_1)\omega_1}{(s + b/2m)^2 + \omega_1^2}. \tag{15.158}$$

By Eq. (15.153),

$$X(t) = X_0 e^{-(b/2m)t}\left(\cos\omega_1 t + \frac{b}{2m\omega_1}\sin\omega_1 t\right)$$
$$= X_0\frac{\omega_0}{\omega_1}e^{-(b/2m)t}\cos(\omega_1 t - \varphi), \tag{15.159}$$

where

$$\tan\varphi = \frac{b}{2m\omega_1}, \qquad \omega_0^2 = \frac{k}{m}.$$

Of course, as $b \to 0$, this solution goes over to the undamped solution (Section 15.9). ■

RLC Analog

It is worth noting the similarity between this damped simple harmonic oscillation of a mass on a spring and an *RLC* circuit (resistance, inductance, and capacitance) (Fig. 15.11). At any instant the sum of the potential differences around the loop must be zero (Kirchhoff's law, conservation of energy). This gives

$$L\frac{dI}{dt} + RI + \frac{1}{C}\int^t I\,dt = 0. \tag{15.160}$$

Differentiating the current I with respect to time (to eliminate the integral), we have

$$L\frac{d^2I}{dt^2} + R\frac{dI}{dt} + \frac{1}{C}I = 0. \tag{15.161}$$

If we replace $I(t)$ with $X(t)$, L with m, R with b, and C^{-1} with k, then Eq. (15.161) is identical with the mechanical problem. It is but one example of the unification of diverse branches of physics by mathematics. A more complete discussion will be found in Olson's book.[17]

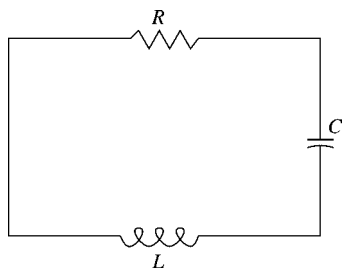

FIGURE 15.11 *RLC* circuit.

[17] H. F. Olson, *Dynamical Analogies*, New York: Van Nostrand (1943).

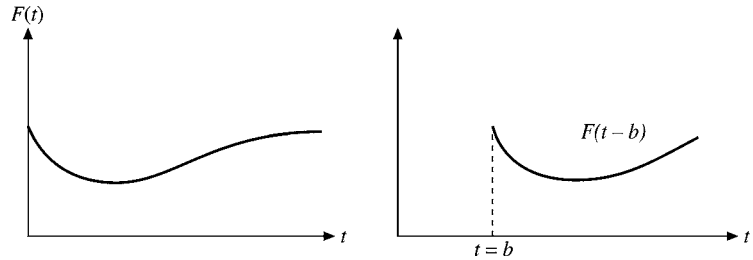

FIGURE 15.12 Translation.

Translation

This time let $f(s)$ be multiplied by e^{-bs}, $b > 0$:

$$e^{-bs} f(s) = e^{-bs} \int_0^\infty e^{-st} F(t)\, dt$$
$$= \int_0^\infty e^{-s(t+b)} F(t)\, dt. \tag{15.162}$$

Now let $t + b = \tau$. Equation (15.162) becomes

$$e^{-bs} f(s) = \int_b^\infty e^{-s\tau} F(\tau - b)\, d\tau$$
$$= \int_0^\infty e^{-s\tau} F(\tau - b) u(\tau - b)\, d\tau, \tag{15.163}$$

where $u(\tau - b)$ is the unit step function. This relation is often called the **Heaviside shifting theorem** (Fig. 15.12).

Since $F(t)$ is assumed to be equal to zero for $t < 0$, $F(\tau - b) = 0$ for $0 \le \tau < b$. Therefore we can extend the lower limit to zero without changing the value of the integral. Then, noting that τ is only a variable of integration, we obtain

$$e^{-bs} f(s) = \mathcal{L}\{F(t - b)\}. \tag{15.164}$$

Example 15.10.2 ELECTROMAGNETIC WAVES

The electromagnetic wave equation with $E = E_y$ or E_z, a transverse wave propagating along the x-axis, is

$$\frac{\partial^2 E(x,t)}{\partial x^2} - \frac{1}{v^2} \frac{\partial^2 E(x,t)}{\partial t^2} = 0. \tag{15.165}$$

Transforming this equation with respect to t, we get

$$\frac{\partial^2}{\partial x^2} \mathcal{L}\{E(x,t)\} - \frac{s^2}{v^2} \mathcal{L}\{E(x,t)\} + \frac{s}{v^2} E(x,0) + \frac{1}{v^2} \frac{\partial E(x,t)}{\partial t}\bigg|_{t=0} = 0. \tag{15.166}$$

If we have the initial condition $E(x,0)=0$ and

$$\left.\frac{\partial E(x,t)}{\partial t}\right|_{t=0}=0,$$

then

$$\frac{\partial^2}{\partial x^2}\mathcal{L}\big\{E(x,t)\big\}=\frac{s^2}{v^2}\mathcal{L}\big\{E(x,t)\big\}. \tag{15.167}$$

The solution (of this **ODE**) is

$$\mathcal{L}\big\{E(x,t)\big\}=c_1e^{-(s/v)x}+c_2e^{+(s/v)x}. \tag{15.168}$$

The "constants" c_1 and c_2 are obtained by additional boundary conditions. They are constant with respect to x but may depend on s. If our wave remains finite as $x\to\infty$, $\mathcal{L}\{E(x,t)\}$ will also remain finite. Hence $c_2=0$. If $E(0,t)$ is denoted by $F(t)$, then $c_1=f(s)$ and

$$\mathcal{L}\big\{E(x,t)\big\}=e^{-(s/v)x}f(s). \tag{15.169}$$

From the translation property (Eq. (15.164)) we find immediately that

$$E(x,t)=\begin{cases}F\big(t-\frac{x}{v}\big), & t\ge\frac{x}{v},\\ 0, & t<\frac{x}{v}.\end{cases} \tag{15.170}$$

Differentiation and substitution into Eq. (15.165) verifies Eq. (15.170). Our solution represents a wave (or pulse) moving in the positive x-direction with velocity v. Note that for $x>vt$ the region remains undisturbed; the pulse has not had time to get there. If we had wanted a signal propagated along the negative x-axis, c_1 would have been set equal to 0 and we would have obtained

$$E(x,t)=\begin{cases}F\big(t+\frac{x}{v}\big), & t\ge-\frac{x}{v},\\ 0, & t<-\frac{x}{v},\end{cases} \tag{15.171}$$

a wave along the negative x-axis. ■

Derivative of a Transform

When $F(t)$, which is at least piecewise continuous, and s are chosen so that $e^{-st}F(t)$ converges exponentially for large s, the integral

$$\int_0^\infty e^{-st}F(t)\,dt$$

is uniformly convergent and may be differentiated (under the integral sign) with respect to s. Then

$$f'(s)=\int_0^\infty(-t)e^{-st}F(t)\,dt=\mathcal{L}\big\{-tF(t)\big\}. \tag{15.172}$$

Continuing this process, we obtain

$$f^{(n)}(s)=\mathcal{L}\big\{(-t)^nF(t)\big\}. \tag{15.173}$$

All the integrals so obtained will be uniformly convergent because of the decreasing exponential behavior of $e^{-st}F(t)$.

This same technique may be applied to generate more transforms. For example,

$$\mathcal{L}\{e^{kt}\} = \int_0^\infty e^{-st}e^{kt}\,dt = \frac{1}{s-k}, \qquad s > k. \tag{15.174}$$

Differentiating with respect to s (or with respect to k), we obtain

$$\mathcal{L}\{te^{kt}\} = \frac{1}{(s-k)^2}, \qquad s > k. \tag{15.175}$$

Example 15.10.3 BESSEL'S EQUATION

An interesting application of a differentiated Laplace transform appears in the solution of Bessel's equation with $n = 0$. From Chapter 11 we have

$$x^2y''(x) + xy'(x) + x^2y(x) = 0. \tag{15.176}$$

Dividing by x and substituting $t = x$ and $F(t) = y(x)$ to agree with the present notation, we see that the Bessel equation becomes

$$tF''(t) + F'(t) + tF(t) = 0. \tag{15.177}$$

We need a regular solution, in particular, $F(0) = 1$. From Eq. (15.177) with $t = 0$, $F'(+0) = 0$. Also, we assume that our unknown $F(t)$ has a transform. Transforming and using Eqs. (15.123), (15.124), and (15.172), we have

$$-\frac{d}{ds}\left[s^2f(s) - s\right] + sf(s) - 1 - \frac{d}{ds}f(s) = 0. \tag{15.178}$$

Rearranging Eq. (15.178), we obtain

$$\left(s^2+1\right)f'(s) + sf(s) = 0, \tag{15.179}$$

or

$$\frac{df}{f} = -\frac{s\,ds}{s^2+1}, \tag{15.180}$$

a first-order ODE. By integration,

$$\ln f(s) = -\tfrac{1}{2}\ln\left(s^2+1\right) + \ln C, \tag{15.181}$$

which may be rewritten as

$$f(s) = \frac{C}{\sqrt{s^2+1}}. \tag{15.182}$$

To make use of Eq. (15.108), we expand $f(s)$ in a series of negative powers of s, convergent for $s > 1$:

$$\begin{aligned} f(s) &= \frac{C}{s}\left(1 + \frac{1}{s^2}\right)^{-1/2} \\ &= \frac{C}{s}\left[1 - \frac{1}{2s^2} + \frac{1\cdot 3}{2^2\cdot 2!s^4} - \cdots + \frac{(-1)^n(2n)!}{(2^n n!)^2 s^{2n}} + \cdots\right]. \end{aligned} \tag{15.183}$$

Inverting, term by term, we obtain

$$F(t) = C\sum_{n=0}^{\infty} \frac{(-1)^n t^{2n}}{(2^n n!)^2}. \tag{15.184}$$

When C is set equal to 1, as required by the initial condition $F(0) = 1$, $F(t)$ is just $J_0(t)$, our familiar Bessel function of order zero. Hence

$$\mathcal{L}\{J_0(t)\} = \frac{1}{\sqrt{s^2+1}}. \tag{15.185}$$

Note that we assumed $s > 1$. The proof for $s > 0$ is left as a problem.

It is worth noting that this application was successful and relatively easy because we took $n = 0$ in Bessel's equation. This made it possible to divide out a factor of x (or t). If this had not been done, the terms of the form $t^2F(t)$ would have introduced a second derivative of $f(s)$. The resulting equation would have been no easier to solve than the original one.

When we go beyond linear ODEs with constant coefficients, the Laplace transform may still be applied, but there is no guarantee that it will be helpful.

The application to Bessel's equation, $n \neq 0$, will be found in the references. Alternatively, we can show that

$$\mathcal{L}\{J_n(at)\} = \frac{a^{-n}(\sqrt{s^2+a^2}-s)^n}{\sqrt{s^2+a^2}} \tag{15.186}$$

by expressing $J_n(t)$ as an infinite series and transforming term by term. ■

Integration of Transforms

Again, with $F(t)$ at least piecewise continuous and x large enough so that $e^{-xt}F(t)$ decreases exponentially (as $x \to \infty$), the integral

$$f(x) = \int_0^{\infty} e^{-xt}F(t)\,dt \tag{15.187}$$

is uniformly convergent with respect to x. This justifies reversing the order of integration in the following equation:

$$\begin{aligned}\int_s^b f(x)\,dx &= \int_s^b dx \int_0^{\infty} dt\, e^{-xt}F(t)\\ &= \int_0^{\infty} \frac{F(t)}{t}\left(e^{-st} - e^{-bt}\right)dt,\end{aligned} \tag{15.188}$$

on integrating with respect to x. The lower limit s is chosen large enough so that $f(s)$ is within the region of uniform convergence. Now letting $b \to \infty$, we have

$$\int_s^{\infty} f(x)\,dx = \int_0^{\infty} \frac{F(t)}{t}e^{-st}dt = \mathcal{L}\left\{\frac{F(t)}{t}\right\}, \tag{15.189}$$

provided that $F(t)/t$ is finite at $t = 0$ or diverges less strongly than t^{-1} (so that $\mathcal{L}\{F(t)/t\}$ will exist).

Limits of Integration — Unit Step Function

The actual limits of integration for the Laplace transform may be specified with the (Heaviside) unit step function

$$u(t-k) = \begin{cases} 0, & t < k \\ 1, & t > k. \end{cases}$$

For instance,

$$\mathcal{L}\{u(t-k)\} = \int_k^\infty e^{-st}\,dt = \frac{1}{s}e^{-ks}.$$

A rectangular pulse of width k and unit height is described by $F(t) = u(t) - u(t-k)$. Taking the Laplace transform, we obtain

$$\mathcal{L}\{u(t) - u(t-k)\} = \int_0^k e^{-st}\,dt = \frac{1}{s}\left(1 - e^{-ks}\right).$$

The unit step function is also used in Eq. (15.163) and could be invoked in Exercise 15.10.13.

Exercises

15.10.1 Solve Eq. (15.154), which describes a damped simple harmonic oscillator for $X(0) = X_0$, $X'(0) = 0$, and

(a) $b^2 = 4$ km (critically damped),
(b) $b^2 > 4$ km (overdamped).

ANS. (a) $X(t) = X_0 e^{-(b/2m)t}\left(1 + \dfrac{b}{2m}t\right)$.

15.10.2 Solve Eq. (15.154), which describes a damped simple harmonic oscillator for $X(0) = 0$, $X'(0) = v_0$, and

(a) $b^2 < 4$ km (underdamped),
(b) $b^2 = 4$ km (critically damped),
(c) $b^2 > 4$ km (overdamped).

ANS. (a) $X(t) = \dfrac{v_0}{\omega_1}e^{-(b/2m)t}\sin\omega_1 t$,
(b) $X(t) = v_0 t e^{-(b/2m)t}$.

15.10.3 The motion of a body falling in a resisting medium may be described by

$$m\frac{d^2X(t)}{dt^2} = mg - b\frac{dX(t)}{dt}$$

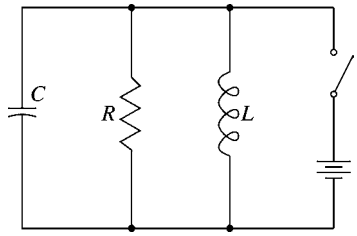

FIGURE 15.13 Ringing circuit.

when the retarding force is proportional to the velocity. Find $X(t)$ and $dX(t)/dt$ for the initial conditions

$$X(0) = \left.\frac{dX}{dt}\right|_{t=0} = 0.$$

15.10.4 **Ringing circuit**. In certain electronic circuits, resistance, inductance, and capacitance are placed in the plate circuit in parallel (Fig. 15.13). A constant voltage is maintained across the parallel elements, keeping the capacitor charged. At time $t = 0$ the circuit is disconnected from the voltage source. Find the voltages across the parallel elements R, L, and C as a function of time. Assume R to be large.
Hint. By Kirchhoff's laws

$$I_R + I_C + I_L = 0 \qquad \text{and} \qquad E_R = E_C = E_L,$$

where

$$E_R = I_R R, \qquad E_L = L\frac{dI_L}{dt}$$

$$E_C = \frac{q_0}{C} + \frac{1}{C}\int_0^t I_C\, dt,$$

$$q_0 = \text{initial charge of capacitor.}$$

With the DC impedance of $L = 0$, let $I_L(0) = I_0$, $E_L(0) = 0$. This means $q_0 = 0$.

15.10.5 With $J_0(t)$ expressed as a contour integral, apply the Laplace transform operation, reverse the order of integration, and thus show that

$$\mathcal{L}\{J_0(t)\} = (s^2+1)^{-1/2}, \qquad \text{for } s > 0.$$

15.10.6 Develop the Laplace transform of $J_n(t)$ from $\mathcal{L}\{J_0(t)\}$ by using the Bessel function recurrence relations.
Hint. Here is a chance to use mathematical induction.

15.10.7 A calculation of the magnetic field of a circular current loop in circular cylindrical coordinates leads to the integral

$$\int_0^\infty e^{-kz} k J_1(ka)\, dk, \qquad \Re(z) \geq 0.$$

Show that this integral is equal to $a/(z^2 + a^2)^{3/2}$.

15.10.8 The electrostatic potential of a point charge q at the origin in circular cylindrical coordinates is

$$\frac{q}{4\pi\varepsilon_0}\int_0^\infty e^{-kz}J_0(k\rho)\,dk = \frac{q}{4\pi\varepsilon_0}\cdot\frac{1}{(\rho^2+z^2)^{1/2}}, \qquad \Re(z)\geq 0.$$

From this relation show that the Fourier cosine and sine transforms of $J_0(k\rho)$ are

(a) $$\sqrt{\frac{\pi}{2}}F_c\{J_0(k\rho)\} = \int_0^\infty J_0(k\rho)\cos k\zeta\,dk = \begin{cases}(\rho^2-\zeta^2)^{-1/2}, & \rho>\zeta,\\ 0, & \rho<\zeta.\end{cases}$$

(b) $$\sqrt{\frac{\pi}{2}}F_s\{J_0(k\rho)\} = \int_0^\infty J_0(k\rho)\sin k\zeta\,dk = \begin{cases}0, & \rho>\zeta,\\ (\rho^2-\zeta^2)^{-1/2}, & \rho<\zeta.\end{cases}$$

Hint. Replace z by $z+i\zeta$ and take the limit as $z\to 0$.

15.10.9 Show that

$$\mathcal{L}\{I_0(at)\} = (s^2-a^2)^{-1/2}, \qquad s>a.$$

15.10.10 Verify the following Laplace transforms:

(a) $$\mathcal{L}\{j_0(at)\} = \mathcal{L}\left\{\frac{\sin at}{at}\right\} = \frac{1}{a}\cot^{-1}\left(\frac{s}{a}\right),$$

(b) $\mathcal{L}\{n_0(at)\}$ does not exist,

(c) $$\mathcal{L}\{i_0(at)\} = \mathcal{L}\left\{\frac{\sinh at}{at}\right\} = \frac{1}{2a}\ln\frac{s+a}{s-a} = \frac{1}{a}\coth^{-1}\left(\frac{s}{a}\right),$$

(d) $\mathcal{L}\{k_0(at)\}$ does not exist.

15.10.11 Develop a Laplace transform solution of Laguerre's equation

$$tF''(t)+(1-t)F'(t)+nF(t)=0.$$

Note that you need a derivative of a transform and a transform of derivatives. Go as far as you can with n; then (and only then) set $n=0$.

15.10.12 Show that the Laplace transform of the Laguerre polynomial $L_n(at)$ is given by

$$\mathcal{L}\{L_n(at)\} = \frac{(s-a)^n}{s^{n+1}}, \qquad s>0.$$

15.10.13 Show that

$$\mathcal{L}\{E_1(t)\} = \frac{1}{s}\ln(s+1), \qquad s>0,$$

where

$$E_1(t) = \int_t^\infty \frac{e^{-\tau}}{\tau}\,d\tau = \int_1^\infty \frac{e^{-xt}}{x}dx.$$

$E_1(t)$ is the exponential-integral function.

15.10.14 (a) From Eq. (15.189) show that

$$\int_0^\infty f(x)\,dx = \int_0^\infty \frac{F(t)}{t}dt,$$

provided the integrals exist.

(b) From the preceding result show that

$$\int_0^\infty \frac{\sin t}{t}dt = \frac{\pi}{2},$$

in agreement with Eqs. (15.122) and (7.56).

15.10.15 (a) Show that

$$\mathcal{L}\left\{\frac{\sin kt}{t}\right\} = \cot^{-1}\left(\frac{s}{k}\right).$$

(b) Using this result (with $k = 1$), prove that

$$\mathcal{L}\{\text{si}(t)\} = -\frac{1}{s}\tan^{-1} s,$$

where

$$\text{si}(t) = -\int_t^\infty \frac{\sin x}{x}dx, \qquad \text{the sine integral.}$$

15.10.16 If $F(t)$ is periodic (Fig. 15.14) with a period a so that $F(t + a) = F(t)$ for all $t \geq 0$, show that

$$\mathcal{L}\{F(t)\} = \frac{\int_0^a e^{-st}F(t)\,dt}{1 - e^{-as}},$$

with the integration now over only the **first period** of $F(t)$.

15.10.17 Find the Laplace transform of the square wave (period a) defined by

$$F(t) = \begin{cases} 1, & 0 < t < \frac{a}{2} \\ 0, & \frac{a}{2} < t < a. \end{cases}$$

ANS. $f(s) = \dfrac{1}{s} \cdot \dfrac{1 - e^{-as/2}}{1 - e^{-as}}$.

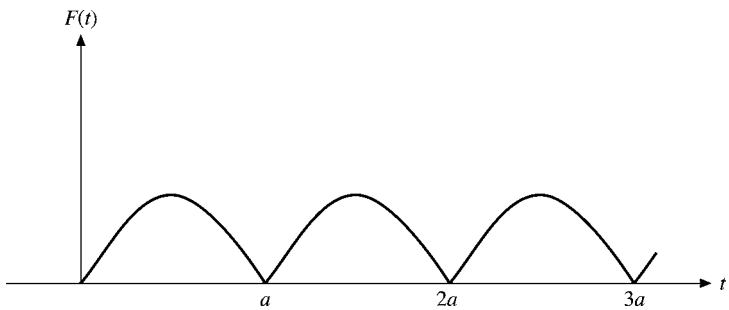

FIGURE 15.14 Periodic function.

15.10.18 Show that

(a) $\mathcal{L}\{\cosh at \cos at\} = \dfrac{s^3}{s^4 + 4a^4}$,

(b) $\mathcal{L}\{\cosh at \sin at\} = \dfrac{as^2 + 2a^3}{s^4 + 4a^4}$,

(c) $\mathcal{L}\{\sinh at \cos at\} = \dfrac{as^2 - 2a^3}{s^4 + 4a^4}$,

(d) $\mathcal{L}\{\sinh at \sin at\} = \dfrac{2a^2 s}{s^4 + 4a^4}$.

15.10.19 Show that

(a) $\mathcal{L}^{-1}\left\{\left(s^2 + a^2\right)^{-2}\right\} = \dfrac{1}{2a^3}\sin at - \dfrac{1}{2a^2} t \cos at$,

(b) $\mathcal{L}^{-1}\left\{s\left(s^2 + a^2\right)^{-2}\right\} = \dfrac{1}{2a} t \sin at$,

(c) $\mathcal{L}^{-1}\left\{s^2\left(s^2 + a^2\right)^{-2}\right\} = \dfrac{1}{2a}\sin at + \dfrac{1}{2} t \cos at$,

(d) $\mathcal{L}^{-1}\left\{s^3\left(s^2 + a^2\right)^{-2}\right\} = \cos at - \dfrac{a}{2} t \sin at$.

15.10.20 Show that

$$\mathcal{L}\left\{\left(t^2 - k^2\right)^{-1/2} u(t - k)\right\} = K_0(ks).$$

Hint. Try transforming an integral representation of $K_0(ks)$ into the Laplace transform integral.

15.10.21 The Laplace transform

$$\int_0^\infty e^{-xs} x J_0(x)\, dx = \frac{s}{(s^2 + 1)^{3/2}}$$

may be rewritten as

$$\frac{1}{s^2}\int_0^\infty e^{-y} y J_0\left(\frac{y}{s}\right) dy = \frac{s}{(s^2 + 1)^{3/2}},$$

which is in Gauss–Laguerre quadrature form. Evaluate this integral for $s = 1.0, 0.9, 0.8, \ldots$, decreasing s in steps of 0.1 until the relative error rises to 10 percent. (The effect of decreasing s is to make the integrand oscillate more rapidly per unit length of y, thus decreasing the accuracy of the numerical quadrature.)

15.10.22 (a) Evaluate

$$\int_0^\infty e^{-kz} k J_1(ka)\, dk$$

by the Gauss–Laguerre quadrature. Take $a = 1$ and $z = 0.1(0.1)1.0$.

(b) From the analytic form, Exercise 15.10.7, calculate the absolute error and the relative error.

15.11 CONVOLUTION (FALTUNGS) THEOREM

One of the most important properties of the Laplace transform is that given by the convolution, or Faltungs, theorem.[18] We take two transforms,

$$f_1(s) = \mathcal{L}\{F_1(t)\} \qquad \text{and} \qquad f_2(s) = \mathcal{L}\{F_2(t)\}, \tag{15.190}$$

and multiply them together. To avoid complications when changing variables, we hold the upper limits finite:

$$f_1(s)\,f_2(s) = \lim_{a\to\infty} \int_0^a e^{-sx} F_1(x)\,dx \int_0^{a-x} e^{-sy} F_2(y)\,dy. \tag{15.191}$$

The upper limits are chosen so that the area of integration, shown in Fig. 15.15a, is the shaded triangle, not the square. If we integrate over a square in the xy-plane, we have a parallelogram in the tz-plane, which simply adds complications. This modification is permissible because the two integrands are assumed to decrease exponentially. In the limit $a \to \infty$, the integral over the unshaded triangle will give zero contribution. Substituting $x = t - z$, $y = z$, the region of integration is mapped into the triangle shown in Fig. 15.15b. To verify the mapping, map the vertices: $t = x + y$, $z = y$. Using Jacobians to transform the element of area, we have

$$dx\,dy = \begin{vmatrix} \dfrac{\partial x}{\partial t} & \dfrac{\partial y}{\partial t} \\ \dfrac{\partial x}{\partial z} & \dfrac{\partial y}{\partial z} \end{vmatrix} dt\,dz = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} dt\,dz \tag{15.192}$$

or $dx\,dy = dt\,dz$. With this substitution Eq. (15.191) becomes

$$\begin{aligned} f_1(s)\,f_2(s) &= \lim_{a\to\infty} \int_0^t e^{-st} \int_0^t F_1(t-z) F_2(z)\,dz\,dt \\ &= \mathcal{L}\left\{ \int_0^t F_1(t-z) F_2(z)\,dz \right\}. \end{aligned} \tag{15.193}$$

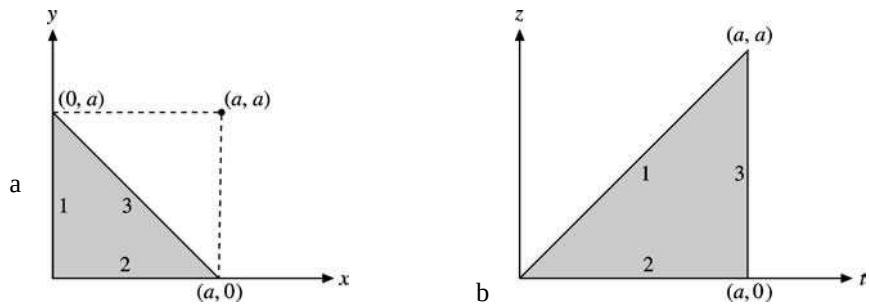

FIGURE 15.15
Change of variables,
(a) xy-plane (b) zt-plane.

[18] An alternate derivation employs the Bromwich integral (Section 15.12). This is Exercise 15.12.3.

For convenience this integral is represented by the symbol

$$\int_0^t F_1(t-z)F_2(z)\,dz \equiv F_1 * F_2 \tag{15.194}$$

and referred to as the **convolution**, closely analogous to the Fourier convolution (Section 15.5). If we substitute $w = t - z$, we find

$$F_1 * F_2 = F_2 * F_1, \tag{15.195}$$

showing that the relation is symmetric.

Carrying out the inverse transform, we also find

$$\mathcal{L}^{-1}\{f_1(s)f_2(s)\} = \int_0^t F_1(t-z)F_2(z)\,dz. \tag{15.196}$$

This can be useful in the development of new transforms or as an alternative to a partial fraction expansion. One immediate application is in the solution of integral equations (Section 16.2). Since the upper limit, t, is variable, this Laplace convolution is useful in treating Volterra integral equations. The Fourier convolution with fixed (infinite) limits would apply to Fredholm integral equations.

Example 15.11.1 DRIVEN OSCILLATOR WITH DAMPING

As one illustration of the use of the convolution theorem, let us return to the mass m on a spring, with damping and a driving force $F(t)$. The equation of motion ((15.129) or (15.154)) now becomes

$$mX''(t) + bX'(t) + kX(t) = F(t). \tag{15.197}$$

Initial conditions $X(0) = 0$, $X'(0) = 0$ are used to simplify this illustration, and the transformed equation is

$$ms^2x(s) + bs\,x(s) + kx(s) = f(s), \tag{15.198}$$

or

$$x(s) = \frac{f(s)}{m}\frac{1}{(s + b/2m)^2 + \omega_1^2}, \tag{15.199}$$

where $\omega_1^2 \equiv k/m - b^2/4m^2$, as before.

By the convolution theorem (Eq. (15.193) or (15.196)),

$$X(t) = \frac{1}{m\omega_1}\int_0^t F(t-z)e^{-(b/2m)z}\sin\omega_1 z\,dz. \tag{15.200}$$

If the force is impulsive, $F(t) = P\delta(t)$,[19]

$$X(t) = \frac{P}{m\omega_1}e^{-(b/2m)t}\sin\omega_1 t. \tag{15.201}$$

[19] Note that $\delta(t)$ lies **inside** the interval $[0, t]$.

P represents the momentum transferred by the impulse, and the constant P/m takes the place of an initial velocity $X'(0)$.

If $F(t) = F_0 \sin \omega t$, Eq. (15.200) may be used, but a partial fraction expansion is perhaps more convenient. With

$$f(s) = \frac{F_0 \omega}{s^2 + \omega^2}$$

Eq. (15.199) becomes

$$\begin{aligned} x(s) &= \frac{F_0 \omega}{m} \cdot \frac{1}{s^2 + \omega^2} \cdot \frac{1}{(s + b/2m)^2 + \omega_1^2} \\ &= \frac{F_0 \omega}{m} \left[\frac{a's + b'}{s^2 + \omega^2} + \frac{c's + d'}{(s + b/2m)^2 + \omega_1^2} \right]. \end{aligned} \tag{15.202}$$

The coefficients a', b', c', and d' are independent of s. Direct calculation shows

$$\begin{aligned} -\frac{1}{a'} &= \frac{b}{m}\omega^2 + \frac{m}{b}\left(\omega_0^2 - \omega^2\right)^2, \\ -\frac{1}{b'} &= -\frac{m}{b}\left(\omega_0^2 - \omega^2\right)\left[\frac{b}{m}\omega^2 + \frac{m}{b}\left(\omega_0^2 - \omega^2\right)^2 \right]. \end{aligned}$$

Since c' and d' will lead to exponentially decreasing terms (transients), they will be discarded here. Carrying out the inverse operation, we find for the steady-state solution

$$X(t) = \frac{F_0}{[b^2\omega^2 + m^2(\omega_0^2 - \omega^2)^2]^{1/2}} \sin(\omega t - \varphi), \tag{15.203}$$

where

$$\tan \varphi = \frac{b\omega}{m(\omega_0^2 - \omega^2)}.$$

Differentiating the denominator, we find that the amplitude has a maximum when

$$\omega^2 = \omega_0^2 - \frac{b^2}{2m^2} = \omega_1^2 - \frac{b^2}{4m^2}. \tag{15.204}$$

This is the resonance condition.[20] At resonance the amplitude becomes $F_0/b\omega_1$, showing that the mass m goes into infinite oscillation at resonance if damping is neglected ($b = 0$). It is worth noting that we have had three different characteristic frequencies:

$$\omega_2^2 = \omega_0^2 - \frac{b^2}{2m^2},$$

resonance for forced oscillations, with damping;

$$\omega_1^2 = \omega_0^2 - \frac{b^2}{4m^2},$$

[20] The amplitude (squared) has the typical resonance denominator, the Lorentz line shape, Exercise 15.3.9.

free oscillation frequency, with damping; and

$$\omega_0^2 = \frac{k}{m},$$

free oscillation frequency, no damping. They coincide only if the damping is zero. ■

Returning to Eqs. (15.197) and (15.199), Eq. (15.197) is our ODE for the response of a dynamical system to an arbitrary driving force. The final response clearly depends on both the driving force and the characteristics of our system. This dual dependence is separated in the transform space. In Eq. (15.199) the transform of the response (output) appears as the product of two factors, one describing the driving force (input) and the other describing the dynamical system. This latter part, which modifies the input and yields the output, is often called a **transfer function**. Specifically, $[(s+b/2m)^2+\omega_1^2]^{-1}$ is the transfer function corresponding to this damped oscillator. The concept of a transfer function is of great use in the field of servomechanisms. Often the characteristics of a particular servomechanism are described by giving its transfer function. The convolution theorem then yields the output signal for a particular input signal.

Exercises

15.11.1 From the convolution theorem show that

$$\frac{1}{s}f(s) = \mathcal{L}\left\{\int_0^t F(x)\,dx\right\},$$

where $f(s) = \mathcal{L}\{F(t)\}$.

15.11.2 If $F(t) = t^a$ and $G(t) = t^b,\ a > -1,\ b > -1$:

(a) Show that the convolution

$$F * G = t^{a+b+1}\int_0^1 y^a(1-y)^b\,dy.$$

(b) By using the convolution theorem, show that

$$\int_0^1 y^a(1-y)^b\,dy = \frac{a!b!}{(a+b+1)!}.$$

This is the Euler formula for the beta function (Eq. (8.59a)).

15.11.3 Using the convolution integral, calculate

$$\mathcal{L}^{-1}\left\{\frac{s}{(s^2+a^2)(s^2+b^2)}\right\}, \qquad a^2 \neq b^2.$$

15.11.4 An undamped oscillator is driven by a force $F_0 \sin\omega t$. Find the displacement as a function of time. Notice that it is a linear combination of two simple harmonic motions, one with the frequency of the driving force and one with the frequency ω_0 of the free oscillator. (Assume $X(0) = X'(0) = 0$.)

$$\text{ANS. } X(t) = \frac{F_0/m}{\omega^2-\omega_0^2}\left(\frac{\omega}{\omega_0}\sin\omega_0 t - \sin\omega t\right).$$

Other exercises involving the Laplace convolution appear in Section 16.2.

15.12 INVERSE LAPLACE TRANSFORM

Bromwich Integral

We now develop an expression for the inverse Laplace transform $\mathcal{L}^{-1}$ appearing in the equation

$$F(t) = \mathcal{L}^{-1}\{f(s)\}. \tag{15.205}$$

One approach lies in the Fourier transform, for which we know the inverse relation. There is a difficulty, however. Our Fourier transformable function had to satisfy the Dirichlet conditions. In particular, we required that

$$\lim_{\omega\to\infty} G(\omega) = 0 \tag{15.206}$$

so that the infinite integral would be well defined.[21] Now we wish to treat functions $F(t)$ that may diverge exponentially. To surmount this difficulty, we extract an exponential factor, $e^{\gamma t}$, from our (possibly) divergent Laplace function and write

$$F(t) = e^{\gamma t} G(t). \tag{15.207}$$

If $F(t)$ diverges as $e^{\alpha t}$, we require γ to be greater than α so **that** $G(t)$ **will be convergent**. Now, with $G(t) = 0$ for $t < 0$ and otherwise suitably restricted so that it may be represented by a Fourier integral (Eq. (15.20)),

$$G(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{iut}\,du \int_0^{\infty} G(v)e^{-iuv}\,dv. \tag{15.208}$$

Using Eq. (15.207), we may rewrite (15.208) as

$$F(t) = \frac{e^{\gamma t}}{2\pi}\int_{-\infty}^{\infty} e^{iut}\,du \int_0^{\infty} F(v)e^{-\gamma v}e^{-iuv}\,dv. \tag{15.209}$$

Now, with the change of variable,

$$s = \gamma + iu, \tag{15.210}$$

the integral over v is thrown into the form of a Laplace transform,

$$\int_0^{\infty} f(v)e^{-sv}\,dv = f(s); \tag{15.211}$$

[21] If delta functions are included, $G(\omega)$ may be a cosine. Although this does not satisfy Eq. (15.206), $G(\omega)$ is still bounded.

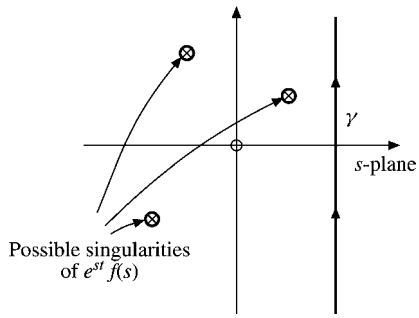

FIGURE 15.16 Singularities of $e^{st} f(s)$.

s is now a complex variable, and $\Re(s) \geq \gamma$ to guarantee convergence. Notice that the Laplace transform has mapped a function specified on the positive real axis onto the complex plane, $\Re(s) \geq \gamma$.[22]

With γ as a constant, $ds = i\,du$. Substituting Eq. (15.211) into Eq. (15.209), we obtain

$$F(t) = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} e^{st} f(s)\, ds. \tag{15.212}$$

Here is our **inverse transform**. We have rotated the line of integration through 90° (by using $ds = i\,du$). The path has become an infinite vertical line in the complex plane, the constant γ having been chosen so that all the singularities of $f(s)$ are on the left-hand side (Fig. 15.16).

Equation (15.212), our inverse transformation, is usually known as the **Bromwich integral**, although sometimes it is referred to as the **Fourier–Mellin theorem** or **Fourier–Mellin integral**. This integral may now be evaluated by the regular methods of contour integration (Chapter 7). If $t > 0$, the contour may be closed by an infinite semicircle in the left half-plane. Then by the residue theorem (Section 7.1)

$$F(t) = \Sigma(\text{residues included for } \Re(s) < \gamma). \tag{15.213}$$

Possibly this means of evaluation with $\Re(s)$ ranging through negative values seems paradoxical in view of our previous requirement that $\Re(s) \geq \gamma$. The paradox disappears when we recall that the requirement $\Re(s) \geq \gamma$ was imposed to guarantee convergence of the Laplace transform integral that defined $f(s)$. Once $f(s)$ is obtained, we may then proceed to exploit its properties as an analytical function in the complex plane wherever we choose.[23] In effect we are employing analytic continuation to get $\mathcal{L}\{F(t)\}$ in the left half-plane, exactly as the recurrence relation for the factorial function was used to extend the Euler integral definition (Eq. (8.5)) to the left half-plane.

Perhaps a pair of examples may clarify the evaluation of Eq. (15.212).

[22] For a derivation of the inverse Laplace transform using only real variables, see C. L. Bohn and R. W. Flynn, Real variable inversion of Laplace transforms: An application in plasma physics. *Am. J. Phys.* **46**: 1250 (1978).

[23] In numerical work $f(s)$ may well be available only for discrete real, positive values of s. Then numerical procedures are indicated. See Krylov and Skoblya in the Additional Reading.

Example 15.12.1 INVERSION VIA CALCULUS OF RESIDUES

If $f(s) = a/(s^2 - a^2)$, then

$$e^{st} f(s) = \frac{ae^{st}}{s^2 - a^2} = \frac{ae^{st}}{(s+a)(s-a)}. \tag{15.214}$$

The residues may be found by using Exercise 6.6.1 or various other means. The first step is to identify the singularities, the poles. Here we have one simple pole at $s = a$ and another simple pole at $s = -a$. By Exercise 6.6.1, the residue at $s = a$ is $(\frac{1}{2})e^{at}$ and the residue at $s = -a$ is $(-\frac{1}{2})e^{-at}$. Then

$$\text{Residues} = \left(\tfrac{1}{2}\right)\left(e^{at} - e^{-at}\right) = \sinh at = F(t), \tag{15.215}$$

in agreement with Eq. (15.105). ■

Example 15.12.2

If

$$f(s) = \frac{1 - e^{-as}}{s},$$

then $e^{s(t-a)}$ grows exponentially for $t < a$ on the semicircle in the left-hand s-plane, so contour integration and the residue theorem are not applicable. However, we can evaluate the integral explicitly as follows. We let $\gamma \to 0$ and substitute $s = iy$, so

$$F(t) = \frac{1}{2\pi i}\int_{\gamma - i\infty}^{\gamma + i\infty} e^{st} f(s) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[e^{iyt} - e^{iy(t-a)}\right]\frac{dy}{y}. \tag{15.216}$$

Using the Euler identity, only the sines survive that are odd in y and we obtain

$$F(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\left[\frac{\sin ty}{y} - \frac{\sin(t-a)y}{y}\right]. \tag{15.217}$$

If $k > 0$, then $\int_0^\infty \frac{\sin ky}{y} dy$ gives $\pi/2$, and it gives $-\pi/2$ if $k < 0$. As a consequence, $F(t) = 0$ if $t > a > 0$ and if $t < 0$. If $0 < t < a$, then $F(t) = 1$. This can be written compactly in terms of the Heaviside unit step function $u(t)$ as follows:

$$F(t) = u(t) - u(t-a) = \begin{cases} 0, & t < 0, \\ 1, & 0 < t < a, \\ 0, & t > a, \end{cases} \tag{15.218}$$

a step function of unit height and length a (Fig. 15.17). ■

Two general comments may be in order. First, these two examples hardly begin to show the usefulness and power of the Bromwich integral. It is always available for inverting a complicated transform when the tables prove inadequate.

Second, this derivation is not presented as a rigorous one. Rather, it is given more as a plausibility argument, although it can be made rigorous. The determination of the inverse transform is somewhat similar to the solution of a differential equation. It makes

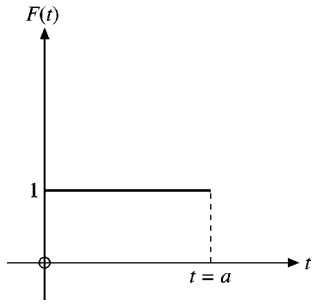

FIGURE 15.17
Finite-length step function
$u(t) - u(t - a)$.

little difference how you get the solution. Guess at it if you want. The solution can always be checked by substitution back into the original differential equation. Similarly, $F(t)$ can (and, to check on careless errors, should) be checked by determining whether, by Eq. (15.99),

$$\mathcal{L}\{F(t)\} = f(s).$$

Two alternate derivations of the Bromwich integral are the subjects of Exercises 15.12.1 and 15.12.2.

As a final illustration of the use of the Laplace inverse transform, we have some results from the work of Brillouin and Sommerfeld (1914) in electromagnetic theory.

Example 15.12.3 VELOCITY OF ELECTROMAGNETIC WAVES IN A DISPERSIVE MEDIUM

The group velocity u of traveling waves is related to the phase velocity v by the equation

$$u = v - \lambda \frac{dv}{d\lambda}. \tag{15.219}$$

Here λ is the wavelength. In the vicinity of an absorption line (resonance), $dv/d\lambda$ may be sufficiently negative so that $u > c$ (Fig. 15.18). The question immediately arises whether a signal can be transmitted faster than c, the velocity of light in vacuum. This question, which assumes that such a group velocity is meaningful, is of fundamental importance to the theory of special relativity.

We need a solution to the wave equation

$$\frac{\partial^2 \psi}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 \psi}{\partial t^2}, \tag{15.220}$$

corresponding to a harmonic vibration starting at the origin at time zero. Since our medium is dispersive, v is a function of the angular frequency. Imagine, for instance, a plane wave, angular frequency ω, incident on a shutter at the origin. At $t = 0$ the shutter is (instantaneously) opened, and the wave is permitted to advance along the positive x-axis.

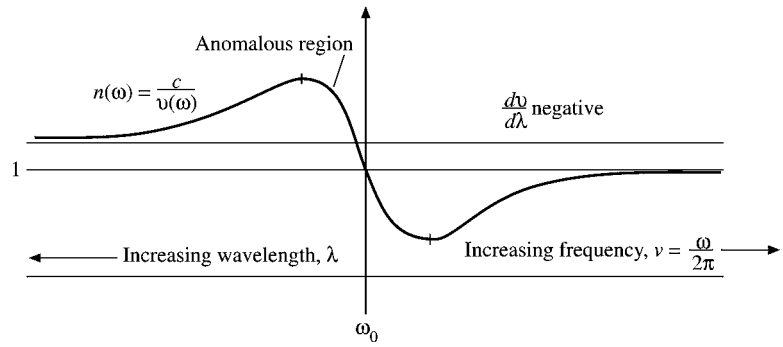

FIGURE 15.18 Optical dispersion.

Let us then build up a solution starting at $x = 0$. It is convenient to use the Cauchy integral formula, Eq. (6.43),

$$\psi(0,t) = \frac{1}{2\pi i}\oint \frac{e^{-izt}}{z - z_0}dz = e^{-iz_0t}$$

(for a contour encircling $z = z_0$ in the positive sense). Using $s = -iz$ and $z_0 = \omega$, we obtain

$$\psi(0,t) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \frac{e^{st}}{s + i\omega}ds = \begin{cases} 0, & t < 0, \\ e^{-i\omega t}, & t > 0. \end{cases} \tag{15.221}$$

To be complete, the loop integral is along the vertical line $\Re(s) = \gamma$ **and** an infinite semicircle, as shown in Fig. 15.19. The location of the infinite semicircle is chosen so that the integral over it vanishes. This means a semicircle in the left half-plane for $t > 0$ and the residue is enclosed. For $t < 0$ we pick the right half-plane and no singularity is enclosed. The fact that this is just the Bromwich integral may be verified by noting that

$$F(t) = \begin{cases} 0, & t < 0, \\ e^{-i\omega t}, & t > 0 \end{cases} \tag{15.222}$$

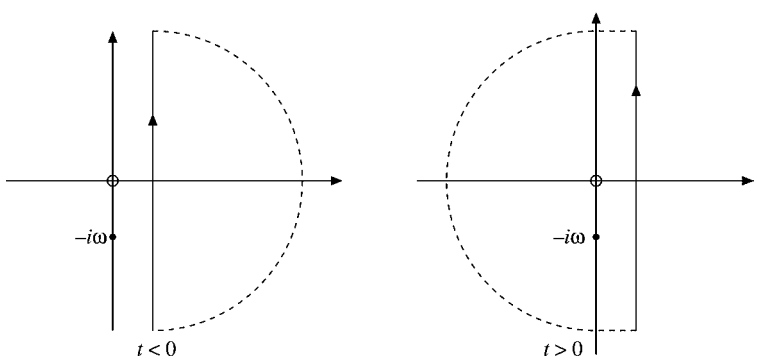

FIGURE 15.19 Possible closed contours.

and applying the Laplace transform. The transformed function $f(s)$ becomes

$$f(s) = \frac{1}{s + i\omega}. \tag{15.223}$$

Our Cauchy–Bromwich integral provides us with the time dependence of a signal leaving the origin at $t = 0$. To include the space dependence, we note that

$$e^{s(t-x/v)}$$

satisfies the wave equation. With this as a clue, we replace t by $t - x/v$ and write a solution:

$$\psi(x,t) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \frac{e^{s(t-x/v)}}{s + i\omega} ds. \tag{15.224}$$

It was seen in the derivation of the Bromwich integral that our variable s replaces the ω of the Fourier transformation. Hence the wave velocity v may become a function of s, that is, $v(s)$. Its particular form need not concern us here. We need only the property $v \le c$ and

$$\lim_{|s|\to\infty} v(s) = \text{constant}, \ c. \tag{15.225}$$

This is suggested by the asymptotic behavior of the curve on the right side of Fig. 15.18.[24]

Evaluating Eq. (15.225) by the calculus of residues, we may close the path of integration by a semicircle in the right half-plane, provided

$$t - \frac{x}{c} < 0.$$

Hence

$$\psi(x,t) = 0, \qquad t - \frac{x}{c} < 0, \tag{15.226}$$

which means that the velocity of our signal cannot exceed the velocity of light in the vacuum, c. This simple but very significant result was extended by Sommerfeld and Brillouin to show just how the wave advanced in the dispersive medium. ■

Summary — Inversion of Laplace Transform

- Direct use of tables, Table 15.2, and references; use of partial fractions (Section 15.8) and the operational theorems of Table 15.1.
- Bromwich integral, Eq. (15.212), and the calculus of residues.
- Numerical inversion, see the Additional Readings.

[24] Equation (15.225) follows rigorously from the theory of anomalous dispersion. See also the Kronig–Kramers optical dispersion relations of Section 7.2.

Table 15.1 Laplace Transform Operations

	Operations	Equation
1. Laplace transform	$f(s) = \mathcal{L}\{F(t)\} = \int_0^\infty e^{-st} F(t)\, dt$	(15.99)
2. Transform of derivative	$sf(s) - F(+0) = \mathcal{L}\{F'(t)\}$	(15.123)
	$s^2 f(s) - sF(+0) - F'(+0) = \mathcal{L}\{F''(t)\}$	(15.124)
3. Transform of integral	$\frac{1}{s} f(s) = \mathcal{L}\left\{\int_0^t F(x)\, dx\right\}$	(Exercise 15.11.1)
4. Substitution	$f(s-a) = \mathcal{L}\{e^{at} F(t)\}$	(15.152)
5. Translation	$e^{-bs} f(s) = \mathcal{L}\{F(t-b)\}$	(15.164)
6. Derivative of transform	$f^{(n)}(s) = \mathcal{L}\{(-t)^n F(t)\}$	(15.173)
7. Integral of transform	$\int_s^\infty f(x)\, dx = \mathcal{L}\left\{\frac{F(t)}{t}\right\}$	(15.189)
8. Convolution	$f_1(s) f_2(s) = \mathcal{L}\left\{\int_0^t F_1(t-z) F_2(z)\, dz\right\}$	(15.193)
9. Inverse transform, Bromwich integral	$\frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} e^{st} f(s)\, ds = F(t)$	(15.212)

Exercises

15.12.1 Derive the Bromwich integral from Cauchy's integral formula.
Hint. Apply the inverse transform $\mathcal{L}^{-1}$ to

$$f(s) = \frac{1}{2\pi i} \lim_{\alpha \to \infty} \int_{\gamma - i\alpha}^{\gamma + i\alpha} \frac{f(z)}{s - z} dz,$$

where $f(z)$ is analytic for $\Re(z) \geq \gamma$.

15.12.2 Starting with

$$\frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} e^{st} f(s)\, ds,$$

show that by introducing

$$f(s) = \int_0^\infty e^{-sz} F(z)\, dz,$$

we can convert one integral into the Fourier representation of a Dirac delta function. From this derive the inverse Laplace transform.

15.12.3 Derive the Laplace transformation convolution theorem by use of the Bromwich integral.

15.12.4 Find

$$\mathcal{L}^{-1}\left\{\frac{s}{s^2 - k^2}\right\}$$

(a) by a partial fraction expansion.
(b) Repeat, using the Bromwich integral.

Table 15.2 Laplace Transforms

$f(s)$	$F(t)$	Limitation	Equation
1. 1	$\delta(t)$	Singularity at $+0$	(15.141)
2. $\frac{1}{s}$	1	$s>0$	(15.102)
3. $\frac{n!}{s^{n+1}}$	t^n	$s>0$ $n>-1$	(15.108)
4. $\frac{1}{s-k}$	e^{kt}	$s>k$	(15.103)
5. $\frac{1}{(s-k)^2}$	te^{kt}	$s>k$	(15.175)
6. $\frac{s}{s^2-k^2}$	$\cosh kt$	$s>k$	(15.105)
7. $\frac{k}{s^2-k^2}$	$\sinh kt$	$s>k$	(15.105)
8. $\frac{s}{s^2+k^2}$	$\cos kt$	$s>0$	(15.107)
9. $\frac{k}{s^2+k^2}$	$\sin kt$	$s>0$	(15.107)
10. $\frac{s-a}{(s-a)^2+k^2}$	$e^{at}\cos kt$	$s>a$	(15.153)
11. $\frac{k}{(s-a)^2+k^2}$	$e^{at}\sin kt$	$s>a$	(15.153)
12. $\frac{s^2-k^2}{(s^2+k^2)^2}$	$t\cos kt$	$s>0$	(Exercise 15.10.19)
13. $\frac{2ks}{(s^2+k^2)^2}$	$t\sin kt$	$s>0$	(Exercise 15.10.19)
14. $(s^2+a^2)^{-1/2}$	$J_0(at)$	$s>0$	(15.185)
15. $(s^2-a^2)^{-1/2}$	$I_0(at)$	$s>a$	(Exercise 15.10.9)
16. $\frac{1}{a}\cot^{-1}\left(\frac{s}{a}\right)$	$j_0(at)$	$s>0$	(Exercise 15.10.10)
17. $\frac{1}{2a}\ln\frac{s+a}{s-a}$ $\frac{1}{a}\coth^{-1}\left(\frac{s}{a}\right)$	$i_0(at)$	$s>a$	(Exercise 15.10.10)
18. $\frac{(s-a)^n}{s^{n+1}}$	$L_n(at)$	$s>0$	(Exercise 15.10.12)
19. $\frac{1}{s}\ln(s+1)$	$E_1(x)=-Ei(-x)$	$s>0$	(Exercise 15.10.13)
20. $\frac{\ln s}{s}$	$-\ln t-\gamma$	$s>0$	(Exercise 15.12.9)

A more extensive table of Laplace transforms appears in Chapter 29 of AMS-55 (see footnote 4 in Chapter 5 for the reference).

15.12.5 Find

$$\mathcal{L}^{-1}\left\{\frac{k^2}{s(s^2+k^2)}\right\}$$

(a) by using a partial fraction expansion.
(b) Repeat using the convolution theorem.
(c) Repeat using the Bromwich integral.

ANS. $F(t) = 1 - \cos kt$.

15.12.6 Use the Bromwich integral to find the function whose transform is $f(s) = s^{-1/2}$. Note that $f(s)$ has a branch point at $s = 0$. The negative x-axis may be taken as a cut line.

ANS. $F(t) = (\pi t)^{-1/2}$.

15.12.7 Show that

$$\mathcal{L}^{-1}\left\{\left(s^2+1\right)^{-1/2}\right\} = J_0(t)$$

by evaluation of the Bromwich integral.

Hint. Convert your Bromwich integral into an integral representation of $J_0(t)$. Figure 15.20 shows a possible contour.

15.12.8 Evaluate the inverse Laplace transform

$$\mathcal{L}^{-1}\left\{\left(s^2-a^2\right)^{-1/2}\right\}$$

by each of the following methods:

(a) Expansion in a series and term-by-term inversion.
(b) Direct evaluation of the Bromwich integral.
(c) Change of variable in the Bromwich integral: $s = (a/2)(z + z^{-1})$.

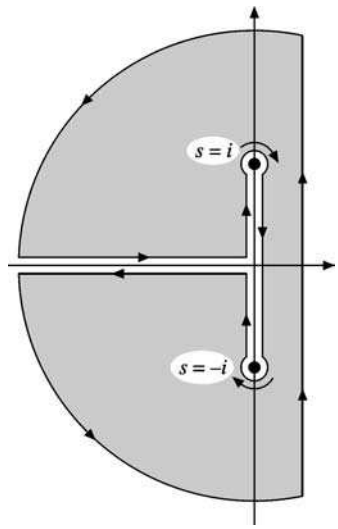

FIGURE 15.20 A possible contour for the inversion of $J_0(t)$.

15.12.9 Show that

$$\mathcal{L}^{-1}\left\{\frac{\ln s}{s}\right\} = -\ln t - \gamma,$$

where $\gamma = 0.5772\ldots$, the Euler–Mascheroni constant.

15.12.10 Evaluate the Bromwich integral for

$$f(s) = \frac{s}{(s^2 + a^2)^2}.$$

15.12.11 **Heaviside expansion theorem**. If the transform $f(s)$ may be written as a ratio

$$f(s) = \frac{g(s)}{h(s)},$$

where $g(s)$ and $h(s)$ are analytic functions, $h(s)$ having simple, isolated zeros at $s = s_i$, show that

$$F(t) = \mathcal{L}^{-1}\left\{\frac{g(s)}{h(s)}\right\} = \sum_i \frac{g(s_i)}{h'(s_i)} e^{s_i t}.$$

Hint. See Exercise 6.6.2.

15.12.12 Using the Bromwich integral, invert $f(s) = s^{-2}e^{-ks}$. Express $F(t) = \mathcal{L}^{-1}\{f(s)\}$ in terms of the (shifted) unit step function $u(t-k)$.

ANS. $F(t) = (t-k)u(t-k)$.

15.12.13 You have a Laplace transform:

$$f(s) = \frac{1}{(s+a)(s+b)}, \qquad a \neq b.$$

Invert this transform by each of three methods:

(a) Partial fractions and use of tables.
(b) Convolution theorem.
(c) Bromwich integral.

ANS. $F(t) = \dfrac{e^{-bt} - e^{-at}}{a-b}$, $a \neq b$.

Additional Readings

Champeney, D. C., *Fourier Transforms and Their Physical Applications*. New York: Academic Press (1973). Fourier transforms are developed in a careful, easy-to-follow manner. Approximately 60% of the book is devoted to applications of interest in physics and engineering.

Erdelyi, A.,W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Tables of Integral Transforms*, 2 vols. New York: McGraw–Hill (1954). This text contains extensive tables of Fourier sine, cosine, and exponential transforms, Laplace and inverse Laplace transforms, Mellin and inverse Mellin transforms, Hankel transforms, and other, more specialized integral transforms.

Hanna, J. R., *Fourier Series and Integrals of Boundary Value Problems*. Somerset, NJ: Wiley (1990). This book is a broad treatment of the Fourier solution of boundary value problems. The concepts of convergence and completeness are given careful attention.

Jeffreys, H., and B. S. Jeffreys, *Methods of Mathematical Physics*, 3rd ed. Cambridge, UK: Cambridge University Press (1972).

Krylov, V. I., and N. S. Skoblya, *Handbook of Numerical Inversion of Laplace* Transform. Jerusalem: Israel Program for Scientific Translations (1969).

Lepage, W. R., *Complex Variables and the Laplace Transform for Engineers*. New York: McGraw-Hill (1961); New York: Dover (1980). A complex variable analysis that is carefully developed and then applied to Fourier and Laplace transforms. It is written to be read by students, but intended for the serious student.

McCollum, P. A., and B. F. Brown, *Laplace Transform Tables and Theorems*. New York: Holt, Rinehart and Winston (1965).

Miles, J. W., *Integral Transforms in Applied Mathematics*. Cambridge, UK: Cambridge University Press (1971). This is a brief but interesting and useful treatment for the advanced undergraduate. It emphasizes applications rather than abstract mathematical theory.

Papoulis, A., *The Fourier Integral and Its Applications*. New York: McGraw-Hill (1962). This is a rigorous development of Fourier and Laplace transforms and has extensive applications in science and engineering.

Roberts, G. E., and H. Kaufman, *Table of Laplace Transforms*. Philadelphia: Saunders (1966).

Sneddon, I. N., *Fourier Transforms*. New York: McGraw-Hill (1951), reprinted, Dover (1995). A detailed comprehensive treatment, this book is loaded with applications to a wide variety of fields of modern and classical physics.

Sneddon, I. H., *The Use of Integral Transforms*. New York: McGraw-Hill (1972). Written for students in science and engineering in terms they can understand, this book covers all the integral transforms mentioned in this chapter as well as in several others. Many applications are included.

Van der Pol, B., and H. Bremmer, *Operational Calculus Based on the Two-sided Laplace Integral*, 3rd ed. Cambridge, UK: Cambridge University Press (1987). Here is a development based on the integral range $-\infty$ to $+\infty$, rather than the useful 0 to ∞. Chapter V contains a detailed study of the Dirac delta function (impulse function).

Wolf, K. B., *Integral Transforms in Science and Engineering*. New York: Plenum Press (1979). This book is a very comprehensive treatment of integral transforms and their applications.

CHAPTER 16

INTEGRAL EQUATIONS

16.1 INTRODUCTION

With the exception of the integral transforms of the last chapter, we have been considering relations between the unknown function $\varphi(x)$ and one or more of its derivatives. We now proceed to investigate equations containing the unknown function within an integral. As with differential equations, we shall confine our attention to linear relations, linear integral equations. Integral equations are classified in two ways:

- If the **limits of integration are fixed**, we call the equation a **Fredholm** equation; if **one limit is variable**, it is a **Volterra** equation.
- If the **unknown function** appears **only under the integral** sign, we label it **first kind**. If it appears both **inside and outside** the integral, it is labeled **second kind**.

Definitions

Symbolically, we have a **Fredholm equation of the first kind**,

$$f(x) = \int_a^b K(x,t)\varphi(t)\,dt; \tag{16.1}$$

the **Fredholm equation of the second kind**, with λ being the eigenvalue,

$$\varphi(x) = f(x) + \lambda \int_a^b K(x,t)\varphi(t)\,dt; \tag{16.2}$$

the **Volterra equation of the first kind**,

$$f(x) = \int_a^x K(x,t)\varphi(t)\,dt; \tag{16.3}$$

and the **Volterra equation of the second kind**,

$$\varphi(x) = f(x) + \int_a^x K(x,t)\varphi(t)\,dt. \tag{16.4}$$

In all four cases $\varphi(t)$ is the unknown function. $K(x,t)$, which we call the **kernel**, and $f(x)$ are assumed to be known. When $f(x) = 0$, the equation is said to be **homogeneous**.

Why do we bother about integral equations? After all, the differential equations have done a rather good job of describing our physical world so far. There are several reasons for introducing integral equations here.

We have placed considerable emphasis on the solution of differential equations **subject to particular boundary conditions**. For instance, the boundary condition at $r = 0$ determines whether the Neumann function $N_n(r)$ is present when Bessel's equation is solved. The boundary condition for $r \to \infty$ determines whether the $I_n(r)$ is present in our solution of the modified Bessel equation. The integral equation relates the unknown function not only to its values at neighboring points (derivatives) but also to its values throughout a region, including the boundary. In a very real sense the boundary conditions are built into the integral equation rather than imposed at the final stage of the solution. It can be seen in Section 10.5, where kernels are constructed, that the form of the kernel depends on the values on the boundary. The integral equation, then, is compact and may turn out to be a more convenient or powerful form than the differential equation. Mathematical problems such as existence, uniqueness, and completeness may often be handled more easily and elegantly in integral form. Finally, whether or not we like it, there are some problems, such as some diffusion and transport phenomena, that cannot be represented by differential equations. If we wish to solve such problems, we are forced to handle integral equations. Finally, an integral equation may also appear as a matter of deliberate choice based on convenience or the need for the mathematical power of an integral equation formulation.

Example 16.1.1 Momentum Representation in Quantum Mechanics

The Schrödinger equation (in ordinary space representation) is

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r}) + V(\mathbf{r})\psi(\mathbf{r}) = E\psi(\mathbf{r}), \tag{16.5}$$

or

$$\left(\nabla^2 + a^2\right)\psi(\mathbf{r}) = v(\mathbf{r})\psi(\mathbf{r}), \tag{16.6}$$

where

$$a^2 = \frac{2m}{\hbar^2}E, \qquad v(\mathbf{r}) = \frac{2m}{\hbar^2}V(\mathbf{r}). \tag{16.7}$$

If we generalize Eq. (16.6) to

$$\left(\nabla^2 + a^2\right)\psi(\mathbf{r}) = \int v(\mathbf{r},\mathbf{r}')\psi(\mathbf{r}')\,d^3r', \tag{16.8}$$

then, for the special case of

$$v(\mathbf{r},\mathbf{r}') = v(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}'), \tag{16.9}$$

a local interaction, Eq. (16.8) reduces to Eq. (16.6). Consider the Fourier transform pair ψ and Ψ (compare footnote 9 in Section 15.6):

$$\Psi(\mathbf{k}) = \frac{1}{(2\pi)^{3/2}} \int \psi(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}}\, d^3r, \qquad \psi(\mathbf{r}) = \frac{1}{(2\pi)^{3/2}} \int \Psi(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}}\, d^3k, \tag{16.10}$$

with the abbreviation $\mathbf{p}$ for momentum so that

$$\frac{\mathbf{p}}{\hbar} = \mathbf{k} \qquad \text{(wave number)}. \tag{16.11}$$

Multiplying Eq. (16.8) by the plane-wave $e^{-i\mathbf{k}\cdot\mathbf{r}}$, we obtain

$$\int e^{-i\mathbf{k}\cdot\mathbf{r}} \big(\nabla^2 + a^2\big)\psi(\mathbf{r})\, d^3r = \int d^3r\, e^{-i\mathbf{k}\cdot\mathbf{r}} \int v(\mathbf{r}, \mathbf{r}')\psi(\mathbf{r}')\, d^3r'. \tag{16.12}$$

Note that the ∇^2 on the left operates only on the $\psi(\mathbf{r})$. Integrating the left-hand side by parts and substituting Eq. (16.10) for $\psi(\mathbf{r}')$ on the right, we get

$$\begin{aligned} \int \big(-k^2 + a^2\big)\psi(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} d^3r &= (2\pi)^{3/2}\big(-k^2 + a^2\big)\Psi(\mathbf{k}) \\ &= \frac{1}{(2\pi)^{3/2}} \iiint v(\mathbf{r}, \mathbf{r}')\Psi(\mathbf{k}') e^{-i(\mathbf{k}\cdot\mathbf{r} - \mathbf{k}'\cdot\mathbf{r}')}\, d^3r'\, d^3r\, d^3k'. \end{aligned} \tag{16.13}$$

If we use

$$f(\mathbf{k}, \mathbf{k}') = \frac{1}{(2\pi)^{3/2}} \iint v(\mathbf{r}, \mathbf{r}') e^{-i(\mathbf{k}\cdot\mathbf{r} - \mathbf{k}'\cdot\mathbf{r}')} d^3r'\, d^3r, \tag{16.14}$$

Eq. (16.13) becomes

$$\big(-k^2 + a^2\big)\Psi(\mathbf{k}) = \int f(\mathbf{k}, \mathbf{k}')\Psi(\mathbf{k}')\, d^3k', \tag{16.15}$$

a Fredholm equation of the second kind in which the parameter a^2 corresponds to the eigenvalue.

For our special but important case of local interaction, application of Eq. (16.9) leads to

$$f(\mathbf{k}, \mathbf{k}') = f(\mathbf{k} - \mathbf{k}'). \tag{16.16}$$

This is our momentum representation, equivalent to an ordinary static interaction potential in coordinate space. Our momentum wave function $\Psi(\mathbf{k})$ satisfies the integral equation Eq. (16.15). It must be emphasized that all through here we have assumed that the required Fourier integrals exist. For a harmonic oscillator potential, $V(\mathbf{r}) = r^2$, the required integrals would not exist. Equation (16.10) would lead to divergent oscillations and we would have no Eq. (16.15). ■

Transformation of a Differential Equation into an Integral Equation

Often we find that we have a choice. The physical problem may be represented by a differential or an integral equation. Let us assume that we have the differential equation and wish to transform it into an integral equation. Starting with a **linear** second-order ODE

$$y'' + A(x)y' + B(x)y = g(x) \tag{16.17}$$

with initial conditions

$$y(a) = y_0, \qquad y'(a) = y_0',$$

we integrate to obtain

$$y'(x) = -\int_a^x A(t)y'(t)\,dt - \int_a^x B(t)y(t)\,dt + \int_a^x g(t)\,dt + y_0'. \tag{16.18}$$

Integrating the first integral on the right by parts yields

$$y'(x) = -Ay(x) - \int_a^x (B - A')y(t)\,dt + \int_a^x g(t)\,dt + A(a)y_0 + y_0'. \tag{16.19}$$

Notice how the initial conditions are being absorbed into our new version. Integrating a second time, we obtain

$$\begin{aligned} y(x) = &-\int_a^x Ay\,dx - \int_a^x du \int_a^u \big[B(t) - A'(t)\big]y(t)\,dt \\ &+ \int_a^x du \int_a^u g(t)\,dt + \big[A(a)y_0 + y_0'\big](x - a) + y_0. \end{aligned} \tag{16.20}$$

To transform this equation into a neater form, we use the relation

$$\int_a^x du \int_a^u f(t)\,dt = \int_a^x (x - t)f(t)\,dt. \tag{16.21}$$

This may be verified by differentiating both sides. Since the derivatives are equal, the original expressions can differ only by a constant. Letting $x \to a$, the constant vanishes and Eq. (16.21) is established. Applying it to Eq. (16.20), we obtain

$$\begin{aligned} y(x) = &-\int_a^x \big\{A(t) + (x - t)\big[B(t) - A'(t)\big]\big\}y(t)\,dt \\ &+ \int_a^x (x - t)g(t)\,dt + \big[A(a)y_0 + y_0'\big](x - a) + y_0. \end{aligned} \tag{16.22}$$

If we now introduce the abbreviations

$$\begin{aligned} K(x,t) &= (t - x)\big[B(t) - A'(t)\big] - A(t), \\ f(x) &= \int_a^x (x - t)g(t)\,dt + \big[A(a)y_0 + y_0'\big](x - a) + y_0, \end{aligned} \tag{16.23}$$

Eq. (16.22) becomes

$$y(x) = f(x) + \int_a^x K(x,t)y(t)\,dt, \tag{16.24}$$

which is a Volterra equation of the second kind. This reformulation as a Volterra integral equation offers certain advantages in investigating questions of existence and uniqueness.

Example 16.1.2 LINEAR OSCILLATOR EQUATION

As an illustration, consider the linear oscillator equation

$$y'' + \omega^2 y = 0 \tag{16.25}$$

with

$$y(0) = 0, \qquad y'(0) = 1.$$

This yields (compare with Eq. (16.17))

$$A(x) = 0, \qquad B(x) = \omega^2, \qquad g(x) = 0.$$

Substituting into Eq. (16.22) (or Eqs. (16.23) and (16.24)), we find that the integral equation becomes

$$y(x) = x + \omega^2 \int_0^x (t - x)y(t)\,dt. \tag{16.26}$$

- This integral equation, Eq. (16.26), is equivalent to the original differential equation plus the initial conditions.

A check shows that each form is indeed satisfied by $y(x) = (1/\omega)\sin\omega x$. ■

Let us reconsider the linear oscillator equation (16.25) but now with the boundary conditions

$$y(0) = 0, \qquad y(b) = 0.$$

Since $y'(0)$ is not given, we must modify the procedure. The first integration gives

$$y' = -\omega^2 \int_0^x y\,dx + y'(0). \tag{16.27}$$

Integrating a second time and again using Eq. (16.21), we have

$$y = -\omega^2 \int_0^x (x - t)y(t)\,dt + y'(0)x. \tag{16.28}$$

To eliminate the unknown $y'(0)$, we now impose the condition $y(b) = 0$. This gives

$$\omega^2 \int_0^b (b - t)y(t)\,dt = by'(0). \tag{16.29}$$

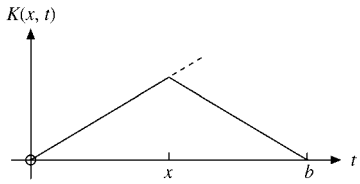

FIGURE 16.1

Substituting this back into Eq. (16.28), we obtain

$$y(x) = -\omega^2 \int_0^x (x-t)y(t)\,dt + \omega^2 \frac{x}{b}\int_0^b (b-t)y(t)\,dt. \tag{16.30}$$

Now let us break the interval $[0, b]$ into two intervals, $[0, x]$ and $[x, b]$. Since

$$\frac{x}{b}(b-t) - (x-t) = \frac{t}{b}(b-x), \tag{16.31}$$

we find

$$y(x) = \omega^2 \int_0^x \frac{t}{b}(b-x)y(t)\,dt + \omega^2 \int_x^b \frac{x}{b}(b-t)y(t)\,dt. \tag{16.32}$$

Finally, if we define a kernel (Fig. 16.1)

$$K(x,t) = \begin{cases} \dfrac{t}{b}(b-x), & t < x, \\ \dfrac{x}{b}(b-t), & t > x, \end{cases} \tag{16.33}$$

we have

$$y(x) = \omega^2 \int_0^b K(x,t)y(t)\,dt, \tag{16.34}$$

a homogeneous Fredholm equation of the second kind.

Our new kernel, $K(x,t)$, has some interesting properties.

1. It is symmetric, $K(x,t) = K(t,x)$.
2. It is continuous, in the sense that
$$\frac{t}{b}(b-x)\Big|_{t=x} = \frac{x}{b}(b-t)\Big|_{t=x}.$$
3. Its derivative with respect to t is **discontinuous**. As t increases through the point $t = x$, there is a discontinuity of -1 in $\partial K(x,t)/\partial t$.

According to these properties in Section 9.7 we identify $K(x,t)$ as a Green's function.

1. In the transformation of a linear, second-order ODE into an integral equation, the initial or boundary conditions play a decisive role. If we have **initial** conditions (only one end of our interval), the differential equation transforms into a Volterra integral equation. For the case of the linear oscillator equation with **boundary** conditions (both ends

of our interval), the differential equation leads to a Fredholm integral equation with a kernel that will be a Green's function.

2. Note that the reverse transformation (integral equation to differential equation) is not always possible. There exist integral equations for which no corresponding differential equation is known.

Exercises

16.1.1 Starting with the ODE, integrate twice and derive the Volterra integral equation corresponding to

(a) $y''(x) - y(x) = 0; \quad y(0) = 0, \quad y'(0) = 1.$

$$\text{ANS. } y = \int_0^x (x-t)y(t)\,dt + x.$$

(b) $y''(x) - y(x) = 0; \quad y(0) = 1, \quad y'(0) = -1.$

$$\text{ANS. } y = \int_0^x (x-t)y(t)\,dt - x + 1.$$

Check your results with Eq. (16.23).

16.1.2 Derive a Fredholm integral equation corresponding to

$$y''(x) - y(x) = 0, \qquad y(1) = 1, \qquad y(-1) = 1,$$

(a) by integrating twice,
(b) by forming the Green's function.

$$\text{ANS. } y(x) = 1 - \int_{-1}^{1} K(x,t)y(t)\,dt,$$

$$K(x,t) = \begin{cases} \frac{1}{2}(1-x)(t+1), & x > t, \\ \frac{1}{2}(1-t)(x+1), & x < t. \end{cases}$$

16.1.3 (a) Starting with the given answers of Exercise 16.1.1, differentiate and recover the original ODEs **and the boundary conditions**.
(b) Repeat for Exercise 16.1.2.

16.1.4 The general second-order linear ODE with constant coefficients is

$$y''(x) + a_1 y'(x) + a_2 y(x) = 0.$$

Given the boundary conditions

$$y(0) = y(1) = 0,$$

integrate twice and develop the integral equation

$$y(x) = \int_0^1 K(x,t)y(t)\,dt,$$

with

$$K(x,t) = \begin{cases} a_2 t(1-x) + a_1(x-1), & t < x, \\ a_2 x(1-t) + a_1 x, & x < t. \end{cases}$$

Note that $K(x,t)$ is symmetric and continuous if $a_1 = 0$. How is this related to self-adjointness of the ODE?

16.1.5 Verify that $\int_a^x \int_a^x f(t)\,dt\,dx = \int_a^x (x-t) f(t)\,dt$ for all $f(t)$ (for which the integrals exist).

16.1.6 Given $\varphi(x) = x - \int_0^x (t-x)\varphi(t)\,dt$, solve this integral equation by converting it to an ODE (plus boundary conditions) and solving the ODE (by inspection).

16.1.7 Show that the homogeneous Volterra equation of the second kind

$$\psi(x) = \lambda \int_0^x K(x,t)\psi(t)\,dt$$

has no solution (apart from the trivial $\psi = 0$).
Hint. Develop a Maclaurin expansion of $\psi(x)$. Assume $\psi(x)$ and $K(x,t)$ are differentiable with respect to x as needed.

16.2 Integral Transforms, Generating Functions

Analogous to differentiation, linear ODEs are solved in Chapter 9. Analogous to integration, there is no general method available for solving integral equations. However, certain special cases may be treated with our **integral transforms** (Chapter 15). For convenience these are listed here. If

$$\psi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ixt} \varphi(t)\,dt,$$

then

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ixt} \psi(t)\,dt \qquad \text{(Fourier)}. \tag{16.35}$$

If

$$\psi(x) = \int_0^{\infty} e^{-xt} \varphi(t)\,dt,$$

then

$$\varphi(x) = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} e^{xt} \psi(t)\,dt \qquad \text{(Laplace)}. \tag{16.36}$$

If

$$\psi(x) = \int_0^{\infty} t^{x-1} \varphi(t)\,dt,$$

then

$$\varphi(x) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} x^{-t}\psi(t)\,dt \qquad \text{(Mellin)}. \tag{16.37}$$

If

$$\psi(x) = \int_0^\infty t\varphi(t)J_\nu(xt)\,dt,$$

then

$$\varphi(x) = \int_0^\infty t\psi(t)J_\nu(xt)\,dt \qquad \text{(Hankel)}. \tag{16.38}$$

Actually the usefulness of the integral transform technique extends a bit beyond these four rather specialized forms.

Example 16.2.1 FOURIER TRANSFORM SOLUTION

Let us consider a Fredholm equation of the first kind with a kernel of the general type $k(x-t)$,

$$f(x) = \int_{-\infty}^\infty k(x-t)\varphi(t)\,dt, \tag{16.39}$$

in which $\varphi(t)$ is our unknown function. **Assuming that the needed transforms exist**, we apply the Fourier convolution theorem (Section 15.5) to obtain

$$f(x) = \int_{-\infty}^\infty K(\omega)\Phi(\omega)e^{-i\omega x}\,d\omega. \tag{16.40}$$

The functions $K(\omega)$, $\Phi(\omega)$, and $F(\omega)$ are the Fourier transforms of $k(x)$, $\varphi(x)$, and $f(x)$, respectively. Taking the Fourier transform of both sides of Eq. (16.40), by Eq. (16.35) we have

$$K(\omega)\Phi(\omega) = \frac{1}{2\pi}\int_{-\infty}^\infty f(x)e^{i\omega x}\,dx = \frac{F(\omega)}{\sqrt{2\pi}}. \tag{16.41}$$

Then

$$\Phi(\omega) = \frac{1}{\sqrt{2\pi}}\cdot\frac{F(\omega)}{K(\omega)}, \tag{16.42}$$

and, using the inverse Fourier transform, we have

$$\varphi(x) = \frac{1}{2\pi}\int_{-\infty}^\infty \frac{F(\omega)}{K(\omega)}e^{-i\omega x}\,d\omega. \tag{16.43}$$

For a rigorous justification of this result one can follow Morse and Feshbach (see the Additional Readings) (1953) across complex planes. An extension of this transformation solution appears as Exercise 16.2.1. ■

Example 16.2.2 GENERALIZED ABEL EQUATION, CONVOLUTION THEOREM

The generalized Abel equation is

$$f(x) = \int_0^x \frac{\varphi(t)}{(x-t)^\alpha} dt, \qquad 0 < \alpha < 1, \qquad \text{with } \begin{cases} f(x) \text{ known,} \\ \varphi(t) \text{ unknown.} \end{cases} \tag{16.44}$$

Taking the Laplace transform of both sides of this equation, we obtain

$$\mathcal{L}\{f(x)\} = \mathcal{L}\left\{\int_0^x \frac{\varphi(t)}{(x-t)^\alpha} dt\right\} = \mathcal{L}\{x^{-\alpha}\}\mathcal{L}\{\varphi(x)\}, \tag{16.45}$$

the last step following by the Laplace convolution theorem (Section 15.11). Then

$$\mathcal{L}\{\varphi(x)\} = \frac{s^{1-\alpha}\mathcal{L}\{f(x)\}}{(-\alpha)!}. \tag{16.46}$$

Dividing by s,[1] we obtain

$$\frac{1}{s}\mathcal{L}\{\varphi(x)\} = \frac{s^{-\alpha}\mathcal{L}\{f(x)\}}{(-\alpha)!} = \frac{\mathcal{L}\{x^{\alpha-1}\}\mathcal{L}\{f(x)\}}{(\alpha-1)!(-\alpha)!}. \tag{16.47}$$

Combining the factorials (Eq. (8.32)) and applying the Laplace convolution theorem again, we discover that

$$\frac{1}{s}\mathcal{L}\{\varphi(x)\} = \frac{\sin\pi\alpha}{\pi}\mathcal{L}\left\{\int_0^x \frac{f(t)}{(x-t)^{1-\alpha}} dt\right\}. \tag{16.48}$$

Inverting with the aid of Exercise 15.11.1, we get

$$\int_0^x \varphi(t)\, dt = \frac{\sin\pi\alpha}{\pi} \int_0^x \frac{f(t)}{(x-t)^{1-\alpha}} dt, \tag{16.49}$$

and finally, by differentiating,

$$\varphi(x) = \frac{\sin\pi\alpha}{\pi} \frac{d}{dx} \int_0^x \frac{f(t)}{(x-t)^{1-\alpha}} dt. \tag{16.50}$$

■

Generating Functions

Occasionally, the reader may encounter integral equations that involve generating functions. Suppose we have the admittedly special case

$$f(x) = \int_{-1}^{1} \frac{\varphi(t)}{(1-2xt+x^2)^{1/2}} dt, \qquad -1 \le x \le 1. \tag{16.51}$$

We notice two important features:

1. $(1-2xt+x^2)^{-1/2}$ generates the Legendre polynomials.
2. $[-1, 1]$ is the orthogonality interval for the Legendre polynomials.

[1] $s^{1-\alpha}$ does not have an inverse for $0 < \alpha < 1$.

If we now expand the denominator (property 1) and assume that our unknown $\varphi(t)$ may be written as a series of these same Legendre polynomials,

$$f(x) = \int_{-1}^{1} \sum_{n=0}^{\infty} a_n P_n(t) \sum_{r=0}^{\infty} P_r(t) x^r \, dt. \tag{16.52}$$

Utilizing the orthogonality of the Legendre polynomials (property 2), we obtain

$$f(x) = \sum_{r=0}^{\infty} \frac{2a_r}{2r+1} x^r. \tag{16.53}$$

We may identify the a_n by differentiating n times and then setting $x = 0$:

$$f^{(n)}(0) = n! \frac{2}{2n+1} a_n. \tag{16.54}$$

Hence

$$\varphi(t) = \sum_{n=0}^{\infty} \frac{2n+1}{2} \frac{f^{(n)}(0)}{n!} P_n(t). \tag{16.55}$$

Similar results may be obtained with the other generating functions (compare Exercise 7.1.6).

- This technique of expanding in a series of special functions is always available. It is worth a try whenever the expansion is possible (and convenient) and the interval is appropriate.

Exercises

16.2.1 The kernel of a Fredholm equation of the second kind,

$$\varphi(x) = f(x) + \lambda \int_{-\infty}^{\infty} K(x,t)\varphi(t)\, dt,$$

is of the form $k(x-t)$.[2] Assuming that the required transforms exist, show that

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{F(t)e^{-ixt}\, dt}{1 - \sqrt{2\pi}\lambda K(t)}.$$

$F(t)$ and $K(t)$ are the Fourier transforms of $f(x)$ and $k(x)$, respectively.

16.2.2 The kernel of a Volterra equation of the first kind,

$$f(x) = \int_{0}^{x} K(x,t)\varphi(t)\, dt,$$

[2] This kernel and a range $0 \le x < \infty$ are the characteristics of integral equations of the Wiener–Hopf type. Details will be found in Chapter 8 of Morse and Feshbach (1953); see the Additional Readings.

has the form $k(x-t)$. Assuming that the required transforms exist, show that

$$\varphi(x) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \frac{F(s)}{K(s)} e^{xs}\, ds.$$

$F(s)$ and $K(s)$ are the Laplace transforms of $f(x)$ and $k(x)$, respectively.

16.2.3 The kernel of a Volterra equation of the second kind,

$$\varphi(x) = f(x) + \lambda \int_0^x K(x,t)\varphi(t)\, dt,$$

has the form $k(x-t)$. Assuming that the required transforms exist, show that

$$\varphi(x) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \frac{F(s)}{1-\lambda K(s)} e^{xs}\, ds.$$

16.2.4 Using the Laplace transform solution (Exercise 16.2.3), solve

(a) $\varphi(x) = x + \displaystyle\int_0^x (t-x)\varphi(t)\, dt.$

ANS. $\varphi(x) = \sin x.$

(b) $\varphi(x) = x - \displaystyle\int_0^x (t-x)\varphi(t)\, dt.$

ANS. $\varphi(x) = \sinh x.$

Check your results by substituting back into the original integral equations.

16.2.5 Reformulate the equations of Example 16.2.1 (Eqs. (16.39) to (16.43)), using Fourier cosine transforms.

16.2.6 Given the Fredholm integral equation,

$$e^{-x^2} = \int_{-\infty}^{\infty} e^{-(x-t)^2}\varphi(t)\, dt,$$

apply the Fourier convolution technique of Example 16.2.1 to solve for $\varphi(t)$.

16.2.7 Solve Abel's equation,

$$f(x) = \int_0^x \frac{\varphi(t)}{(x-t)^\alpha} dt, \qquad 0 < \alpha < 1,$$

by the following method:

(a) Multiply both sides by $(z-x)^{\alpha-1}$ and integrate with respect to x over the range $0 \le x \le z$.
(b) Reverse the order of integration and evaluate the integral on the right-hand side (with respect to x) by the beta function.

Note.

$$\int_t^z \frac{dx}{(z-x)^{1-\alpha}(x-t)^{\alpha}} = B(1-\alpha,\alpha) = (-\alpha)!(\alpha-1)! = \frac{\pi}{\sin\pi\alpha}.$$

16.2.8 Given the generalized Abel equation with $f(x) = 1$,

$$1 = \int_0^x \frac{\varphi(t)}{(x-t)^{\alpha}} dt, \qquad 0 < \alpha < 1,$$

solve for $\varphi(t)$ and verify that $\varphi(t)$ is a solution of the given equation.

ANS. $\varphi(t) = \dfrac{\sin\pi\alpha}{\pi} t^{\alpha-1}$.

16.2.9 A Fredholm equation of the first kind has a kernel $e^{-(x-t)^2}$:

$$f(x) = \int_{-\infty}^{\infty} e^{-(x-t)^2} \varphi(t)\, dt.$$

Show that the solution is

$$\varphi(x) = \frac{1}{\sqrt{\pi}} \sum_{\pi=0}^{\infty} \frac{f^{(n)}(0)}{2^n n!} H_n(x),$$

in which $H_n(x)$ is an nth-order Hermite polynomial.

16.2.10 Solve the integral equation

$$f(x) = \int_{-1}^{1} \frac{\varphi(t)}{(1-2xt+x^2)^{1/2}} dt, \qquad -1 \le x \le 1,$$

for the unknown function $\varphi(t)$ if
(a) $f(x) = x^{2s}$, (b) $f(x) = x^{2s+1}$.

ANS. (a) $\varphi(t) = \dfrac{4s+1}{2} P_{2s}(t)$, (b) $\varphi(t) = \dfrac{4s+3}{2} P_{2s+1}(t)$.

16.2.11 A Kirchhoff diffraction theory analysis of a laser leads to the integral equation

$$v(\mathbf{r}_2) = \gamma \iint K(\mathbf{r}_1, \mathbf{r}_2) v(\mathbf{r}_1)\, dA.$$

The unknown, $v(\mathbf{r}_1)$, gives the geometric distribution of the radiation field over one mirror surface; the range of integration is over the surface of that mirror. For square confocal spherical mirrors the integral equation becomes

$$v(x_2, y_2) = \frac{-i\gamma e^{ikb}}{\lambda b} \int_{-a}^{a}\int_{-a}^{a} e^{-(ik/b)(x_1x_2+y_1y_2)} v(x_1, y_1)\, dx_1\, dy_1,$$

in which b is the centerline distance between the laser mirrors. This can be put in a somewhat simpler form by the substitutions

$$\frac{kx_i^2}{b} = \xi_i^2, \qquad \frac{ky_i^2}{b} = \eta_i^2, \qquad \text{and} \qquad \frac{ka^2}{b} = \frac{2\pi a^2}{\lambda b} = \alpha^2.$$

(a) Show that the variables separate and we get two integral equations.

(b) Show that the new limits, $\pm\alpha$, may be approximated by $\pm\infty$ for a mirror dimension $a \gg \lambda$.
(c) Solve the resulting integral equations.

16.3 Neumann Series, Separable (Degenerate) Kernels

Many and probably most integral equations cannot be solved by the specialized integral transform techniques of the preceding section. Here we develop three rather general techniques for solving integral equations. The first, due largely to Neumann, Liouville, and Volterra, develops the unknown function $\varphi(x)$ as a power series in λ, where λ is a given constant. The method is applicable whenever the series converges.

The second method is somewhat restricted because it requires that the two variables appearing in the kernel $K(x,t)$ be separable. However, there are two major rewards: (1) The relation between an integral equation and a set of simultaneous linear algebraic equations is shown explicitly, and (2) the method leads to eigenvalues and eigenfunctions—in close analogy to Section 3.5.

Third, a technique for numerical solution of Fredholm equations of both the first and second kind is outlined. The problem posed by ill-conditioned matrices is emphasized.

Neumann Series

We solve a linear integral equation of the second kind by successive approximations; our integral equation is the Fredholm equation,

$$\varphi(x) = f(x) + \lambda \int_a^b K(x,t)\varphi(t)\,dt, \tag{16.56}$$

in which $f(x) \neq 0$. If the upper limit of the integral is a variable (Volterra equation), the following development will still hold, but with minor modifications. Let us try (there is no **guarantee** that it will work) to approximate our unknown function by

$$\varphi(x) \approx \varphi_0(x) = f(x). \tag{16.57}$$

This choice is not mandatory. If you can make a better guess, go ahead and guess. The choice here is equivalent to saying that the integral or the constant λ is small. To improve this first crude approximation, we feed $\varphi_0(x)$ back into the integral, Eq. (16.56), and get

$$\varphi_1(x) = f(x) + \lambda \int_a^b K(x,t) f(t)\,dt. \tag{16.58}$$

Repeating this process of substituting the new $\varphi_n(x)$ back into Eq. (16.56), we develop the sequence

$$\varphi_2(x) = f(x) + \lambda \int_a^b K(x,t_1) f(t_1)\,dt_1 + \lambda^2 \int_a^b \int_a^b K(x,t_1)K(t_1,t_2) f(t_2)\,dt_2\,dt_1 \tag{16.59}$$

and

$$\varphi_n(x) = \sum_{i=0}^{n} \lambda^i u_i(x), \tag{16.60}$$

where

$$\begin{aligned} u_0(x) &= f(x), \\ u_1(x) &= \int_a^b K(x,t_1) f(t_1)\, dt_1, \\ u_2(x) &= \int_a^b \int_a^b K(x,t_1) K(t_1,t_2) f(t_2)\, dt_2\, dt_1, \\ u_n(x) &= \iint \cdots \int K(x,t_1) K(t_1,t_2) \cdots K(t_{n-1},t_n) \cdot f(t_n)\, dt_n \cdots dt_1. \end{aligned} \tag{16.61}$$

We expect that our solution $\varphi(x)$ will be

$$\varphi(x) = \lim_{n\to\infty} \varphi_n(x) = \lim_{n\to\infty} \sum_{i=0}^{n} \lambda^i u_i(x), \tag{16.62}$$

provided that our infinite series converges. We may conveniently check the convergence by the Cauchy ratio test, Section 5.2, noting that

$$|\lambda^n u_n(x)| \le |\lambda|^n \cdot |f|_{\max} \cdot |K|_{\max}^n \cdot |b-a|^n, \tag{16.63}$$

using $|f|_{\max}$ to represent the **maximum** value of $|f(x)|$ in the interval $[a,b]$ and $|K|_{\max}$ to represent the maximum value of $|K(x,t)|$ in its domain in the x, t-plane. We have convergence if

$$|\lambda| \cdot |K|_{\max} \cdot |b-a| < 1. \tag{16.64}$$

Note that $\lambda|u_n(\max)|$ is being used as a **comparison** series. If it converges, our actual series must converge. If this condition is not satisfied, we may or may not have convergence. A more sensitive test is required. Of course, even if the Neumann series diverges, there still may be a solution obtainable by another method.

To see what has been done with this iterative manipulation, we may find it helpful to rewrite the Neumann series solution, Eq. (16.59), in operator form. We start by rewriting Eq. (16.56) as

$$\varphi = \lambda K \varphi + f,$$

where K represents the **integral operator** $\int_a^b K(x,t)[\ ]\, dt$. Solving for φ, we obtain

$$\varphi = (1 - \lambda K)^{-1} f.$$

Binomial expansion leads to Eq. (16.59). The **convergence** of the Neumann series is a demonstration that the inverse operator $(1-\lambda K)^{-1}$ exists.

Example 16.3.1 NEUMANN SERIES SOLUTION

To illustrate the Neumann method, we consider the integral equation

$$\varphi(x) = x + \frac{1}{2}\int_{-1}^{1}(t-x)\varphi(t)\,dt. \tag{16.65}$$

To start the Neumann series, we take

$$\varphi_0(x) = x. \tag{16.66}$$

Then

$$\varphi_1(x) = x + \frac{1}{2}\int_{-1}^{1}(t-x)t\,dt = x + \frac{1}{2}\left(\frac{1}{3}t^3 - \frac{1}{2}t^2x\right)\Big|_{-1}^{1} = x + \frac{1}{3}.$$

Substituting $\varphi_1(x)$ back into Eq. (16.65), we get

$$\varphi_2(x) = x + \frac{1}{2}\int_{-1}^{1}(t-x)t\,dt + \frac{1}{2}\int_{-1}^{1}(t-x)\frac{1}{3}dt = x + \frac{1}{3} - \frac{x}{3}.$$

Continuing this process of substituting back into Eq. (16.65), we obtain

$$\varphi_3(x) = x + \frac{1}{3} - \frac{x}{3} - \frac{1}{3^2},$$

and by induction

$$\varphi_{2n}(x) = x + \sum_{s=1}^{n}(-1)^{s-1}3^{-s} - x\sum_{s=1}^{n}(-1)^{s-1}3^{-s}. \tag{16.67}$$

Letting $n \to \infty$, we get

$$\varphi(x) = \frac{3}{4}x + \frac{1}{4}. \tag{16.68}$$

This solution can (and should) be checked by substituting back into the original equation, Eq. (16.65). ■

It is interesting to note that our series converged easily even though Eq. (16.64) is **not** satisfied in this particular case. Actually Eq. (16.64) is a rather crude upper bound on λ. It can be shown that a necessary and sufficient condition for the convergence of our series solution is that $|\lambda| < |\lambda_e|$, where λ_e is the eigenvalue of smallest magnitude of the corresponding homogeneous equation [$f(x) = 0$)]. For this particular example $\lambda_e = \sqrt{3}/2$. Clearly, $\lambda = \frac{1}{2} < \lambda_e = \sqrt{3}/2$.

One approach to the calculation of time-dependent perturbations in quantum mechanics starts with the integral equation for the evolution operator

$$U(t,t_0) = 1 - \frac{i}{\hbar}\int_{t_0}^{t} V(t_1)U(t_1,t_0)\,dt_1. \tag{16.69a}$$

Iteration leads to

$$U(t,t_0) = 1 - \frac{i}{\hbar}\int_{t_0}^{t} V(t_1)\,dt_1 + \left(\frac{i}{\hbar}\right)^2\int_{t_0}^{t}\int_{t_0}^{t_1} V(t_1)V(t_2)\,dt_2\,dt_1 + \cdots. \tag{16.69b}$$

The evolution operator is obtained as a series of multiple integrals of the perturbing potential $V(t)$, closely analogous to the Neumann series, Eq. (16.60). For $V = V_0$, independent of t, the evolution operator becomes (see Exercise 3.4.13, replace $t \to \Delta t$, and construct U from products of $T(t + \Delta t, t)$ as in Eq. (4.26))

$$U(t_1, t_0) = \exp\left[-\frac{i}{\hbar}(t - t_0)V_0\right].$$

A second and similar relationship between the Neumann series and quantum mechanics appears when the Schrödinger wave equation for scattering is reformulated as an integral equation. The first term in a Neumann series solution is the incident (unperturbed) wave. The second term is the first-order Born approximation, Eq. (9.203b) of Section 9.7.

The Neumann method may also be applied to Volterra integral equations of the second kind, Eq. (16.4) or Eq. (16.56) with the fixed upper limit, b, replaced by a variable, x. In the Volterra case the Neumann series converges for all λ as long as the kernel is square integrable.

Separable Kernel

The technique of replacing our integral equation by simultaneous algebraic equations may also be used whenever our kernel $K(x, t)$ is separable, in the sense that

$$K(x, t) = \sum_{j=1}^{n} M_j(x) N_j(t), \tag{16.70}$$

where n, the upper limit of the sum, is **finite**. Such kernels are sometimes called **degenerate**. Our class of separable kernels includes all polynomials and many of the elementary transcendental functions; that is,

$$\cos(t - x) = \cos t \cos x + \sin t \sin x. \tag{16.70a}$$

If Eq. (16.70) is satisfied, substitution into the Fredholm equation of the second kind, Eq. (16.2), yields

$$\varphi(x) = f(x) + \lambda \sum_{j=1}^{n} M_j(x) \int_a^b N_j(t)\varphi(t)\, dt, \tag{16.71}$$

interchanging integration and summation. Now, the integral with respect to t is a constant,

$$\int_a^b N_j(t)\varphi(t)\, dt = c_j. \tag{16.72}$$

Hence Eq. (16.71) becomes

$$\varphi(x) = f(x) + \lambda \sum_{j=1}^{n} c_j M_j(x). \tag{16.73}$$

This gives us $\varphi(x)$, our solution, once the constants c_i have been determined. Equation (16.73) further tells us the form of $\varphi(x)$: $f(x)$, plus a linear combination of the x-dependent factors of the separable kernel.

We may find c_i by multiplying Eq. (16.73) by $N_i(x)$ and integrating to eliminate the x-dependence. Use of Eq. (16.72) yields

$$c_i = b_i + \lambda \sum_{j=1}^{n} a_{ij} c_j, \tag{16.74}$$

where

$$b_i = \int_a^b N_i(x) f(x)\, dx, \qquad a_{ij} = \int_a^b N_i(x) M_j(x)\, dx. \tag{16.75}$$

It is perhaps helpful to write Eq. (16.74) in matrix form, with $\mathsf{A} = (a_{ij})$:

$$\mathbf{b} = \mathbf{c} - \lambda \mathsf{A}\mathbf{c} = (1 - \lambda \mathsf{A})\mathbf{c}, \tag{16.76a}$$

or[3]

$$\mathbf{c} = (1 - \lambda \mathsf{A})^{-1}\mathbf{b}. \tag{16.76b}$$

Equation (16.76a) is equivalent to a set of simultaneous linear algebraic equations

$$\begin{aligned} (1 - \lambda a_{11})c_1 - \lambda a_{12}c_2 - \lambda a_{13}c_3 - \cdots &= b_1, \\ -\lambda a_{21}c_1 + (1 - \lambda a_{22})c_2 - \lambda a_{23}c_3 - \cdots &= b_2, \\ -\lambda a_{31}c_1 - \lambda a_{32}c_2 + (1 - \lambda a_{33})c_3 - \cdots &= b_3, \qquad \text{and so on.} \end{aligned} \tag{16.77}$$

If our integral equation is homogeneous, $[f(x) = 0]$, then $\mathbf{b} = 0$. To get a solution, we set the determinant of the coefficients of c_i equal to zero,

$$|1 - \lambda \mathsf{A}| = 0, \tag{16.78}$$

exactly as in Section 3.5. The roots of Eq. (16.78) yield our eigenvalues. Substituting into $(1 - \lambda \mathsf{A})\mathbf{c} = 0$, we find the c_i, and then Eq. (16.73) gives our solution.

Example 16.3.2

To illustrate this technique for determining eigenvalues and eigenfunctions of the homogeneous Fredholm equation, we consider the case

$$\varphi(x) = \lambda \int_{-1}^{1} (t + x)\varphi(t)\, dt. \tag{16.79}$$

Here (compare with Eqs. (16.71) and (16.77))

$$\begin{aligned} M_1 &= 1, \qquad M_2(x) = x, \\ N_1(t) &= t, \qquad N_2 = 1. \end{aligned}$$

Equation (16.75) yields

$$a_{11} = a_{22} = 0, \qquad a_{12} = \tfrac{2}{3}, \qquad a_{21} = 2; \qquad b_1 = 0 = b_2.$$

[3]Notice the similarity to the operator form of the Neumann series.

Equation (16.78), our secular equation, becomes

$$\begin{vmatrix} 1 & -\dfrac{2\lambda}{3} \\ -2\lambda & 1 \end{vmatrix} = 0. \tag{16.80}$$

Expanding, we obtain

$$1 - \frac{4\lambda^2}{3} = 0, \qquad \lambda = \pm\frac{\sqrt{3}}{2}. \tag{16.81}$$

Substituting the eigenvalues $\lambda = \pm\sqrt{3}/2$ into Eq. (16.76), we have

$$c_1 \mp \frac{c_2}{\sqrt{3}} = 0. \tag{16.82}$$

Finally, with a choice of $c_1 = 1$, Eq. (16.73) gives

$$\varphi_1(x) = \frac{\sqrt{3}}{2}(1 + \sqrt{3}x), \qquad \lambda = \frac{\sqrt{3}}{2}, \tag{16.83}$$

$$\varphi_2(x) = -\frac{\sqrt{3}}{2}(1 - \sqrt{3}x), \qquad \lambda = -\frac{\sqrt{3}}{2}. \tag{16.84}$$

Since our equation is homogeneous, the normalization of $\varphi(x)$ is arbitrary. ■

If the kernel is not separable in the sense of Eq. (16.70), there is still the possibility that it may be approximated by a kernel that is separable. Then we can get the exact solution of an approximate equation, an equation that approximates the original equation. The solution of the separable approximate kernel problem can then be checked by substituting back into the original, unseparable kernel problem.

Numerical Solution

There is extensive literature on the numerical solution of integral equations, and much of it concerns special techniques for certain situations. One method of fair generality is the replacement of the single integral equation by a set of simultaneous algebraic equations. And again matrix techniques are invoked. This simultaneous algebraic equation–matrix approach is applied here to two different cases. For the homogeneous Fredholm equation of the second kind this method works well. For the Fredholm equation of the first kind the method is a disaster. First we deal with the disaster.

We consider the Fredholm integral equation of the first kind,

$$f(x) = \int_a^b K(x,t)\varphi(t)\,dt, \tag{16.84a}$$

with $f(x)$ and $K(x,t)$ known and $\varphi(t)$ unknown. The integral can be evaluated (in principle) by quadrature techniques. For maximum accuracy the Gaussian method is recommended (if the kernel is continuous and has continuous derivatives). The numerical quadrature replaces the integral by a summation,

$$f(x_i) = \sum_{k=1}^{n} A_k K(x_i, t_k)\varphi(t_k), \tag{16.84b}$$

with A_k the quadrature coefficients. We abbreviate $f(x_i)$ as f_i, $\varphi(t_k)$ as φ_k, and $A_k K(x_i, t_k)$ as B_{ik}. In effect we are changing from a function description to a vector–matrix description, with the n components of the vector (f_i) defined as the values of the function at the n discrete points $[f(x_i)]$. Equation (16.84b) becomes

$$f_i = \sum_{k=1}^{n} B_{ik}\varphi_k,$$

a matrix equation. Inverting (B_{ik}), we obtain

$$\varphi(x_k) = \varphi_k = \sum_{k=1}^{n} B_{ki}^{-1} f_i, \tag{16.84c}$$

and Eq. (16.84a) is solved — in principle. In practice, the quadrature coefficient–kernel matrix is often "ill-conditioned" (with respect to inversion). This means that in the inversion process small (numerical) errors are multiplied by large factors. In the inversion process all significant figures may be lost and Eq. (16.84c) becomes numerical nonsense.

This disaster should not be entirely unexpected. Integration is essentially a smoothing operation. $f(x)$ is relatively insensitive to local variation of $\varphi(t)$. Conversely, $\varphi(t)$ may be exceedingly sensitive to small changes in $f(x)$. Small errors in $f(x)$ or in B^{-1} are magnified and accuracy disappears. This same behavior shows up in attempts to invert Laplace transforms numerically.

When the quadrature–matrix technique is applied to the integral equation eigenvalue problem, the symmetric kernel, homogeneous Fredholm equation of the second kind,[4]

$$\lambda\varphi(x) = \int_a^b K(x,t)\varphi(t)\,dt, \tag{16.84d}$$

the technique is far more successful. Replacing the integral by a set of simultaneous algebraic equations (numerical quadrature), we have

$$\lambda\varphi_i = \sum_{k=1}^{n} A_k K_{ik}\varphi_k, \tag{16.84e}$$

with $\varphi_i = \varphi(x_i)$, as before. The points x_i, $i = 1, 2, \ldots, n$, are taken to be the same (numerically) as t_k, $k = 1, 2, \ldots, n$, so K_{ik} will be symmetric. The system is symmetrized by multiplying by $A_i^{1/2}$ so that

$$\lambda\left(A_i^{1/2}\varphi_i\right) = \sum_{k=1}^{n}\left(A_i^{1/2}K_{ik}A_k^{1/2}\right)\left(A_k^{1/2}\varphi_k\right). \tag{16.84f}$$

Replacing $A_i^{1/2}\varphi_i$ by ψ_i and $A_i^{1/2}K_{ik}A_k^{1/2}$ by S_{ik}, we obtain

$$\lambda\psi = \mathsf{S}\psi, \tag{16.84g}$$

with S symmetric (since the kernel $K(x,t)$ was assumed symmetric). Of course, ψ has components $\psi_i = \psi(x_i)$. Equation (16.84g) is our matrix eigenvalue equation, Eq. (3.136).

[4]The eigenvalue λ has been written on the left side, multiplying the eigenfunction, as is customary in matrix analysis (Section 3.5). In this form λ will take on a **maximum value**.

The eigenvalues are readily obtained by calling a canned eigenroutine.[5] For kernels such as those of Exercise 16.3.15 and using a 10-point Gauss–Legendre quadrature, the eigenroutine determines the largest eigenvalue to within about 0.5 percent for the cases where the kernel has discontinuities in its derivatives. If the derivatives are continuous, the accuracy is much better.

Linz[6] has described an interesting variational refinement in the determination of $\lambda_{\max}$ to high accuracy. The key to his method is Exercise 17.8.7. The components of the eigenfunction vector are obtained from Eq. (16.84d) with $\varphi(t_k)$ now known and $\varphi_i = \varphi(x_i)$ generated as required. (The x_i are no longer tied to the t_k.)

Exercises

16.3.1 Using the Neumann series, solve

(a) $\varphi(x) = 1 - 2\displaystyle\int_0^x t\varphi(t)\,dt,$

(b) $\varphi(x) = x + \displaystyle\int_0^x (t-x)\varphi(t)\,dt,$

(c) $\varphi(x) = x - \displaystyle\int_0^x (t-x)\varphi(t)\,dt.$

ANS. (a) $\varphi(x) = e^{-x^2}$.

16.3.2 Solve the equation

$$\varphi(x) = x + \frac{1}{2}\int_{-1}^{1} (t+x)\varphi(t)\,dt$$

by the separable kernel method. Compare with the Neumann method solution of Section 16.3.

ANS. $\varphi(x) = \frac{1}{2}(3x-1)$.

16.3.3 Find the eigenvalues and eigenfunctions of

$$\varphi(x) = \lambda\int_{-1}^{1} (t-x)\varphi(t)\,dt.$$

16.3.4 Find the eigenvalues and eigenfunctions of

$$\varphi(x) = \lambda\int_0^{2\pi} \cos(x-t)\varphi(t)\,dt.$$

ANS. $\lambda_1 = \lambda_2 = \frac{1}{\pi}$, $\qquad \varphi(x) = A\cos x + B\sin x$.

[5] See W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*, 2nd ed., Cambridge, UK: Cambridge University Press (1992), Chapter 11, for details, references, and computer codes. The symbolic software *Mathematica* and *Maple* also include matrix functions for computing eigenvalues and eigenvectors.

[6] P. Linz, On the numerical computation of eigenvalues and eigenvectors of symmetric integral equations. *Math. Comput.* **24**: 905 (1970).

16.3.5 Find the eigenvalues and eigenfunctions of

$$y(x) = \lambda \int_{-1}^{1} (x-t)^2 y(t)\, dt.$$

Hint. This problem may be treated by the separable kernel method or by a Legendre expansion.

16.3.6 If the separable kernel technique of this section is applied to a Fredholm equation of the first kind (Eq. (16.1)), show that Eq. (16.76) is replaced by

$$\mathbf{c} = \mathsf{A}^{-1}\mathbf{b}.$$

In general the solution for the unknown $\varphi(t)$ is **not** unique.

16.3.7 Solve

$$\psi(x) = x + \int_0^1 (1 + xt)\psi(t)\, dt$$

by each of the following methods:

(a) the Neumann series technique,
(b) the separable kernel technique,
(c) educated guessing.

16.3.8 Use the separable kernel technique to show that

$$\psi(x) = \lambda \int_0^{\pi} \cos x \sin t\, \psi(t)\, dt$$

has **no** solution (apart from the trivial $\psi = 0$). Explain this result in terms of separability and symmetry.

16.3.9 Solve

$$\varphi(x) = 1 + \lambda^2 \int_0^x (x-t)\varphi(t)\, dt$$

by each of the following methods:

(a) reduction to an ODE (find the boundary conditions),
(b) the Neumann series,
(c) the use of Laplace transforms.

ANS. $\varphi(x) = \cosh \lambda x$.

16.3.10 (a) In Eq. (16.69a) take $V = V_0$, independent of t. Without using Eq. (16.69b), show that Eq. (16.69a) leads directly to

$$U(t - t_0) = \exp\left[-\frac{i}{\hbar}(t - t_0)V_0\right].$$

(b) Repeat for Eq. (16.69b) without using Eq. (16.69a).

16.3.11 Given $\varphi(x) = \lambda \int_0^1 (1 + xt)\varphi(t)\,dt$, solve for the eigenvalues and the eigenfunctions by the separable kernel technique.

16.3.12 Knowing the form of the solutions can be a great advantage, for the integral equation

$$\varphi(x) = \lambda \int_0^1 (1 + xt)\varphi(t)\,dt,$$

assume $\varphi(x)$ to have the form $1 + bx$. Substitute into the integral equation. Integrate and solve for b and λ.

16.3.13 The integral equation

$$\varphi(x) = \lambda \int_0^1 J_0(\alpha xt)\varphi(t)\,dt, \qquad J_0(\alpha) = 0,$$

is approximated by

$$\varphi(x) = \lambda \int_0^1 \left[1 - x^2t^2\right]\varphi(t)\,dt.$$

Find the minimum eigenvalue λ and the corresponding eigenfunction $\varphi(t)$ of the approximate equation.

ANS. $\lambda_{\min} = 1.112486, \quad \varphi(x) = 1 - 0.303337x^2.$

16.3.14 You are given the integral equation

$$\varphi(x) = \lambda \int_0^1 \sin \pi xt\varphi(t)\,dt.$$

Approximate the kernel by

$$K(x,t) = 4xt(1 - xt) \approx \sin \pi xt.$$

Find the positive eigenvalue and the corresponding eigenfunction for the approximate integral equation.
Note. For $K(x,t) = \sin \pi xt$, $\lambda = 1.6334$.

ANS. $\lambda = 1.5678, \quad \varphi(x) = x - 0.6955x^2$
$(\lambda_+ = \sqrt{31} - 4,\ \lambda_- = -\sqrt{31} - 4).$

16.3.15 The equation

$$f(x) = \int_a^b K(x,t)\varphi(t)\,dt$$

has a degenerate kernel $K(x,t) = \sum_{i=1}^n M_i(x)N_i(t)$.

(a) Show that this integral equation has no solution unless $f(x)$ can be written as

$$f(x) = \sum_{i=1}^n f_i M_i(x),$$

with the f_i constants.

(b) Show that to any solution $\varphi(x)$ we may add $\psi(x)$, provided $\psi(x)$ is orthogonal to all $N_i(x)$:

$$\int_a^b N_i(x)\psi(x)\,dx = 0 \qquad \text{for all } i.$$

16.3.16 Using numerical quadrature, convert

$$\varphi(x) = \lambda \int_0^1 J_0(\alpha x t)\varphi(t)\,dt, \qquad J_0(\alpha) = 0,$$

to a set of simultaneous linear equations.

(a) Find the minimum eigenvalue λ.
(b) Determine $\varphi(x)$ at discrete values of x and plot $\varphi(x)$ versus x. Compare with the approximate eigenfunction of Exercise 16.3.13.

ANS. (a) $\lambda_{\text{min}} = 1.14502$.

16.3.17 Using numerical quadrature, convert

$$\varphi(x) = \lambda \int_0^1 \sin \pi x t \varphi(t)\,dt$$

to a set of simultaneous linear equations.

(a) Find the minimum eigenvalue λ.
(b) Determine $\varphi(x)$ at discrete values of x and plot $\varphi(x)$ versus x. Compare with the approximate eigenfunction of Exercise 16.3.14.

ANS. (a) $\lambda_{\text{min}} = 1.6334$.

16.3.18 Given a homogeneous Fredholm equation of the second kind

$$\lambda\varphi(x) = \int_0^1 K(x,t)\varphi(t)\,dt.$$

(a) Calculate the largest eigenvalue λ_0. Use the 10-point Gauss–Legendre quadrature technique. For comparison the eigenvalues listed by Linz are given as λ_{exact}.
(b) Tabulate $\varphi(x_k)$, where the x_k are the 10 evaluation points in $[0, 1]$.
(c) Tabulate the ratio

$$\frac{1}{\lambda_0\varphi(x)} \int_0^1 K(x,t)\varphi(t)\,dt \qquad \text{for } x = x_k.$$

This is the test of whether or not you really have a solution.

(a) $K(x,t) = e^{xt}$.

ANS. $\lambda_{\text{exact}} = 1.35303$.

(b) $K(x,t) = \begin{cases} \frac{1}{2}x(2-t), & x < t, \\ \frac{1}{2}t(2-x), & x > t. \end{cases}$

ANS. $\lambda_{\text{exact}} = 0.24296$.

(c) $K(x,t) = |x-t|$.

ANS. $\lambda_{\text{exact}} = 0.34741$.

(d) $K(x,t) = \begin{cases} x, & x < t, \\ t, & x > t. \end{cases}$

ANS. $\lambda_{\text{exact}} = 0.40528$.

Note. (1) The evaluation points x_i of Gauss–Legendre quadrature for $[-1, 1]$ may be **linearly** transformed into $[0, 1]$,

$$x_i[0,1] = \tfrac{1}{2}\big(x_i[-1,1] + 1\big).$$

Then the weighting factors A_i are reduced in proportion to the length of the interval:

$$A_i[0,1] = \tfrac{1}{2}A_i[-1,1].$$

16.3.19 Using the matrix variational technique of Exercise 17.8.7, refine your calculation of the eigenvalue of Exercise 16.3.18(c) [$K(x,t) = |x-t|$]. Try a 40×40 matrix.
Note. Your matrix should be symmetric so that the (unknown) eigenvectors will be orthogonal.

ANS. (40-point Gauss–Legendre quadrature) 0.34727.

16.4 Hilbert–Schmidt Theory

Symmetrization of Kernels

This is the development of the properties of linear integral equations (Fredholm type) with symmetric kernels:

$$K(x,t) = K(t,x). \tag{16.85}$$

Before plunging into the theory, we note that some important nonsymmetric kernels can be symmetrized. If we have the equation

$$\varphi(x) = f(x) + \lambda \int_a^b K(x,t)\rho(t)\varphi(t)\,dt, \tag{16.86}$$

the total kernel is actually $K(x,t)\rho(t)$, clearly not symmetric if $K(x,t)$ alone is symmetric. However, if we multiply Eq. (16.86) by $\sqrt{\rho(x)}$ and substitute

$$\sqrt{\rho(x)}\varphi(x) = \psi(x), \tag{16.87}$$

we obtain

$$\psi(x) = \sqrt{\rho(x)}\, f(x) + \lambda \int_a^b \left[K(x,t)\sqrt{\rho(x)\rho(t)}\right]\psi(t)\, dt, \tag{16.88}$$

with a symmetric total kernel $K(x,t)\sqrt{\rho(x)\rho(t)}$. We shall meet $\rho(x)$ later as a positive weighting factor in this integral equation Sturm–Liouville theory.

Orthogonal Eigenfunctions

We now focus on the homogeneous Fredholm equation of the second kind:

$$\varphi(x) = \lambda \int_a^b K(x,t)\varphi(t)\, dt. \tag{16.89}$$

We assume that the kernel $K(x,t)$ is symmetric and real. Perhaps one of the first questions we might ask about the equation is: "Does it make sense?" or more precisely, "Does an eigenvalue λ satisfying this equation exist?" With the aid of the Schwarz and Bessel inequalities, Chapter 10 and Courant and Hilbert (Chapter III, Section 4 — see the Additional Readings) show that if $K(x,t)$ is continuous, there is at least one such eigenvalue and possibly an infinite number of them.

We show that the eigenvalues, λ, are real and that the corresponding eigenfunctions, $\varphi_i(x)$, are orthogonal. Let λ_i, λ_j be two **different** eigenvalues and $\varphi_i(x)$, $\varphi_j(x)$ be the corresponding eigenfunctions. Equation (16.89) then becomes

$$\varphi_i(x) = \lambda_i \int_a^b K(x,t)\varphi_i(t)\, dt, \tag{16.90a}$$

$$\varphi_j(x) = \lambda_j \int_a^b K(x,t)\varphi_j(t)\, dt. \tag{16.90b}$$

If we multiply Eq. (16.90a) by $\lambda_j\varphi_j(x)$ and Eq. (16.90b) by $\lambda_i\varphi_i(x)$ and then integrate with respect to x, the two equations become[7]

$$\lambda_j \int_a^b \varphi_i(x)\varphi_j(x)\, dx = \lambda_i\lambda_j \int_a^b\int_a^b K(x,t)\varphi_i(t)\varphi_j(x)\, dt\, dx, \tag{16.91a}$$

$$\lambda_i \int_a^b \varphi_i(x)\varphi_j(x)\, dx = \lambda_i\lambda_j \int_a^b\int_a^b K(x,t)\varphi_j(t)\varphi_i(x)\, dt\, dx, \tag{16.91b}$$

Since we have demanded that $K(x,t)$ by symmetric, Eq. (16.91b) may be rewritten as

$$\lambda_i \int_a^b \varphi_i(x)\varphi_j(x)\, dx = \lambda_i\lambda_j \int_a^b\int_a^b K(x,t)\varphi_i(t)\varphi_j(x)\, dt\, dx. \tag{16.92}$$

Subtracting Eq. (16.92) from Eq. (16.91a), we obtain

$$(\lambda_j - \lambda_i) \int_a^b \varphi_i(x)\varphi_j(x)\, dx = 0. \tag{16.93}$$

[7] We assume that the necessary integrals exist. For an example of a simple pathological case, see Exercise 16.4.3.

This has the same form as Eq. (10.34) in the Sturm–Liouville theory. Since $\lambda_i \neq \lambda_j$,

$$\int_a^b \varphi_i(x)\varphi_j(x)\,dx = 0, \qquad i \neq j, \tag{16.94}$$

proving orthogonality. Note that with a real symmetric kernel, no complex conjugates are involved in Eq. (16.94). For the self-adjoint or Hermitian kernel, see Exercise 16.4.1.

If the eigenvalue λ_i is degenerate,[8] the eigenfunctions for that particular eigenvalue may be orthogonalized by the Gram–Schmidt method (Section 10.3). Our orthogonal eigenfunctions may, of course, be normalized, and we assume that this has been done. The result is

$$\int_a^b \varphi_i(x)\varphi_j(x)\,dx = \delta_{ij}. \tag{16.95}$$

To demonstrate that the λ_i are real, we need to admit complex conjugates. Taking the complex conjugate of Eq. (16.90a), we have

$$\varphi_i^*(x) = \lambda_i^* \int_a^b K(x,t)\varphi_i^*(t)\,dt, \tag{16.96}$$

provided the kernel $K(x,t)$ is real. Now, using Eq. (16.96) instead of Eq. (16.90b), we see that the analysis leads to

$$(\lambda_i^* - \lambda_i)\int_a^b \varphi_i^*(x)\varphi_i(x)\,dx = 0. \tag{16.97}$$

This time the integral cannot vanish (unless we have the trivial solution, $\varphi_i(x) = 0$) and

$$\lambda_i^* = \lambda_i, \tag{16.98}$$

or λ_i, our eigenvalue, is real.

This is the **third** time we have passed this way, first with Hermitian matrices, then with Sturm–Liouville (self-adjoint) ODEs, and now with Hilbert–Schmidt integral equations. The correspondence between the Hermitian matrices and the self-adjoint ODEs shows up in physics as the two outstanding formulations of quantum mechanics — the Heisenberg matrix approach and the Schrödinger differential operator approach. In Section 17.8 and Exercise 17.7.6 we shall explore further the correspondence between the Hilbert–Schmidt symmetric kernel integral equations and the Sturm–Liouville self-adjoint differential equations.

The eigenfunctions of our integral equations form a complete set,[9] in the sense that any function $g(x)$ that can be generated by the integral

$$g(x) = \int K(x,t)h(t)\,dt, \tag{16.99}$$

[8] If more than one distinct eigenfunction corresponds to the same eigenvalue (satisfying Eq. (16.89)), that eigenvalue is said to be degenerate (see Chapters 3 and 4).

[9] For a proof of this statement, see Courant and Hilbert (1953), Chapter III, Section 5, in the Additional Readings.

in which $h(t)$ is any piecewise continuous function, can be represented by a series of eigenfunctions,

$$g(x) = \sum_{n=1}^{\infty} a_n \varphi_n(x). \tag{16.100}$$

The series converges uniformly and absolutely.

Let us extend this to the kernel $K(x,t)$ by asserting that

$$K(x,t) = \sum_{n=1}^{\infty} a_n \varphi_n(t), \tag{16.101}$$

and $a_n = a_n(x)$. Substituting into the original integral equation (Eq. (16.89)) and using the orthogonality integral, we obtain

$$\varphi_i(x) = \lambda_i a_i(x). \tag{16.102}$$

Therefore for our homogeneous Fredholm equation of the second kind, the kernel may be expressed in terms of the eigenfunctions and eigenvalues by

$$K(x,t) = \sum_{n=1}^{\infty} \frac{\varphi_n(x)\varphi_n(t)}{\lambda_n} \qquad \text{(zero not an eigenvalue).} \tag{16.103}$$

Here we have a bilinear expansion, a linear expansion in $\varphi_n(x)$ and linear in $\varphi_n(t)$. Similar bilinear expansions appear in Section 9.7. It is possible that the expansion given by Eq. (16.101) may not exist. As an illustration of the sort of pathological behavior that may occur, you are invited to apply this analysis to

$$\varphi(x) = \lambda \int_0^{\infty} e^{-xt} \varphi(t)\, dt$$

(compare Exercise 16.4.3).

It should be emphasized that this Hilbert–Schmidt theory is concerned with the establishment of properties of the eigenvalues (real) and eigenfunctions (orthogonality, completeness), properties that may be of great interest and value. The Hilbert–Schmidt theory does **not solve** the homogeneous integral equation for us any more than the Sturm–Liouville theory of Chapter 10 solved the ODEs. The solutions of the integral equation come from Sections 16.2 and 16.3 (including numerical analysis).

Nonhomogeneous Integral Equation

We need a solution of the nonhomogeneous equation

$$\varphi(x) = f(x) + \lambda \int_a^b K(x,t)\varphi(t)\, dt. \tag{16.104}$$

Let us assume that the solutions of the corresponding homogeneous integral equation are known:

$$\varphi_n(x) = \lambda_n \int_a^b K(x,t)\varphi_n(t)\, dt, \tag{16.105}$$

the solution $\varphi_n(x)$ corresponding to the eigenvalue λ_n. We expand both $\varphi(x)$ and $f(x)$ in terms of this set of eigenfunctions:

$$\varphi(x) = \sum_{n=1}^{\infty} a_n \varphi_n(x) \qquad (a_n \text{ unknown}), \tag{16.106}$$

$$f(x) = \sum_{n=1}^{\infty} b_n \varphi_n(x) \qquad (b_n \text{ known}). \tag{16.107}$$

Substituting into Eq. (16.104), we obtain

$$\sum_{n=1}^{\infty} a_n \varphi_n(x) = \sum_{n=1}^{\infty} b_n \varphi_n(x) + \lambda \int_a^b K(x,t) \sum_{n=1}^{\infty} a_n \varphi_n(t)\, dt. \tag{16.108}$$

By interchanging the order of integration and summation, we may evaluate the integral by Eq. (16.105), and we get

$$\sum_{n=1}^{\infty} a_n \varphi_n(x) = \sum_{n=1}^{\infty} b_n \varphi_n(x) + \lambda \sum_{n=1}^{\infty} \frac{a_n \varphi_n(x)}{\lambda_n}. \tag{16.109}$$

If we multiply by $\varphi_i(x)$ and integrate from $x = a$ to $x = b$, the orthogonality of our eigenfunctions leads to

$$a_i = b_i + \lambda \frac{a_i}{\lambda_i}. \tag{16.110}$$

This can be rewritten as

$$a_i = b_i + \frac{\lambda}{\lambda_i - \lambda} b_i, \tag{16.111}$$

which brings us to our solution

$$\varphi(x) = f(x) + \lambda \sum_{i=1}^{\infty} \frac{\int_a^b f(t)\varphi_i(t)\, dt}{\lambda_i - \lambda} \varphi_i(x). \tag{16.112}$$

Here it is assumed that the eigenfunctions $\varphi_i(x)$ are normalized to unity. **Note that if** $f(x) = 0$ **there is no solution unless** $\lambda = \lambda_i$. This means that our homogeneous equation has no solution (except the trivial $\varphi(x) = 0$) unless λ is an eigenvalue, λ_i.

In the event that λ for the nonhomogeneous equation (16.104) is equal to one of the eigenvalues λ_p of the homogeneous equation, our solution (Eq. (16.112)) blows up. To repair the damage we return to Eq. (16.110) and give the value

$$a_p = b_p + \lambda_p \frac{a_p}{\lambda_p} = b_p + a_p \tag{16.113}$$

special attention. Clearly, a_p drops out and is no longer determined by b_p, whereas $b_p = 0$. This implies that $\int f(x)\varphi_p(x)\, dx = 0$; that is, $f(x)$ is orthogonal to the eigenfunction $\varphi_p(x)$. If this is **not** the case, we have no solution.

Equation (16.111) still holds for $i \neq p$, so we multiply by $\varphi_i(x)$ and sum over $i\,(i \neq p)$ to obtain

$$\varphi(x) = f(x) + a_p\varphi_p + \lambda_p \sum_{\substack{i=1\\ i\neq p}}^{\infty} \frac{\int_a^b f(t)\varphi_i(t)\,dt}{\lambda_i - \lambda_p}\varphi_i(x). \tag{16.114}$$

In this solution the a_p remains as an undetermined constant.[10]

Exercises

16.4.1 In the Fredholm equation

$$\varphi(x) = \lambda \int_a^b K(x,t)\varphi(t)\,dt$$

the kernel $K(x,t)$ is self-adjoint or Hermitian:

$$K(x,t) = K^*(t,x).$$

Show that

(a) the eigenfunctions are orthogonal, in the sense that

$$\int_a^b \varphi_m^*(x)\varphi_n(x)\,dx = 0, \qquad m \neq n\ (\lambda_m \neq \lambda_n),$$

(b) the eigenvalues are real.

16.4.2 Solve the integral equation

$$\varphi(x) = x + \frac{1}{2}\int_{-1}^{1} (t+x)\varphi(t)\,dt$$

(compare Exercise 16.3.2) by the Hilbert–Schmidt method.
Note. The application of the Hilbert–Schmidt technique here is somewhat like using a shotgun to kill a mosquito, especially when the equation can be solved quickly by expanding in Legendre polynomials.

16.4.3 Solve the Fredholm integral equation

$$\varphi(x) = \lambda \int_0^{\infty} e^{-xt}\varphi(t)\,dt.$$

Note. A series expansion of the kernel e^{-xt} would permit a separable kernel-type solution (Section 16.3), except that the series is infinite. This suggests an infinite number of eigenvalues and eigenfunctions. If you stop with

$$\varphi(x) = x^{-1/2}, \qquad \lambda = \pi^{-1/2},$$

[10] This is like the inhomogeneous linear ODE. We may add to its solution any constant times a solution of the corresponding homogeneous ODE.

you will have missed most of the solutions. Show that the normalization integrals of the eigenfunctions do **not** exist. A basic reason for this anomalous behavior is that the range of integration is infinite, making this a "singular" integral equation.

16.4.4 Given

$$y(x) = x + \lambda \int_0^1 xty(t)\,dt.$$

(a) Determine $y(x)$ as a Neumann series.
(b) Find the range of λ for which your Neumann series solution is convergent. Compare with the value obtained from

$$|\lambda| \cdot |K|_{\max} < 1.$$

(c) Find the eigenvalue and the eigenfunction of the corresponding homogeneous integral equation.
(d) By the separable kernel method show that the solution is

$$y(x) = \frac{3x}{3-\lambda}.$$

(e) Find $y(x)$ by the Hilbert–Schmidt method.

16.4.5 In Exercise 16.3.4,

$$K(x,t) = \cos(x-t).$$

The (unnormalized) eigenfunctions are $\cos x$ and $\sin x$.

(a) Show that there is a function $h(t)$ such that $K(x,s)$, considered as a function of s alone, may be written as

$$K(x,s) = \int_0^{2\pi} K(s,t)h(t)\,dt.$$

(b) Show that $K(x,t)$ may be expanded as

$$K(x,t) = \sum_{n=1}^{2} \frac{\varphi_n(x)\varphi_n(t)}{\lambda_n}.$$

16.4.6 The integral equation $\varphi(x) = \lambda \int_0^1 (1+xt)\varphi(t)\,dt$ has eigenvalues $\lambda_1 = 0.7889$ and $\lambda_2 = 15.211$ and eigenfunctions $\varphi_1 = 1 + 0.5352x$ and $\varphi_2 = 1 - 1.8685x$.

(a) Show that these eigenfunctions are orthogonal over the interval $[0, 1]$.
(b) Normalize the eigenfunctions to unity.
(c) Show that

$$K(x,t) = \frac{\varphi_1(x)\varphi_1(t)}{\lambda_1} + \frac{\varphi_2(x)\varphi_2(t)}{\lambda_2}.$$

ANS. (b) $\varphi_1(x) = 0.7831 + 0.4191x$
$\varphi_2(x) = 1.8403 - 3.4386x.$

16.4.7 An alternate form of the solution to the nonhomogeneous integral equation, Eq. (16.104), is

$$\varphi(x) = \sum_{i=1}^{\infty} \frac{b_i \lambda_i}{\lambda_i - \lambda} \varphi_i(x).$$

(a) Derive this form without using Eq. (16.112).
(b) Show that this form and Eq. (16.112) are equivalent.

16.4.8 (a) Show that the eigenfunctions of Exercise 16.3.5 are orthogonal.
(b) Show that the eigenfunctions of Exercise 16.3.11 are orthogonal.

Additional Readings

Bocher, M., *An Introduction to the Study of Integral Equations*, Cambridge Tracts in Mathematics and Mathematical Physics, No. 10. New York: Hafner (1960). This is a helpful introduction to integral equations.

Cochran, J. A., *The Analysis of Linear Integral Equations*. New York: McGraw-Hill (1972). This is a comprehensive treatment of linear integral equations which is intended for applied mathematicians and mathematical physicists. It assumes a moderate to high level of mathematical competence on the part of the reader.

Courant, R., and D. Hilbert, *Methods of Mathematical Physics*, Vol.1 (English edition). New York: Interscience (1953). This is one of the classic works of mathematical physics. Originally published in German in 1924, the revised English edition is an excellent reference for a rigorous treatment of integral equations, Green's functions, and a wide variety of other topics on mathematical physics.

Golberg, M. A., ed., *Solution Methods of Integral Equations*. New York: Plenum Press (1979). This is a set of papers from a conference on integral equations. The initial chapter is excellent for up-to-date orientation and a wealth of references.

Kanval, R. P., *Linear Integral Equations*. New York: Academic Press (1971), reprinted, Birkhäuser (1996). This book is a detailed but readable treatment of a variety of techniques for solving linear integral equations.

Morse, P. M., and H. Feshbach, *Methods of Theoretical Physics*. New York: McGraw-Hill (1953). Chapter 7 is a particularly detailed, complete discussion of Green's functions from the point of view of mathematical physics. Note, however, that Morse and Feshbach frequently choose a source of $4\pi\delta(\mathbf{r} - \mathbf{r}')$ in place of our $\delta(\mathbf{r} - \mathbf{r}')$. Considerable attention is devoted to bounded regions.

Muskhelishvili, N. I., *Singular Integral Equations*, 2nd ed., New York: Dover (1992).

Stakgold, I., *Green's Functions and Boundary Value Problems*. New York: Wiley (1979).

CHAPTER 17

CALCULUS OF VARIATIONS

Uses of the Calculus of Variations

We now address problems where we search for a function or curve, rather than a value of some variable, that makes a given quantity stationary, usually an energy or action integral. Because a function is varied, these problems are called **variational**. Variational principles, such as D'Alembert's and Hamilton's, have been developed in classical mechanics, and Lagrangian techniques occur in quantum mechanics and field theory, for example, Fermat's principle of the shortest optical path in electrodynamics. Before plunging into this rather different branch of mathematical physics, let us summarize some of its uses in both physics and mathematics.

1. **In existing physical theories:**
   a. Unification of diverse areas of physics using energy as a key concept.
   b. Convenience in analysis — Lagrange equations, Section 17.3.
   c. Elegant treatment of constraints, Section 17.7.
2. **Starting point for new, complex areas of physics and engineering**. In general relativity the geodesic is taken as the minimum path of a light pulse or the free-fall path of a particle in curved Riemannian space (see geodesics in Section 2.10). Variational principles appear in quantum field theory. Variational principles have been applied extensively in control theory.
3. **Mathematical unification**. Variational analysis provides a proof of the completeness of the Sturm–Liouville eigenfunctions, Chapter 10, and establishes a lower bound for the eigenvalues. Similar results follow for the eigenvalues and eigenfunctions of the Hilbert–Schmidt integral equation, Section 16.4.
4. **Calculation techniques**, Section 17.8. Calculation of the eigenfunctions and eigenvalues of the Sturm–Liouville equation. Integral equation eigenfunctions and eigenvalues may be calculated using numerical quadrature and matrix techniques, Section 16.3.

17.1 A DEPENDENT AND AN INDEPENDENT VARIABLE

Concept of Variation

The calculus of variations involves problems in which the quantity to be minimized (or maximized) appears as a stationary integral, a functional, because a function $y(x, \alpha)$ needs to be determined from a class described by an infinitesimal parameter α. As the simplest case, let

$$J = \int_{x_1}^{x_2} f(y, y_x, x)\, dx. \tag{17.1}$$

Here J is the quantity that takes on a stationary value. Under the integral sign, f is a known function of the indicated variables x and α, as are $y(x, \alpha)$, $y_x(x, \alpha) \equiv \partial y(x, \alpha)/\partial x$, but the dependence of y on x (and α) is not yet known; that is, $y(x)$ is **unknown**. This means that although the integral is from x_1 to x_2, the exact path of integration is not known (Fig. 17.1). We are to choose the path of integration through points (x_1, y_1) and (x_2, y_2) to minimize J. Strictly speaking, we determine stationary values of J: minima, maxima, or saddle points. In most cases of physical interest the stationary value will be a minimum. This problem is considerably more difficult than the corresponding problem of a function $y(x)$ in differential calculus. Indeed, there may be no solution. In differential calculus the minimum is determined by comparing $y(x_0)$ with $y(x)$, where x ranges over neighboring points. Here we assume the existence of an optimum path, that is, an acceptable path for which J is stationary, and then compare J for our (unknown) optimum path with that obtained from neighboring paths. In Fig. 17.1 two possible paths are shown. (There are an infinite number of possibilities.) The difference between these two for a given x is called the variation of y, δy, and is conveniently described by introducing a new function, $\eta(x)$, to define the arbitrary deformation of the path and a scale factor, α, to give the magnitude of the variation. The function $\eta(x)$ is arbitrary except for two restrictions. First,

$$\eta(x_1) = \eta(x_2) = 0, \tag{17.2}$$

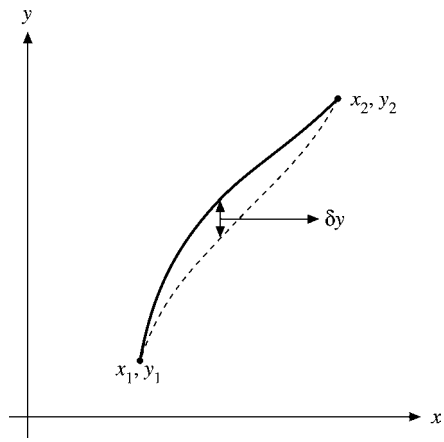

FIGURE 17.1 A varied path.

which means that all varied paths must pass through the fixed endpoints. Second, as will be seen shortly, $\eta(x)$ must be differentiable; that is, we may not use

$$\begin{aligned} \eta(x) &= 1, \qquad x = x_0, \\ &= 0, \qquad x \neq x_0, \end{aligned} \tag{17.3}$$

but we can choose $\eta(x)$ to have a form similar to the functions used to represent the Dirac delta function (Chapter 1) so that $\eta(x)$ differs from zero only over an infinitesimal region.[1] Then, with the path described by α and $\eta(x)$,

$$y(x, \alpha) = y(x, 0) + \alpha\eta(x) \tag{17.4}$$

and

$$\delta y = y(x, \alpha) - y(x, 0) = \alpha\eta(x). \tag{17.5}$$

Let us choose $y(x, \alpha = 0)$ as the unknown path that will minimize J. Then $y(x, \alpha)$ for nonzero α describes a neighboring path. In Eq. (17.1), J is now a function[2] of our parameter α:

$$J(\alpha) = \int_{x_1}^{x_2} f\big[y(x, \alpha), y_x(x, \alpha), x\big]\, dx, \tag{17.6}$$

and our condition for an extreme value is that

$$\left[\frac{\partial J(\alpha)}{\partial \alpha}\right]_{\alpha=0} = 0, \tag{17.7}$$

analogous to the vanishing of the derivative dy/dx in differential calculus.

Now, the α-dependence of the integral is contained in $y(x, \alpha)$ and $y_x(x, \alpha) = (\partial/\partial x)y(x, \alpha)$. Therefore[3]

$$\frac{\partial J(\alpha)}{\partial \alpha} = \int_{x_1}^{x_2} \left[\frac{\partial f}{\partial y}\frac{\partial y}{\partial \alpha} + \frac{\partial f}{\partial y_x}\frac{\partial y_x}{\partial \alpha}\right] dx. \tag{17.8}$$

From Eq. (17.4),

$$\frac{\partial y(x, \alpha)}{\partial \alpha} = \eta(x), \tag{17.9}$$

$$\frac{\partial y_x(x, \alpha)}{\partial \alpha} = \frac{d\eta(x)}{dx}, \tag{17.10}$$

so Eq. (17.8) becomes

$$\frac{\partial J(\alpha)}{\partial \alpha} = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y}\eta(x) + \frac{\partial f}{\partial y_x}\frac{d\eta(x)}{dx}\right) dx. \tag{17.11}$$

[1] Compare H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics*, 3rd ed., Cambridge, UK: Cambridge University Press (1966), Chapter 10, for a more complete discussion of this point.

[2] Technically, J is a **functional of** y, y_x, but a function of α depending on the functions $y(x, \alpha)$ and $y_x(x, \alpha)$: $J[y(x, \alpha), y_x(x, \alpha)]$.

[3] Note that y and y_x are being treated as **independent** variables.

Integrating the second term by parts to get $\eta(x)$ as a common and arbitrary nonvanishing factor, we obtain

$$\int_{x_1}^{x_2} \frac{d\eta(x)}{dx}\frac{\partial f}{\partial y_x}\,dx = \eta(x)\frac{\partial f}{\partial y_x}\bigg|_{x_1}^{x_2} - \int_{x_1}^{x_2} \eta(x)\frac{d}{dx}\frac{\partial f}{\partial y_x}\,dx. \tag{17.12}$$

The integrated part vanishes by Eq. (17.2), and Eq. (17.11) becomes

$$\int_{x_1}^{x_2}\left[\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y_x}\right]\eta(x)\,dx = 0. \tag{17.13}$$

In this form α has been set equal to zero, corresponding to the solution path, and, in effect, is no longer part of the problem.

Occasionally we will see Eq. (17.13) multiplied by $\delta\alpha$, which gives, upon using $\eta(x)\delta\alpha = \delta y$,

$$\int_{x_1}^{x_2}\left(\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y_x}\right)\delta y\,dx = \delta\alpha\left[\frac{\partial J}{\partial \alpha}\right]_{\alpha=0} = \delta J = 0. \tag{17.14}$$

Since $\eta(x)$ is arbitrary, we may choose it to have the same sign as the bracketed expression in Eq. (17.13) whenever the latter differs from zero. Hence the integrand is always nonnegative. Equation (17.13), our condition for the existence of a stationary value, can then be satisfied only if the bracketed term itself is zero almost everywhere. The condition for our stationary value is thus a PDE,[4]

$$\boxed{\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y_x} = 0,} \tag{17.15}$$

known as the Euler equation, which can be expressed in various other forms. Sometimes solutions are missed when they are not twice differentiable, as required by Eq. (17.15). An example is Goldschmidt's discontinuous solution of Section 17.2. It is clear that Eq. (17.15) must be satisfied for J to take on a stationary value, that is, for Eq. (17.14) to be satisfied. Equation (17.15) is necessary, but it is by no means sufficient.[5] Courant and Robbins (1996; see the Additional Readings) illustrate this very nicely by considering the distance over a sphere between points on the sphere, A and B, Fig. 17.2. Path (1), a great circle, is found from Eq. (17.15). But path (2), the remainder of the great circle through points A and B, also satisfies the Euler equation. Path (2) is a maximum, but only if we demand that it be a great circle and then only if we make less than one circuit; that is, path (2) $+n$ complete revolutions is also a solution. If the path is not required to be a great circle, any deviation from (2) will increase the length. This is hardly the property of a local maximum, and that is why it is important to check the properties of solutions of Eq. (17.15) to see if they satisfy the physical conditions of the given problem.

[4] It is important to watch the meaning of $\partial/\partial x$ and d/dx closely. For example, if $f = f[y(x), y_x, x]$,

$$\frac{df}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx} + \frac{\partial f}{\partial y_x}\frac{d^2y}{dx^2}.$$

The first term on the right gives the **explicit** x-dependence. The second and third terms give the **implicit** x-dependence via y and y_x.

[5] For a discussion of sufficiency conditions and the development of the calculus of variations as a part of mathematics, see G. M. Ewing, *Calculus of Variations with Applications*, New York: Norton (1969). Sufficiency conditions are also covered by Sagan (in the Additional Readings at the end of this chapter).

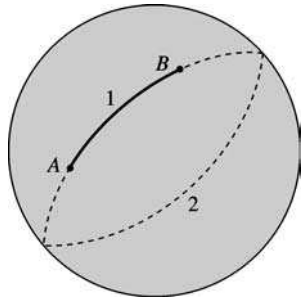

FIGURE 17.2 Stationary paths over a sphere.

Example 17.1.1 OPTICAL PATH NEAR EVENT HORIZON OF A BLACK HOLE

Determine the optical path in an atmosphere where the velocity of light increases in proportion to the height, $v(y) = y/b$, with $b > 0$ some parameter describing the light speed. So $v = 0$ at $y = 0$, which simulates the conditions at the surface of a black hole, called its **event horizon**, where the gravitational force is so strong that the velocity of light goes to zero, thus even trapping light.

Because light takes the shortest time, the variational problem takes the form

$$\Delta t = \int_{t_1}^{t_2} dt = \int \frac{ds}{v} = b \int \frac{\sqrt{dx^2 + dy^2}}{y} dt = \text{minimum}.$$

Here $v = ds/dt = y/b$ is the velocity of light in this environment, the y coordinate being the height. A look at the variational functional suggests choosing y as the independent variable because x does not appear in the integrand. We can bring dy outside the radical and change the role of x and y in J of Eq. (17.1) and the resulting Euler equation. With $x = x(y),\ x' = dx/dy$, we obtain

$$b \int \frac{\sqrt{x'^2 + 1}}{y} dy = \text{minimum},$$

and the Euler equation becomes

$$\frac{\partial f}{\partial x} - \frac{d}{dy} \frac{\partial f}{\partial x'} = 0.$$

Since $\partial f/\partial x = 0$, this can be integrated, giving

$$\frac{x'}{y\sqrt{x'^2 + 1}} = C_1 = \text{const.}, \qquad \text{or} \qquad x'^2 = C_1^2 y^2 \left(x'^2 + 1\right).$$

Separating dx and dy in this first-order ODE we find the integral

$$\int^x dx = \int^y \frac{C_1 y\, dy}{\sqrt{1 - C_1^2 y^2}},$$

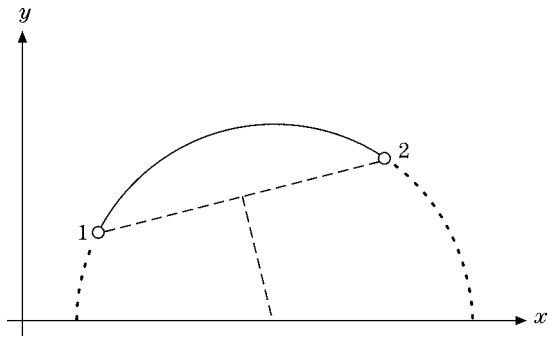

FIGURE 17.3 Circular optical path in medium.

which yields

$$x + C_2 = \frac{-1}{C_1}\sqrt{1 - C_1^2 y^2}, \qquad \text{or} \qquad (x + C_2)^2 + y^2 = \frac{1}{C_1^2}.$$

This is a circular light path with center on the x-axis along the event horizon. (See Fig. 17.3.) This example may be adapted to a mirage (Fata Morgana) in a desert with hot air near the ground and cooler air aloft (the index of refraction changes with height in cool versus hot air), thus changing the velocity law from $v = y/b \to v_0 - y/b$. In this case, the circular light path is no longer convex with center on the x-axis, but becomes concave. ■

Alternate Forms of Euler Equations

One other form (Exercise 17.1.1), which is often useful, is

$$\frac{\partial f}{\partial x} - \frac{d}{dx}\left(f - y_x \frac{\partial f}{\partial y_x}\right) = 0. \tag{17.16}$$

In problems in which $f = f(y, y_x)$, that is, in which x does not appear explicitly, Eq. (17.16) reduces to

$$\frac{d}{dx}\left(f - y_x \frac{\partial f}{\partial y_x}\right) = 0, \tag{17.17}$$

or

$$f - y_x \frac{\partial f}{\partial y_x} = \text{constant}. \tag{17.18}$$

Example 17.1.2 Missing Dependent Variables

Consider the variational problem $\int f(\dot{\mathbf{r}})\, dt =$ minimum. Here $\mathbf{r}$ is absent from the integrand. Therefore the Euler equations become

$$\frac{d}{dt}\frac{\partial f}{\partial \dot{x}} = 0, \qquad \frac{d}{dt}\frac{\partial f}{\partial \dot{y}} = 0, \qquad \frac{d}{dt}\frac{\partial f}{\partial \dot{z}} = 0,$$

with $\mathbf{r} = (x, y, z)$, so $f_{\dot{\mathbf{r}}} = \mathbf{c} = \text{const}$. Solving these three equations for the three unknowns $\dot{x}, \dot{y}, \dot{z}$ yields $\dot{\mathbf{r}} = \mathbf{c}_1 = \text{const}$. Integrating this constant velocity gives $\mathbf{r} = \mathbf{c}_1 t + \mathbf{c}_2$. The solutions are straight lines, despite the general nature of the function f.

A physical example illustrating this case is the propagation of light in a crystal, where the velocity of light depends on the (crystal) directions but not on the location in the crystal, because a crystal is an anisotropic **homogeneous medium**. The variational problem

$$\int \frac{ds}{v} = \int \frac{\sqrt{\dot{\mathbf{r}}^2}}{v(\dot{\mathbf{r}})}\, dt = \text{minimum}$$

has the form of our example. Note that t need not be the time, but it parameterizes the light path. ■

Exercises

17.1.1 For $dy/dx \equiv y_x \neq 0$, show the equivalence of the two forms of Euler's equation:

$$\frac{\partial f}{\partial x} - \frac{d}{dx}\frac{\partial f}{\partial y_x} = 0$$

and

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(f - y_x \frac{\partial f}{\partial y_x}\right) = 0.$$

17.1.2 Derive Euler's equation by expanding the integrand of

$$J(\alpha) = \int_{x_1}^{x_2} f\big[y(x,\alpha), y_x(x,\alpha), x\big]\, dx$$

in powers of α, using a Taylor (Maclaurin) expansion with y and y_x as the two variables (Section 5.6).

Note. The stationary condition is $\partial J(\alpha)/\partial\alpha = 0$, evaluated at $\alpha = 0$. The terms quadratic in α may be useful in establishing the nature of the stationary solution (maximum, minimum, or saddle point).

17.1.3 Find the Euler equation corresponding to Eq. (17.15) if $f = f(y_{xx}, y_x, y, x)$.

$$\text{ANS. } \frac{d^2}{dx^2}\left(\frac{\partial f}{\partial y_{xx}}\right) - \frac{d}{dx}\left(\frac{\partial f}{\partial y_x}\right) + \frac{\partial f}{\partial y} = 0,$$
$$\eta(x_1) = \eta(x_2) = 0, \quad \eta_x(x_1) = \eta_x(x_2) = 0.$$

17.1.4 The integrand $f(y, y_x, x)$ of Eq. (17.1) has the form

$$f(y, y_x, x) = f_1(x, y) + f_2(x, y) y_x.$$

(a) Show that the Euler equation leads to

$$\frac{\partial f_1}{\partial y} - \frac{\partial f_2}{\partial x} = 0.$$

(b) What does this imply for the dependence of the integral J upon the choice of path?

17.1.5 Show that the condition that

$$J = \int f(x, y)\, dx$$

has a stationary value

(a) leads to $f(x, y)$ independent of y and
(b) yields no information about any x-dependence.

We get no (continuous, differentiable) solution. To be a meaningful variational problem, dependence on y or higher derivatives is essential.
Note. The situation will change when constraints are introduced (compare Exercise 17.7.7).

17.2 APPLICATIONS OF THE EULER EQUATION

Example 17.2.1 STRAIGHT LINE

Perhaps the simplest application of the Euler equation is in the determination of the shortest distance between two points in the Euclidean xy-plane. Since the element of distance is

$$ds = \left[(dx)^2 + (dy)^2\right]^{1/2} = \left[1 + y_x^2\right]^{1/2} dx, \tag{17.19}$$

the distance J may be written as

$$J = \int_{x_1, y_1}^{x_2, y_2} ds = \int_{x_1}^{x_2} \left[1 + y_x^2\right]^{1/2} dx. \tag{17.20}$$

Comparison with Eq. (17.1) shows that

$$f(y, y_x, x) = \left(1 + y_x^2\right)^{1/2}. \tag{17.21}$$

Substituting into Eq. (17.16), we obtain

$$-\frac{d}{dx}\left[\frac{1}{(1 + y_x^2)^{1/2}}\right] = 0, \tag{17.22}$$

or

$$\frac{1}{(1 + y_x^2)^{1/2}} = C, \qquad \text{a constant.} \tag{17.23}$$

This is satisfied by

$$y_x = a, \qquad \text{a second constant,} \tag{17.24}$$

and

$$y = ax + b, \tag{17.25}$$

which is the familiar equation for a straight line. The constants a and b are chosen so that the line passes through the two points (x_1, y_1) and (x_2, y_2). Hence the Euler equation

predicts that the shortest[6] distance between two fixed points in Euclidean space is a straight line. ■

The generalization of this in curved four-dimensional space–time leads to the important concept of the geodesic in general relativity (see Section 2.10).

Example 17.2.2 SOAP FILM

As a second illustration (Fig. 17.4), consider two parallel coaxial wire circles to be connected by a surface of minimum area that is generated by revolving a curve $y(x)$ about the x-axis. The curve is required to pass through fixed endpoints (x_1, y_1) and (x_2, y_2). The variational problem is to choose the curve $y(x)$ so that the area of the resulting surface will be a minimum.

For the element of area shown in Fig. 17.4,

$$dA = 2\pi y\, ds = 2\pi y\left(1 + y_x^2\right)^{1/2} dx. \tag{17.26}$$

The variational equation is then

$$J = \int_{x_1}^{x_2} 2\pi y\left(1 + y_x^2\right)^{1/2} dx. \tag{17.27}$$

Neglecting the 2π, we obtain

$$f(y, y_x, x) = y\left(1 + y_x^2\right)^{1/2}. \tag{17.28}$$

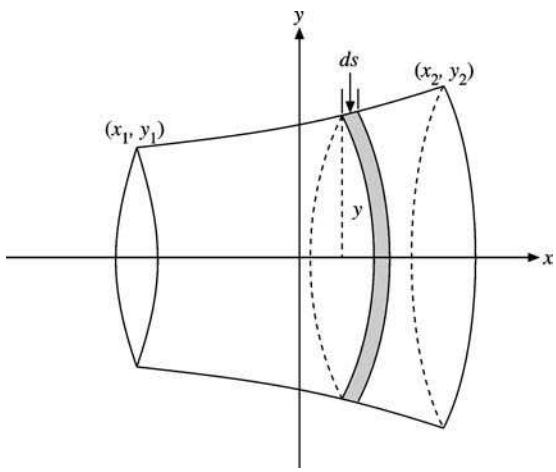

FIGURE 17.4 Surface of rotation — soap film problem.

[6]Technically, we have a stationary value. From the α^2 terms it can be identified as a minimum (Exercise 17.2.2).

Since $\partial f/\partial x = 0$, we may apply Eq. (17.18) directly and get

$$y\left(1+y_x^2\right)^{1/2} - yy_x^2\frac{1}{(1+y_x^2)^{1/2}} = c_1, \tag{17.29}$$

or

$$\frac{y}{(1+y_x^2)^{1/2}} = c_1. \tag{17.30}$$

Squaring, we get

$$\frac{y^2}{1+y_x^2} = c_1^2 \qquad \text{with } c_1^2 \le y_{\min}^2, \tag{17.31}$$

and

$$(y_x)^{-1} = \frac{dx}{dy} = \frac{c_1}{\sqrt{y^2 - c_1^2}}. \tag{17.32}$$

This may be integrated to give

$$x = c_1 \cosh^{-1}\frac{y}{c_1} + c_2. \tag{17.33}$$

Solving for y, we have

$$y = c_1 \cosh\left(\frac{x-c_2}{c_1}\right), \tag{17.34}$$

and again c_1 and c_2 are determined by requiring the hyperbolic cosine to pass through the points (x_1, y_1) and (x_2, y_2). Our "minimum"-area surface is a special case of a catenary of revolution, or a **catenoid**. ■

Soap Film — Minimum Area

This calculus of variations contains many pitfalls for the unwary. (Remember, the Euler equation is a **necessary** condition assuming a **differentiable solution**. The sufficiency conditions are quite involved. See the Additional Readings for details.) Respect for some of these hazards may be developed by considering a specific physical problem, for example, a minimum-area problem with $(x_1, y_1) = (-x_0, 1)$, $(x_2, y_2) = (+x_0, 1)$. The minimum surface is a soap film stretched between the two rings of unit radius at $x = \pm x_0$. The problem is to predict the curve $y(x)$ assumed by the soap film.

By referring to Eq. (17.34), we find that $c_2 = 0$ by the symmetry of the problem about $x = 0$. Then

$$y = c_1 \cosh\left(\frac{x}{c_1}\right), \qquad c_1 \cosh\left(\frac{x_0}{c_1}\right) = 1. \tag{17.34a}$$

If we take $x_0 = \frac{1}{2}$ we obtain a transcendental equation for c_1, viz.

$$1 = c_1 \cosh\left(\frac{1}{2c_1}\right). \tag{17.35}$$

We find that this equation has two solutions: $c_1 = 0.2350$, leading to a "deep" curve, and $c_1 = 0.8483$, leading to a "flat" curve. Which curve is assumed by the soap film? Before answering this question, consider the physical situation with the rings moved apart so that $x_0 = 1$. Then Eq. (17.34a) becomes

$$1 = c_1 \cosh\left(\frac{1}{c_1}\right), \tag{17.36}$$

which has **no real solutions**. The physical significance is that as the unit-radius rings were moved out from the origin, a point was reached at which the soap film could no longer maintain the same horizontal force over each vertical section. Stable equilibrium was no longer possible. The soap film broke (irreversible process) and formed a circular film over each ring (with a total area of $2\pi = 6.2832\ldots$). This is the Goldschmidt discontinuous solution.

The next question is: How large may x_0 be and still give a real solution for Eq. (17.34a)?[7] Letting $c_1^{-1} = p$, Eq. (17.34a) becomes

$$p = \cosh px_0. \tag{17.37}$$

To find $x_{0\max}$ we could solve for x_0 (as in Eq. (17.33)) and then differentiate with respect to p. Finally, with an eye on Fig. 17.5, dx_0/dp would be set equal to zero. Alternatively, direct differentiation of Eq. (17.37) with respect to p yields

$$1 = \left[x_0 + p\frac{dx_0}{dp}\right] \sinh px_0.$$

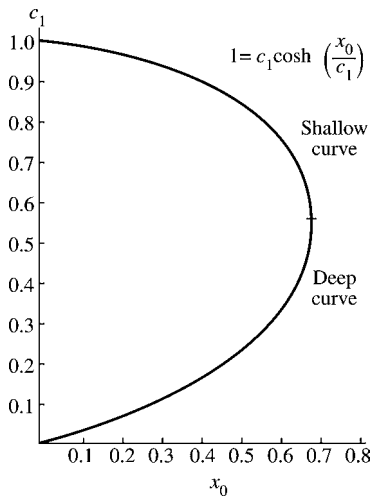

FIGURE 17.5 Solutions of Eq. (17.34a) for unit-radius rings at $x = \pm x_0$.

[7] From a numerical point of view it is easier to invert the problem. Pick a value of c_1 and solve for x_0. Equation (17.34a) becomes $x_0 = c_1 \cosh^{-1}(1/c_1)$. This has numerical solutions in the range $0 < c_1 \leq 1$.

The requirement that dx_0/dp vanish leads to

$$1 = x_0 \sinh px_0. \tag{17.38}$$

Equations (17.37) and (17.38) may be combined to form

$$px_0 = \coth px_0, \tag{17.39}$$

with the root

$$px_0 = 1.1997. \tag{17.40}$$

Substituting into Eq. (17.37) or (17.38), we obtain

$$p = 1.810, \qquad c_1 = 0.5524 \tag{17.41}$$

and

$$x_{0\,\max} = 0.6627. \tag{17.42}$$

Returning to the question of the solution of Eq. (17.35) that describes the soap film, let us calculate the area corresponding to each solution. We have

$$\begin{aligned} A &= 4\pi \int_0^{x_0} y\left(1 + y_x^2\right)^{1/2} dx = \frac{4\pi}{c_1} \int_0^{x_0} y^2\, dx \qquad \text{(by Eq. (17.30))} \\ &= 4\pi c_1 \int_0^{x_0} \left(\cosh \frac{x}{c_1}\right)^2 dx = \pi c_1^2 \left[\sinh\left(\frac{2x_0}{c_1}\right) + \frac{2x_0}{c_1}\right]. \end{aligned} \tag{17.43}$$

For $x_0 = \frac{1}{2}$, Eq. (17.35) leads to

$$\begin{aligned} c_1 &= 0.2350 \rightarrow A = 6.8456, \\ c_1 &= 0.8483 \rightarrow A = 5.9917, \end{aligned}$$

showing that the former can at most be only a local minimum. A more detailed investigation (compare Bliss, *Calculus of Variations*, Chapter IV) shows that this surface is not even a local minimum. For $x_0 = \frac{1}{2}$, the soap film will be described by the flat curve

$$y = 0.8483 \cosh\left(\frac{x}{0.8483}\right). \tag{17.44}$$

This flat or shallow catenoid (catenary of revolution) will be an absolute minimum for $0 \le x_0 < 0.528$. However, for $0.528 < x < 0.6627$ its area is greater than that of the Goldschmidt discontinuous solution (6.2832) and it is only a relative minimum (Fig. 17.6).

For an excellent discussion of both the mathematical problems and experiments with soap films, we refer to Courant and Robbins (1996) in the Additional Readings at the end of the chapter.

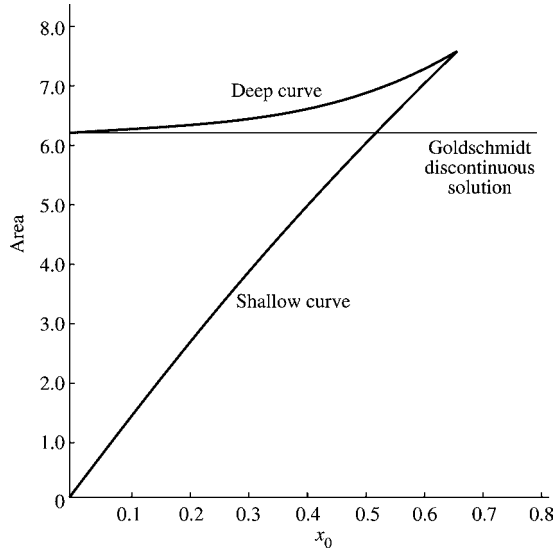

FIGURE 17.6 Catenoid area (unit-radius rings at $x = \pm x_0$).

Exercises

17.2.1 A soap film is stretched across the space between two rings of unit radius centered at $\pm x_0$ on the x-axis and perpendicular to the x-axis. Using the solution developed in Section 17.2, set up the transcendental equations for the condition that x_0 is such that the area of the curved surface of rotation equals the area of the two rings (Goldschmidt discontinuous solution). Solve for x_0 (Fig. 17.7).

17.2.2 In Example 17.2.1, expand $J[y(x, \alpha)] - J[y(x, 0)]$ in powers of α. The term linear in α leads to the Euler equation and to the straight-line solution, Eq. (17.25). Investigate

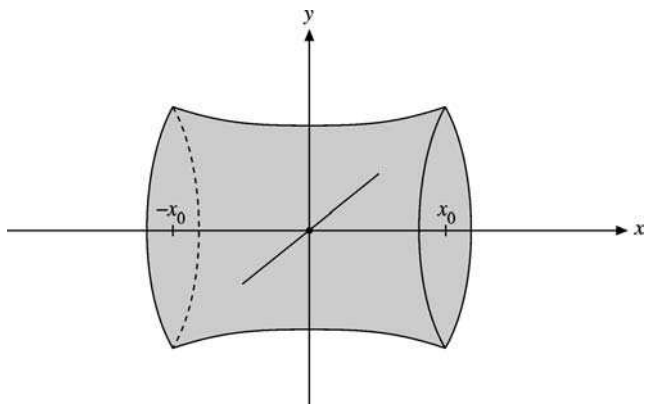

FIGURE 17.7 Surface of rotation.

the α^2 term and show that the stationary value of J, the straight-line distance, is a **minimum**.

17.2.3 (a) Show that the integral

$$J = \int_{x_1}^{x_2} f(y, y_x, x)\, dx, \qquad \text{with } f = y(x),$$

has **no** extreme values.

(b) If $f(y, y_x, x) = y^2(x)$, find a discontinuous solution similar to the Goldschmidt solution for the soap film problem.

17.2.4 Fermat's principle of optics states that a light ray will follow the path $y(x)$ for which

$$\int_{x_1, y_1}^{x_2, y_2} n(y, x)\, ds$$

is a minimum when n is the index of refraction. For $y_2 = y_1 = 1$, $-x_1 = x_2 = 1$, find the ray path if

(a) $n = e^y$, (b) $n = a(y - y_0),\ y > y_0$.

17.2.5 A frictionless particle moves from point A on the surface of the Earth to point B by sliding through a tunnel. Find the differential equation to be satisfied if the transit time is to be a minimum.

Note. Assume the Earth to be nonrotating sphere of uniform density.

ANS. (Eq. (17.15)): $r_{\varphi\varphi}(r^3 - ra^2) + r_\varphi^2(2a^2 - r^2) + a^2r^2 = 0,$

$r(\varphi = 0) = r_0,\ r_\varphi(\varphi = 0) = 0,\ r(\varphi = \varphi_A) = a,\ r(\varphi = \varphi_B) = a.$

Eq. (17.18): $r_\varphi^2 = \frac{a^2r^2}{r_0^2} \cdot \frac{r^2 - r_0^2}{a^2 - r^2}$. The solution of these equations is a hypocycloid, generated by a circle of radius $\frac{1}{2}(a - r_0)$ rolling inside the circle of radius a. You might like to show that the transit time is

$$t = \pi \frac{(a^2 - r_0^2)^{1/2}}{(ag)^{1/2}}.$$

For details see P. W. Cooper, *Am. J. Phys.* **34**: 68 (1966); G. Veneziano *et al., ibid.*, pp. 701–704.

17.2.6 A ray of light follows a straight-line path in a first homogeneous medium, is refracted at an interface, and then follows a new straight-line path in the second medium. Use Fermat's principle of optics to derive Snell's law of refraction:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2.$$

Hint. Keep the points (x_1, y_1) and (x_2, y_2) fixed and vary x_0 to satisfy Fermat (Fig. 17.8). This is **not** an Euler equation problem. (The light path is not differentiable at x_0.)

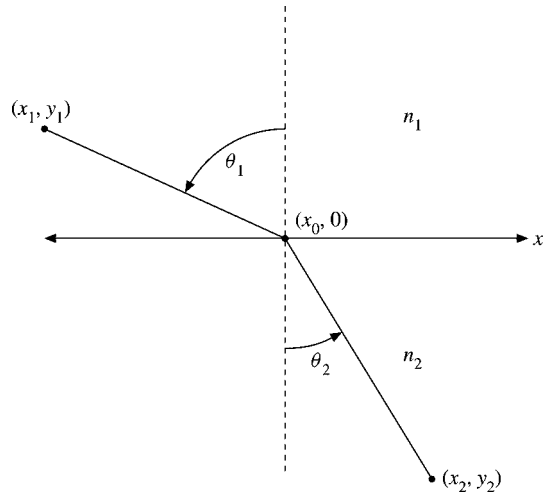

FIGURE 17.8 Snell's law.

17.2.7 A second soap film configuration for the unit-radius rings at $x = \pm x_0$ consists of a circular disk, radius a, in the $x = 0$ plane and two catenoids of revolution, one joining the disk and each ring. One catenoid may be described by

$$y = c_1 \cosh\left(\frac{x}{c_1} + c_3\right).$$

(a) Impose boundary conditions at $x = 0$ and $x = x_0$.
(b) Although not necessary, it is convenient to require that the catenoids form an angle of 120° where they join the central disk. Express this third boundary condition in mathematical terms.
(c) Show that the total area of catenoids plus central disk is

$$A = c_1^2\left[\sinh\left(\frac{2x_0}{c_1} + 2c_3\right) + \frac{2x_0}{c_1}\right].$$

Note. Although this soap film configuration is physically realizable and stable, the area is larger than that of the simple catenoid for all ring separations for which both films exist.

$$\text{ANS. (a)} \begin{cases} 1 = c_1 \cosh\left(\dfrac{x_0}{c_1} + c_3\right) \\ a = c_1 \cosh c_3, \end{cases} \qquad \text{(b)} \ \frac{dy}{dx} = \tan 30^\circ = \sinh c_3.$$

17.2.8 For the soap film described in Exercise 17.2.7, find (numerically) the maximum value of x_0.
Note. This calls for a pocket calculator with hyperbolic functions or a table of hyperbolic cotangents.

ANS. $x_{0\max} = 0.4078$.

17.2.9 Find the root of $px_0 = \coth px_0$ (Eq. (17.39)) and determine the corresponding values of p and x_0 (Eqs. (17.41) and (17.42)). Calculate your values to five significant figures.

17.2.10 For the two-ring soap film problem of this section calculate and tabulate x_0, p, p^{-1}, and A, the soap film area for $px_0 = 0.00(0.02)1.30$.

17.2.11 Find the value of x_0 (to five significant figures) that leads to a soap film area, Eq. (17.43), equal to 2π, the Goldschmidt discontinuous solution.

ANS. $x_0 = 0.52770$.

17.2.12 Find the curve of quickest descent from $(0, 0)$ to (x_0, y_0) for a particle sliding under gravity and without friction. Show that the ratio of times taken by the particle along a straight line joining the two points compared to along the curve of quickest descent is $(1 + 4/\pi^2)^{1/2}$.

Hint. Take y to increase downwards. Apply Eq. (17.18) to obtain $y_x^2 = (1 - c^2 y)/c^2 y$, where c is an integration constant. Then make the substitution $y = (\sin^2 \varphi/2)/c^2$ to parametrize the cycloid and take $(x_0, y_0) = (\pi/2c^2, 1/c^2)$.

17.3 SEVERAL DEPENDENT VARIABLES

Our original variational problem, Eq. (17.1), may be generalized in several respects. In this section we consider the integrand f to be a function of several **dependent** variables $y_1(x), y_2(x), y_3(x), \ldots$, all of which depend on x, the independent variable. In Section 17.4 f again will contain only one unknown function y, but y will be a function of several independent variables (over which we integrate). In Section 17.5 these two generalizations are combined. In Section 17.7 the stationary value is restricted by one or more constraints.

For more than one dependent variable, Eq. (17.1) becomes

$$J = \int_{x_1}^{x_2} f\big[y_1(x), y_2(x), \ldots, y_n(x), y_{1x}(x), y_{2x}(x), \ldots, y_{nx}(x), x\big]\, dx. \tag{17.45}$$

As in Section 17.1, we determine the extreme value of J by comparing neighboring paths. Let

$$y_i(x, \alpha) = y_i(x, 0) + \alpha \eta_i(x), \qquad i = 1, 2, \ldots, n, \tag{17.46}$$

with the η_i independent of one another but subject to the restrictions discussed in Section 17.1. By differentiating Eq. (17.45) with respect to α and setting $\alpha = 0$, since Eq. (17.7) still applies, we obtain

$$\int_{x_1}^{x_2} \sum_i \left(\frac{\partial f}{\partial y_i} \eta_i + \frac{\partial f}{\partial y_{ix}} \eta_{ix} \right) dx = 0, \tag{17.47}$$

the subscript x denoting partial differentiation with respect to x; that is, $y_{ix} = \partial y_i/\partial x$, and so on. Again, each of the terms $(\partial f/\partial y_{ix})\eta_{ix}$ is integrated by parts. The integrated part vanishes and Eq. (17.47) becomes

$$\int_{x_1}^{x_2} \sum_i \left(\frac{\partial f}{\partial y_i} - \frac{d}{dx} \frac{\partial f}{\partial y_{ix}} \right) \eta_i \, dx = 0. \tag{17.48}$$

Since the η_i are arbitrary and **independent** of one another,[8] each of the terms in the sum must vanish **independently**. We have

$$\frac{\partial f}{\partial y_i} - \frac{d}{dx}\frac{\partial f}{\partial(\partial y_i/\partial x)} = 0, \qquad i = 1, 2, \ldots, n, \tag{17.49}$$

a whole set of Euler equations, each of which must be satisfied for an extreme value.

Hamilton's Principle

The most important application of Eq. (17.45) occurs when the integrand f is taken to be a Lagrangian L. The Langrangian (for nonrelativistic systems; see Exercise 17.3.5 for a relativistic particle) is defined as the **difference** of kinetic and potential energies of a system:

$$L \equiv T - V. \tag{17.50}$$

Using time as an independent variable instead of x and $x_i(t)$ as the dependent variables, we get

$$x \to t, \qquad y_i \to x_i(t), \qquad y_{ix} \to \dot{x}_i(t);$$

$x_i(t)$ is the location and $\dot{x}_i = dx_i/dt$ is the velocity of particle i as a function of time. The equation $\delta J = 0$ is then a mathematical statement of Hamilton's principle of classical mechanics,

$$\delta \int_{t_1}^{t_2} L(x_1, x_2, \ldots, x_n, \dot{x}_1, \dot{x}_2, \ldots, \dot{x}_n; t)\, dt = 0. \tag{17.51}$$

In words, Hamilton's principle asserts that the motion of the system from time t_1 to t_2 is such that the time integral of the Lagrangian L, or action, has a stationary value. The resulting Euler equations are usually called the Lagrangian equations of motion,

$$\boxed{\frac{d}{dt}\frac{\partial L}{\partial \dot{x}_i} - \frac{\partial L}{\partial x_i} = 0.} \tag{17.52}$$

These Lagrangian equations can be derived from Newton's equations of motion, and Newton's equations can be derived from Lagrange's. The two sets of equations are equally "fundamental."

The Lagrangian formulation has advantages over the conventional Newtonian laws. Whereas Newton's equations are vector equations, we see that Lagrange's equations involve only scalar quantities. The coordinates $x_1, x_2, \ldots$ need not be any standard set of coordinates or lengths. They can be selected to match the conditions of the physical problem. The Lagrange equations are invariant with respect to the choice of coordinate system. Newton's equations (in component form) are not manifestly invariant. Exercise 2.5.10 shows what happens to $\mathbf{F} = m\mathbf{a}$ resolved in spherical polar coordinates.

[8] For example, we could set $\eta_2 = \eta_3 = \eta_4 = \cdots = 0$, eliminating all but one term of the sum, and then treat η_1 exactly as in Section 17.1.

Exploiting the concept of energy, we may easily extend the Lagrangian formulation from mechanics to diverse fields, such as electrical networks and acoustical systems. Extensions to electromagnetism appear in the exercises. The result is a unity of otherwise-separate areas of physics. In the development of new areas, the quantization of Lagrangian particle mechanics provided a model for the quantization of electromagnetic fields and led to the gauge theory of quantum electrodynamics.

One of the most valuable advantages of the Hamilton principle — Lagrange equation formulation — is the ease in seeing a relation between a symmetry and a conservation law. As an example, let $x_i = \varphi$, an azimuthal angle. If our Lagrangian is independent of φ (that is, if φ is an ignorable coordinate), there are two consequences: (1) the conservation or invariance of a component of angular momentum and (2) from Eq. (17.52) $\partial L/\partial\dot{\varphi} =$ constant. Similarly, invariance under translation leads to conservation of linear momentum. Noether's theorem is a generalization of this invariance (symmetry) — the conservation law relation.

Example 17.3.1 MOVING PARTICLE — CARTESIAN COORDINATES

Consider Eq. (17.50), which describes one particle with kinetic energy

$$T = \tfrac{1}{2}m\dot{x}^2 \tag{17.53}$$

and potential energy $V(x)$, in which, as usual, the force is given by the negative gradient of the potential,

$$F(x) = -\frac{dV(x)}{dx}. \tag{17.54}$$

From Eq. (17.52),

$$\frac{d}{dt}(m\dot{x}) - \frac{\partial(T-V)}{\partial x} = m\ddot{x} - F(x) = 0, \tag{17.55}$$

which is Newton's second law of motion. ■

Example 17.3.2 MOVING PARTICLE — CIRCULAR CYLINDRICAL COORDINATES

Now let us describe a moving particle in cylindrical coordinates of the xy-plane, that is, $z = 0$. The kinetic energy is

$$T = \tfrac{1}{2}m\left(\dot{x}^2 + \dot{y}^2\right) = \tfrac{1}{2}m\left(\dot{\rho}^2 + \rho^2\dot{\varphi}^2\right), \tag{17.56}$$

and we take $V = 0$ for simplicity.

The transformation of $\dot{x}^2 + \dot{y}^2$ into circular cylindrical coordinates could be carried out by taking $x(\rho, \varphi)$ and $y(\rho, \varphi)$, Eq. (2.28), and differentiating with respect to time and squaring. It is much easier to interpret $\dot{x}^2 + \dot{y}^2$ as v^2 and just write down the components of $\mathbf{v}$ as $\hat{\boldsymbol{\rho}}(ds_\rho/dt) = \hat{\boldsymbol{\rho}}\dot{\rho}$, and so on. (The ds_ρ is an increment of **length**, ρ changing by $d\rho$, φ remaining constant. See Sections 2.1 and 2.4.)

The Lagrangian equations yield

$$\frac{d}{dt}(m\dot{\rho}) - m\rho\dot{\varphi}^2 = 0, \qquad \frac{d}{dt}\left(m\rho^2\dot{\varphi}\right) = 0. \tag{17.57}$$

The second equation is a statement of conservation of angular momentum. The first may be interpreted as radial acceleration[9] equated to centrifugal force. In this sense the centrifugal force is a real force. It is of some interest that this interpretation of centrifugal force as a real force is supported by the general theory of relativity. ■

Exercises

17.3.1 (a) Develop the equations of motion corresponding to $L = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2)$.
(b) In what sense do your solutions minimize the integral $\int_{t_1}^{t_2} L\,dt$?
Compare the result for your solution with $x = \text{const.}$, $y = \text{const.}$

17.3.2 From the Lagrangian equations of motion, Eq. (17.52), show that a system in stable equilibrium has a minimum potential energy.

17.3.3 Write out the Lagrangian equations of motion in terms of a rotating coordinate system and identify the (a) centrifugal force potential and (b) Coriolis force.

17.3.4 The spherical pendulum consists of a mass on a wire of length l, free to move in polar angle θ and azimuth angle φ (Fig. 17.9).

(a) Set up the Lagrangian for this physical system.
(b) Develop the Lagrangian equations of motion.

17.3.5 Show that the Lagrangian

$$L = m_0c^2\left(1 - \sqrt{1 - \frac{v^2}{c^2}}\right) - V(\mathbf{r})$$

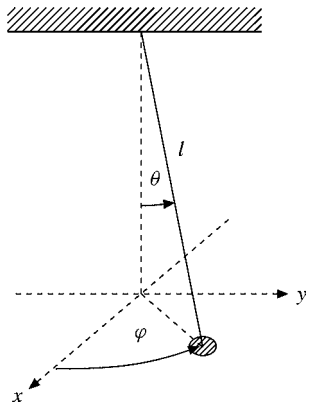

FIGURE 17.9 Spherical pendulum.

[9]Here is a second method of attacking Exercise 2.4.8.

leads to a relativistic form of Newton's second law of motion,

$$\frac{d}{dt}\left(\frac{m_0 v_i}{\sqrt{1-v^2/c^2}}\right) = F_i,$$

in which the force components are $F_i = -\partial V/\partial x_i$.

17.3.6 The Lagrangian for a particle with charge q in an electromagnetic field described by scalar potential φ and vector potential $\mathbf{A}$ is

$$L = \tfrac{1}{2}mv^2 - q\varphi + q\mathbf{A}\cdot\mathbf{v}.$$

Find the equation of motion of the charged particle.
Hint. $(d/dt)A_j = \partial A_j/\partial t + \sum_i (\partial A_j/\partial x_i)\dot{x}_i$. The dependence of the force fields $\mathbf{E}$ and $\mathbf{B}$ upon the potentials φ and $\mathbf{A}$ is developed in Section 1.13 (compare Exercise 1.13.10).

ANS. $m\ddot{x}_i = q[\mathbf{E} + \mathbf{v}\times\mathbf{B}]_i$.

17.3.7 Consider a system in which the Lagrangian is given by

$$L(q_i, \dot{q}_i) = T(q_i, \dot{q}_i) - V(q_i),$$

where q_i and $\dot{q}_i$ represent sets of variables. The potential energy V is independent of velocity and neither T nor V has any explicit time dependence.

(a) Show that

$$\frac{d}{dt}\left(\sum_j \dot{q}_j \frac{\partial L}{\partial \dot{q}_j} - L\right) = 0.$$

(b) The constant quantity

$$\sum_j \dot{q}_j \frac{\partial L}{\partial \dot{q}_j} - L$$

defines the Hamiltonian H. Show that under the preceding assumed conditions, $H = T + V$, the total energy.

Note. The kinetic energy T is a quadratic function of the $\dot{q}_i$.

17.4 SEVERAL INDEPENDENT VARIABLES

Sometimes the integrand f of Eq. (17.1) will contain one unknown function, u, that is a function of several independent variables, $u = u(x, y, z)$, for the three-dimensional case, for example. Equation (17.1) becomes

$$J = \iiint f[u, u_x, u_y, u_z, x, y, z]\, dx\, dy\, dz, \tag{17.58}$$

$u_x = \partial u/\partial x$, and so on. The variational problem is to find the function $u(x, y, z)$ for which J is stationary,

$$\delta J = \delta\alpha \frac{\partial J}{\partial \alpha}\bigg|_{\alpha=0} = 0. \tag{17.59}$$

Generalizing Section 17.1, we let

$$u(x, y, z, \alpha) = u(x, y, z, 0) + \alpha\eta(x, y, z), \tag{17.60}$$

where $u(x, y, z, \alpha = 0)$ represents the (unknown) function for which Eq. (17.59) is satisfied, whereas again $\eta(x, y, z)$ is the arbitrary deviation that describes the varied function $u(x, y, z, \alpha)$. This deviation $\eta(x, y, z)$ is required to be differentiable and to vanish at the endpoints. Then from Eq. (17.60),

$$u_x(x, y, z, \alpha) = u_x(x, y, z, 0) + \alpha\eta_x, \tag{17.61}$$

and similarly for u_y and u_z.

Differentiating the integral Eq. (17.58) with respect to the parameter α and then setting $\alpha = 0$, we obtain

$$\left.\frac{\partial J}{\partial \alpha}\right|_{\alpha=0} = \iiint \left(\frac{\partial f}{\partial u}\eta + \frac{\partial f}{\partial u_x}\eta_x + \frac{\partial f}{\partial u_y}\eta_y + \frac{\partial f}{\partial u_z}\eta_z\right) dx\, dy\, dz = 0. \tag{17.62}$$

Again, we integrate each of the terms $(\partial f/\partial u_i)\eta_i$ by parts. The integrated part vanishes at the endpoints (because the deviation η is required to go to zero at the endpoints) and

$$\iiint \left(\frac{\partial f}{\partial u} - \frac{\partial}{\partial x}\frac{\partial f}{\partial u_x} - \frac{\partial}{\partial y}\frac{\partial f}{\partial u_y} - \frac{\partial}{\partial z}\frac{\partial f}{\partial u_z}\right)\eta(x, y, z)\, dx\, dy\, dz = 0.^{10} \tag{17.63}$$

Since the variation $\eta(x, y, z)$ is arbitrary, the term in large parentheses is set equal to zero. This yields the Euler equation for (three) independent variables,

$$\frac{\partial f}{\partial y} - \frac{\partial}{\partial x}\frac{\partial f}{\partial u_x} - \frac{\partial}{\partial y}\frac{\partial f}{\partial u_y} - \frac{\partial}{\partial z}\frac{\partial f}{\partial u_z} = 0. \tag{17.64}$$

Example 17.4.1 LAPLACE'S EQUATION

An example of this sort of variational problem is provided by electrostatics. The energy of an electrostatic field is

$$\text{energy density} = \tfrac{1}{2}\varepsilon\mathbf{E}^2, \tag{17.65}$$

in which $\mathbf{E}$ is the usual electrostatic force field. In terms of the static potential φ,

$$\text{energy density} = \tfrac{1}{2}\varepsilon(\boldsymbol{\nabla}\varphi)^2. \tag{17.66}$$

Now let us impose the requirement that the electrostatic energy (associated with the field) in a given volume be a minimum. (Boundary conditions on $\mathbf{E}$ and φ must still be satisfied.) We have the volume integral[11]

$$J = \iiint (\boldsymbol{\nabla}\varphi)^2\, dx\, dy\, dz = \iiint \left(\varphi_x^2 + \varphi_y^2 + \varphi_z^2\right) dx\, dy\, dz. \tag{17.67}$$

[10] Recall that $\partial/\partial x$ is a partial derivative, where y and z are held constant. But $\partial/\partial x$ also acts on **implicit** x-dependence as well as on **explicit** x-dependence. In this sense, for example,

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial u_x}\right) = \frac{\partial^2 f}{\partial x \partial u_x} + \frac{\partial^2 f}{\partial u \partial u_x}u_x + \frac{\partial^2 f}{\partial u_x^2}u_{xx} + \frac{\partial^2 f}{\partial u_y \partial u_x}u_{xy} + \frac{\partial^2 f}{\partial u_z \partial u_x}u_{xz}.$$

[11] The subscript x indicates the x-partial derivative, not an x-component.

With

$$f(\varphi, \varphi_x, \varphi_y, \varphi_z, x, y, z) = \varphi_x^2 + \varphi_y^2 + \varphi_z^2, \tag{17.68}$$

the function φ replacing the u of Eq. (17.64), Euler's equation (Eq. (17.64)) yields

$$-2(\varphi_{xx} + \varphi_{yy} + \varphi_{zz}) = 0, \tag{17.69}$$

or

$$\nabla^2 \varphi(x, y, z) = 0, \tag{17.70}$$

which is Laplace's equation of electrostatics.

Closer investigation shows that this stationary value is indeed a minimum. Thus the demand that the field energy be minimized leads to Laplace's PDE. ■

Exercises

17.4.1 The Lagrangian for a vibrating string (small-amplitude vibrations) is

$$L = \int \left(\tfrac{1}{2}\rho u_t^2 - \tfrac{1}{2}\tau u_x^2\right) dx,$$

where ρ is the (constant) linear mass density and τ is the (constant) tension. The x-integration is over the length of the string. Show that application of Hamilton's principle to the Lagrangian density (the integrand), now with two independent variables, leads to the classical wave equation

$$\frac{\partial^2 u}{\partial x^2} = \frac{\rho}{\tau}\frac{\partial^2 u}{\partial t^2}.$$

17.4.2 Show that the stationary value of the total energy of the electrostatic field of Example 17.4.1 is a **minimum**.
Hint. Use Eq. (17.61) and investigate the α^2 terms.

17.5 Several Dependent and Independent Variables

In some cases our integrand f contains more than one dependent variable and more than one independent variable. Consider

$$f = f\big[p(x, y, z), p_x, p_y, p_z, q(x, y, z), q_x, q_y, q_z, r(x, y, z), r_x, r_y, r_z, x, y, z\big]. \tag{17.71}$$

We proceed as before with

$$\begin{aligned} p(x, y, z, \alpha) &= p(x, y, z, 0) + \alpha\xi(x, y, z), \\ q(x, y, z, \alpha) &= q(x, y, z, 0) + \alpha\eta(x, y, z), \\ r(x, y, z, \alpha) &= r(x, y, z, 0) + \alpha\zeta(x, y, z), \qquad \text{and so on.} \end{aligned} \tag{17.72}$$

Keeping in mind that ξ, η, and ζ are independent of one another, as were the η_i in Section 17.3, the same differentiation and then integration by parts leads to

$$\frac{\partial f}{\partial p} - \frac{\partial}{\partial x}\frac{\partial f}{\partial p_x} - \frac{\partial}{\partial y}\frac{\partial f}{\partial p_y} - \frac{\partial}{\partial z}\frac{\partial f}{\partial p_z} = 0, \tag{17.73}$$

with similar equations for functions q and r. Replacing $p, q, r, \ldots$ with y_i and $x, y, z, \ldots$ with x_y, we can put Eq. (17.73) in a more compact form:

$$\frac{\partial f}{\partial y_i} - \sum_j \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial y_{ij}}\right) = 0, \qquad i = 1, 2, \ldots, \tag{17.73a}$$

in which

$$y_{ij} \equiv \frac{\partial y_i}{\partial x_j}.$$

An application of Eq. (17.73) appears in Section 17.7.

Relation to Physics

The calculus of variations as developed so far provides an elegant description of a wide variety of physical phenomena. The physics includes classical mechanics in Section 17.3; relativistic mechanics, Exercise 17.3.5; electrostatics, Example 17.4.1; and electromagnetic theory in Exercise 17.5.1. The convenience should not be minimized, but at the same time we should be aware that in these cases the calculus of variations has only provided an alternate description of what was already known. The situation does change with incomplete theories.

- If the basic physics is not yet known, a postulated variational principle can be a useful starting point.

Exercise

17.5.1 The Lagrangian (per unit volume) of an electromagnetic field with a charge density ρ is given by

$$\mathcal{L} = \frac{1}{2}\left(\varepsilon_0 \mathbf{E}^2 - \frac{1}{\mu_0}\mathbf{B}^2\right) - \rho\varphi + \rho\mathbf{v}\cdot\mathbf{A}.$$

Show that Lagrange's equations lead to two of Maxwell's equations. (The remaining two are a consequence of the definition of **E** and **B** in terms of **A** and φ.) This Lagrange density comes from a scalar expression in Section 4.6.
Hint. Take A_1, A_2, A_3, and φ as **dependent** variables, x, y, z, and t as **independent** variables. **E** and **B** are given in terms of **A** and φ by Eq. (4.142) and Eq. (1.88).

17.6 LAGRANGIAN MULTIPLIERS

In this section the concept of a constraint is introduced. To simplify the treatment, the constraint appears as a simple function rather than as an integral. In this section we are not concerned with the calculus of variations, but in Section 17.7 the constraints, with our newly developed Lagrangian multipliers, are incorporated into the calculus of variations.

Consider a function of three independent variables, $f(x, y, z)$. For the function f to be a maximum (or extreme),[12]

$$df = 0. \tag{17.74}$$

The necessary and sufficient condition for this is

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = \frac{\partial f}{\partial z} = 0, \tag{17.75}$$

in which

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz. \tag{17.76}$$

Often in physical problems the variables x, y, z are subject to constraints so that they are no longer all independent. It is possible, at least in principle, to use each constraint to eliminate one variable and to proceed with a new and smaller set of independent variables.

The use of Lagrangian multipliers is an alternate technique that may be applied when this elimination of variables is inconvenient or undesirable. Let our **equation of constraint** be

$$\varphi(x, y, z) = 0, \tag{17.77}$$

from which $z(x, y)$ may be extracted if x, y are taken as the independent coordinates. Returning to Eq. (17.74), Eq. (17.75) no longer follows because there are now only two independent variables, so dz is no longer arbitrary. From the total differential $d\varphi = 0$, we then obtain

$$-\frac{\partial \varphi}{\partial z}dz = \frac{\partial \varphi}{\partial x}dx + \frac{\partial \varphi}{\partial y}dy \tag{17.78}$$

and therefore

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \lambda\left(\frac{\partial \varphi}{\partial x}dx + \frac{\partial \varphi}{\partial x}dx\right), \qquad \lambda = -\frac{f_z}{\varphi_z},$$

assuming $\varphi_z = \frac{\partial \varphi}{\partial z} \neq 0$. Thus, we may add Eq. (17.76) and a multiple of Eq. (17.78) to obtain

$$df + \lambda\, d\varphi = \left(\frac{\partial f}{\partial x} + \lambda\frac{\partial \varphi}{\partial x}\right)dx + \left(\frac{\partial f}{\partial y} + \lambda\frac{\partial \varphi}{\partial y}\right)dy + \left(\frac{\partial f}{\partial z} + \lambda\frac{\partial \varphi}{\partial z}\right)dz = 0. \tag{17.79}$$

In other words, our Lagrangian multiplier λ is chosen so that

$$\frac{\partial f}{\partial z} + \lambda\frac{\partial \varphi}{\partial z} = 0, \tag{17.80}$$

[12] Including a saddle point.

assuming that $\partial\varphi/\partial z \neq 0$. Equation (17.79) now becomes

$$\left(\frac{\partial f}{\partial x}+\lambda\frac{\partial\varphi}{\partial x}\right)dx+\left(\frac{\partial f}{\partial y}+\lambda\frac{\partial\varphi}{\partial y}\right)dy=0. \tag{17.81}$$

However, now dx and dy are arbitrary and the quantities in parentheses must vanish:

$$\frac{\partial f}{\partial x}+\lambda\frac{\partial\varphi}{\partial x}=0, \qquad \frac{\partial f}{\partial y}+\lambda\frac{\partial\varphi}{\partial y}=0. \tag{17.82}$$

When Eqs. (17.80) and (17.82) are satisfied, $df=0$ and f is an extremum. Notice that there are now four unknowns: x, y, z, and λ. The fourth equation is, of course, the constraint Eq. (17.77). We want only x, y, and z, so λ need not be determined. For this reason λ is sometimes called **Lagrange's undetermined multiplier**. This method will fail if all the coefficients of λ vanish at the extremum, $\partial\varphi/\partial x$, $\partial\varphi/\partial y$, $\partial\varphi/\partial z=0$. It is then impossible to solve for λ.

Note that from the form of Eqs. (17.80) and (17.82), we could identify f as the function taking an extreme value subject to φ, the constraint, or we could identify f as the constraint and φ as the function.

If we have a **set of constraints** φ_k, then Eqs. (17.80) and (17.82) become

$$\frac{\partial f}{\partial x_i}+\sum_k \lambda_k\frac{\partial\varphi_k}{\partial x_i}=0, \qquad i=1,2,\ldots,n,$$

with a separate Lagrange multiplier λ_k for each φ_k.

Example 17.6.1 PARTICLE IN A BOX

As an example of the use of Lagrangian multipliers, consider the quantum mechanical problem of a particle (mass m) in a box. The box is a rectangular parallelepiped with sides a, b, and c. The ground-state energy of the particle is given by

$$E=\frac{h^2}{8m}\left(\frac{1}{a^2}+\frac{1}{b^2}+\frac{1}{c^2}\right). \tag{17.83}$$

We seek the shape of the box that will minimize the energy E, subject to constraint that the volume is constant,

$$V(a,b,c)=abc=k. \tag{17.84}$$

With $f(a,b,c)=E(a,b,c)$ and $\varphi(a,b,c)=abc-k=0$, we obtain

$$\frac{\partial E}{\partial a}+\lambda\frac{\partial\varphi}{\partial a}=-\frac{h^2}{4ma^3}+\lambda bc=0. \tag{17.85}$$

Also,

$$-\frac{h^2}{4mb^3}+\lambda ac=0, \qquad -\frac{h^2}{4mc^3}+\lambda ab=0.$$

Multiplying the first of these expressions by a, the second by b, and the third by c, we have

$$\lambda abc = \frac{h^2}{4ma^2} = \frac{h^2}{4mb^2} = \frac{h^2}{4mc^2}. \tag{17.86}$$

Therefore our solution is

$$a = b = c, \qquad \text{a cube.} \tag{17.87}$$

Notice that λ has not been determined but follows from Eq. (17.86). ■

Example 17.6.2 CYLINDRICAL NUCLEAR REACTOR

A further example is provided by the nuclear reactor theory. Suppose a (thermal) nuclear reactor is to have the shape of a right circular cylinder of radius R and height H. Neutron diffusion theory supplies a constraint:

$$\varphi(R, H) = \left(\frac{2.4048}{R}\right)^2 + \left(\frac{\pi}{H}\right)^2 = \text{constant.}^{13} \tag{17.88}$$

We wish to minimize the volume of the reactor vessel,

$$f(R, H) = \pi R^2 H. \tag{17.89}$$

Application of Eq. (17.82) leads to

$$\begin{aligned} \frac{\partial f}{\partial R} + \lambda \frac{\partial \varphi}{\partial R} &= 2\pi RH - 2\lambda \frac{(2.4048)^2}{R^3} = 0, \\ \frac{\partial f}{\partial H} + \lambda \frac{\partial \varphi}{\partial H} &= \pi R^2 - 2\lambda \frac{\pi^2}{H^3} = 0. \end{aligned} \tag{17.90}$$

By multiplying the first of these equations by $R/2$ and the second by H, we obtain

$$\pi R^2 H = \lambda \frac{(2.4048)^2}{R^2} = \lambda \frac{2\pi^2}{H^2}, \tag{17.91}$$

or height

$$H = \frac{\sqrt{2}\pi R}{2.4048} = 1.847R, \tag{17.92}$$

for the minimum-volume right-circular cylindrical reactor.

Strictly speaking, we have found only an extremum. Its identification as a minimum follows from a consideration of the original equations. ■

[13] 2.4048 . . . is the lowest root of Bessel function $J_0(R)$ (compare Section 11.1).

Exercises

The following problems are to be solved by using Lagrangian multipliers.

17.6.1 The ground-state energy of a quantum particle of mass m in a pillbox (right-circular cylinder) is given by

$$E = \frac{\hbar^2}{2m}\left(\frac{(2.4048)^2}{R^2} + \frac{\pi^2}{H^2}\right),$$

in which R is the radius and H is the height of the pillbox. Find the ratio of R to H that will minimize the energy for a fixed volume.

17.6.2 Find the ratio of R (radius) to H (height) that will minimize the total surface area of a right-circular cylinder of fixed volume.

17.6.3 The U.S. Post Office limits first class mail to Canada to a total of 36 inches, length plus girth. Using a Lagrange multiplier, find the maximum volume and the dimensions of a (rectangular parallelepiped) package subject to this constraint.

17.6.4 A thermal nuclear reactor is subject to the constraint

$$\varphi(a,b,c) = \left(\frac{\pi}{a}\right)^2 + \left(\frac{\pi}{b}\right)^2 + \left(\frac{\pi}{c}\right)^2 = B^2, \qquad \text{a constant.}$$

Find the ratios of the sides of the rectangular parallelepiped reactor of minimum volume.

ANS. $a = b = c$, cube.

17.6.5 For a lens of focal length f, the object distance p and the image distance q are related by $1/p + 1/q = 1/f$. Find the minimum object–image distance $(p + q)$ for fixed f. Assume real object and image (p and q both positive).

17.6.6 You have an ellipse $(x/a)^2 + (y/b)^2 = 1$. Find the inscribed rectangle of maximum-area. Show that the ratio of the area of the maximum-area rectangle to the area of the ellipse is $2/\pi = 0.6366$.

17.6.7 A rectangular parallelepiped is inscribed in an ellipsoid of semiaxes a, b, and c. Maximize the volume of the inscribed rectangular parallelepiped. Show that the ratio of the maximum volume to the volume of the ellipsoid is $2/\pi\sqrt{3} \approx 0.367$.

17.6.8 A **deformed** sphere has a radius given by $r = r_0\{\alpha_0 + \alpha_2 P_2(\cos\theta)\}$, where $\alpha_0 \approx 1$ and $|\alpha_2| \ll |\alpha_0|$. From Exercise 12.5.16 the area and volume are

$$A = 4\pi r_0^2 \alpha_0^2 \left\{1 + \frac{4}{5}\left(\frac{\alpha_2}{\alpha_0}\right)^2\right\}, \qquad V = \frac{4\pi r_0^3}{3} a_0^3 \left\{1 + \frac{3}{5}\left(\frac{\alpha_2}{\alpha_0}\right)^2\right\}.$$

Terms of order α_2^3 have been neglected.

(a) With the constraint that the enclosed volume be held constant, that is, $V = 4\pi r_0^3/3$, show that the bounding surface of minimum area is a sphere ($\alpha_0 = 1$, $\alpha_2 = 0$).

(b) With the constraint that the area of the bounding surface be held constant, that is, $A = 4\pi r_0^2$, show that the enclosed volume is a maximum when the surface is a sphere.

17.6.9 Find the maximum value of the directional derivative of $\varphi(x, y, z)$,

$$\frac{d\varphi}{ds} = \frac{\partial\varphi}{\partial x}\cos\alpha + \frac{\partial\varphi}{\partial y}\cos\beta + \frac{\partial\varphi}{\partial z}\cos\gamma,$$

subject to the constraint

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1.$$

$$\text{ANS. } \left(\frac{d\varphi}{ds}\right) = |\nabla\varphi|.$$

Note concerning the following exercises: In a quantum mechanical system there are g_i distinct quantum states between energies E_i and $E_i + dE_i$. The problem is to describe how n_i particles are distributed among these states subject to two constraints:

(a) fixed number of particles,

$$\sum_i n_i = n.$$

(b) fixed total energy,

$$\sum_i n_i E_i = E.$$

17.6.10 For identical particles obeying the Pauli exclusion principle, the probability of a given arrangement is

$$W_{FD} = \prod_i \frac{g_i!}{n_i!(g_i - n_i)!}.$$

Show that maximizing W_{FD}, subject to a fixed number of particles and fixed total energy, leads to

$$n_i = \frac{g_i}{e^{\lambda_1 + \lambda_2 E_i} + 1}.$$

With $\lambda_1 = -E_0/kT$ and $\lambda_2 = 1/kT$, this yields Fermi–Dirac statistics.
Hint. Try working with $\ln W$ and using Stirling's formula, Section 8.3. The justification for **differentiation** with respect to n_i is that we are dealing here with a large number of particles, $\Delta n_i/n_i \ll 1$.

17.6.11 For identical particles but no restriction on the number in a given state, the probability of a given arrangement is

$$W_{BE} = \prod_i \frac{(n_i + g_i - 1)!}{n_i!(g_i - 1)!}.$$

Show that maximizing W_{BE}, subject to a fixed number of particles and fixed total energy, leads to

$$n_i = \frac{g_i}{e^{\lambda_1 + \lambda_2 E_i} - 1}.$$

With $\lambda_1 = -E_0/kT$ and $\lambda_2 = 1/kT$, this yields Bose–Einstein statistics.
Note. Assume that $g_i \gg 1$.

17.6.12 Photons satisfy W_{BE} and the constraint that total energy is constant. They clearly do **not** satisfy the fixed-number constraint. Show that eliminating the fixed-number constraint leads to the foregoing result but with $\lambda_1 = 0$.

17.7 VARIATION WITH CONSTRAINTS

As in the preceding sections, we seek the path that will make the integral

$$J = \int f\left(y_i, \frac{\partial y_i}{\partial x_j}, x_j\right) dx_j \tag{17.93}$$

stationary. This is the general case in which x_j represents a set of independent variables and y_i a set of dependent variables. Again,

$$\delta J = 0. \tag{17.94}$$

Now, however, we introduce one or more constraints. This means that the y_i are no longer independent of each other. Not all the η_i may be varied arbitrarily, and Eqs. (17.62) and (17.73a) would not apply. The constraint may have the form

$$\varphi_k(y_i, x_j) = 0, \tag{17.95}$$

as in Section 17.6. In this case we may multiply by a function of x_j, say, $\lambda_k(x_j)$, and integrate over the same range as in Eq. (17.93) to obtain

$$\int \lambda_k(x_j)\varphi_k(y_i, x_j)\, dx_j = 0. \tag{17.96}$$

Then clearly

$$\delta \int \lambda_k(x_j)\varphi_k(y_i, x_j)\, dx_j = 0. \tag{17.97}$$

Alternatively, the constraint may appear in the form of an integral

$$\int \varphi_k(y_i, \partial y_i/\partial x_j, x_j)\, dx_j = \text{constant}. \tag{17.98}$$

We may introduce any **constant** Lagrangian multiplier, and again Eq. (17.97) follows—now with λ a constant.

In either case, by adding Eqs. (17.94) and (17.97), possibly with more than one constraint, we obtain

$$\delta \int \left[f\left(y_i, \frac{\partial y_i}{\partial x_j}, x_j\right) + \sum_k \lambda_k \varphi_k(y_i, x_j) \right] dx_j = 0. \tag{17.99}$$

The Lagrangian multiplier λ_k may depend on x_j when $\varphi(y_i, x_j)$ is given in the form of Eq. (17.95).

Treating the entire integrand as a new function,

$$g\left(y_i, \frac{\partial y_i}{\partial x_j}, x_j\right),$$

we obtain

$$g\left(y_i, \frac{\partial y_i}{\partial x_j}, x_j\right) = f + \sum_k \lambda_k \varphi_k. \tag{17.100}$$

If we have N y_i $(i = 1, 2, \ldots, N)$ and m constraints $(k = 1, 2, \ldots m)$, then $N - m$ of the η_i may be taken as arbitrary. For the remaining m η_i, the λ may, in principle, be chosen so that the remaining Euler–Lagrange equations are satisfied, completely analogous to Eq. (17.80). The result is that our composite function g must satisfy the usual Euler–Lagrange equations,

$$\frac{\partial g}{\partial y_i} - \sum_j \frac{\partial}{\partial x_j} \frac{\partial g}{(\partial y_i / \partial x_j)} = 0, \tag{17.101}$$

with one such equation for each dependent variable y_i (compare Eqs. (17.64) and (17.73)). These Euler equations and the equations of constraint are then solved simultaneously to find the function yielding a stationary value.

Lagrangian Equations

In the absence of constraints, Lagrange's equations of motion (Eq. (17.52)) were found to be[14]

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0,$$

with t (time) the one independent variable and $q_i(t)$ (particle positions) a set of dependent variables. Usually the generalized coordinates q_i are chosen to eliminate the forces of constraint, but this is not necessary and not always desirable. In the presence of (holonomic) constraints, $\varphi_k = 0$, Hamilton's principle is

$$\delta \int \left[L(q_i, \dot{q}_i, t) + \sum_k \lambda_k(t) \varphi_k(q_i, t) \right] dt = 0, \tag{17.102}$$

and the constrained Lagrangian equations of motion are

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = \sum_k a_{ik} \lambda_k. \tag{17.103}$$

Usually $\varphi_k = \varphi_k(q_i, t)$, independent of the generalized velocities $\dot{q}_i$. In this case the coefficient a_{ik} is given by

$$a_{ik} = \frac{\partial \varphi_k}{\partial q_i}. \tag{17.104}$$

Then $a_{ik}\lambda_k$ (no summation) represents the force of the kth constraint in the q_i-direction, appearing in Eq. (17.103) in exactly the same way as $-\partial V / \partial q_i$.

[14]The symbol q is customary in classical mechanics. It serves to emphasize that the variable is not necessarily a Cartesian variable (and not necessarily a length).

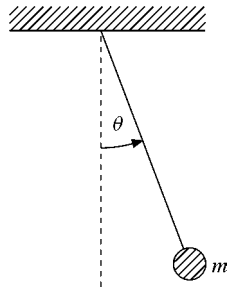

FIGURE 17.10
Simple pendulum.

Example 17.7.1 SIMPLE PENDULUM

To illustrate, consider the simple pendulum, a mass m constrained by a wire of length l to swing in an arc (Fig. 17.10). In the absence of the one constraint

$$\varphi_1 = r - l = 0 \tag{17.105}$$

there are two generalized coordinates r and θ (motion in vertical plane). The Lagrangian is

$$L = T - V = \tfrac{1}{2}m\left(\dot{r}^2 + r^2\dot{\theta}^2\right) + mgr\cos\theta, \tag{17.106}$$

taking the potential V to be zero when the pendulum is horizontal, $\theta = \pi/2$. By Eq. (17.103) the equations of motion are

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{r}} - \frac{\partial L}{\partial r} = \lambda_1, \qquad \frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}} - \frac{\partial L}{\partial \theta} = 0 \qquad (a_{r1} = 1,\ a_{\theta 1} = 0), \tag{17.107}$$

or

$$\begin{aligned} \frac{d}{dt}(m\dot{r}) - mr\dot{\theta}^2 - mg\cos\theta &= \lambda_1, \\ \frac{d}{dt}\left(mr^2\dot{\theta}\right) + mgr\sin\theta &= 0. \end{aligned} \tag{17.108}$$

Substituting in the equation of constraint ($r = l, \dot{r} = 0$), we have

$$ml\dot{\theta}^2 + mg\cos\theta = -\lambda_1, \qquad ml^2\ddot{\theta} + mgl\sin\theta = 0. \tag{17.109}$$

The second equation may be solved for $\theta(t)$ to yield simple harmonic motion if the amplitude is small ($\sin\theta \sim \theta$), whereas the first equation expresses the tension in the wire in terms of θ and $\dot{\theta}$.

Note that since the equation of constraint, Eq. (17.105), is in the form of Eq. (17.95), the Lagrange multiplier λ may be (and here is) a function of t (or of θ). ■

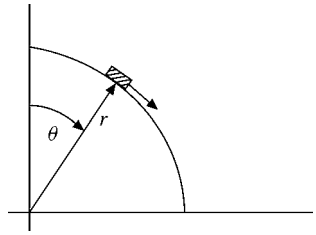

FIGURE 17.11 A particle sliding on a cylindrical surface.

Example 17.7.2 SLIDING OFF A LOG

Closely related to this is the problem of a particle sliding on a cylindrical surface. The object is to find the critical angle θ_c at which the particle flies off from the surface. This critical angle is the angle at which the radial force of constraint goes to zero (Fig. 17.11). We have

$$L = T - V = \tfrac{1}{2}m\left(\dot{r}^2 + r^2\dot{\theta}^2\right) - mgr\cos\theta \tag{17.110}$$

and the one equation of constraint,

$$\varphi_1 = r - l = 0. \tag{17.111}$$

Proceeding as in Example 17.7.1 with $a_{r1} = 1$,

$$\begin{aligned} m\ddot{r} - mr\dot{\theta}^2 + mg\cos\theta &= \lambda_1(\theta), \\ mr^2\ddot{\theta} + 2mr\dot{r}\dot{\theta} - mgr\sin\theta &= 0, \end{aligned} \tag{17.112}$$

in which the constraining force $\lambda_1(\theta)$ is a function of the angle θ.[15] Since $r = l, \ddot{r} = \dot{r} = 0$, Eq. (17.112) reduces to

$$-ml\dot{\theta}^2 + mg\cos\theta = \lambda_1(\theta), \tag{17.113a}$$

$$ml^2\ddot{\theta} - mgl\sin\theta = 0. \tag{17.113b}$$

Differentiating Eq. (17.113a) with respect to time and remembering that

$$\frac{df(\theta)}{dt} = \frac{df(\theta)}{d\theta}\dot{\theta}, \tag{17.114}$$

we obtain

$$-2ml\ddot{\theta} - mg\sin\theta = \frac{d\lambda_1(\theta)}{d\theta}. \tag{17.115}$$

[15]Note that λ_1 is the **radial** force exerted by the cylinder on the particle. Consideration of the physical problem shows that λ_1 must depend on the angle θ. We permitted $\lambda = \lambda(t)$. Now we are replacing the time dependence by an (unknown) angular dependence using $\theta = \theta(t)$.

Using Eq. (17.113b) to eliminate the $\ddot{\theta}$ term and then integrating, we have

$$\lambda_1(\theta) = 3mg\cos\theta + C. \tag{17.116}$$

Since

$$\lambda_1(0) = mg, \tag{17.117}$$

$$C = -2mg. \tag{17.118}$$

The particle m will stay on the surface as long as the force of constraint is nonnegative, that is, as long as the surface has to push outward on the particle:

$$\lambda_1(\theta) = 3mg\cos\theta - 2mg \geq 0. \tag{17.119}$$

The critical angle lies where $\lambda_1(\theta_c) = 0$, the force of constraint going to zero. From Eq. (17.119),

$$\cos\theta_c = \tfrac{2}{3}, \qquad \text{or} \qquad \theta_c = 48°11' \tag{17.120}$$

from the vertical. At this angle (neglecting all friction) our particle takes off.

It must be admitted that this result can be obtained more easily by considering a varying centripetal force furnished by the radial component of the gravitational force. The example was chosen to illustrate the use of Lagrange's undetermined multiplier without confusing the reader with a complicated physical system. ■

Example 17.7.3 THE SCHRÖDINGER WAVE EQUATION

As a final illustration of a constrained minimum, let us find the Euler equations for a quantum mechanical problem

$$\delta \iiint \psi^*(x, y, z) H \psi(x, y, z)\, dx\, dy\, dz = 0, \tag{17.121}$$

with the normalization constraint

$$\iiint \psi^* \psi\, dx\, dy\, dz = 1. \tag{17.122}$$

Equation (17.121) is a statement that the energy of the system is stationary, H being the quantum mechanical Hamiltonian for a particle of mass m, a differential operator,

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V(x, y, z). \tag{17.123}$$

Equation (17.122) is a bound-state constraint, ψ is the usual wave function, a dependent variable, and ψ^*, its complex conjugate, is treated as a **second**[16] dependent variable.

The integrand in Eq. (17.121) involves **second** derivatives, which can be converted to first derivatives by integrating by parts:

$$\int \psi^* \frac{\partial^2 \psi}{\partial x^2}\, dx = \psi^* \frac{\partial \psi}{\partial x}\bigg| - \int \frac{\partial \psi^*}{\partial x}\frac{\partial \psi}{\partial x}\, dx. \tag{17.124}$$

We assume either periodic boundary conditions (as in the Sturm–Liouville theory, Chapter 10) or that the volume of integration is so large that ψ and ψ^* vanish rapidly

[16] Compare Section 6.1.

enough[17] at the boundary. Then the integrated part vanishes and Eq. (17.121) may be rewritten as

$$\delta \iiint \left[\frac{\hbar^2}{2m} \nabla\psi^* \cdot \nabla\psi + V\psi^*\psi \right] dx\, dy\, dz = 0. \tag{17.125}$$

The function g of Eq. (17.100) is

$$\begin{aligned} g &= \frac{\hbar^2}{2m} \nabla\psi^* \cdot \nabla\psi + V\psi^*\psi - \lambda\psi^*\psi \\ &= \frac{\hbar^2}{2m} (\psi_x^*\psi_x + \psi_y^*\psi_y + \psi_z^*\psi_z) + V\psi^*\psi - \lambda\psi^*\psi, \end{aligned} \tag{17.126}$$

again using the subscript x to denote $\partial/\partial x$. For $y_i = \psi^*$, Eq. (17.101) becomes

$$\frac{\partial g}{\partial \psi^*} - \frac{\partial}{\partial x}\frac{\partial g}{\partial \psi_x^*} - \frac{\partial}{\partial y}\frac{\partial g}{\partial \psi_y^*} - \frac{\partial}{\partial z}\frac{\partial g}{\partial \psi_z^*} = 0.$$

This yields

$$V\psi - \lambda\psi - \frac{\hbar^2}{2m}(\psi_{xx} + \psi_{yy} + \psi_{zz}) = 0,$$

or

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V\psi = \lambda\psi. \tag{17.127}$$

Reference to Eq. (17.123) enables us to identify λ physically as the energy of the quantum mechanical system. With this interpretation, Eq. (17.127) is the celebrated Schrödinger wave equation. ■

This variational approach is more than just a matter of academic curiosity. It provides a very powerful method of obtaining approximate solutions of the wave equation (Rayleigh–Ritz variational method, Section 17.8).

Exercises

17.7.1 A particle, mass m, is on a frictionless horizontal surface. It is constrained to move so that $\theta = \omega t$ (rotating radial arm, no friction). With the initial conditions

$$t = 0, \qquad r = r_0, \qquad \dot{r} = 0,$$

(a) find the radial positions as a function of time.

ANS. $r(t) = r_0 \cosh \omega t$.

(b) find the force exerted on the particle by the constraint.

ANS. $F^{(c)} = 2m\dot{r}\omega = 2mr_0\omega^2 \sinh \omega t$.

[17] For example, $\lim_{r\to\infty} r\psi(r) = 0$.

17.7.2 A point mass m is moving over a flat, horizontal, frictionless plane. The mass is constrained by a string to move radially inward at a constant rate. Using plane polar coordinates (ρ, φ), $\rho = \rho_0 - kt$,

(a) Set up the Lagrangian.
(b) Obtain the constrained Lagrange equations.
(c) Solve the φ-dependent Lagrange equation to obtain $\omega(t)$, the angular velocity. What is the physical significance of the constant of integration that you get from your "free" integration?
(d) Using the $\omega(t)$ from part (b), solve the ρ-dependent (constrained) Lagrange equation to obtain $\lambda(t)$. In other words, explain what is happening to the **force** of constraint as $\rho \to 0$.

17.7.3 A flexible cable is suspended from two fixed points. The length of the cable is fixed. Find the curve that will minimize the total gravitational potential energy of the cable.

ANS. Hyperbolic cosine.

17.7.4 A fixed volume of water is rotating in a cylinder with constant angular velocity ω. Find the curve of the water surface that will minimize the total potential energy of the water in the combined gravitational-centrifugal force field.

ANS. Parabola.

17.7.5 (a) Show that for a fixed-length perimeter the figure with maximum area is a circle.
(b) Show that for a fixed area the curve with minimum perimeter is a circle.
Hint. The radius of curvature R is given by

$$R = \frac{(r^2 + r_\theta^2)^{3/2}}{r r_{\theta\theta} - 2r_\theta^2 - r^2}.$$

Note. The problems of this section, variation subject to constraints, are often called *isoperimetric*. The term arose from problems of maximizing area subject to a fixed perimeter — as in Exercise 17.7.5(a).

17.7.6 Show that requiring J, given by

$$J = \int_a^b \left[p(x) y_x^2 - q(x) y^2\right] dx,$$

to have a stationary value subject to the normalizing condition

$$\int_a^b y^2 w(x)\, dx = 1$$

leads to the Sturm–Liouville equation of Chapter 10:

$$\frac{d}{dx}\left(p\frac{dy}{dx}\right) + qy + \lambda w y = 0.$$

Note. The boundary condition

$$p y_x y\Big|_a^b = 0$$

is used in Section 10.1 in establishing the Hermitian property of the operator.

17.7.7 Show that requiring J, given by

$$J = \int_a^b \int_a^b K(x,t)\varphi(x)\varphi(t)\,dx\,dt,$$

to have a stationary value subject to the normalizing condition

$$\int_a^b \varphi^2(x)\,dx = 1$$

leads to the Hilbert–Schmidt integral equation, Eq. (16.89).
Note. The kernel $K(x,t)$ is symmetric.

17.8 RAYLEIGH–RITZ VARIATIONAL TECHNIQUE

Exercise 17.7.6 opens up a relation between the calculus of variations and eigenfunction–eigenvalue problems. We may rewrite the expression of Exercise 17.7.6 as

$$F\big[y(x)\big] = \frac{\int_a^b (py_x^2 - qy^2)\,dx}{\int_a^b y^2 w\,dx}, \tag{17.128}$$

in which the constraint appears in the denominator as a normalizing condition. After the unconstrained minimum of F has been found, y can be normalized without changing the stationary value of F because stationary values of J correspond to stationary values of F. Then from Exercise 17.7.6, when $y(x)$ is such that J and F take on a stationary value, the optimum function $y(x)$ satisfies the Sturm–Liouville equation

$$\frac{d}{dx}\left(p\frac{dy}{dx}\right) + qy + \lambda wy = 0, \tag{17.129}$$

with λ the eigenvalue (**not** a Lagrangian multiplier). Integrating the first term in the numerator of Eq. (17.128) by parts and using the **boundary condition**,

$$py_x y\Big|_a^b = 0, \tag{17.130}$$

we obtain

$$F\big[y(x)\big] = -\int_a^b y\left\{\frac{d}{dx}\left(p\frac{dy}{dx}\right) + qy\right\}dx \Big/ \int_a^b y^2 w\,dx. \tag{17.131}$$

Then substituting in Eq. (17.129), the stationary values of $F[y(x)]$ are given by

$$F\big[y(x)\big] = \lambda_n, \tag{17.132}$$

with λ_n the eigenvalue corresponding to the eigenfunction y_n. Equation (17.132) with F given by either Eq. (17.128) or Eq. (17.131) forms the basis of the Rayleigh–Ritz method for the computation of eigenfunctions and eigenvalues.

Ground State Eigenfunction

Suppose that we seek to compute the ground-state eigenfunction y_0 and eigenvalue[18] λ_0 of some complicated atomic or nuclear system. The classical example, for which no exact solution exists, is the helium atom problem. The eigenfunction y_0 is **unknown**, but we shall assume we can make a pretty good guess at an approximate function y, so mathematically we may write[19]

$$y = y_0 + \sum_{i=1}^{\infty} c_i y_i. \tag{17.133}$$

The c_i are small quantities. (How small depends on how good our guess was.) The y_i are orthonormalized eigenfunctions (also unknown), and therefore our trial function y is not normalized.

Substituting the approximate function y into Eq. (17.131) and noting that

$$\int_a^b y_i \left\{ \frac{d}{dx}\left(p\frac{dy_j}{dx} \right) + q y_i \right\} dx = -\lambda_i \delta_{ij}, \tag{17.134}$$

$$F\big[y(x)\big] = \frac{\lambda_0 + \sum_{i=1}^{\infty} c_i^2 \lambda_i}{1 + \sum_{i=1}^{\infty} c_i^2}. \tag{17.135}$$

Here we have taken the eigenfunctions to be orthonormal — since they are solutions of the Sturm–Liouville equation, Eq. (17.129). We also assume that y_0 is nondegenerate. Now, if we replace $\sum_i c_i^2 \lambda_i \to \sum_i c_i^2 \lambda_0 + \sum_i c_i^2 (\lambda_i - \lambda_0)$ we obtain

$$F\big[y(x)\big] = \lambda_0 + \frac{\sum_{i=1}^{\infty} c_i^2 (\lambda_i - \lambda_0)}{1 + \sum_{i=1}^{\infty} c_i^2}. \tag{17.136}$$

Equation (17.136) contains two important results.

- Whereas the error in the eigenfunction y was $\mathcal{O}(c_i)$, the error in λ is only $\mathcal{O}(c_i^2)$. Even a poor approximation of the eigenfunctions may yield an accurate calculation of the eigenvalue.
- If λ_0 is the lowest eigenvalue (ground state), then since $\lambda_i - \lambda_0 > 0$,

  $$F\big[y(x)\big] = \lambda \geq \lambda_0, \tag{17.137}$$

  or our approximation is always on the high side and becoming lower, converging on λ_0 as our approximate eigenfunction y improves ($c_i \to 0$). Note that Eq. (17.137) is a direct consequence of Eq. (17.135). More directly, $F[y(x)]$ in Eq. (17.135) is a positively weighted average of the λ_i and, therefore, must be no smaller than the smallest λ_i, to wit, λ_0. In practical problems in quantum mechanics, y often depends on parameters that may be varied to minimize F and thereby improve the estimate of the ground-state energy λ_0. This is the "variational method" discussed in quantum mechanics texts.

[18] This means that λ_0 is the lowest eigenvalue. It is clear from Eq. (17.128) that if $p(x) \geq 0$ and $q(x) \leq 0$ (compare Table 10.1), then $F[y(x)]$ has a lower bound and this lower bound is nonnegative. Recall from Section 10.1 that $w(x) \geq 0$.

[19] We are guessing at the form of the function. The normalization is irrelevant.

Example 17.8.1 VIBRATING STRING

A vibrating string, clamped at $x = 0$ and 1, satisfies the eigenvalue equation

$$\frac{d^2y}{dx^2} + \lambda y = 0 \tag{17.138}$$

and the boundary condition $y(0) = y(1) = 0$. For this simple example we recognize immediately that $y_0(x) = \sin \pi x$ (unnormalized) and $\lambda_0 = \pi^2$. But let us try out the Rayleigh–Ritz technique.

With one eye on the boundary conditions, we try

$$y(x) = x(1 - x). \tag{17.139}$$

Then with $p = 1$ and $w = 1$, Eq. (17.128) yields

$$F[y(x)] = \frac{\int_0^1 (1 - 2x)^2\, dx}{\int_0^1 x^2(1 - x)^2\, dx} = \frac{1/3}{1/30} = 10. \tag{17.140}$$

This result, $\lambda = 10$, is a fairly good approximation (1.3% error)[20] of $\lambda_0 = \pi^2 = 9.8696$. You may have noted that $y(x)$, Eq. (17.139), is not normalized to unity. The denominator in $F[y(x)]$ compensates for the lack of unit normalization. F may also be calculated from Eq. (17.131) since Eq. (17.130) is satisfied by y from Eq. (17.139).

In the usual scientific calculation the eigenfunction would be improved by introducing more terms and adjustable parameters, such as

$$y = x(1 - x) + a_2x^2(1 - x)^2. \tag{17.141}$$

It is convenient to have the additional terms orthogonal, but it is not necessary. The parameter a_2 is adjusted to **minimize** $F[y(x)]$. In this case, choosing $a_2 = 1.1353$ drives $F[y(x)]$ down to 9.8697, very close to the correct eigenvalue value. ■

Exercises

17.8.1 From Eq. (17.128) develop in detail the argument when $\lambda \geq 0$ or $\lambda < 0$. Explain the circumstances under which $\lambda = 0$, and illustrate with several examples.

17.8.2 An unknown function satisfies the differential equation

$$y'' + \left(\frac{\pi}{2}\right)^2 y = 0$$

and the boundary conditions

$$y(0) = 1, \qquad y(1) = 0.$$

[20] The closeness of the fit may be checked by a Fourier sine expansion (compare Exercise 14.2.3 over the half-interval $[0, 1]$ or, equivalently, over the interval $[-1, 1]$, with $y(x)$ taken to be odd). Because of the even symmetry relative to $x = 1/2$, only odd n terms appear:

$$y(x) = x(1 - x) = \left(\frac{8}{\pi^3}\right)\left[\sin \pi x + \frac{\sin 3\pi x}{3^3} + \frac{\sin 5\pi x}{5^3} + \cdots\right].$$

(a) Calculate the approximation

$$\lambda = F[y_{\text{trial}}]$$

for

$$y_{\text{trial}} = 1 - x^2.$$

(b) Compare with the exact eigenvalue.

ANS. (a) $\lambda = 2.5$, (b) $\lambda/\lambda_{\text{exact}} = 1.013$.

17.8.3 In Exercise 17.8.2 use a trial function

$$y = 1 - x^n.$$

(a) Find the value of n that will minimize $F[y_{\text{trial}}]$.
(b) Show that the optimum value of n drives the ratio $\lambda/\lambda_{\text{exact}}$ down to 1.003.

ANS. (a) $n = 1.7247$.

17.8.4 A quantum mechanical particle in a sphere (Example 11.7.1) satisfies

$$\nabla^2\psi + k^2\psi = 0,$$

with $k^2 = 2mE/\hbar^2$. The boundary condition is that $\psi(r = a) = 0$, where a is the radius of the sphere. For the ground state [where $\psi = \psi(r)$] try an approximate wave function

$$\psi_a(r) = 1 - \left(\frac{r}{a}\right)^2$$

and calculate an approximate eigenvalue k_a^2.
Hint. To determine $p(r)$ and $w(r)$, put your equation in self-adjoint form (in spherical polar coordinates).

ANS. $k_a^2 = \dfrac{10.5}{a^2}$, $\quad k_{\text{exact}}^2 = \dfrac{\pi^2}{a^2}$.

17.8.5 The wave equation for the quantum mechanical oscillator may be written as

$$\frac{d^2\psi(x)}{dx^2} + \left(\lambda - x^2\right)\psi(x) = 0,$$

with $\lambda = 1$ for the ground state (Eq. (13.18)). Take

$$\psi_{\text{trial}} = \begin{cases} 1 - \frac{x^2}{a^2}, & x^2 \le a^2 \\ 0, & x^2 > a^2 \end{cases}$$

for the ground-state wave function (with a^2 an adjustable parameter) and calculate the corresponding ground-state energy. How much error do you have?
Note. Your parabola is really not a very good approximation to a Gaussian exponential. What improvements can you suggest?

17.8.6 The Schrödinger equation for a central potential may be written as

$$\mathcal{L}u(r) + \frac{\hbar^2 l(l+1)}{2Mr^2}u(r) = Eu(r).$$

The $l(l+1)$ term, the angular momentum barrier, comes from splitting off the angular dependence (Section 9.3). Treating this term as a perturbation, use your variational technique to show that $E > E_0$, where E_0 is the energy eigenvalue of $\mathcal{L}u_0 = E_0 u_0$ corresponding to $l = 0$. This means that the minimum energy state will have $l = 0$, zero angular momentum.
Hint. You can expand $u(r)$ as $u_0(r) + \sum_{i=1}^{\infty} c_i u_i$, where $\mathcal{L}u_i = E_i u_i$, $E_i > E_0$.

17.8.7 In the matrix eigenvector, eigenvalue equation

$$\mathsf{A}\mathbf{r}_i = \lambda_i \mathbf{r}_i,$$

where λ is an $n \times n$ Hermitian matrix. For simplicity, assume that its n real eigenvalues (Section 3.5) are distinct, λ_1 being the largest. If $\mathbf{r}$ is an approximation to $\mathbf{r}_1$,

$$\mathbf{r} = \mathbf{r}_1 + \sum_{i=2}^{n} \delta_i \, \mathbf{r}_i,$$

show that

$$\frac{\mathbf{r}^\dagger \mathsf{A}\mathbf{r}}{\mathbf{r}^\dagger \mathbf{r}} \leq \lambda_1$$

and that the error in λ_1 is of the order $|\delta_i|^2$. Take $|\delta_i| \ll 1$.
Hint. The n $\mathbf{r}_i$ form a complete orthogonal set spanning the n-dimensional (complex) space.

17.8.8 The variational solution of Example 17.8.1 may be refined by taking $y = x(1-x) + a_2 x^2(1-x)^2$. Using the numerical quadrature, calculate $\lambda_{\text{approx}} = F[y(x)]$, Eq. (17.128), for a fixed value of a_2. Vary a_2 to minimize λ. Calculate the value of a_2 that minimizes λ and calculate λ itself, both to five significant figures. Compare your eigenvalue λ with π^2.

Additional Readings

Bliss, G. A., *Calculus of Variations*. The Mathematical Association of America. LaSalle, IL: Open Court Publishing Co. (1925). As one of the older texts, this is still a valuable reference for details of problems such as minimum-area problems.

Courant, R., and H. Robbins, *What Is Mathematics?* 2nd ed. New York: Oxford University Press (1996). Chapter VII contains a fine discussion of the calculus of variations, including soap film solutions to minimum-area problems.

Lanczos, C., *The Variational Principles of Mechanics*, 4th ed. Toronto: University of Toronto Press (1970), reprinted, Dover (1986). This book is a very complete treatment of variational principles and their applications to the development of classical mechanics.

Sagan, H., *Boundary and Eigenvalue Problems in Mathematical Physics*. New York: Wiley (1961), reprinted, Dover (1989). This delightful text could also be listed as a reference for Sturm–Liouville theory, Legendre and Bessel functions, and Fourier Series. Chapter 1 is an introduction to the calculus of variations, with applications to mechanics. Chapter 7 picks up the calculus of variations again and applies it to eigenvalue problems.

Sagan, H., *Introduction to the Calculus of Variations*. New York: McGraw-Hill (1969), reprinted, Dover (1983). This is an excellent introduction to the modern theory of the calculus of variations, which is more sophisticated and complete than his 1961 text. Sagan covers sufficiency conditions and relates the calculus of variations to problems of space technology.

Weinstock, R., *Calculus of Variations*. New York: McGraw-Hill (1952); New York: Dover (1974). A detailed, systematic development of the calculus of variations and applications to Sturm–Liouville theory and physical problems in elasticity, electrostatics, and quantum mechanics.

Yourgrau, W., and S. Mandelstam, *Variational Principles in Dynamics and Quantum Theory*, 3rd ed. Philadelphia: Saunders (1968); New York: Dover (1979). This is a comprehensive, authoritative treatment of variational principles. The discussions of the historical development and the many metaphysical pitfalls are of particular interest.

This page intentionally left blank

CHAPTER 18

NONLINEAR METHODS AND CHAOS

Our mind would lose itself in the complexity of the world if that complexity were not harmonious; like the short–sighted, it would only see the details, and would be obliged to forget each of these details before examining the next, because it would be incapable of taking in the whole. The only facts worthy of our attention are those which introduce order into this complexity and so make it accessible to us.

HENRI POINCARÉ

18.1 INTRODUCTION

The origin of nonlinear dynamics goes back to the work of the renowned French mathematician Henri Poincaré on celestial mechanics at the turn of the twentieth century. Classical mechanics is, in general, nonlinear in its dependence on the coordinates of the particles and the velocities, one example being vibrations with a nonlinear restoring force. The Navier–Stokes equations are nonlinear, which makes hydrodynamics difficult to handle. For almost four centuries however, following the lead of Galileo, Newton, and others, physicists have focused on predictable, effectively linear responses of classical systems, which usually have linear and nonlinear properties.

Poincaré was the first to understand the possibility of completely irregular, or "chaotic," behavior of solutions of nonlinear differential equations that are characterized by an extreme sensitivity to initial conditions: Given slightly different initial conditions, from errors in measurements for example, solutions can grow exponentially apart with time, so the system soon becomes effectively unpredictable, or "chaotic." This property of chaos, often called the "butterfly" effect, will be discussed in Section 18.3. Since the rediscovery of this effect by Lorenz in meteorology in the early 1960s, the field of nonlinear dynamics has grown tremendously. Thus, nonlinear dynamics and chaos theory now have entered the mainstream of physics.

Numerous examples of nonlinear systems have been found to display irregular behavior. Surprisingly, order, in the sense of quantitative similarities as universal properties, or other regularities may arise spontaneously in chaos; a first example. Feigenbaum's universal numbers α and δ will appear in Section 18.2. Dynamical chaos is not a rare phenomenon but is ubiquitous in nature. It includes irregular shapes of clouds, coast lines, and other landscapes, which are examples of fractals, to be discussed in Section 18.3, and turbulent flow of fluids, water dripping from a faucet, and the weather, of course. The damped, driven pendulum is among the simplest systems displaying chaotic motion.

Necessary conditions for chaotic motion in dynamical systems described by **first-order** differential equations are

- at least three dynamical variables, and
- one or more nonlinear terms coupling two or several of them.

As in classical mechanics, the space of the time-dependent dynamical variables of a system of coupled differential equations is called its **phase space**. In such deterministic systems, trajectories in phase space are not allowed to cross. If they did, the system would have a choice at each intersection and would not be deterministic. In two dimensions such nonlinear systems allow only for fixed points. An example is a damped pendulum, whose second derivative, $\ddot{\theta} = f(\dot{\theta}, \theta)$, can be written as two first-order derivatives, $\omega = \dot{\theta}$, $\dot{\omega} = f(\omega, \theta)$, involving just two dynamic variables, $\omega(t)$ and $\theta(t)$. In the undamped case, there will only be periodic motion and equilibrium points. With three or more dynamic variables (for example, **damped**, **driven** pendulum, written as first-order coupled ODEs again), more complicated nonintersecting trajectories are possible. These include chaotic motion and are called **deterministic chaos**.

A central theme in chaos is the evolution of **complex** forms from the repetition of **simple** but **nonlinear** operations; this is being recognized as a **fundamental organizing principle of nature**. While nonlinear differential equations are a natural place in physics for chaos to occur, the mathematically simpler iteration of nonlinear functions provides a quicker entry to chaos theory, which we will pursue first in Section 18.2. In this context, chaos already arises in certain nonlinear functions of a **single** variable.

18.2 THE LOGISTIC MAP

The nonlinear one-dimensional iteration, or difference equation,

$$x_{n+1} = \mu x_n(1 - x_n), \qquad x_n \in [0, 1]; \qquad 1 < \mu < 4, \tag{18.1}$$

is called the **logistic map**. It is patterned after the nonlinear differential equation $dx/dt = \mu x(1 - x)$, used by P. F. Verhulst in 1845 to model the development of a breeding population whose generations do not overlap. The density of the population at time n is x_n. The linear term simulates the birth rate and the nonlinear term the death rate of the species in a constant environment controlled by the parameter μ.

The quadratic function $f_\mu(x) = \mu x(1 - x)$ is chosen because it has one maximum in the interval $[0, 1]$ and is zero at the endpoints, $f_\mu(0) = 0 = f_\mu(1)$. The maximum at $x_m = 1/2$

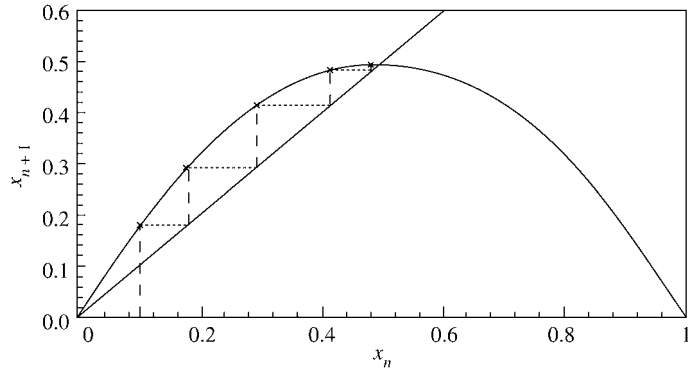

FIGURE 18.1 Cycle $(x_0, x_1, \ldots)$ for the logistic map for $\mu = 2$, starting value $x_0 = 0.1$ and attractor $x^* = 1/2$.

is determined from $f'(x) = 0$, that is,

$$f'_\mu(x_m) = \mu(1 - 2x_m) = 0, \qquad x_m = \frac{1}{2}, \tag{18.2}$$

where $f_\mu(1/2) = \mu/4$.

- Varying the single parameter μ controls a rich and complex behavior, including one-dimensional chaos, as we shall see. More parameters or additional variables are hardly necessary at this point to increase the complexity. In a rather qualitative sense the simple **logistic** map of Eq. (18.1) is representative of many dynamical systems in biology, chemistry, and physics.

Figure (18.1) shows a plot of $f_\mu(x) = \mu x(1 - x)$ along with the diagonal and a series of points $(x_0, x_1, \ldots)$ called a **cycle**. To construct a cycle for a fixed value of μ $(= 2$ in Fig. 18.1), we choose some $x_0 \in [0, 1]$ $[x_0 = 0.1$ in Eq. (18.1)]. The vertical line through x_0 intersects the curve $f_\mu(x)$ at $x_1 = f_\mu(x_0)$ $(= 0.18$ in Fig. 18.1). Proceeding horizontally from x_1 leads us to x_1 on the diagonal. Going vertically from the abscissa x_1 gives $x_2 = f_\mu(x_1)$ on the curve $(x_2 = 0.2952$ in Fig. 18.1), etc. That is, straight vertical lines show the intersections with the curve f_μ and horizontal lines convert $f_\mu(x_i) = x_{i+1}$ to the next abscissa.

For any initial value x_0 with $0 < x_0 < 1$, the x_i converge toward the fixed point x^*, or **attractor** $[= (0.5, 0.5)$ in Fig. 18.1]:

$$\boxed{f_\mu(x^*) = \mu x^*(1 - x^*) = x^*,} \qquad \text{i.e., } x^* = 1 - \frac{1}{\mu}. \tag{18.3}$$

The interval $(0, 1)$ defines a **basin of attraction** for the fixed point x^*. The attractor x^* is **stable** provided the slope $|f'_\mu(x^*)| = |2 - \mu| < 1$, or $1 < \mu < 3$. This can be seen from a Taylor expansion of an iteration near the attractor:

$$x_{n+1} = f_\mu(x_n) = f_\mu(x^*) + f'_\mu(x^*)(x_n - x^*) + \cdots, \qquad \text{i.e., } \frac{x_{n+1} - x^*}{x_n - x^*} = f'_\mu(x^*),$$

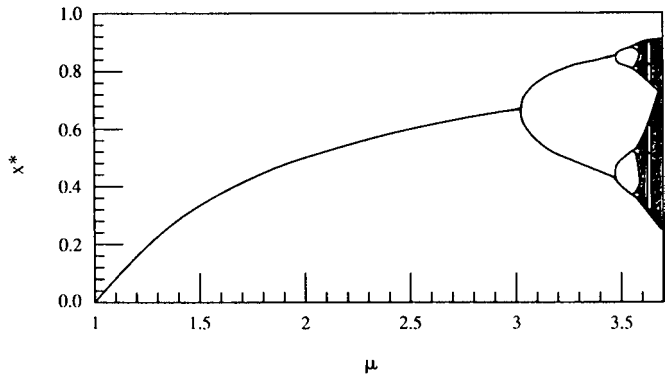

FIGURE 18.2 Part of the bifurcation plot for the logistic map: fixed points x^* versus μ.

upon dropping all higher-order terms. Thus, if $|f'_\mu(x^*)| < 1$, the next iterate, x_{n+1}, lies closer to x^* than does x_n, implying convergence to and stability of the fixed point. However, if $|f'_\mu(x^*)| > 1$, x_{n+1} moves farther from x^* than does x_n implying divergence and instability. Given the continuity of f'_μ in μ, the fixed point and its properties persist when the parameter (here μ) is slightly varied.

For $\mu > 1$ and $x_0 < 0$ or $x_0 > 1$, it is easy to verify graphically or analytically that the $x_i \to -\infty$. The origin, $x = 0$, is a **repellent** fixed point since $f'_\mu(0) = \mu > 1$ and the iterates move away from it. Since $f'_\mu(1) = -\mu$, the point $x = 1$ is a repellor for $\mu > 1$.

When

$$f'_\mu(x^*) = \mu(1 - 2x^*) = 2 - \mu = -1$$

is reached for $\mu = 3$, **two fixed points occur**, shown as the two branches in Fig. 18.2, as μ increases beyond the value 3. They can be located by solving

$$x_2^* = f_\mu\big(f_\mu(x_2^*)\big) = \mu^2 x_2^*(1 - x_2^*)\big[1 - \mu x_2^*(1 - x_2^*)\big]$$

for x_2^*. Here it is convenient to abbreviate $f^{(1)}(x) = f_\mu(x)$, $f^{(2)}(x) = f_\mu(f_\mu(x))$ for the second iterate, etc. Now we drop the common x_2^* and then reduce the remaining third-order polynomial to second-order by recalling that a fixed point of f_μ is also a fixed point of $f^{(2)}$ because $f_\mu(f_\mu(x^*)) = f_\mu(x^*) = x^*$. So $x_2^* = x^*$ is one solution. Factoring out the quadratic polynomial we obtain

$$\begin{aligned} 0 &= \mu^2\big[1 - (\mu + 1)x_2^* + 2(x_2^*)^2 - \mu(x_2^*)^3\big] - 1 \\ &= (\mu - 1 - \mu x_2^*)\big[\mu + 1 - \mu(\mu + 1)x_2^* + \mu^2(x_2^*)^2\big]. \end{aligned}$$

The roots of the quadratic polynomial are

$$x_2^* = \frac{1}{2\mu}\big(\mu + 1 \pm \sqrt{(\mu + 1)(\mu - 3)}\big),$$

which are the two branches in Fig. 18.2 for $\mu > 3$ starting at $x_2^* = 2/3$. This shows that both fixed points bifurcate at the same value of μ. Each x_2^* is a point of period 2 and invariant

under two iterations of the map f_μ. The iterates oscillate between both branches of fixed points x_2^*. A point x_n is defined as a **periodic point of period** n for f_μ if $f^{(n)}(x_0) = x_0$, but $f^{(i)}(x_0) \neq x_0$ for $0 < i < n$. Thus, for $3 < \mu < 3.45$ (see Fig. 18.2) the stable attractor **bifurcates**, or splits, into two fixed points x_2^*. The bifurcation for $\mu = 3$, where the doubling occurs, is called a **pitchfork** bifurcation because of its characteristic (rounded Y-) shape. A bifurcation is a sudden change in the evolution of the system, such as a splitting of one curve into two curves.

As μ increases beyond 3, the derivative $df^{(2)}/dx$ decreases from unity to -1. For $\mu = 1 + \sqrt{6} \sim 3.44949$, which can be derived from

$$\left.\frac{df^{(2)}}{dx}\right|_{x=x^*} = -1,\ f^{(2)}(x^*) = x^*,$$

each branch of fixed points bifurcates again, so $x_4^* = f^{(4)}(x_4^*)$, that is, has period 4. For $\mu = 1 + \sqrt{6}$ these are $x_4^* = 0.43996$ and $x_4^* = 0.849938$.

With increasing period doublings it becomes impossible to obtain analytic solutions. The iterations are better done numerically on a programmable pocket calculator or a personal computer, whose rapid improvements (computer-driven graphics, in particular) and wide distribution since the 1970s has accelerated the development of chaos theory. The sequence of bifurcations continues with ever longer periods until we reach $\mu_\infty = 3.5699456\ldots$, where an infinite number of bifurcations occur. Near bifurcation points, fluctuations, rounding errors in initial conditions, etc., play an increasing role because the system has to choose between two possible branches and becomes much more sensitive to small perturbations. In the present case the x_n never repeat. The bands of fixed points x^* begin forming a continuum (shown dark in Fig. 18.2); this is where chaos starts. This increasing period doubling is the route to chaos for the logistic map that is characterized by a universal constant δ, called a **Feigenbaum number**. If the first bifurcation occurs at $\mu_1 = 3$, the second at $\mu_2 = 3.45, \ldots$, then the ratio of spacings between the μ_n converges to δ:

$$\boxed{\lim_{n\to\infty} \frac{\mu_n - \mu_{n-1}}{\mu_{n+1} - \mu_n} = \delta = 4.66920161\ldots.} \tag{18.4}$$

From the bifurcation plot in Fig. 18.2 we obtain

$$\frac{\mu_2 - \mu_1}{\mu_3 - \mu_2} = \frac{3.45 - 3.00}{3.54 - 3.45} = 5.0$$

as a first approximation for the dimensionless δ. The corresponding critical-period-$2n$ points x_n^* lead to another universal and dimensionless quantity:

$$\boxed{\lim_{n\to\infty} \frac{x_n^* - x_{n-1}^*}{x_{n+1}^* - x_n^*} = \alpha = 2.5029\ldots.} \tag{18.5}$$

Again reading off Fig. 18.2's approximate values for x_n^* we obtain

$$\frac{0.44 - 0.67}{0.37 - 0.44} = 3.3$$

as a first approximation for α.

The Feigenbaum number δ is universal for the route to chaos via period doublings for all maps with a **quadratic** maximum similar to the logistic map. It is an example of order in chaos. Experience shows that its validity is even wider, including two-dimensional (dissipative) systems and twice continuously differentiable functions with subharmonic bifurcations.[1] When the maps behave like $|x - x_m|^{1+\varepsilon}$ near their maximum x_m for some ε between 0 and 1, the Feigenbaum number will depend on the exponent ε; thus $\delta(\varepsilon)$ varies between $\delta(1)$ given in Eq. (18.4) for quadratic maps to $\delta(0) = 2$ for $\varepsilon = 0$.[2]

Exercises

18.2.1 Show that $x^* = 1$ is a nontrivial fixed point of the map $x_{n+1} = x_n \exp[r(1 - x_n)]$ with a slope $1 - r$, so that the equilibrium is stable if $0 < r < 2$.

18.2.2 Draw a bifurcation diagram for the exponential map of Exercise 18.2.1 for $r > 1.9$.

18.2.3 Determine fixed points of the cubic map $x_{n+1} = ax_n^3 + (1 - a)x_n$ for $0 < a < 4$ and $0 < x_n < 1$.

18.2.4 Write the time-delayed logistical map $x_{n+1} = \mu x_n(1 - x_{n-1})$ as a two-dimensional map $x_{n+1} = \mu x_n(1 - y_n)$, $y_{n+1} = x_n$, and determine some of its fixed points.

18.2.5 Show that the second bifurcation for the logistical map that leads to cycles of period 4 is located at $\mu = 1 + \sqrt{6}$.

18.2.6 Construct a nonlinear iteration function with Feigenbaum δ in the interval $2 < \delta < 4.6692\ldots$.

18.2.7 Determine the Feigenbaum δ for (a) the exponential map of Exercise 18.2.1, (b) some cubic map of Exercise 18.2.3, (c) the time-delayed logistic map of Exercise 18.2.4.

18.2.8 Repeat Exercise 18.2.7 for Feigenbaum's α instead of δ.

18.2.9 Find numerically the first four points μ for period doubling of the logistic map, and then obtain the first two approximations to the Feigenbaum δ. Compare with Fig. 18.2 and Eq. (18.4).

18.2.10 Find numerically the values μ where the cycle of period 1, 3, 4, 5, 6 begins and then where it becomes unstable.

Check values. For period $\begin{matrix} 3,\ \mu = 3.8284, \\ 4,\ \mu = 3.9601, \\ 5,\ \mu = 3.7382, \\ 6,\ \mu = 3.6265. \end{matrix}$

18.2.11 Repeat Exercise 18.2.9 for Feigenbaum's α.

[1] More details and computer codes for the logistic map are given by G. L. Baker and J. P. Gollub, *Chaotic Dynamics: An Introduction*, Cambridge, UK: Cambridge University Press (1990).

[2] For other maps and a discussion of the fascinating history how chaos became again a hot research topic, see D. Holton and R. M. May in *The Nature of Chaos* (T. Mullin, ed.), Oxford, UK: Clarendon Press (1993), Section 5, p. 95; and Gleick's *Chaos* (1987)—see the Additional Readings.

18.3 SENSITIVITY TO INITIAL CONDITIONS AND PARAMETERS

Lyapunov Exponents

In Section 18.2 we described how, as we approach the period-doubling accumulation parameter value $\mu_\infty = 3.5699\ldots$ from below, the period $n+1$ of cycles $(x_0, x_1, \ldots, x_n)$ with $x_{n+1} = x_0$ gets longer. It is also easy to check that the distances

$$d_n = |f^{(n)}(x_0 + \varepsilon) - f^{(n)}(x_0)| \tag{18.6}$$

grow as well for small $\varepsilon > 0$. From experience with chaotic behavior we find that this distance increases exponentially with $n \to \infty$; that is, $d_n/\varepsilon = e^{\lambda n}$, or

$$\lambda = \frac{1}{n} \ln\left(\frac{|f^{(n)}(x_0 + \varepsilon) - f^{(n)}(x_0)|}{\varepsilon}\right), \tag{18.7}$$

where λ is a **Lyapunov exponent** for the cycle. For $\varepsilon \to 0$ we may rewrite Eq. (18.7) in terms of derivatives as

$$\lambda = \frac{1}{n} \ln\left|\frac{df^{(n)}(x_0)}{dx}\right| = \frac{1}{n} \sum_{i=0}^{n} \ln|f'(x_i)|, \tag{18.8}$$

using the chain rule of differentiation for $df^{(n)}(x)/dx$, where

$$\frac{df^{(2)}(x_0)}{dx} = \frac{df_\mu}{dx}\bigg|_{x=f_\mu(x_0)} \frac{df_\mu}{dx}\bigg|_{x=x_0} = f'_\mu(x_1) f'_\mu(x_0) \tag{18.9}$$

and $f'_\mu = df_\mu/dx$, etc. Our Lyapunov exponent has been calculated at the point x_0, and Eq. (18.8) is exact for one-dimensional maps.

As a measure of the sensitivity of the system to changes in initial conditions, one point is not enough to determine λ in higher-dimensional dynamical systems in general, where the motion often is bounded, so the d_n cannot go to ∞. In such cases, we repeat the procedure for several points on the trajectory and average over them. This way, we obtain the **average Lyapunov exponent** for the sample. This average value is often called and taken as the Lyapunov exponent.

The Lyapunov exponent λ is a quantitative measure of chaos: A one-dimensional iterated function similar to the logistic map has **chaotic** cycles $(x_0, x_1, \ldots)$ for the parameter μ **if the average Lyapunov exponent is positive** for that value of μ. Any such initial point x_0 is called a **strange** or **chaotic attractor** (the shaded region in Fig. 18.2). For cycles of finite period, λ is negative. This is the case for $\mu < 3$, for $\mu < \mu_\infty$, and even in the periodic window at $\mu \sim 3.627$ inside the chaotic region of Fig. 18.2. At bifurcation points, $\lambda = 0$. For $\mu > \mu_\infty$ the Lyapunov exponent is positive, except in the periodic windows, where $\lambda < 0$, and λ grows with μ. In other words, the system becomes more chaotic as the control parameter μ increases.

In the chaos region of the logistic map there is a scaling law for the average Lyapunov exponent (we do not derive it),

$$\lambda(\mu) = \lambda_0(\mu - \mu_\infty)^{\ln 2/\ln \delta}, \tag{18.10}$$

where $\ln 2/\ln\delta \sim 0.445$, δ is the universal Feigenbaum number of Section 18.2, and λ_0 is a constant. This relation (18.10) is reminiscent of a physical observable at a (second-order) phase transition. The exponent in Eq. (18.10) is a universal number; the Lyapunov exponent plays the role of an **order parameter**, while $\mu - \mu_\infty$ is the analog of $T - T_c$, where T_c is the **critical** temperature at which the phase transition occurs.

Fractals

In dissipative chaotic systems (but rarely in conservative Hamiltonian systems) often new geometric objects with intricate shapes appear that are called **fractals** because of their noninteger dimension. Fractals are irregular geometric objects whose dimension is typically not integral and that exist at many scales, so their smaller parts resemble their larger parts. Intuitively a fractal is a set which is (approximately) **self-similar** under magnification. A set of attracting points with noninteger dimension is called a **strange** attractor.

We need a quantitative measure of dimensionality in order to describe fractals. Unfortunately, there are several definitions with usually different numerical values, none of which has yet become a standard. For strictly self-similar, sets, one measure suffices. More complicated (for instance, only approximately self-similar) sets require more measures for their complete description. The simplest is the **box-counting dimension**, due to Kolmogorov and Hausdorff. For a one-dimensional set, we cover the curve by line segments of length R. In two dimensions the boxes are squares of area R^2, in three dimensions cubes of volume R^3, etc. Then we count the number $N(R)$ of boxes needed to cover the set. Letting R go to zero we expect N to scale as $N(R) \sim R^{-d}$. Taking the logarithm the **box-counting dimension** is defined as

$$d \equiv \lim_{R\to 0} \frac{\ln N(R)}{\ln R}. \tag{18.11}$$

For example, in a two-dimensional space a single point is covered by one square, so $\ln N(R) = 0$ and $d = 0$. A finite set of isolated points also has dimension $d = 0$. For a differentiable curve of length L, $N(R) \sim L/R$ as $R \to 0$, so $d = 1$ from Eq. (18.11), as expected.

Let us now construct a more irregular set, the **Koch** curve. We start with a line segment of unit length in Fig. 18.3 and remove the middle third. Then we replace it with two segments of length $1/3$, which form a triangle in Fig. 18.3. We iterate this procedure with each segment ad infinitum. The resulting Koch curve is infinitely long and is nowhere differentiable because of the infinitely many discontinuous changes of slope. At the nth step each line segment has length $R_n = 3^{-n}$ and there are $N(R_n) = 4^n$ segments. Hence its dimension is $d = \ln 4/\ln 3 = 1.26\ldots$, which is more than a curve but less than a surface. Because the Koch curve results from iteration of the first step, it is strictly self-similar.

For the logistic map the box-counting dimension at a period–doubling accumulation point μ_∞ is $0.5388\ldots$, which is a universal number for iterations of functions in one variable with a quadratic maximum. To see roughly how this comes about, consider the pairs of line segments originating from successive bifurcation points for a given parameter μ in the chaos regime (see Fig. 18.2). Imagine removing the interior space from the chaotic bands. When we go to the next bifurcation, the relevant scale parameter is $\alpha = 2.5029\ldots$ from Eq. (18.5). Suppose we need 2^n line segments of length R to cover 2^n bands. In the

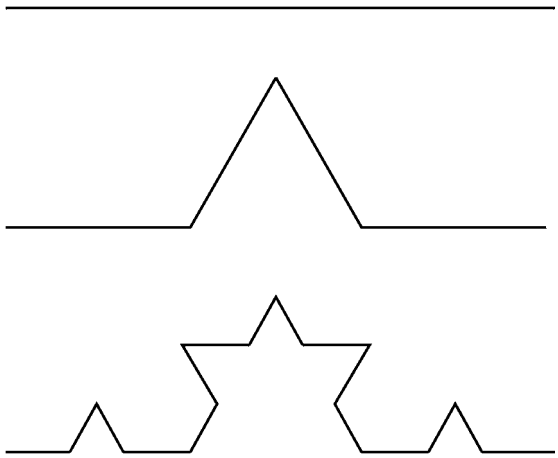

FIGURE 18.3 Construction of the Koch curve by iterations.

next stage then we need 2^{n+1} segments of length R/α to cover the bands. This yields a dimension $d = -\ln(2^n/2^{n+1})/\ln\alpha = 0.4498\ldots$. This crude estimate can be improved by taking into account that the width between neighboring pairs of line segments differs by $1/\alpha$ (see Fig. 18.2). The improved estimate, 0.543, is closer to $0.5388\ldots$. This example suggests that when the fractal set does not have a simple self-similar structure, then the box-counting dimension depends on the box-construction method.

Finally, we turn to the beautiful fractals that are surprisingly easy to generate and whose color pictures had considerable impact. For complex $c = a + ib$, the corresponding quadratic complex map involving the complex variable $z = x + iy$,

$$z_{n+1} = z_n^2 + c, \tag{18.12}$$

looks deceptively simple, but the equivalent two-dimensional map in terms of the real variables

$$x_{n+1} = x_n^2 - y_n^2 + a, \qquad y_{n+1} = 2x_n y_n + b \tag{18.13}$$

reveals already more of its complexity. This map forms the basis for some of Mandelbrot's beautiful multicolor fractal pictures (we refer the reader to Mandelbrot (1988) and Peitgen and Richter (1986) in the Additional Readings), and it has been found to generate rather intricate shapes for various $c \neq 0$. For example, the **Julia set** of a map $z_{n+1} = F(z_n)$ is defined as the set of all its repelling fixed or periodic points. Thus it forms the boundary between initial conditions of a two-dimensional iterated map leading to iterates that diverge and those that stay within some finite region of the complex plane. For the case $c = 0$ and $F(z) = z^2$, the Julia set can be shown to be just a circle about the origin of the complex plane. Yet, just by adding a constant $c \neq 0$, the Julia set becomes fractal. For instance, for $c = -1$ one finds a fractal necklace with infinitely many loops (see Devaney (1989) in the Additional Readings).

While the Julia set is drawn in the complex plane, the **Mandelbrot set** is constructed in the two-dimensional parameter space $c = (a, b) = a + bi$. It is constructed as follows.

Starting from the initial value $z_0 = 0 = (0, 0)$ one searches Eq. (18.12) for parameter values c so that the iterated $\{z_n\}$ do **not** diverge to ∞. Each color outside the fractal boundary of the Mandelbrot set represents a given number of iterations m, say, needed for the z_n to go beyond a specified absolute (real) value $R, |z_m| > R > |z_{m-1}|$. For real parameter value $c = a$, the resulting map, $x_{n+1} = x_n^2 + a$, is equivalent to the logistic map with period-doubling bifurcations (see Section 18.2) as a increases on the real axis inside the Mandelbrot set.

Exercises

18.3.1 Use a programmable pocket calculator (or a personal computer with BASIC or FORTRAN or symbolic software such as Mathematica or Maple) to obtain the iterates x_i of an initial $0 < x_0 < 1$ and $f'_\mu(x_i)$ for the logistic map. Then calculate the Lyapunov exponent for cycles of period $2, 3, \ldots$ of the logistic map for $2 < \mu < 3.7$. Show that for $\mu < \mu_\infty$ the Lyapunov exponent λ is 0 at bifurcation points and negative elsewhere, while for $\mu > \mu_\infty$ it is positive except in periodic windows.
Hint. See Fig. 9.3 of Hilborn (1994) in the Additional Readings.

18.3.2 Consider the map $x_{n+1} = F(x_n)$ with

$$F(x) = \begin{cases} a + bx, & x < 1, \\ c + dx, & x > 1, \end{cases}$$

for $b > 0$ and $d < 0$. Show that its Lyapunov exponent is positive when $b > 1,\ d < -1$. Plot a few iterations in the (x_{n+1}, x_n) plane.

18.4 NONLINEAR DIFFERENTIAL EQUATIONS

In Section 18.1 we mentioned nonlinear differential equations (abbreviated as NDEs) as the natural place in physics for chaos to occur, but continued with the simpler iteration of nonlinear functions of one variable (maps). Here we briefly address the much broader area of NDEs and the far greater complexity in the behavior of their solutions. However, maps and systems of solutions of NDEs are closely related. The latter can often be analyzed in terms of discrete maps. One prescription is the so-called **Poincaré section** of a system of NDE solutions. Placing a plane transverse into a trajectory (of a solution of a NDE), it intersects the plane in a series of points at increasing discrete times, for example, in Fig. 18.4 $(x(t_1), y(t_1)) = (x_1, y_1), (x_2, y_2), \ldots$, which are recorded and graphically or numerically analyzed for fixed points, period-doubling bifurcations, etc. This method is useful when solutions of NDEs are obtained numerically in computer simulations so that one can generate Poincaré sections at various locations and with different orientations, with further analysis leading to two-dimensional iterated maps

$$x_{n+1} = F_1(x_n, y_n), \qquad y_{n+1} = F_2(x_n, y_n) \tag{18.14}$$

stored by the computer. Extracting the functions F_j analytically or graphically is not always easy, though.

Let us start with a few classical examples of NDEs. In Chapter 9 we have already discussed the soliton solution of the nonlinear Korteweg–de Vries PDE, Eq. (9.11).

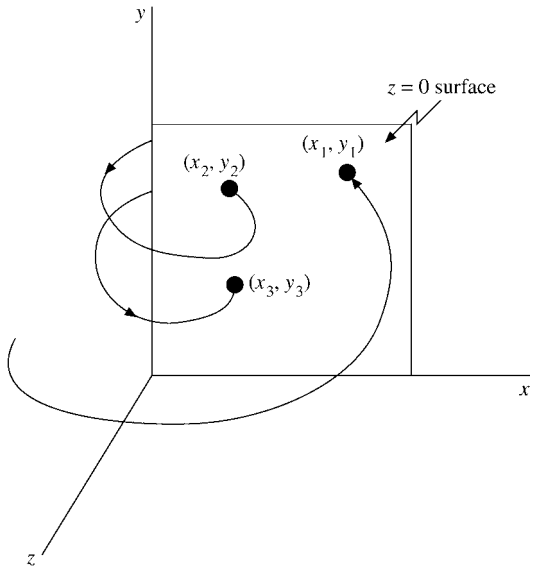

FIGURE 18.4 Schematic of a Poincaré section.

Exercise

18.4.1 For the damped harmonic oscillator

$$\ddot{x} + 2a\dot{x} + x = 0,$$

consider the Poincaré section $\{x > 0, y = \dot{x} = 0\}$. Take $0 < a \ll 1$ and show that the map is given by $x_{n+1} = bx_n$ with $b < 1$. Find an estimate for b.

Bernoulli and Riccati Equations

Bernoulli equations are also nonlinear, having the form

$$y'(x) = p(x)y(x) + q(x)\big[y(x)\big]^n, \tag{18.15}$$

where p and q are real functions and $n \neq 0, 1$ to exclude first-order linear ODEs. If we substitute

$$u(x) = \big[y(x)\big]^{1-n}, \tag{18.16}$$

then Eq. (18.15) becomes a first-order linear ODE,

$$u' = (1-n)y^{-n}y' = (1-n)\big[p(x)u(x) + q(x)\big], \tag{18.17}$$

which we can solve as described in Section 9.2.

Riccati equations are quadratic in $y(x)$:

$$y' = p(x)y^2 + q(x)y + r(x), \tag{18.18}$$

where $p \neq 0$ to exclude linear ODEs and $r \neq 0$ to exclude Bernoulli equations. There is no general method for solving Riccati equations. However, when a special solution $y_0(x)$ of Eq. (18.18) is known by a guess or inspection, then one can write the general solution in the form $y = y_0 + u$, with u satisfying the Bernoulli equation

$$u' = pu^2 + (2py_0 + q)u, \tag{18.19}$$

because substitution of $y = y_0 + u$ into Eq. (18.18) removes $r(x)$ from Eq. (18.18).

Just as for Riccati equations there are no general methods for obtaining exact solutions of other nonlinear ODEs. It is more important to develop methods for finding the qualitative behavior of solutions. In Chapter 9 we mentioned that power-series solutions of ODEs exist except (possibly) at regular or essential singularities, which are directly given by local analysis of the coefficient functions of the ODE. Such local analysis provides us with the asymptotic behavior of solutions as well.

Fixed and Movable Singularities, Special Solutions

Solutions of NDEs also have such singular points, independent of the initial or boundary conditions and called **fixed singularities**. In addition they may have **spontaneous**, or **movable**, singularities that vary with the initial or boundary conditions. They complicate the (asymptotic) analysis of NDEs. This point is illustrated by a comparison of the linear ODE

$$y' + \frac{y}{x-1} = 0, \tag{18.20}$$

which has the obvious regular singularity at $x = 1$, with the NDE $y' = y^2$. Both have the same solution with initial condition $y(0) = 1$, namely, $y(x) = 1/(1-x)$. For $y(0) = 2$, though, the pole in the (obvious, but check) solution $y(x) = 2/(1-2x)$ of the NDE has moved to $x = 1/2$.

For a second-order ODE we have a complete description of (the asymptotic behavior of) its solutions when (that of) two linearly independent solutions are known. For NDEs there may still be **special solutions** whose asymptotic behavior is not obtainable from two independent solutions. This is another **characteristic** property of NDEs, which we illustrate again by an example.

The general solution of the NDE $y'' = yy'/x$ is given by

$$y(x) = 2c_1 \tan(c_1 \ln x + c_2) - 1, \tag{18.21}$$

where c_i are integration constants. An obvious (check it) special solution is $y = c_3 =$ constant, which cannot be obtained from Eq. (18.20) for any choice of the parameters c_1, c_2. Note that using the substitution $x = e^t$, $Y(t) = y(e^t)$ so that $x\,dy/dx = dY/dt$, we obtain the ODE $Y'' = Y'(Y+1)$. This ODE can be integrated once to give $Y' = \frac{1}{2}Y^2 + Y + c$ with $c = 2(c_1^2 + 1/4)$ an integration constant, and again according to Section 9.2 to lead to the solution of Eq. (18.21).

Autonomous Differential Equations

Differential equations that do not explicitly contain the independent variable, taken to be the time t here, are called **autonomous**. Verhulst's NDE $\dot{y} = dy/dt = \mu y(1-y)$, which we encountered briefly in Section 18.2 as motivation for the logistic map, is a special case of this wide and important class of ODEs.[3] For one dependent variable $y(t)$ they can be written as

$$\dot{y} = f(y), \tag{18.22a}$$

and for several dependent variables as a system

$$\dot{y}_i = f_i(y_1, y_2, \ldots, y_n), \qquad i = 1, 2, \ldots, n, \tag{18.22b}$$

with sufficiently differentiable functions f, f_i. A solution of Eq. (18.22b) is a curve or trajectory $y(t)$ for $n = 1$ and in general a trajectory $(y_1(t), y_2(t), \ldots, y_n(t))$ in an n-dimensional (so-called) **phase space**. As discussed already in Section 18.1, two trajectories cannot cross because of the uniqueness of the solutions of ODEs. Clearly, solutions of the algebraic system

$$f_i(y_1, y_2, \ldots, y_n) = 0 \tag{18.23}$$

are special points in phase space, where the position vector $(y_1, y_2, \ldots, y_n)$ does not move on the trajectory; they are called **critical** (or **fixed**) **points**. It turns out that a local analysis of solutions near critical points leads to an understanding of the global behavior of the solutions. First let us look at a simple example.

For Verhulst's ODE, $f(y) = \mu y(1-y) = 0$ gives $y = 0$ and $y = 1$ as the critical points. For the logistic map, $y = 0$ and $y = 1$ are repellent fixed points because $df/dy(0) = \mu$ at $y = 0$ and $df/dy(1) = -\mu$ at $y = 1$ for $\mu > 1$. A local analysis near $y = 0$ suggests neglecting the y^2 term and solving $\dot{y} = \mu y$ instead. Integrating $\int dy/y = \mu t + \ln c$ gives the solution $y(t) = ce^{\mu t}$, which diverges as $t \to \infty$, so $y = 0$ is a repellent critical point. (Note that for $\mu < 0$ of the logistic map the critical point $y = 0$ would be attracting, leading to a converging $y \sim e^{\mu t}$ solution.) Similarly at $y = 1$, $\int dy/(1-y) = \mu t - \ln c$ leads to $y(t) = 1 - ce^{-\mu t} \to 1$ for $t \to \infty$. Hence $y = 1$ is an attracting critical point. Because the ODE is separable, its general solution is given by

$$\int \frac{dy}{y(1-y)} = \int dy \left[\frac{1}{y} + \frac{1}{1-y}\right] = \ln \frac{y}{1-y} = \mu t + \ln c.$$

Hence $y(t) = ce^{\mu t}/(1 + ce^{\mu t})$ for $t \to \infty$ converges to 1, thus confirming the local analysis. This example motivates us to look next at the properties of fixed points in more detail. For an arbitrary function f, it is easy to see that

- in **one dimension**, fixed points y_i with $f(y_i) = 0$ divide the y-axis into dynamically separate intervals because, given an initial value in one of the intervals, the trajectory $y(t)$ will stay there, for it cannot go beyond either fixed point where $\dot{y} = 0$.

[3] Solutions of nonautonomous equations can be much more complicated.

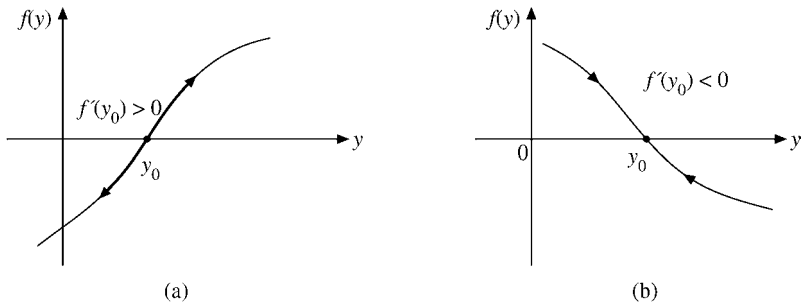

FIGURE 18.5 Fixed points: (a) repellor, (b) sink.

If $f'(y_0) > 0$ at the fixed point y_0 where $f(y_0) = 0$, then at $y_0 + \varepsilon$ for $\varepsilon > 0$ sufficiently small, $\dot{y} = f'(y_0)\varepsilon + \mathcal{O}(\varepsilon^2) > 0$ in a neighborhood to the right of y_0, so the trajectory $y(t)$ keeps moving to the right, away from the fixed point y_0. To the left of y_0, $\dot{y} = -f'(y_0)\varepsilon + \mathcal{O}(\varepsilon^2) < 0$, so the trajectory moves away from the fixed point here as well. Hence,

- a fixed point [with $f(y_0) = 0$] at y_0 with $f'(y_0) > 0$, as shown in Fig. 18.5a, repels trajectories; that is, all trajectories move away from the critical point: $\cdots \leftarrow \cdot \rightarrow \cdots$; it is a **repellor**. Similarly, we see that
- a fixed point at y_0 with $f'(y_0) < 0$, as shown in Fig. 18.5b, attracts trajectories; that is, all trajectories converge toward the critical point y_0: $\cdots \rightarrow \cdot \leftarrow \cdots$; it is a **sink or node**.

Let us now consider the remaining case when also $f'(y_0) = 0$.

Let us assume $f''(y_0) > 0$. Then at $y_0 + \varepsilon$ to the right of fixed point y_0, $\dot{y} = f''(y_0)\varepsilon^2/2 + \mathcal{O}(\varepsilon^3) > 0$, so the trajectory moves away from the fixed point there, while to the left it moves closer to y_0. In other words, we have a **saddle point**. For $f''(y_0) < 0$, the sign of $\dot{y}$ is reversed, so we deal again with a saddle point with the motion to the right of y_0 toward the fixed point and at left away from it. Let us summarize the local behavior of trajectories near such a fixed point y_0: We have a

- a saddle point at y_0 when $f(y_0) = 0$, and $f'(y_0) = 0$, as shown in Fig. 18.6a,b corresponding to the cases where (a) $f''(y_0) > 0$ and trajectories on one side of the critical point, converge toward it and diverge from it on the other side: $\cdots \rightarrow \cdot \rightarrow \cdots$; and (b) $f''(y_0) < 0$. Here the direction is simply reversed compared to (a). Figure 18.6(c) shows the cases where $f''(y_0) = 0$.

So far we have ignored the additional dependence of $f(y)$ on one or more parameters, such as μ for the logistic map. When a critical point maintains its properties qualitatively as we adjust a parameter slightly, we call it **structurally stable**. This is reasonable because structurally unstable objects are unlikely to occur in reality because noise and other neglected degrees of freedom act as perturbations on the system that effectively prevent such unstable points from being observed. Let us now look at fixed points from this point of view. Upon varying such a control parameter slightly we deform the function f, or we

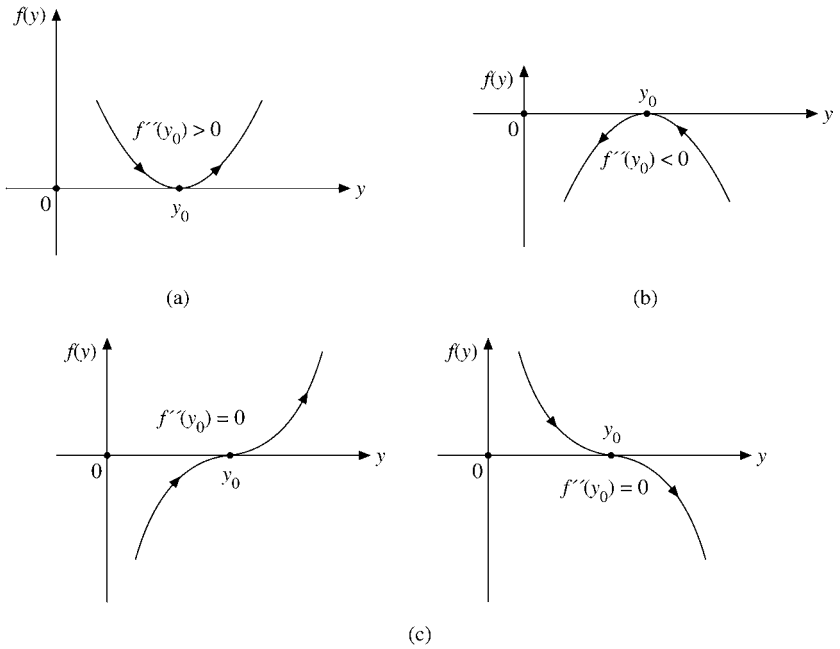

FIGURE 18.6 Saddle points.

may just shift f up or down or sideways in Fig. 18.5 a bit. This will move a little the location y_0 of the fixed point with $f(y_0) = 0$, but maintain the sign of $f'(y_0)$. Thus, both **sinks and repellors are stable**, while a saddle point in general is not. For example, shifting f in Fig. 18.6a down a bit creates two fixed points, one a sink and the other a repellor, and removes the saddle point. Since two conditions must be satisfied at a saddle point, they are less common and important, being unstable with respect to variations of parameters. However, they mark the border between different types of dynamics and are useful and meaningful for the global analysis of the dynamics. We are now ready to consider the richer, but more complicated, higher-dimensional cases.

Local and Global Behavior in Higher Dimensions

In two or more dimensions we start the local analysis at a fixed point $(y_1^0, y_2^0, \ldots)$ with $\dot{y}_i = f_i(y_1^0, y_2^0, \ldots) = 0$ using the same Taylor expansion of the f_i in Eq. (18.22b) as for the one-dimensional case. Retaining only the first-order derivatives, this approach linearizes the coupled NDEs of Eq. (18.22b) and reduces their solution to linear algebra as follows. We abbreviate the constant derivatives at the fixed point as a matrix F with elements

$$f_{ij} \equiv \left.\frac{\partial f_i}{\partial y_j}\right|_{(y_1^0, y_2^0, \cdots)}. \tag{18.24}$$

In contrast to the standard linear algebra in Chapter 3, however, F is neither symmetric nor Hermitian in general. As a result, its eigenvalues may not be real. If we shift the fixed point to the origin and call the shifted coordinates $x_i = y_i - y_i^0$, then the coupled NDEs of Eq. (18.22b) become

$$\dot{x}_i = \sum_j f_{ij} x_j, \tag{18.25}$$

that is, coupled linear ODEs with constant coefficients. We solve Eq. (18.25) with the standard exponential Ansatz,

$$x_i(t) = \sum_j c_{ij} e^{\lambda_j t}, \tag{18.26}$$

with constant exponents λ_j and a constant matrix C of coefficients c_{ij}, so $\mathbf{c}_j = (c_{ij},\ i = 1, 2, \ldots)$ forms the jth column vector of C. Substituting Eq. (18.26) into Eq. (18.25) yields a linear combination of exponential functions,

$$\sum_j c_{ij} \lambda_j e^{\lambda_j t} = \sum_{j,k} f_{ik} c_{kj} e^{\lambda_j t}, \tag{18.27}$$

which are independent if $\lambda_i \neq \lambda_j$. This is the general case on which we focus, while degeneracies where two or more λ are equal require special treatment similar to saddle points in one dimension. Comparing coefficients of exponential functions with the same exponent yields the linear eigenvalue equations

$$\sum_k f_{ik} c_{kj} = \lambda_j c_{ij}, \qquad \text{or} \qquad \mathsf{F}\mathbf{c}_j = \lambda_j \mathbf{c}_j. \tag{18.28}$$

A nontrivial solution comprising the eigenvalue λ_j and eigenvector $\mathbf{c}_j$ of the homogeneous linear equations (18.28) requires λ_j to be a root of the secular equation (compare with Section 3.5):

$$\det(\mathsf{F} - \lambda \cdot 1) = 0. \tag{18.29}$$

Equation (18.28) means that C diagonalizes F, so we can write Eq. (18.28) also as

$$\mathsf{C}^{-1}\mathsf{F}\mathsf{C} = [\lambda_1, \lambda_2, \ldots]. \tag{18.30}$$

In the new but in general nonorthogonal coordinates ξ_j, defined as $\mathsf{C}\boldsymbol{\xi} = \mathbf{x}$, we have a fixed point for each direction ξ_j, as $\dot{\xi}_j = \lambda_j \xi_j$, where the λ_j play the role of $f'(y_0)$ in the one-dimensional case. The λ are **characteristic exponents** and complex numbers in general. This is seen by substituting $\mathbf{x} = \mathsf{C}\boldsymbol{\xi}$ into Eq. (18.25) in conjunction with Eqs. (18.28) and (18.30). Thus, this solution represents the independent combination of one-dimensional fixed points, one for each component of $\boldsymbol{\xi}$ and each independent of the other components. In two dimensions for $\lambda_1 < 0$ and $\lambda_2 < 0$, then, we have a sink in all directions. When both λ are greater than 0, we have repellor in all directions.

Example 18.4.1 STABLE SINK

The coupled ODEs

$$\dot{x} = -x, \qquad \dot{y} = -x - 3y$$

have an equilibrium point at the origin. The solutions have the form

$$x(t) = c_{11}e^{\lambda_1 t}, \qquad y(t) = c_{21}e^{\lambda_1 t} + c_{22}e^{\lambda_2 t},$$

so the eigenvalue $\lambda_1 = -1$ results from $\lambda_1 c_{11} = -c_{11}$, and the solution is $x = c_{11}e^{-t}$. The determinant of Eq. (18.29),

$$\begin{vmatrix} -1-\lambda & 0 \\ -1 & -3-\lambda \end{vmatrix} = (1+\lambda)(3+\lambda) = 0,$$

yields the eigenvalues $\lambda_1 = -1$, $\lambda_2 = -3$. Because both are negative we have a stable sink at the origin. The ODE for y gives the linear relations

$$\lambda_1 c_{21} = -c_{11} - 3c_{21} = -c_{21}, \qquad \lambda_2 c_{22} = -3c_{22},$$

from which we infer $2c_{21} = -c_{11}$, or $c_{21} = -c_{11}/2$. Because the general solution will contain two constants, it is given by

$$x(t) = c_{11}e^{-t}, \qquad y(t) = -\frac{c_{11}}{2}e^{-t} + c_{22}e^{-3t}.$$

As the time $t \to \infty$, we have $y \sim -x/2$ and $x \to 0$ and $y \to 0$, while for $t \to -\infty$, $y \sim x^3$ and $x, y \to \pm\infty$. The motion toward the sink is indicated by arrows in Fig. 18.7. To find the orbit, we eliminate the independent variable, t, and find the cubics:

$$y = -\frac{x}{2} + \frac{c_{22}}{c_{11}^3}x^3.$$

■

When both λ are greater than 0, we have repellor. In this case the motion is away from the fixed point. However, when the λ have different signs, we have a saddle point, that is, a combination of a sink in one dimension and a repellor in the other. This type of behavior generalizes to higher dimensions.

Example 18.4.2 SADDLE POINT

The coupled ODEs

$$\dot{x} = -2x - y, \qquad \dot{y} = -x + 2y$$

have a fixed point at the origin. The solutions have the form

$$x(t) = c_{11}e^{\lambda_1 t} + c_{12}e^{\lambda_2 t}, \qquad y(t) = c_{21}e^{\lambda_1 t} + c_{22}e^{\lambda_2 t}.$$

The eigenvalues $\lambda = \pm\sqrt{5}$ are determined from $\lambda^2 - 5 = 0$.

$$\begin{vmatrix} -2-\lambda & -1 \\ -1 & 2-\lambda \end{vmatrix}$$

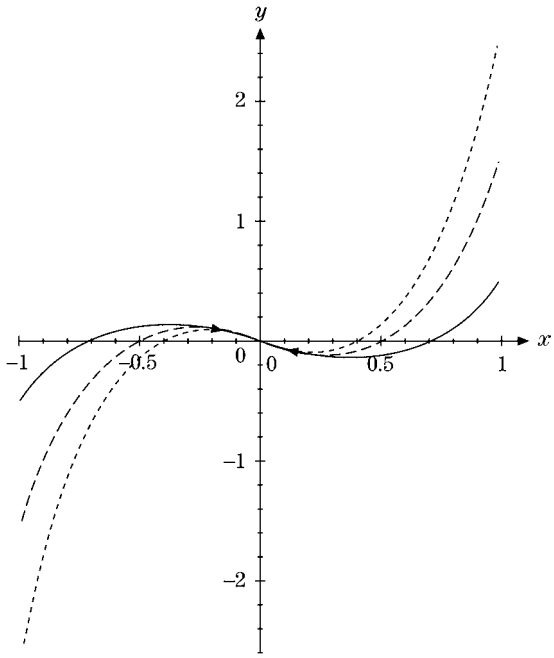

FIGURE 18.7 Stable sink.

Substituting the general solutions into the ODEs yields the linear equations

$$\lambda_1 c_{11} = -2c_{11} - c_{21} = \sqrt{5}c_{11}, \qquad \lambda_1 c_{21} = -c_{11} + 2c_{21} = \sqrt{5}c_{21},$$
$$\lambda_2 c_{12} = -2c_{12} - c_{22} = -\sqrt{5}c_{12}, \qquad \lambda_2 c_{22} = -c_{12} + 2c_{22} = -\sqrt{5}c_{22},$$

or

$$(\sqrt{5}+2)c_{11} = -c_{21}, \qquad (\sqrt{5}-2)c_{12} = c_{22},$$
$$(\sqrt{5}-2)c_{21} = -c_{11}, \qquad (\sqrt{5}+2)c_{22} = c_{12},$$

so $c_{21} = -(2+\sqrt{5})c_{11}$, $c_{22} = (\sqrt{5}-2)c_{12}$. The family of solutions depends on two parameters, c_{11}, c_{12}. For large time $t \to \infty$, the positive exponent prevails and $y \sim -(\sqrt{5}+2)x$, while for $t \to -\infty$ we have $y = (\sqrt{5}-2)x$. These straight lines are the asymptotes of the orbits. Because $-(\sqrt{5}+2)(\sqrt{5}-2) = -1$ they are orthogonal. We find the orbits by eliminating the independent variable, t, as follows. Substituting the c_{2j} we write

$$y = -2x - \sqrt{5}\left(c_{11}e^{\sqrt{5}t} - c_{12}e^{-\sqrt{5}t}\right), \qquad \text{so} \qquad \frac{y+2x}{\sqrt{5}} = -c_{11}e^{\sqrt{5}t} + c_{12}e^{-\sqrt{5}t}.$$

Now we add and subtract the solution $x(t)$ to get

$$\frac{1}{\sqrt{5}}(y+2x) + x = 2c_{12}e^{-\sqrt{5}t}, \qquad \frac{1}{\sqrt{5}}(y+2x) - x = -2c_{11}e^{\sqrt{5}t},$$

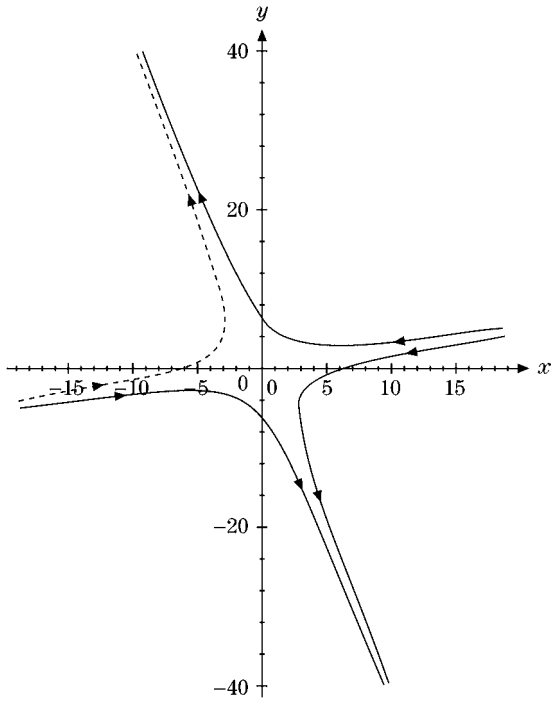

FIGURE 18.8 Saddle point.

which we multiply to obtain

$$\frac{1}{5}(y+2x)^2 - x^2 = -4c_{12}c_{11} = \text{const.}$$

The resulting quadratic form, $y^2 + 4xy - x^2 = \text{const.}$, is a hyperbola because of the negative sign. The hyperbola is rotated in the sense that its asymptotes are not aligned with the x, y-axes (Fig. 18.8). Its orientation is given by the direction of the asymptotes that we found earlier. Alternatively we could find the direction of minimal distance from the origin, proceeding as follows. We set the x and y derivatives of $f + \Lambda g \equiv x^2 + y^2 + \Lambda(y^2 + 4xy - x^2)$ equal to zero, where Λ is the Lagrange multiplier for the hyperbolic constraint. The four branches of hyperbolas correspond to the different signs of the parameters c_{11} and c_{12}. Figure 18.8 is plotted for the cases $c_{11} = \pm 1$, $c_{12} = \pm 2$. ■

However, a new kind of behavior arises for a pair of complex conjugate eigenvalues $\lambda_{1,2} = \rho \pm i\kappa$. If we write the complex solutions $\xi_{1,2} = \exp(\rho t \pm i\kappa t)$ in real variables $\xi_+ = (\xi_1 + \xi_2)/2, \xi_- = (\xi_1 - \xi_2)/2i$ upon using the Euler identity $\exp(ix) = \cos x + i \sin x$ (see Section 6.1),

$$\xi_+ = \exp(\rho t)\cos(\kappa t), \qquad \xi_- = \exp(\rho t)\sin(\kappa t) \tag{18.31}$$

describe a trajectory that spirals inward to the fixed point at the origin for $\rho < 0$, a **spiral node**, and spirals away from the fixed point for $\rho > 0$, a **spiral repellor**.

Example 18.4.3 SPIRAL FIXED POINT

The coupled ODEs

$$\dot{x} = -x + 3y, \quad \dot{y} = (1 + \lambda)x + (\lambda - 3)y = 0,$$

have a fixed point at the origin and solutions of the form
or $\lambda^2 - \lambda + 7 = 0$. The eigenvalues are complex conjugate, $\lambda = 1/2 \pm i\sqrt{}$

$$x(t) = c_{11}e^{\lambda_1 t} + c_{12}e^{\lambda_2 t}, \qquad y(t) = c_{21}e^{\lambda_1 t} + c_{22}e^{\lambda_2 t}.$$

The exponents $\lambda_{1,2}$ are solutions of

$$\begin{vmatrix} -1-\lambda & 3 \\ -3 & 2-\lambda \end{vmatrix}$$

$\overline{27}/2$, so we deal with a spiral fixed point at the origin (a repellor because $1/2 > 0$). Substituting the general solutions into the ODEs yields the linear equations

$$\lambda_1 c_{11} = -c_{11} + 3c_{21}, \qquad \lambda_1 c_{21} = -3c_{11} + 2c_{21},$$
$$\lambda_2 c_{12} = -c_{12} + 3c_{22}, \qquad \lambda_2 c_{22} = -3c_{12} + 2c_{22},$$

or

$$(\lambda_1 + 1)c_{11} = 3c_{21}, \qquad (\lambda_1 - 2)c_{21} = -3c_{11},$$
$$(\lambda_2 + 1)c_{12} = 3c_{22}, \qquad (\lambda_2 - 2)c_{22} = -3c_{12},$$

which, using the values of $\lambda_{1,2}$, imply the family of curves

$$x(t) = e^{t/2}\left(c_{11}e^{i\sqrt{27}t/2} + c_{12}e^{-i\sqrt{27}t/2}\right),$$
$$y(t) = \frac{x}{2} + \frac{\sqrt{27}}{6}e^{t/2}i\left(c_{11}e^{i\sqrt{27}t/2} - c_{12}e^{-i\sqrt{27}t/2}\right),$$

which depends on two parameters, $c_{11},\ c_{12}$. To simplify we can separate real and imaginary parts of $x(t)$ and $y(t)$ using the Euler identity $e^{ix} = \cos x + i\sin x$. It is equivalent, but more convenient, to choose $c_{11} = c_{12} = c/2$ and rescale $t \to 2t$, so with the Euler identity we have

$$x(t) = ce^t\cos(\sqrt{27}t), \qquad y(t) = \frac{x}{2} - \frac{\sqrt{27}}{6}ce^t\sin(\sqrt{27}t).$$

Here we can eliminate t and find the orbit

$$x^2 + \frac{4}{3}\left(y - \frac{x}{2}\right)^2 = \left(ce^t\right)^2.$$

For fixed t this is the positive definite quadratic form $x^2 - xy + y^2 = \text{const.}$, that is, an ellipse. But there is no ellipse in the solutions because t is not fixed. Nonetheless, it is useful to find its orientation. We proceed as follows. With Λ the Lagrange multiplier for the elliptical constraint we seek the directions of maximal and minimal distance from the origin, forming

$$f(x, y) + \Lambda g(x, y) \equiv x^2 + y^2 + \Lambda\left(x^2 - xy + y^2\right)$$

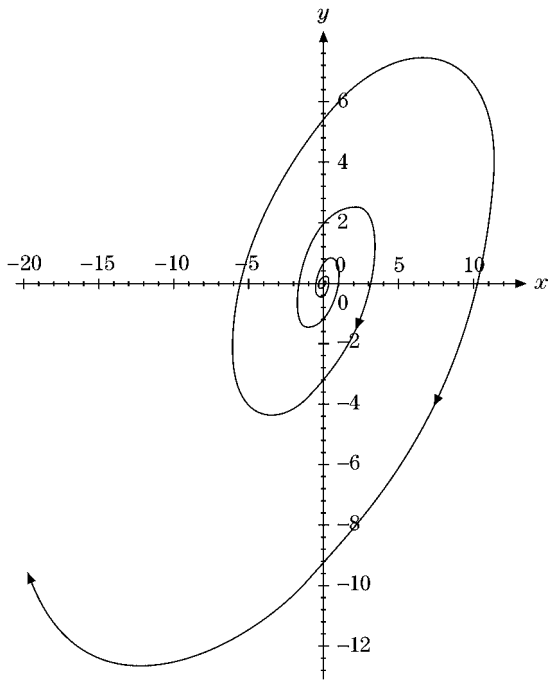

FIGURE 18.9 Spiral point.

and setting

$$\frac{\partial(f+\Lambda g)}{\partial x} = 2x + 2\Lambda x - \Lambda y = 0, \qquad \frac{\partial(f+\Lambda g)}{\partial y} = 2y + 2\Lambda y - \Lambda x = 0.$$

From

$$2(\Lambda+1)x = \Lambda y, \qquad 2(\Lambda+1)y = \Lambda x$$

we obtain the directions

$$\frac{x}{y} = \frac{\Lambda}{2(\Lambda+1)} = \frac{2(\Lambda+1)}{\Lambda},$$

or $\Lambda^2 + \frac{8}{3}\Lambda + \frac{4}{3} = 0$. This yields the values $\Lambda = -2/3, -2$ and the directions $y = \pm x$. In other words, our ellipse is centered at the origin and rotated by 45°. As we vary the independent variable, t, the size of the ellipse changes, so we get the rotated spiral shown in Fig. 18.9 for $c = 1$. ■

In the special case when $\rho = 0$ in Eq. (18.31), the circular trajectory is called a **cycle**. When trajectories near it are attracted as time goes on, it is called a **limit cycle**, representing periodic motion for autonomous systems.

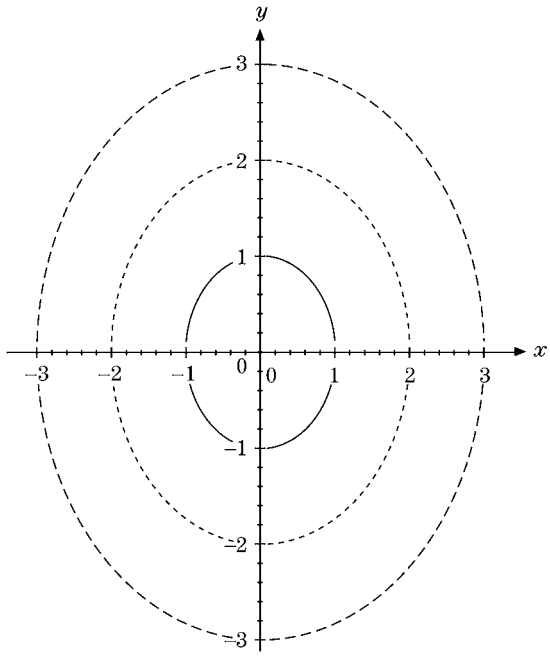

FIGURE 18.10 Center.

Example 18.4.4 CENTER OR CYCLE

The undamped linear harmonic oscillator ODE $\ddot{x} + \omega^2 x = 0$ can be written as two coupled ODEs:

$$\dot{x} = -\omega y, \qquad \dot{y} = \omega x.$$

Integrating the resulting ODE $\dot{x}x + \dot{y}y = 0$ yields the circular orbits $x^2 + y^2 = \text{const.}$, which define a center at the origin and are shown in Fig. 18.10. The solutions can be parameterized as $x = R\cos t$, $y = R\sin t$, where R is the radius parameter. They correspond to the complex conjugate eigenvalues $\lambda_{1,2} = \pm i\omega$. We can check them if we write the general solution as

$$x(t) = c_{11}e^{\lambda_1 t} + c_{12}e^{\lambda_2 t}, \qquad y(t) = c_{21}e^{\lambda_1 t} + c_{22}e^{\lambda_2 t}.$$

Then the eigenvalues follow from

$$\begin{vmatrix} -\lambda & -\omega \\ \omega & -\lambda \end{vmatrix}$$

■

Another classic attractor is quasiperiodic motion, such as the trajectory

$$x(t) = A_1 \sin(\omega_1 t + b_1) + A_2 \sin(\omega_2 t + b_2), \tag{18.32}$$

where the ratio ω_1/ω_2 is an irrational number. Such combined oscillations occur as solutions of a damped anharmonic oscillator (Van der Pol nonautonomous system)

$$\ddot{x} + 2\gamma\dot{x} + \omega_2^2 x + \beta x^3 = f\cos(\omega_1 t). \tag{18.33}$$

In three dimensions, when there is a positive characteristic exponent and an attracting complex conjugate pair in the other directions, we have a spiral saddle point as a new feature. Conversely, a negative characteristic exponent in conjunction with a repelling pair also gives rise to a spiral saddle point, where trajectories spiral out in two dimensions but are attracted in a third direction.

In general, when some form of damping (or dissipation of energy) is present, the transients decay and the system settles either in equilibrium, that is, a single point, or in periodic or quasiperiodic motion. Chaotic motion in dissipative systems is now recognized as a fourth state, and its attractors are often called **strange**. In dissipative systems, initial conditions are not important because trajectories end up on some attractor. They are crucial in Hamiltonian systems. In nonintegrable Hamiltonian systems, chaos may also occur, and then it is called **conservative** chaos. We refer to Chapter 8 of Hilborn (1994) in the Additional Readings for this more complicated topic.

For the driven damped pendulum when trajectories near a center (closed orbit) are attracted to it as time goes on, this closed orbit is defined as a **limit cycle**, representing periodic motion for autonomous systems. A damped pendulum usually spirals into the origin (the position at rest); that is, the origin is a spiral fixed point in its phase space. When we turn on a driving force, then the system formally becomes nonautonomous, because of its explicit time dependence, but also more interesting. In this case, we can call the explicit time in a sinusoidal driving force a new variable, φ, where ω_0 is a fixed rate, in the equation of motion

$$\dot{\omega} + \gamma\omega + \sin\theta = f\sin\varphi, \qquad \omega = \dot{\theta}, \qquad \varphi = \omega_0 t.$$

Then we increase the dimension of our phase space by 1 (adding one variable, φ) because $\dot{\varphi} = \omega_0 = \text{const.}$, but we keep the coupled ODEs autonomous. This driven damped pendulum has trajectories that cross a closed orbit in phase space and spiral back to it; it is called **limit cycle**. This happens for a range of strength f of the driving force, the control parameter of the system. As we increase f, the phase space trajectories go through several neighboring limit cycles and eventually become aperiodic and chaotic. Such closed limit cycles are called **Hopf bifurcations** of the pendulum on its road to chaos, after the mathematician E. Hopf, who generalized Poincaré's results on such bifurcations to higher dimensions of phase space.

Such spiral sinks or saddle points cannot occur in one dimension, but we might ask if they are stable when they occur in higher dimensions. An answer is given by the Poincaré–Bendixson theorem, which says that either trajectories (in the finite region to be specified in a moment) are attracted to a fixed point as time goes on, or they approach a limit cycle provided the relevant two-dimensional subsystem stays inside a finite region, that is, does not diverge there as $t \to \infty$. For a proof we refer to Hirsch and Smale (1974) and Jackson (1989) in the Additional Readings.

In general, when some form of damping is present, the transients decay and the system settles either in equilibrium, that is, a single point, or in periodic or quasiperiodic motion. Chaotic motion is now recognized as a fourth state, and its attractors are often **strange**.

Exercise

18.4.2 Show that the (Rössler) coupled ODEs

$$\dot{x}_1 = -x_2 - x_3, \qquad \dot{x}_2 = x_1 + a_1 x_2, \qquad \dot{x}_3 = a_2 + (x_1 - a_3)x_3$$

(a) have two fixed points for $a_2 = 2$, $a_3 = 4$, and $0 < a_1 < 2$,
(b) have a spiral repellor at the origin, and
(c) have a spiral chaotic attractor for $a_1 = 0.398$.

Dissipation in Dynamical Systems

Dissipative forces often involve velocities, that is, first-order time derivatives, such as friction (for example, for the damped oscillator). Let us look for a measure of dissipation, that is, how a small area $A = c_{1,2}\Delta\xi_1\Delta\xi_2$ at a fixed point shrinks or expands, first in two dimensions for simplicity. Here $c_{1,2} \equiv \sin(\hat{\xi_1}, \hat{\xi_2})$, involving the sine of the characteristic directions, is a time-independent angular factor that takes into account the nonorthogonality of the characteristic directions $\hat{\xi_1}$ and $\hat{\xi_2}$ of Eq. (18.28). If we take the time derivative of A and use $\dot{\xi}_j = \lambda_j \xi_j$ of the characteristic coordinates, implying $\dot{\Delta}\xi_j = \lambda_j \Delta\xi_j$, we obtain, to lowest order in the $\Delta\xi_j$,

$$\dot{A} = c_{1,2}[\Delta\xi_1\lambda_2\Delta\xi_2 + \Delta\xi_2\lambda_1\Delta\xi_1] = c_{1,2}\Delta\xi_1\Delta\xi_2(\lambda_1 + \lambda_2). \tag{18.34}$$

In the limit $\Delta\xi_j \to 0$, we find from Eq. (18.34) that the rate

$$\frac{\dot{A}}{A} = \lambda_1 + \lambda_2 = \text{trace}(\mathsf{F}) = \boldsymbol{\nabla}\cdot\mathbf{f}|_{\mathbf{y}_0}, \tag{18.35}$$

with $\mathbf{f} = (f_1, f_2)$ the vector of time evolution functions of Eq. (18.22b). Note that the time-independent sine of the angle between ξ_1 and ξ_2 drops out of the rate. The generalization to higher dimensions is obvious. Moreover, in n dimensions,

$$\text{trace}(\mathsf{F}) = \sum_i \lambda_i. \tag{18.36}$$

This trace formula follows from the invariance of the secular polynomial in Eq. (18.29) under a linear transformation, $\mathsf{C}\boldsymbol{\xi} = \mathbf{x}$ in particular, and it is a result of its determinental form using the product theorem for determinants (see Section 3.2), viz.

$$\begin{aligned}\det(\mathsf{F} - \lambda\cdot 1) &= \big[\det(\mathsf{C})\big]^{-1}\det(\mathsf{F} - \lambda\cdot 1)\big[\det(\mathsf{C})\big] = \det\big(\mathsf{C}^{-1}(\mathsf{F} - \lambda\cdot 1)\mathsf{C}\big)\\ &= \det\big(\mathsf{C}^{-1}\mathsf{F}\mathsf{C} - \lambda\cdot 1\big) = \prod_{i=1}^{n}(\lambda_i - \lambda).\end{aligned} \tag{18.37}$$

Here the product form comes about by substituting Eq. (18.30). Now, trace(F) is the coefficient of $(-\lambda)^{n-1}$ upon expanding $\det(\mathsf{F} - \lambda\cdot 1)$ in powers of λ, while it is $\sum_i \lambda_i$ from the product form $\prod_i(\lambda_i - \lambda)$, which proves Eq. (18.36). Clearly, according to Eqs. (18.35) and (18.36),

- it is the sign (and more precisely the trace) of the characteristic exponents of the derivative matrix at the fixed point that determines whether there is expansion or shrinkage of areas and volumes in higher dimensions near a critical point.

In summary then, Eq. (18.35) states that dissipation requires $\nabla \cdot \mathbf{f}(\mathbf{y}) \neq 0$, where $\dot{y}_j = f_j$, and does not occur in Hamiltonian systems where $\nabla \cdot \mathbf{f} = 0$.

Moreover, in two or more dimensions, there are the following global possibilities:

- The trajectory may describe a closed orbit (cycle).
- The trajectory may approach a closed orbit (spiraling inward or outward toward the orbit) as $t \to \infty$. In this case we have a limit cycle.

The local behavior of a trajectory near a critical point is also more varied in general than in one dimension: At a stable critical point all trajectories may approach the critical point along straight lines or spiral inward (toward the **spiral node**) or may follow a more complicated path. If all time-reversed trajectories move toward the critical point in spirals as $t \to -\infty$, then the critical point is a divergent spiral point, or **spiral repellor**. When some trajectories approach the critical point while others move away from it, then it is called a **saddle point**. When all trajectories form closed orbits about the critical point, it is called a **center**.

Bifurcations in Dynamical Systems

A bifurcation is a sudden change in dynamics for specific parameter values, such as the birth of a node–repellor pair of fixed points or their disappearance upon adjusting a control parameter; that is, the motions before and after the bifurcation are topologically different. At a bifurcation point, not only are solutions unstable when one or more parameters are changed slightly, but the character of the bifurcation in phase space or in the parameter manifold may change. Thus we are dealing with fairly sudden events of nonlinear dynamics. Rather sudden changes from regular to random behavior of trajectories are characteristic of bifurcations, as is sensitive dependence on initial conditions: Nearby initial conditions can lead to very different long-term behavior. If a bifurcation does not change qualitatively with parameter adjustments, it is called **structurally stable**. Note that structurally unstable bifurcations are unlikely to occur in reality because noise and other neglected degrees of freedom act as perturbations on the system that effectively eliminate unstable bifurcations from our view. Bifurcations (such as doublings in maps) are important as one among many routes to chaos. Others are sudden changes in trajectories associated with several critical points called **global bifurcations**. Often they involve changes in basins of attraction and/or other global structures. The theory of global bifurcations is fairly complicated and is still in its infancy at present.

Bifurcations that are linked to sudden changes in the qualitative behavior of dynamical systems at a single fixed point are called **local bifurcations**. More specifically, a change in stability occurs in parameter space where the real part of a characteristic exponent of the fixed point alters its sign, that is, moves from attracting to repelling trajectories, or vice versa. The **center–manifold theorem** says that at a local bifurcation only those degrees

of freedom matter that are involved with characteristic exponents going to zero: $\Re\lambda_i = 0$. Locating the set of these points is the first step in a bifurcation analysis. Another step consists in cataloguing the types of bifurcations in dynamical systems, to which we turn next.

The conventional **normal forms** of dynamical equations represent a start in classifying bifurcations. For systems with one parameter (that is, a one-dimensional center manifold) we write the general case of NDE as follows:

$$\dot{x} = \sum_{j=0}^{\infty} a_j^{(0)} x^j + c \sum_{j=0}^{\infty} a_j^{(1)} x^j + c^2 \sum_{j=0}^{\infty} a_j^{(2)} x^j + \cdots , \tag{18.38}$$

where the superscript on the $a^{(m)}$ denotes the power of the parameter c they are associated with. One-dimensional iterated nonlinear maps such as the logistic map of Section 18.2 (which occur in Poincaré sections) of nonlinear dynamical systems can be classified similarly, viz.

$$x_{n+1} = \sum_{j=0}^{\infty} a_j^{(0)} x_n^j + c \sum_{j=0}^{\infty} a_j^{(1)} x_n^j + c^2 \sum_{j=0}^{\infty} a_j^{(2)} x_n^j + \cdots . \tag{18.39}$$

Thus, one of the simplest NDEs with a bifurcation is

$$\dot{x} = x^2 - c, \tag{18.40}$$

which corresponds to all $a_j^{(m)} = 0$ except for $a_0^{(1)} = -1$ and $a_2^{(0)} = 1$. For $c > 0$, there are two fixed points (recall, $\dot{x} = 0$) $x_\pm = \pm\sqrt{c}$ with characteristic exponents $2x_\pm$, so x_- is a node and x_+ is a repellor. For $c < 0$ there are no fixed points. Therefore, as $c \to 0$ the fixed point pair disappears suddenly; that is, the parameter value $c = 0$ is a repellor-node bifurcation that is structurally unstable. This complex map (with $c \to -c$) generates the fractal Julia and Mandelbrot sets discussed in Section 18.3.

A pitchfork bifurcation occurs for the undamped (nondissipative and special case of the Duffing) oscillator with a cubic anharmonicity

$$\ddot{x} + ax + bx^3 = 0, \qquad b > 0. \tag{18.41}$$

It has a continuous frequency spectrum and is, among others, a model for a ball bouncing between two walls. When the control parameter $a > 0$, there is only one fixed point, at $x = 0$, a node, while for $a < 0$ there are two more nodes, at $x_\pm = \pm\sqrt{-a/b}$. Thus, we have a pitchfork bifurcation of a node at the orgin into a saddle point at the origin and two nodes, at $x_\pm \neq 0$. In terms of a potential formulation, $V(x) = ax^2/2 + bx^4/4$ is a single well for $a > 0$ but a double well (with a maximum at $x = 0$) for $a < 0$.

When a pair of complex conjugate characteristic exponents $\rho \pm i\kappa$ crosses from a spiral node ($\rho < 0$) to a repelling spiral ($\rho > 0$) and periodic motion (limit cycle) emerges, then we call the qualitative change a **Hopf bifurcation**. They occur in the quasiperiodic route to chaos that will be discussed in the next section, on chaos.

In a global analysis we piece together the motions near various critical points, such as nodes and bifurcations, to bundles of trajectories that flow more or less together in two dimensions. (This geometric view is the current mode of analyzing solutions of dynamical systems.) But this flow is no longer collective in the case of three dimensions, where they diverge from each other in general, because chaotic motion is possible that typically fills the plane of a Poincaré section with points.

Chaos in Dynamical Systems

Our previous summaries of intricate and complicated features of dynamical systems due to nonlinearities in one and two dimensions do not include chaos, although some of them, such as bifurcations, sometimes are precursors to chaos. In three- or more-dimensional NDEs, chaotic motion may occur, often when a constant of the motion (an energy integral for NDEs defined by a Hamiltonian, for example) restricts the trajectories to a finite volume in phase space and when there are no critical points. Another characteristic signal for chaos is when for each trajectory there are nearby ones, some of which move away from it, while others approach it with increasing time. The notion of exponential divergence of nearby trajectories is made quantitative by the Lyapunov exponent λ (see Section 18.3 for more details) of iterated maps of Poincaré sections associated with the dynamical system. If two nearby trajectories are at a distance d_0 at time $t = 0$ but diverge with a distance $d(t)$ at a later time t, then $d(t) \approx d_0 e^{\lambda t}$ holds. Thus, by analyzing the series of points, that is, iterated maps generated on Poincaré sections, one can study routes to chaos of three-dimensional dynamical systems. This is the key method for studying chaos. As one varies the location and orientation of the Poincaré plane, a fixed point on it often is recognized to originate from a limit cycle in the three-dimensional phase space whose structural stability can be checked there. For example, attracting limit cycles show up as nodes in Poincaré sections, repelling limit cycles as repellors of Poincaré maps, and saddle cycles as saddle points of associated Poincaré maps.

Three or more dimensions of phase space are required for chaos to occur because of the interplay of the necessary conditions we just discussed, viz.

- bounded trajectories (are often the case for Hamiltonian systems),
- exponential divergence of nearby trajectories (is guaranteed by positive Lyapunov exponents of corresponding Poincaré maps),
- no intersection of trajectories.

The last condition is obeyed by deterministic systems in particular, as we discussed in Section 18.1. A surprising feature of chaos, mentioned in Section 18.1, is how prevalent it is and how universal the routes to chaos often are, despite the overwhelming variety of NDEs.

An example for spatially complex patterns in classical mechanics is the planar pendulum, whose one-dimensional equation of motion

$$I\frac{d\theta}{dt} = L, \qquad \frac{dL}{dt} = -lmg\sin\theta \tag{18.42}$$

is nonlinear in the dynamic variable $\theta(t)$. Here I is the moment of inertia, l is the distance to the center of mass, m is the mass, and g is the gravitational acceleration constant. When all parameters in Eq. (18.42) are constant in time and space, then the solutions are given in terms of elliptic integrals (see Section 5.8) and no chaos exists. However, a pendulum under a periodic external force can exhibit chaotic dynamics, for example, for the Lagrangian

$$L = \frac{m}{2}\dot{\mathbf{r}}^2 - mg(l - z),\ (x - x_0)^2 + y^2 + z^2 = l^2, \tag{18.43}$$

$$x_0 = \varepsilon l\cos\omega t. \tag{18.44}$$

(See Moon (1992) in the Additional Readings.)

Good candidates for chaos are multiple well potential problems,

$$\frac{d^2\mathbf{r}}{dt^2} + \nabla V(\mathbf{r}) = F\left(\mathbf{r}, \frac{d\mathbf{r}}{dt}, t\right), \tag{18.45}$$

where F represents dissipative and/or driving forces. Another classic example is rigid-body rotation, whose nonlinear three-dimensional Euler equations are familiar, viz.

$$\begin{aligned} \frac{d}{dt} I_1\omega_1 &= (I_2 - I_3)\omega_2\omega_3 + M_1, \\ \frac{d}{dt} I_2\omega_2 &= (I_3 - I_1)\omega_1\omega_3 + M_2, \\ \frac{d}{dt} I_3\omega_3 &= (I_1 - I_2)\omega_1\omega_2 + M_3. \end{aligned} \tag{18.46}$$

Here the I_j are the principal moments of inertia and $\boldsymbol{\omega}$ is the angular velocity with components ω_j about the body-fixed principal axes. Even free rigid-body rotation can be chaotic, for its nonlinear couplings and three-dimensional form satisfy all requirements for chaos to occur (see Section 18.1). A rigid-body example of chaos in our solar system is the chaotic tumbling of Hyperion, one of Saturn's moons that is highly nonspherical. It is a world where the Saturn rise and set is so irregular as to be unpredictable. Another is Halley's comet, whose orbit is perturbed by Jupiter and Saturn. In general, when three or more celestial bodies interact gravitationally, stochastic dynamics are possible. Note, though, that computer simulations over large time intervals are required to ascertain chaotic dynamics in the solar system. For more details on chaos in such conservative Hamiltonian systems we refer to Chapter 8 of Hilborn (1994) in the Additional Readings.

Exercise

18.4.3 Construct a Poincaré map for the Duffing oscillator in Eq. (18.41).

Routes to Chaos in Dynamical Systems

Let us now look at some routes to chaos. The period-doubling route to chaos is exemplified by the logistic map in Section 18.2, and the universal Feigenbaum numbers α, δ are its quantitative features, along with Lyapunov exponents. It is common in dynamical systems. It may begin with limit cycle (periodic) motion that shows up as a fixed point in a Poincaré section. The limit cycle may have originated in a bifurcation from a node or some other fixed point. As a control parameter changes, the fixed point of the Poincaré map splits into two points; that is, the limit cycle has a characteristic exponent going through zero from attracting to repelling, say. The periodic motion now has a period twice as long as before, etc. We refer to Chapter 11 of Barger and Olsson (1995) in the Additional Readings for period-doubling plots of Poincaré sections for the Duffing equation (18.41) with a periodic external force. Another example for period doubling is a forced oscillator with friction (see Helleman in Cvitanovic (1989) in the Additional Readings).

The quasiperiodic route to chaos is also quite common in dynamical systems, for example, starting from a time-independent node, a fixed point. If we adjust a control parameter, the system undergoes a Hopf bifurcation to the periodic motion corresponding to a limit cycle in phase space. With further change of the control parameter, a second frequency appears. If the frequency ratio is an irrational number, the trajectories are quasiperiodic, eventually covering the surface of a torus in phase space; that is, quasiperiodic orbits never close or repeat. Further changes of the control parameter may lead to a third frequency or directly to chaotic motion. Bands of chaotic motion can alternate with quasiperiodic motion in parameter space. An example for such a dynamic system is a periodically driven pendulum.

A third route to chaos goes via intermittency, where the dynamical system switches between two qualitatively different motions at fixed control parameters. For example, at the beginning, periodic motion alternates with an occasional burst of chaotic motion. With a change of the control parameter, the chaotic bursts typically lengthen until, eventually, no periodic motion remains. The chaotic parts are irregular and do not resemble each other, but one needs to check for a positive Lyapunov exponent to demonstrate chaos. Intermittencies of various types are common features of turbulent states in fluid dynamics. The Lorenz coupled NDEs also show intermittency.

Exercise

18.4.4 Plot the intermittency region of the logistic map at $\mu = 3.8319$. What is the period of the cycles? What happens at $\mu = 1 + 2\sqrt{2}$?

ANS. There is a tangent bifurcation to period 3 cycles.

Additional Readings

Amann, H., *Ordinary Differential Equations: An Introduction To Nonlinear Analysis*. New York: de Gruyter (1990).

Baker, G. L., and J. P. Gollub, *Chaotic Dynamics: An Introduction*, 2nd ed. Cambridge, UK: Cambridge University Press (1996).

Barger, V. D., and M. G. Olsson, *Classical Mechanics*, 2nd ed. New York: McGraw-Hill (1995).

Bender, C. M., and S. A. Orszag, *Advanced Mathematical Methods For Scientists and Engineers*. New York: McGraw-Hill (1978), Chapter 4 in particular.

Bergé, P., Y. Pomeau, and C. Vidal, *Order within Chaos*. New York: Wiley (1987).

Cvitanovic, P., ed., *Universality in Chaos*, 2nd ed. Bristol, UK: Adam Hilger (1989).

Devaney, R. L., *An Introduction to Chaotic Dynamical Systems*. Menlo Park, CA: Benjamin/Cummings; 2nd ed., Perseus (1989).

Earnshaw, J. C., and D. Haughey, Lyapunov exponents for pedestrians. *Am. J. Phys.* **61**: 401 (1993).

Gleick, J., *Chaos*. New York: Penguin Books (1987).

Hilborn, R. C., *Chaos and Nonlinear Dynamics*. New York: Oxford University Press (1994).

Hirsch, M. W., and S. Smale, *Differential Equations, Dynamical Systems, and Linear Algebra*. New York: Academic Press (1974).

Infeld, E., and G. Rowlands, *Nonlinear Waves, Solitons and Chaos*. Cambridge, UK: Cambridge University Press (1990).

Jackson, E. A., *Perspectives of Nonlinear Dynamics*. Cambridge, UK: Cambridge University Press (1989).

Jordan, D. W., and P. Smith, *Nonlinear Ordinary Differential Equations*, 2nd ed. Oxford, UK: Oxford University Press (1987).

Lyapunov, A. M., *The General Problem of the Stability of Motion*. Bristol, PA: Taylor & Francis (1992).

Mandelbrot, B. B., *The Fractal Geometry of Nature*. San Francisco: W. H. Freeman, reprinted (1988).

Moon, F. C., *Chaotic and Fractal Dynamics*. New York: Wiley (1992).

Peitgen, H.-O., and P. H. Richter, *The Beauty of Fractals*. New York: Springer (1986).

Sachdev, P. L., *Nonlinear Differential Equations and their Applications*. New York: Marcel Dekker (1991).

Tufillaro, N. B., T. Abbott, and J. Reilly, *An Experimental Approach to Nonlinear Dynamics and Chaos*. Redwood City, CA: Addison-Wesley (1992).

CHAPTER 19

PROBABILITY

Probabilities arise in many problems dealing with random events or large numbers of particles defining random variables. An event is called **random** if it is practically impossible to predict from the initial state. This includes those cases where we have merely incomplete information about initial states and/or the dynamics, as in statistical mechanics, where we may know the energy of the system that corresponds to very many possible microscopic configurations, preventing us from predicting individual outcomes. Often the average properties of many similar events are predictable, as in quantum theory. This is why probability theory can be and has been developed.

Random variables are involved when data depend on chance, such as weather reports and stock prices. The theory of probability describes mathematical models of chance processes in terms of probability distributions of random variables that describe how some "random events" are more likely than others. In this sense probability is a measure of our ignorance, giving quantitative meaning to qualitative statements such as "It will probably rain tomorrow" and "I'm unlikely to draw the heart queen." Probabilities are of fundamental importance in quantum mechanics and statistical mechanics and are applied in meteorology, economics, games, and many other areas of daily life.

To a mathematician, probabilities are based on axioms, but we will discuss here practical ways of calculating probabilities for random events. Because experiments in the sciences are always subject to errors, theories of errors and their propagation involve probabilities. In statistics we deal with the applications of probability theory to experimental data.

19.1 DEFINITIONS, SIMPLE PROPERTIES

All possible **mutually exclusive**[1] outcomes of an experiment that is subject to chance represent the events (or points) of the **sample space** S. For example, each time we toss a coin we give the trial a number $i = 1, 2, \ldots$ and observe the outcomes x_i. Here the sample

[1] This means that given that one particular event did occur, the others could not have occurred.

consists of two events: heads and tails, and the x_i represent a discrete random variable that takes on one of two values, heads or tails. When two coins are tossed, the sample contains the events two heads, one head and one tail, two tails; the number of heads is a good value to assign to the random variable, so the possible values are 2, 1, and 0. There are four equally probable outcomes, of which one has value 2, two have value 1, and one has value 0. So the probabilities of the three values of the random variable are $1/4$ for two heads (value 2), $1/4$ for no heads (value 0), and $1/2$ for value 1. In other words, we define the theoretical probability P of an event denoted by the point x_i of the sample as

$$P(x_i) \equiv \frac{\text{number of outcomes of event } x_i}{\text{total number of all events}}. \tag{19.1}$$

An experimental definition applies when the total number of events is not well defined (or is difficult to obtain) or equally likely outcomes do not always occur. Then

$$P(x_i) \equiv \frac{\text{number of times event } x_i \text{ occurs}}{\text{total number of trials}} \tag{19.2}$$

is more appropriate. A large, thoroughly mixed pile of black and white sand grains of the same size and in equal proportions is a relevant example, because it is impractical to count them all. But we can count the grains in a small sample volume that we pick. This way we can check that white and black grains turn up with roughly equal probability $1/2$, provided we put back each sample and mix the pile again. It is found that the larger the sample volume, the smaller the spread about $1/2$ will be. The more trials we run, the closer the average occurrence of all trial counts will be to $1/2$. We could even pick single grains and check if the probability $1/4$ of picking two black grains in a row equals that of two white grains, etc. There are lots of statistics questions we can pursue. Thus, piles of colored sand provide for instructive experiments.

The following axioms are self-evident.

- Probabilities satisfy $0 \leq P \leq 1$. Probability 1 means certainty; probability 0 means impossibility.
- The entire sample has probability 1. For example, drawing an arbitrary card has probability 1.
- The probabilities for mutually exclusive events add. The probability for getting one head in two coin tosses is $1/4 + 1/4 = 1/2$ because it is $1/4$ for head first and then tail, plus $1/4$ for tail first and then head.

Example 19.1.1 PROBABILITY FOR A OR B

What is the probability for drawing[2] a club or a jack from a shuffled deck of cards? Because there are 52 cards in a deck, each being equally likely, 13 cards for each suit and 4 jacks, there are 13 clubs including the club jack, and 3 other jacks; that is, there are 16 possible cards out of 52, giving the probability $(13 + 3)/52 = 16/52 = 4/13$. ■

[2] These are examples of non-mutually exclusive events.

If we represent the sample space by a set S of points, then events are subsets $A, B, \ldots$ of S, denoted as $A \subset S$, etc. Two sets A, B are equal if A is contained in B, $A \subset B$, and B is contained in A, $B \subset A$. The **union** $A \cup B$ consists of all points (events) that are in A or B or both (see Fig. 19.1). The **intersection** $A \cap B$ consists of all points, that are in both A and B. If A and B have no common points, their intersection is the **empty set**, $A \cap B = \emptyset$, which has no elements (events). The set of points in A that are not in the intersection of A and B is denoted by $A - A \cap B$, **defining a subtraction of sets**. If we take the club suit in Example 19.1.1 as set A and the four jacks as set B, then their union comprises all clubs and jacks, and their intersection is the club jack only.

Each subset A has its probability $P(A) \geq 0$. In terms of these set theory concepts and notations, the probability laws we just discussed become

$$0 \leq P(A) \leq 1.$$

The entire sample space has $P(S) = 1$. The probability of the union $A \cup B$ of mutually exclusive events is the sum

$$P(A \cup B) = P(A) + P(B), \qquad A \cap B = \emptyset.$$

The **addition rule** for probabilities of arbitrary sets is given by the following theorem.

ADDITION RULE:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{19.3}$$

To prove this, we decompose the union into two mutually exclusive sets $A \cup B = A \cup (B - B \cap A)$, subtracting the intersection of A and B from B before joining them. Their probabilities are $P(A)$, $P(B) - P(B \cap A)$, which we add. We could also have decomposed $A \cup B = (A - A \cap B) \cup B$, from which our theorem follows similarly by adding these probabilities, $P(A \cup B) = [P(A) - P(A \cap B)] + P(B)$. Note that $A \cap B = B \cap A$. (See Fig. 19.1.)

Sometimes the rules and definitions of probabilities that we have discussed so far are not sufficient, however.

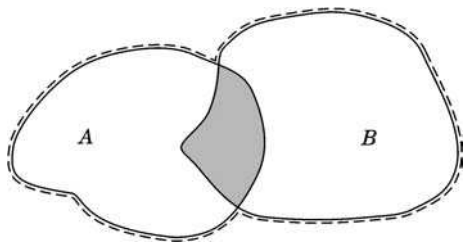

FIGURE 19.1 The shaded area gives the intersection $A \cap B$, corresponding to the A and B events, the dashed line encloses $A \cup B$, corresponding to the A or B events.

Example 19.1.2 CONDITIONAL PROBABILITY

A simple example consists of a box of 10 identical red and 20 identical blue pens, arranged in random order, from which we remove pens successively, that is, without putting them back. Suppose we draw a red pen first, event A. That will happen with probability $P(A) = 10/30 = 1/3$ if the pens are thoroughly mixed up. The conditional probability $P(B|A)$ of drawing a blue pen in the next round, event B, however, will depend on the fact that we drew a red pen in the first round. It is given by $20/29$. There are $10 \cdot 20$ possible sample points (red/blue pen events) in two rounds, and the sample has $30 \cdot 29$ events, so the combined probability is

$$P(A, B) = \frac{10}{30}\frac{20}{29} = \frac{10 \cdot 20}{30 \cdot 29} = \frac{20}{87}.$$ ■

In general, the combined probability $P(A, B)$ that A and B happen (in this order) is given by the product of the probability that A happens, $P(A)$, and the probability that B happens if A does, $P(B|A)$:

$$P(A, B) = P(A)P(B|A). \tag{19.4}$$

In other words, the **conditional probability** $P(B|A)$ is given by the ratio

$$P(B|A) = \frac{P(A, B)}{P(A)}. \tag{19.5}$$

If the conditional probability $P(B|A) = P(B)$ is independent of A, then the events A and B are called **independent**, and the combined probability

$$P(A \cap B) = P(A)P(B) \tag{19.6}$$

is simply the **product of both probabilities**.

Example 19.1.3 SCHOLASTIC APTITUDE TESTS

Colleges and universities rely on the verbal and mathematics SAT scores, among others, as predictors of a student's success in passing courses and graduating. A research university is known to admit mostly students with a combined verbal and mathematics score above 1400 points. The graduation rate is 95%; that is, 5% drop out or transfer elsewhere. Of those who graduate, 97% have an SAT score of more than 1400 points, while 80% of those who drop out have an SAT score below 1400. Suppose a student has an SAT score below 1400. What is his/her probability of graduating?

Let A be the cases having an SAT test score below 1400, B represent those above 1400, mutually exclusive events with $P(A) + P(B) = 1$, and C be those students who graduate. That is, we want to know the conditional probabilities $P(C|A)$ and $P(C|B)$. To apply Eq. (19.5) we need $P(A)$ and $P(B)$. There are 3% of students with scores below 1400 among those who graduate (95%) and 80% of those 5% who do not graduate, so

$$P(A) = 0.03 \cdot 0.95 + \frac{4}{5}0.05 = 0.0685, \qquad P(B) = 0.97 \cdot 0.95 + \frac{0.05}{5} = 0.9315,$$

and also

$$P(C \cap A) = 0.03 \cdot 0.95 = 0.0285 \qquad \text{and} \qquad P(C \cap B) = 0.97 \cdot 0.95 = 0.9215.$$

Here the combined probabilities $P(C, A) = P(C \cap A)$, $P(C, B) = P(C \cap B)$ as C and A (and C and B) are parts of the same sample space. Therefore,

$$P(C|A) = \frac{P(C \cap A)}{P(A)} = \frac{0.0285}{0.0685} \sim 41.6\%,$$
$$P(C|B) = \frac{P(C \cap B)}{P(B)} = \frac{0.9215}{0.9315} \sim 98.9\%;$$

that is, a little less than 42% is the probability for a student with a score below 1400 to graduate at this particular university. ■

As a corollary to the definition of a conditional probability, Eq. (19.5), we compare $P(A|B) = P(A \cap B)/P(B)$ and $P(B|A) = P(A \cap B)/P(A)$, which leads to the following theorem.

BAYES THEOREM:

$$P(A|B) = \frac{P(A)}{P(B)} P(B|A). \tag{19.7}$$

This can be generalized to the following.

THEOREM: *If the random events A_i with probabilities $P(A_i) > 0$ are mutually exclusive and their union represents the entire sample S, then an arbitrary random event $B \subset S$ has the probability*

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B|A_i). \tag{19.8}$$

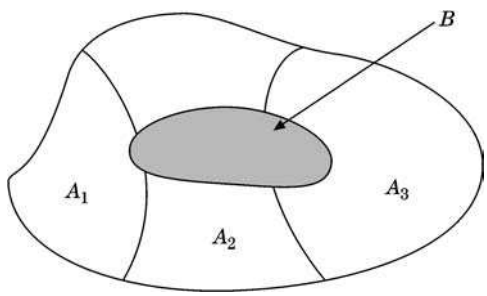

FIGURE 19.2 The shaded area B is composed of mutually exclusive subsets of B belonging also to A_1, A_2, A_3, where the A_i are mutually exclusive.

This decomposition law resembles the expansion of a vector into a basis of unit vectors defining the components of the vector. This relation follows from the obvious decomposition $B = \bigcup_i (B \cap A_i)$, Fig. 19.2, which implies $P(B) = \sum_i P(B \cap A_i)$ for the probabilities because the components $B \cap A_i$ are mutually exclusive. For each i, we know from Eq. (19.5) that $P(B \cap A_i) = P(A_i)P(B|A_i)$, which proves the theorem.

Counting of Permutations and Combinations

Counting particles in samples can help us find probabilities, as in statistical mechanics.

If we have n different molecules, let us ask in how many ways we can arrange them in a row, that is, permute them. This number is defined as the number of their **permutations**. Thus, by definition, the **order matters in permutations**. There are n choices of picking the first molecule, $n-1$ for the second, etc. Altogether there are $n!$ permutations of n **different** molecules or objects.

Generalizing this, suppose there are n people but only $k < n$ chairs to seat them. In how many ways can we seat k people in the chairs? Counting as before, we get

$$n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

for the number of permutations of n different objects, k at a time.

We now consider the number of **combinations** of objects when their **order is irrelevant** by definition. For example, three letters a, b, c can be combined, two letters at a time, in $3 = \frac{3!}{2!}$ ways: ab, ac, bc. If letters can be repeated, then we add the pairs aa, bb, cc and have six combinations. Thus, a **combination** of different particles differs from a permutation in that their **order does not matter**. Combinations occur with repetition (the mathematician's way of treating indistinguishable objects) and without, where no two sets contain the same particles.

The number of different combinations of n particles, k at a time and without repetitions, is given by the binomial coefficient

$$\frac{n(n-1)\cdots(n-k+1)}{k!} = \binom{n}{k}.$$

If repetition is allowed, then the number is

$$\binom{n+k-1}{k}.$$

In the number $n!/(n-k)!$ of permutations of n particles, k at a time, we have to divide out the number $k!$ of permutations of the groups of k particles because their order does not matter in a combination. This proves the first claim. The second one is shown by mathematical induction.

In statistical mechanics, we ask in how many ways we can put n particles in k boxes so that there will be n_i (distinguishable) particles in the ith box, without regard to order in each box, with $\sum_{i=1}^{k} n_i = n$. Counting as before, there are n choices for selecting the first particle, $n-1$ for picking the second, etc., but the $n_1!$ permutations within the first box are

discounted, and $n_2!$ permutations within the second box are disregarded, etc. Therefore the number of combinations is

$$\frac{n!}{n_1!n_2!\cdots n_k!}, \qquad n_1+n_2+\cdots+n_k=n.$$

In statistical mechanics, particles that obey

- Maxwell–Boltzmann (MB) statistics are distinguishable, without restriction on their number in each state;
- Bose–Einstein (BE) statistics are indistinguishable, with no restriction on the number of particles in each quantum state;
- Fermi–Dirac (FD) statistics are indistinguishable, with at most one particle per state.

For example, putting three particles in four boxes, there are 4^3 equally likely arrangements for the MB case, because each particle can be put into any box in four ways, giving a total of 4^3 choices. For BE statistics, the number of combinations with repetitions is $\binom{3+4-1}{3}=\binom{6}{3}$ for the Bose–Einstein case. For FD statistics, it is $\binom{3+1}{3}=\binom{4}{3}$. More generally, for MB statistics the number of distinct arrangements of n particles among k states (boxes) is k^n, for BE statistics it is $\binom{n+k-1}{n}$, and for FD statistics it is $\binom{k}{n}$.

Exercises

19.1.1 A card is drawn from a shuffled deck. (a) What is the probability that it is black, (b) a red nine, (c) or a queen of spades?

19.1.2 Find the probability of drawing two kings from a shuffled deck of cards (a) if the first card is put back before the second is drawn, and (b) if the first card is not put back after being drawn.

19.1.3 When two fair dice are thrown, what is the probability of (a) observing a number less than 4 or (b) a number greater than or equal to 4 but less than 6?

19.1.4 Rolling three fair dice, what is the probability of obtaining six points?

19.1.5 Determine the probability $P(A\cap B\cap C)$ in terms of $P(A)$, $P(B)$, $P(C)$, etc.

19.1.6 Determine directly or by mathematical induction the probability of a distribution of N (Maxwell–Boltzmann) particles in k boxes with N_1 in box 1, N_2 in box $2,\ldots,N_k$ in the kth box for any numbers $N_j\geq 1$ with $N_1+N_2+\cdots+N_k=N$, $k<N$. Repeat this for Fermi–Dirac and Bose–Einstein particles.

19.1.7 Show that $P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)$ $-P(B\cap C)+P(A\cap B\cap C)$.

19.1.8 Determine the probability that a positive integer $n\leq 100$ is divisible by a prime number $p\leq 100$. Verify your result for $p=3,5,7$.

19.1.9 Put two particles obeying Maxwell–Boltzmann (Fermi–Dirac, or Bose–Einstein) statistics in three boxes. How many ways are there in each case?

19.2 RANDOM VARIABLES

Each time we toss a die, we give the trial a number $i = 1, 2, \ldots$ and observe the point $x_i = 1$, or 2, 3, 4, 5, 6 with probability $1/6$. If i denotes the trial number, then x_i is a discrete random variable that takes the discrete values from 1 to 6 with a definite probability $P(x_i) = 1/6$.

Example 19.2.1 DISCRETE RANDOM VARIABLE

If we toss two dice and record the sum of the points shown in each trial, then this sum is also a discrete random variable, which takes on the value 2 when both dice show 1 with probability $(1/6)^2$; the value 3 when one die has 1 and the other 2, hence with probability $(1/6)^2 + (1/6)^2 = 1/18$; the value 4 when both dice have 2 or one has 1 and the other 3, so with probability $(1/6)^2 + (1/6)^2 + (1/6)^2 = 1/12$; the value 5 with probability $4(1/6)^2 = 1/9$; the value 6 with probability $5/36$; the value 7 with the maximum probability, $6(1/6)^2 = 1/6$; up to the value 12 when both dice show 6 points with probability $(1/6)^2$. This probability distribution is symmetric about 7. This symmetry is obvious from Fig. 19.3 and becomes visible algebraically when we write the rising and falling linear parts as

$$P(x) = \frac{x-1}{36} = \frac{6-(7-x)}{36}, \qquad x = 2, 3, \ldots, 7,$$

$$P(x) = \frac{13-x}{36} = \frac{6+(7-x)}{36}, \qquad x = 7, 8, \ldots, 12. \quad \blacksquare$$

In summary, then,

- The different values x_i that a random variable X assumes denote and distinguish the events in the sample space of an experiment; each event occurs by chance with a prob-

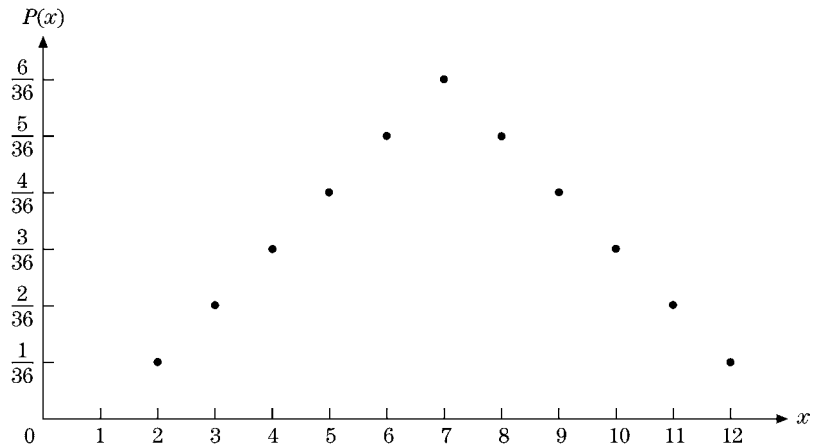

FIGURE 19.3 Probability distribution $P(x)$ of the sum of points when two dice are tossed.

ability $P(X = x_i) = p_i \geq 0$ that is a function of the random variable X. A random variable $X(e_i) = x_i$ is defined on the sample space, that is, for the events $e_i \in S$.

- We define the probability density $f(x)$ of a **continuous random variable** X as

$$P(x \leq X \leq x + dx) = f(x)\,dx; \tag{19.9}$$

that is, $f(x)\,dx$ is the probability that X lies in the interval $x \leq X \leq x + dx$. For $f(x)$ to be a probability density, it has to satisfy $f(x) \geq 0$ and $\int f(x)\,dx = 1$. The generalization to probability distributions depending on several random variables is straightforward. Quantum physics abounds in examples.

Example 19.2.2 CONTINUOUS RANDOM VARIABLE: HYDROGEN ATOM

Quantum mechanics gives the probability $|\psi|^2\,d^3r$ of finding a $1s$ electron in a hydrogen atom in volume[3] d^3r, where $\psi = Ne^{-r/a}$ is the wave function that is normalized to

$$1 = \int |\psi|^2\,dV = 4\pi N^2 \int_0^\infty e^{-2r/a} r^2\,dr = \pi a^3 N^2, \qquad dV = r^2\,dr\,d\cos\theta\,d\varphi$$

being the volume element and a the Bohr radius. The radial integral is found by repeated integration by parts or by rescaling it to the gamma function

$$\int_0^\infty e^{-2r/a} r^2\,dr = \left(\frac{a}{2}\right)^3 \int_0^\infty e^{-x} x^2\,dx = \frac{a^3}{8}\Gamma(3) = \frac{a^3}{4}.$$

Here all points in space constitute the sample and represent three random variables, but the probability density $|\psi|^2$ in this case depends only on the radial variable because of the spherical symmetry of the $1s$ state.

A measure for the size of the H atom is given by the average radial distance of the electron from the proton at the center, which in quantum mechanics is called the **expectation value**:

$$\langle 1s|r|1s\rangle = \int r|\psi|^2\,dV = 4\pi N^2 \int_0^\infty r e^{-2r/a} r^2\,dr = \frac{3}{2}a.$$

We shall define this concept for arbitrary probability distributions shortly. ■

- A random variable that takes only discrete values $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$, respectively, is called a discrete random variable, so $\sum_i p_i = 1$. If an "experiment" or trial is performed, some outcome must occur, with unit probability.
- If the values comprise a continuous range of values $a \leq x \leq b$, then we deal with a continuous random variable, whose probability distribution may or may not be a continuous function as well.

[3]Note that $|\psi|^2 4\pi r^2\,dr$ gives the probability for the electron to be found between r and $r + dr$, at any angle.

When we measure a quantity x n times, obtaining the values x_j, we define the **average value**

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j \tag{19.10}$$

of the trials, also called the **mean** or **expectation value**, where this formula assumes that every observed value x_i is equally likely and occurs with probability $1/n$. This connection is the key link of experimental data with probability theory. This observation and practical experience suggest defining the **mean value for a discrete random variable** X as

$$\langle X \rangle \equiv \sum_i x_i p_i \tag{19.11}$$

and that for a **continuous random variable** characterized by probability density $f(x)$ as

$$\langle X \rangle = \int x f(x)\, dx. \tag{19.12}$$

These are linear averages. Other notations in the literature are $\bar{X}$ and $E(X)$.

The use of the arithmetic mean $\bar{x}$ of n measurements as the average value is suggested by simplicity and plain experience, assuming equal probability for each x_i again. But why do we not consider the geometric mean

$$x_g = (x_1 \cdot x_2 \cdot \cdots \cdot x_n)^{1/n} \tag{19.13}$$

or the harmonic mean x_h determined by the relation

$$\frac{1}{x_h} = \frac{1}{n}\left(\frac{1}{x_1} + \frac{1}{x_1} + \cdots + \frac{1}{x_n}\right) \tag{19.14}$$

or that value $\tilde{x}$ that minimizes the sum of absolute deviations $|x_i - \tilde{x}|$? Here the x_i are taken to increase monotonically. When we plot $O(x) = \sum_{i=1}^{2n+1} |x_i - x|$, as in Fig. 19.4a, for an odd number of points, we realize that it has a minimum at its central value $i = n$,

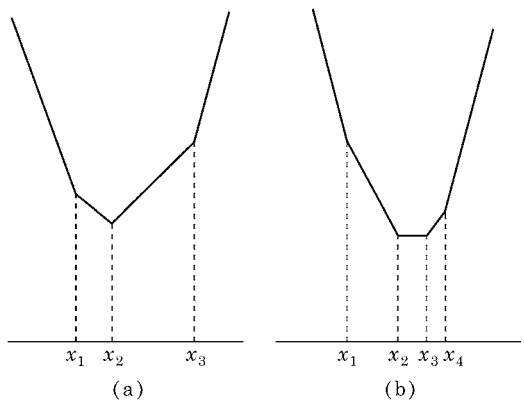

FIGURE 19.4 (a) $\sum_{i=1}^{3} |x_i - x|$ for an odd number of points; (b) $\sum_{i=1}^{4} |x_i - x|$ for an even number of points.

while for an even number of points $E(x) = \sum_{i=1}^{2n} |x_i - x|$ is flat in its central region, as shown in Fig. 19.4b. These properties make these functions unacceptable for determining average values. Instead, when we minimize the sum of quadratic deviations,

$$\sum_{i=1}^{n} (x - x_i)^2 = \text{minimum}, \tag{19.15}$$

setting the derivative equal to zero yields $2\sum_i (x - x_i) = 0$, or

$$x = \frac{1}{n} \sum_i x_i \equiv \bar{x},$$

that is, the arithmetic mean. It has another important property: If we denote by $v_i = x_i - \bar{x}$ the deviations, then $\sum_i v_i = 0$, that is, the sum of positive deviations equals the sum of negative deviations. This principle of minimizing the quadratic sum of deviations, called the **method of least squares**, is due to C. F. Gauss, among others.

How close a fit of the mean value to a set of data points is depends on the spread of the individual measurements from this mean. Again, we reject the average sum of deviations $\sum_{i=1}^{n} |x_i - \bar{x}|/n$ as a measure of the spread because it selects the central measurement as the best value for no good reason. A more appropriate definition of the spread is the average of the deviations from the mean, squared, or **standard deviation**

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2},$$

where the square root is motivated by dimensional analysis.

Example 19.2.3 STANDARD DEVIATION OF MEASUREMENTS

From the measurements $x_1 = 7$, $x_2 = 9$, $x_3 = 10$, $x_4 = 11$, $x_5 = 13$ we extract $\bar{x} = 10$ for the mean value and $\sigma = \sqrt{(9 + 1 + 1 + 9)/4} = 2.2361$ for the standard deviation, or spread, using the experimental formula (19.2) because the probabilities are not known. ■

There is yet another interpretation of the standard variation, in terms of the sum of squares of measurement differences

$$\begin{aligned} \sum_{i<k} (x_i - x_k)^2 &= \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{n} \left(x_i^2 + x_k^2 - 2x_i x_k \right) \\ &= \frac{1}{2} \left(2n^2 \langle x^2 \rangle - 2n^2 \langle x \rangle^2 \right) = n^2 \sigma^2, \end{aligned} \tag{19.16}$$

because by multiplying out the square in the definition of σ^2 we obtain

$$\begin{aligned} \sigma^2 &= \frac{1}{n} \sum_i \left(x_i - \langle x \rangle \right)^2 = \frac{1}{n} \sum_i x_i^2 - \frac{2\langle x \rangle}{n} \sum_i x_i + \langle x \rangle^2 \\ &= \frac{1}{n} \sum_i x_i^2 - \langle x \rangle^2 = \langle x^2 \rangle - \langle x \rangle^2. \end{aligned} \tag{19.17}$$

This formula is often used and widely applied for σ^2.

Now we are ready to generalize the spread in a set of n measurements with equal probability $1/n$ to the **variance** of an arbitrary probability distribution. For a discrete random variable X with probabilities p_i at $X = x_i$ we define the **variance**

$$\sigma^2 = \sum_j \big(x_j - \langle X\rangle\big)^2 p_j, \tag{19.18}$$

and similarly for a continuous probability distribution

$$\sigma^2 = \int_{-\infty}^{\infty} \big(x - \langle X\rangle\big)^2 f(x)\,dx. \tag{19.19}$$

These definitions imply the following.

THEOREM: *If a random variable $Y = aX + b$ is linearly related to X, then we can immediately derive the mean value $\langle Y\rangle = a\langle X\rangle + b$ and variance $\sigma^2(Y) = a^2\sigma^2(X)$ from these definitions.*

We prove this theorem only for a continuous distribution and leave the case of a discrete random variable as an exercise for the reader. For the infinitesimal probability we know that $f(x)\,dx = g(y)\,dy$ with $y = ax + b$, because the linear transformation has to preserve probability, so

$$\langle Y\rangle = \int_{-\infty}^{\infty} y g(y)\,dy = \int_{-\infty}^{\infty} (ax + b) f(x)\,dx = a\langle X\rangle + b,$$

since $\int f(x)\,dx = 1$. For the variance we similarly obtain

$$\begin{aligned}\sigma^2(Y) &= \int_{-\infty}^{\infty} \big(y - \langle Y\rangle\big)^2 g(y)\,dy = \int_{-\infty}^{\infty} \big(ax + b - a\langle X\rangle - b\big)^2 f(x)\,dx \\ &= a^2\sigma^2(X)\end{aligned}$$

after substituting our result for the mean value $\langle Y\rangle$.

Finally we prove the general **Chebychev inequality**[1]

$$P\big(|x - \langle X\rangle| \geq k\sigma\big) \leq \frac{1}{k^2}, \tag{19.20}$$

which demonstrates why the standard deviation serves as a measure of the spread of an arbitrary probability distribution from its mean value $\langle X\rangle$ and shows why experimental or other data are often characterized according to their spread in numbers of standard deviations. We first show the simpler inequality

$$P(Y \geq K) \leq \frac{\langle Y\rangle}{K}$$

for a continuous random variable Y whose values $y \geq 0$. (The proof for a discrete random variable follows along similar lines.) This inequality follows from

$$\begin{aligned}\langle Y\rangle &= \int_0^{\infty} y f(y)\,dy = \int_0^{K} y f(y)\,dy + \int_K^{\infty} y f(y)\,dy \\ &\geq \int_K^{\infty} y f(y)\,dy \geq K \int_K^{\infty} f(y)\,dy = K\,P(Y \geq K).\end{aligned}$$

Next we apply the same method to the positive variance integral

$$\sigma^2 = \int \big(x - \langle X\rangle\big)^2 f(x)\,dx \geq \int_{|x-\langle X\rangle|\geq k\sigma} \big(x - \langle X\rangle\big)^2 f(x)\,dx$$
$$\geq k^2\sigma^2 \int_{|x-\langle X\rangle|\geq k\sigma} f(x)\,dx = k^2\sigma^2 P\big(|x - \langle X\rangle| \geq k\sigma\big),$$

decreasing the right-hand side first by omitting the part of the positive integral with $|x - \langle X\rangle| \leq k\sigma$ and then again by replacing $(x - \langle X\rangle)^2$ in the remaining integral by its lowest limit, $k^2\sigma^2$. This proves the Chebychev inequality. For $k = 3$ we have the conventional three-standard-deviation estimate

$$P\big(|x - \langle X\rangle| \geq 3\sigma\big) \leq \frac{1}{9}. \tag{19.21}$$

It is straightforward to generalize the mean value to higher moments of probability distributions relative to the mean value $\langle X\rangle$:

$$\big\langle \big(X - \langle X\rangle\big)^k \big\rangle = \sum_j \big(x_j - \langle X\rangle\big)^k p_j, \qquad \text{discrete distribution,} \tag{19.22}$$

$$\big\langle \big(X - \langle X\rangle\big)^k \big\rangle = \int_{-\infty}^{\infty} \big(x - \langle X\rangle\big)^k f(x)\,dx, \qquad \text{continuous distribution.}$$

The **moment-generating function**

$$\big\langle e^{tX} \big\rangle = \int e^{tx} f(x)\,dx = 1 + t\langle X\rangle + \frac{t^2}{2!}\langle X^2\rangle + \cdots \tag{19.23}$$

is a weighted sum of the moments of the continuous random variable X upon substituting the Taylor expansion of the exponential functions. So $\langle X\rangle = \frac{d\langle e^{tX}\rangle}{dt}\big|_{t=0}$. Notice that the moments here are not relative to the expectation value; they are called **central moments**. The nth central moment, $\langle X^n\rangle = \frac{d^n\langle e^{tX}\rangle}{dt^n}\big|_{t=0}$, is given by the nth derivative of the moment-generating function at $t = 0$. By a change of the parameter $t \to it$ the moment-generating function is related to the **characteristic function** $\langle e^{itX}\rangle$, which is the often-used Fourier transform of the probability density $f(x)$.

Moreover, mean values, moments, and variance can be defined similarly for probability distributions that depend on several random variables. For simplicity, let us restrict our attention to two continuous random variables X, Y and list the corresponding quantities:

$$\begin{aligned}
\langle X\rangle &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)\,dx\,dy,\\
\langle Y\rangle &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(x,y)\,dx\,dy,
\end{aligned} \tag{19.24}$$

$$\begin{aligned}
\sigma^2(X) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \big(x - \langle X\rangle\big)^2 f(x,y)\,dx\,dy,\\
\sigma^2(Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \big(y - \langle Y\rangle\big)^2 f(x,y)\,dx\,dy.
\end{aligned} \tag{19.25}$$

Two random variables are said to be **independent if the probability density** $f(x, y)$ **factorizes** into a product $f(x)g(y)$ of probability distributions of one random variable each.

The covariance, defined as

$$\operatorname{cov}(X, Y) = \big\langle \big(X - \langle X\rangle\big)\big(Y - \langle Y\rangle\big)\big\rangle, \tag{19.26}$$

is a measure of how much the random variables X, Y are correlated (or related): It is zero for independent random variables because

$$\begin{aligned}\operatorname{cov}(X, Y) &= \int \big(x - \langle X\rangle\big)\big(y - \langle Y\rangle\big) f(x, y)\, dx\, dy \\ &= \int \big(x - \langle X\rangle\big) f(x)\, dx \int \big(y - \langle Y\rangle\big) g(y)\, dy = \big(\langle X\rangle - \langle X\rangle\big)\big(\langle Y\rangle - \langle Y\rangle\big) = 0.\end{aligned}$$

The normalized covariance $\frac{\operatorname{cov}(X,Y)}{\sigma(X)\sigma(Y)}$, which has values between -1 and $+1$, is often called **correlation**.

In order to demonstrate that the correlation is bounded by

$$-1 \le \frac{\operatorname{cov}(X, Y)}{\sigma(X)\sigma(Y)} \le 1$$

we analyze the positive mean value

$$\begin{aligned}Q &= \big\langle \big[u\big(X - \langle X\rangle\big) + v\big(Y - \langle Y\rangle\big)\big]^2 \big\rangle \\ &= u^2 \big\langle \big[X - \langle X\rangle\big]^2 \big\rangle + 2uv \big\langle \big[X - \langle X\rangle\big]\big[Y - \langle Y\rangle\big] \big\rangle + v^2 \big\langle \big[Y - \langle Y\rangle\big]^2 \big\rangle \\ &= u^2\sigma(X)^2 + 2uv \operatorname{cov}(X, Y) + v^2\sigma(Y)^2 \ge 0,\end{aligned} \tag{19.27}$$

where u, v are numbers, not functions. For this quadratic form to be nonnegative, its discriminant must obey $\operatorname{cov}(X, Y)^2 - \sigma(X)^2\sigma(Y)^2 \le 0$, which proves the desired inequality.

The usefulness of the correlation as a quantitative measure is emphasized by the following.

THEOREM: *$P(Y = aX + b) = 1$ is valid if, and only if, the correlation is equal to ± 1.*

This theorem states that a $\pm 100\%$ correlation between X, Y implies not only some functional relation between both random variables but also a **linear relation** between them. We denote by $\langle B|A\rangle$ the expectation value of the conditional probability distribution $P(B|A)$.

To prove this strong correlation property, we apply Bayes' decomposition law (Eq. (19.8)) to the mean value and variance of the random variable Y, assuming first that $P(Y = aX + b) = 1$, so $P(Y \ne aX + b) = 0$. This yields

$$\begin{aligned}\langle Y\rangle &= P(Y = aX + b)\langle Y|Y = aX + b\rangle \\ &\quad + P(Y \ne aX + b)\langle Y|Y \ne aX + b\rangle \\ &= \langle aX + b\rangle = a\langle X\rangle + b, \\ \sigma(Y)^2 &= P(Y = aX + b)\big\langle \big[Y - \langle Y\rangle\big]^2 \big| Y = aX + b\big\rangle \\ &\quad + P(Y \ne aX + b)\big\langle \big[Y - \langle Y\rangle\big]^2 \big| Y \ne aX + b\big\rangle \\ &= \big\langle \big[aX + b - \langle Y\rangle\big]^2 \big\rangle = \big\langle a^2\big[X - \langle X\rangle\big]^2 \big\rangle = a^2\sigma(X)^2,\end{aligned}$$

substituting $\langle Y\rangle = a\langle X\rangle + b$. Similarly we obtain $\text{cov}(X, Y) = a^2\sigma(X)^2$. These results show that the correlation is ± 1.

Conversely, we start from $\text{cov}(X, Y)^2 = \sigma(X)^2\sigma(Y)^2$. Hence the quadratic form in Eq. (19.27) must be zero for (practically) all x for some $(u_0, v_0) \neq (0, 0)$:

$$\left\langle \left[u_0\big(X - \langle X\rangle\big) + v_0\big(Y - \langle Y\rangle\big)\right]^2\right\rangle = 0.$$

Because the argument of this mean value is positive definite, this relationship is satisfied only if $P(u_0(X - \langle X\rangle) + v_0(Y - \langle Y\rangle) = 0) = 1$, which means that Y and X are linearly related.

When we integrate out one random variable, we are left with the probability distribution of the other random variable,

$$F(x) = \int f(x, y)\, dy, \qquad \text{or} \qquad G(y) = \int f(x, y)\, dx, \tag{19.28}$$

and analogously for discrete probability distributions. When one or more random variables are integrated out, the remaining probability distribution is called **marginal,** motivated by the geometric aspects of projection. It is straightforward to show that these marginal distributions satisfy all the requirements of properly normalized probability distributions.

If we are interested in the distribution of the random variable X for a definite value $y = y_0$ of the other random variable, then we deal with a **conditional probability distribution** $P(X = x|Y = y_0)$. The corresponding continuous probability density is $f(x, y_0)$.

Example 19.2.4 Repeated Draws of Cards

When we draw cards repeatedly, we shuffle the deck often because we want to make sure these events stay independent. So we draw the first card at random from a bridge deck containing 52 cards and then put it back at a random place. Now we repeat the process for a second card. Then the deck is reshuffled, etc. We now define the random variables

- $X =$ number of so-called honors, that is, 10s, jacks, queens, kings, or aces;
- $Y =$ number of 2s or 3s.

In a single draw the probability of a 10 to an ace is $a = 5 \cdot 4/52 = 5/13$, and $b = 2 \cdot 4/52 = 2/13$ for two or three to be drawn and $c = (13 - 5 - 2)/13 = 6/13$ for anything else, with $a + b + c = 1$.

In two drawings X and Y can be $x = 0 = y$, when no 10 to ace show up or 2 or 3. This case has probability c^2. In general X, Y have the values 0, 1, and 2, so $0 \leq x + y \leq 2$ because we will have drawn two cards. The probability function of $(X = x,\ Y = y)$ is given by the product of the probabilities of the three possibilities a^x, b^y, c^{2-x-y} times the number of distributions (or permutations) of two cards over the three cases with probabilities a, b, c, which is $2!/[x!y!(2 - x - y)!]$. This number is the coefficient of the power $a^x b^y c^{2-x-y}$ in the generalized binomial expansion of all possibilities in two drawings with

probability 1:

$$1 = (a+b+c)^2 = \sum_{0\le x+y\le 2} \frac{2!}{x!y!(2-x-y)!} a^x b^y c^{2-x-y}$$
$$= a^2 + b^2 + c^2 + 2(ab + ac + bc). \tag{19.29}$$

Hence the probability distribution of our discrete random variables is given by

$$f(X=x, Y=y) = \frac{2!}{x!y!(2-x-y)!}\left(\frac{5}{13}\right)^x \left(\frac{2}{13}\right)^y \left(\frac{6}{13}\right)^{2-x-y},$$
$$x, y = 0, 1, 2; \qquad 0 \le x + y \le 2, \tag{19.30}$$

or more explicitly as

$$f(0,0) = \left(\frac{6}{13}\right)^2, \qquad f(1,0) = 2\cdot\frac{5}{13}\cdot\frac{6}{13} = \frac{60}{13^2},$$
$$f(2,0) = \left(\frac{5}{13}\right)^2, \qquad f(0,1) = 2\frac{2}{13}\cdot\frac{6}{13} = \frac{24}{13^2},$$
$$f(0,2) = \left(\frac{2}{13}\right)^2, \qquad f(1,1) = 2\frac{5}{13}\cdot\frac{2}{13} = \frac{20}{13^2}.$$

The probability distribution is properly normalized according to Eq. (19.29). Its expectation values are given by

$$\langle X\rangle = \sum_{0\le x+y\le 2} x f(x,y) = f(1,0) + f(1,1) + 2f(2,0)$$
$$= \frac{60}{13^2} + \frac{20}{13^2} + 2\left(\frac{5}{13}\right)^2 = \frac{130}{13^2} = \frac{10}{13} = 2a,$$

and

$$\langle Y\rangle = \sum_{0\le x+y\le 2} y f(x,y) = f(0,1) + f(1,1) + 2f(0,2)$$
$$= \frac{24}{13^2} + \frac{20}{13^2} + 2\left(\frac{2}{13}\right)^2 = \frac{52}{13^2} = \frac{4}{13} = 2b,$$

as expected because we are drawing a card two times. The variances are

$$\sigma^2(X) = \sum_{0\le x+y\le 2} \left(x - \frac{10}{13}\right)^2 f(x,y)$$
$$= \left(\frac{10}{13}\right)^2 \big[f(0,0) + f(0,1) + f(0,2)\big]$$
$$+ \left(\frac{3}{13}\right)^2 \big[f(1,0) + f(1,1)\big] + \left(\frac{16}{13}\right)^2 f(2,0)$$
$$= \frac{10^2\cdot 64 + 3^2\cdot 80 + 16^2\cdot 5^2}{13^4} = \frac{4^2\cdot 5\cdot 169}{13^4} = \frac{80}{13^2},$$

$$\begin{aligned}\sigma^2(Y) &= \sum_{0\le x+y\le 2}\left(y-\frac{4}{13}\right)^2 f(x,y)\\ &= \left(\frac{4}{13}\right)^2\big[f(0,0)+f(1,0)+f(2,0)\big]\\ &\quad+\left(\frac{9}{13}\right)^2\big[f(0,1)+f(1,1)\big]+\left(\frac{22}{13}\right)^2 f(0,2)\\ &= \frac{4^2\cdot 11^2+9^2\cdot 44+22^2\cdot 2^2}{13^4}=\frac{11\cdot 4\cdot 169}{13^4}=\frac{44}{13^2}.\end{aligned}$$

It is reasonable that $\sigma^2(Y)<\sigma^2(X)$, because Y takes only two values, 2 and 3, while X varies over the five honors. The covariance is given by

$$\begin{aligned}\operatorname{cov}(X,Y) &= \sum_{0\le x+y\le 2}\left(x-\frac{10}{13}\right)\left(y-\frac{4}{13}\right)f(x,y)\\ &= \frac{10\cdot 4}{13^2}\cdot\frac{6^2}{13^2}-\frac{10\cdot 9}{13^2}\cdot\frac{24}{13^2}-\frac{10\cdot 22}{13^2}\cdot\frac{4}{13^2}-\frac{3\cdot 4}{13^2}\cdot\frac{60}{13^2}\\ &\quad+\frac{3\cdot 9}{13^2}\cdot\frac{20}{13^2}-\frac{16\cdot 4}{13^2}\cdot\frac{5^2}{13^2}=-\frac{20\cdot 169}{13^4}=-\frac{20}{13^2}.\end{aligned}$$

Therefore the correlation of the random variables X, Y is given by

$$\frac{\operatorname{cov}(X,Y)}{\sigma(X)\sigma(Y)}=-\frac{20}{8\sqrt{5\cdot 11}}=-\frac{1}{2}\sqrt{\frac{5}{11}}=-0.3371,$$

which means that there is a small (negative) correlation between these random variables, because if a card is an honor it cannot be a 2 or a 3, and vice versa.

Finally, let us determine the marginal distribution:

$$F(X=x)=\sum_{y=0}^{2} f(x,y), \tag{19.31}$$

or explicitly

$$\begin{aligned}F(0) &= f(0,0)+f(0,1)+f(0,2)=\left(\frac{6}{13}\right)^2+\frac{24}{13^2}+\left(\frac{2}{13}\right)^2=\left(\frac{8}{13}\right)^2,\\ F(1) &= f(1,0)+f(1,1)=\frac{60}{13^2}+\frac{20}{13^2}=\frac{80}{13^2},\\ F(2) &= f(2,0)=\left(\frac{5}{13}\right)^2,\end{aligned}$$

which is properly normalized because

$$F(0)+F(1)+F(2)=\frac{64+80+25}{13^2}=\frac{169}{13^2}=1.$$

Its mean value is given by

$$\langle X\rangle_F = \sum_{x=0}^{2} xF(x) = F(1) + 2F(2) = \frac{80 + 2\cdot 25}{13^2} = \frac{130}{13^2} = \frac{10}{13} = \langle X\rangle,$$

and its variance

$$\sigma_F^2 = \sum_{x=0}^{2}\left(x - \frac{10}{13}\right)^2 F(x) = \left(\frac{10}{13}\right)^2\cdot\left(\frac{8}{13}\right)^2 + \left(\frac{3}{13}\right)^2\frac{80}{13^2} + \left(\frac{16}{13}\right)^2\cdot\left(\frac{5}{13}\right)^2$$
$$= \frac{80\cdot 169}{13^4} = \frac{80}{13^2} = \sigma^2(X).$$

From the definitions it follows that these results hold generally. ■

Finally we address the transformation of two random variables X, Y into $U(X, Y)$, $V(X, Y)$. We treat the continuous case, leaving the discrete case, as an exercise. If

$$u = u(x, y), \qquad v = v(x, y); \qquad x = x(u, v), \qquad y = y(u, v) \tag{19.32}$$

describe the transformation and its inverse, then the probability stays invariant and the integral of the density transforms according to the rules of Jacobians of Chapter 2, so the transformed probability density becomes

$$g(u, v) = f\big(x(u, v), y(u, v)\big)|J|, \tag{19.33}$$

with the Jacobian

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}. \tag{19.34}$$

Example 19.2.5 SUM, PRODUCT, AND RATIO OF RANDOM VARIABLES

Let us consider three examples. **(1) The sum** $Z = X + Y$, where the transformation may be taken to be using

$$x = x, \qquad z = x + y, \qquad J = \begin{vmatrix} \frac{\partial x}{\partial x} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial x} & \frac{\partial y}{\partial z} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix}$$

$$\frac{\partial(z - y)}{\partial x} = 1, \qquad \frac{\partial(z - y)}{\partial z} = 1, \qquad \frac{\partial y}{\partial x} = 0, \qquad \frac{\partial(z - x)}{\partial z} = 1,$$

so the probability is given by

$$F(Z) = \int_{-\infty}^{Z}\int_{-\infty}^{\infty} f(x, z - x)\,dx\,dz. \tag{19.35}$$

If the random variables X, Y are independent with densities f_1, f_2, then

$$F(Z) = \int_{-\infty}^{Z}\int_{-\infty}^{\infty} f_1(x)\,f_2(z - x)\,dx\,dz. \tag{19.36}$$

(2) The product $Z = XY$, taking X, Z as the new variables, leads to the Jacobian

$$J = \begin{vmatrix} 1 & \frac{1}{y} \\ 0 & \frac{1}{x} \end{vmatrix} = \frac{1}{x},$$

using

$$\frac{\partial x}{\partial x} = 1, \qquad \frac{\partial(\frac{z}{y})}{\partial z} = \frac{1}{y}, \qquad \frac{\partial y}{\partial x} = 0, \qquad \frac{\partial(\frac{z}{x})}{\partial z} = \frac{1}{x},$$

so the probability is given by

$$F(Z) = \int_{-\infty}^{Z} \int_{-\infty}^{\infty} f\left(x, \frac{z}{x}\right) \frac{dx}{|x|} dz. \tag{19.37}$$

If the random variables X, Y are independent with densities f_1, f_2, then

using
$$F(Z) = \int_{-\infty}^{Z} \int_{-\infty}^{\infty} f_1(x) f_2\left(\frac{z}{x}\right) \frac{dx}{|x|} dz. \tag{19.38}$$

(3) The ratio $Z = \frac{X}{Y}$, taking Y, Z as the new variables, has the Jacobian

$$J = \begin{vmatrix} z & y \\ 1 & 0 \end{vmatrix}$$

$$\frac{\partial(yz)}{\partial y} = z, \qquad \frac{\partial(yz)}{\partial z} = y, \qquad \frac{\partial y}{\partial y} = 1, \qquad \frac{\partial y}{\partial z} = 0,$$

so the probability is given by

$$F(Z) = \int_{-\infty}^{Z} \int_{-\infty}^{\infty} f(yz, y)|y| \, dy \, dz. \tag{19.39}$$

If the random variables X, Y are independent with densities f_1, f_2, then

$$F(Z) = \int_{-\infty}^{Z} \int_{-\infty}^{\infty} f_1(yz) f_2(y)|y| \, dy \, dz. \tag{19.40}$$

■

Exercises

19.2.1 Show that adding a constant c to a random variable X changes the expectation value $\langle X \rangle$ by that same constant but not the variance. Show also that multiplying a random variable by a constant multiplies both the mean and variance by that constant. Show that the random variable $X - \langle X \rangle$ has mean value zero.

19.2.2 If $\langle X \rangle$, $\langle Y \rangle$ are the average values of two independent random variables X, Y, what is the expectation value of the product $X \cdot Y$?

19.2.3 A velocity $v_j = x_j/t_j$ is measured by recording the distances x_j at the corresponding times t_j. Show that $\bar{x}/\bar{t}$ is a good approximation for the average velocity v, provided all the errors $|x_j - \bar{x}| \ll |\bar{x}|$ and $|t_j - \bar{t}| \ll |\bar{t}|$ are small.

19.2.4 Define the random variable Y in Example 19.2.4 as the number of 4s, 5s, 6s, 7s, 8s, or 9s. Then determine the correlation of the X and Y random variables.

19.2.5 If X and Y are two independent random variables with different probability densities and the function $f(x, y)$ has derivatives of any order, express $\langle f(X, Y)\rangle$ in terms of $\langle X\rangle$ and $\langle Y\rangle$. Develop similarly the covariance and correlation.

19.2.6 Let $f(x, y)$ be the joint probability density of two random variables X, Y. Find the variance $\sigma^2(aX + bY)$, where a, b are constants. What happens when X, Y are independent?

19.2.7 The probability that a particle of an ideal gas travels a small distance dx between collisions is $\sim e^{-x/f}\,dx$, where f is the constant mean free path. Verify that f is the average distance between collisions, and determine the probability of a free path of length $l \geq 3f$.

19.2.8 Determine the probability density for a particle in simple harmonic motion in the interval $-A \leq x \leq A$.
Hint. The probability that the particle is between x and $x + dx$ is proportional to the time it takes to travel across the interval.

19.3 BINOMIAL DISTRIBUTION

Example 19.3.1 REPEATED TOSSES OF DICE

What is the probability of three 6s in four tosses, all trials being independent? Getting one 6 in a single toss of a fair die has probability $a = 1/6$, and anything else has probability $b = 5/6$ with $a + b = 1$. Let the random variable $X = x$ be the number of 6s. In four tosses, $0 \leq x \leq 4$. The probability distribution $f(X)$ is given by the product of the two possibilities, a^x and b^{4-x}, times the number of combinations of four tosses over the two cases with probabilities a, b. This number is the coefficient of the power $a^x b^{4-x}$ in the binomial expansion of all possibilities in four tosses with probability 1:

$$\begin{aligned} 1 = (a+b)^4 &= \sum_{x=0}^{4} \frac{4!}{x!(4-x)!} a^x b^{4-x} \\ &= a^4 + b^4 + 4a^3 b + 4ab^3 + 6a^2 b^2. \end{aligned} \tag{19.41}$$

Hence the probability distribution of our discrete random variable is given by

$$f(X = x) = \frac{4!}{x!(4-x)!} a^x b^{4-x}, \qquad 0 \leq x \leq 4,$$

or more explicitly

$$f(0) = b^4, \qquad f(1) = 4ab^3, \qquad f(2) = 6a^2 b^2, \qquad f(3) = 4a^3 b, \qquad f(4) = a^4.$$

The probability distribution is properly normalized according to Eq. (19.41). The probability of three 6s in four tosses is

$$4a^3b = 4\frac{5/6}{6^3} = \frac{5}{4 \cdot 3^4},$$

fairly small. ■

This case dealt with repeated independent trials, each with two possible outcomes of constant probability p for a hit and $q = 1 - p$ for a miss, and it is typical of many applications, such as defective products, hits or misses of a target, and decays of radioactive atoms. The generalization to $X = x$ successes in n trials is given by the **binomial probability distribution**

$$f(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x} = \binom{n}{x} p^x q^{n-x}, \tag{19.42}$$

using the binomial coefficients (see Chapter 5). This distribution is normalized to the probability 1 of all possibilities in n trials, as can be seen from the binomial expansion

$$1 = (p+q)^n = p^n + np^{n-1}q + \cdots + npq^{n-1} + q^n. \tag{19.43}$$

Figure 19.5 shows typical histograms. The random variable X takes the values $0, 1, 2, \ldots, n$ in discrete steps and can also be viewed as a composition $\sum_i X_i$ of n independent random variables X_i, one for each trial, that have the value 0 for a miss and 1 for a hit. This observation allows us to employ the moment-generating functions

$$\langle e^{tX_i} \rangle = P(X_i = 0) + e^t P(X_i = 1) = q + pe^t \tag{19.44}$$

and

$$\langle e^{tX} \rangle = \prod_i \langle e^{tX_i} \rangle = \left(pe^t + q\right)^n, \tag{19.45}$$

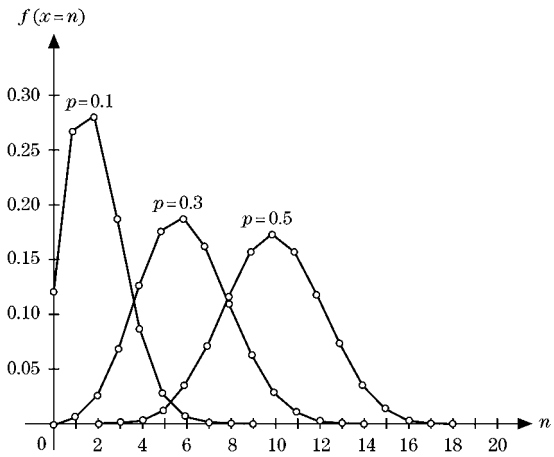

FIGURE 19.5 Binomial probability distributions for $n = 20$ and $p = 0.1, 0.3, 0.5$.

from which the mean values and higher moments can be read off upon differentiating and setting $t = 0$. Using

$$\frac{\partial \langle e^{tX} \rangle}{\partial t} = npe^t \left(pe^t + q\right)^{n-1},$$

$$\left.\frac{\partial \langle e^{tX} \rangle}{\partial t}\right|_{t=0} = \langle X \rangle = \sum_i x_i f(x_i) = np,$$

$$\frac{\partial^2 \langle e^{tX} \rangle}{\partial t^2} = npe^t \left(pe^t + q\right)^{n-1} + n(n-1)p^2 e^{2t} \left(pe^t + q\right)^{n-2},$$

$$\left\langle X^2 \right\rangle = \left.\frac{\partial^2 \langle e^{tX} \rangle}{\partial t^2}\right|_{t=0} = \sum_i x_i^2 f(x_i) = np + n(n-1)p^2,$$

we obtain, with Eq. (19.17),

$$\begin{aligned} \sigma^2(X) &= \left\langle X^2 \right\rangle - \langle X \rangle^2 = np + n(n-1)p^2 - n^2 p^2 \\ &= np(1-p) = npq. \end{aligned} \tag{19.46}$$

Figure 19.5 illustrates these results with peaks at $x = np = 2, 6, 10$, which widen with increasing p.

Exercises

19.3.1 Show that the variable $X = x$ number of heads in n coin tosses is a random variable, and determine its probability distribution. Describe the sample space. What are its mean value, the variance, and the standard deviation? Plot the probability function $f(x) = n!/(x!(n-x)!2^n)$ for $n = 10, 20, 30$ using graphical software.

19.3.2 Plot the binomial probability function for the probabilities $p = 1/6,\ q = 5/6$ and $n = 6$ throws of a die.

19.3.3 A hardware company knows that the probability of mass-producing nails includes a small probability $p = 0.03$ of defective nails (without a sharp tip usually). What is the probability of finding more than two defective nails in its commercial box of 100 nails?

19.3.4 Four cards are drawn from a shuffled bridge deck. What is the probability that they are all red? that they are all hearts? that they are honors? Compare the probabilities when the cards are put back at random places, or not.

19.3.5 Show that for the binomial distribution of Eq. (19.42) the most probable value of x is np.

19.4 POISSON DISTRIBUTION

The Poisson distribution typically occurs in situations involving an event repeated at a constant rate of probability, thereby depleting the population. The decay of a radioactive

sample is a case in point because, once a particle decays, it does not decay again. If the observation time dt is small enough so that the emission of two or more particles is negligible, then the probability that one particle (He^4 in α decay or an electron in β decay) is emitted is $\mu\,dt$ with constant μ and $\mu\,dt \ll 1$. We can set up a recursion relation for the probability $P_n(t)$ of observing n counts during a time interval t. For $n > 0$ the probability $P_n(t+dt)$ is composed of two mutually exclusive events that (i) n particles are emitted in the time t, none in dt, and (ii) $n-1$ particles are emitted in time t, one in dt. Therefore

$$P_n(t+dt) = P_n(t)P_0(dt) + P_{n-1}(t)P_1(dt).$$

Here we substitute the probability of observing one particle, $P_1(dt) = \mu\,dt$, and no particle, $P_0(dt) = 1 - P_1(dt)$, in time dt. This yields

$$P_n(t+dt) = P_n(t)(1-\mu\,dt) + P_{n-1}(t)\mu\,dt.$$

So, after rearranging and dividing by dt, we get

$$\frac{dP_n(t)}{dt} = \frac{P_n(t+dt) - P_n(t)}{dt} = \mu P_{n-1}(t) - \mu P_n(t). \tag{19.47}$$

For $n = 0$ this differential recursion relation simplifies, because there is no particle in times t and dt giving

$$\frac{dP_0(t)}{dt} = -\mu P_0(t). \tag{19.48}$$

The ODE says that particles have a constant decay probability and decay removes them from the distribution. This ODE integrates to $P_0(t) = e^{-\mu t}$ if the probability that no particle is emitted during a zero time interval $P_0(0) = 1$ is used. Here $P_0(0) = 1$ means no decay takes place at $t \le 0$.

Now we go back to Eq. (19.47) for $n = 1$,

$$\dot{P}_1 = \mu\left(e^{-\mu t} - P_1\right), \qquad P_1(0) = 0, \tag{19.49}$$

and solve the homogeneous equation, which is the same for P_1 as Eq. (19.48). This yields $P_1(t) = \mu_1 e^{-\mu t}$. Then we solve the inhomogeneous ODE (Eq. (19.49)) by varying the constant μ_1 to find $\dot{\mu}_1 = \mu$, so $P_1(t) = \mu t e^{-\mu t}$. The general solution is

$$P_n(t) = \frac{(\mu t)^n}{n!} e^{-\mu t}, \tag{19.50}$$

as may be confirmed by substitution into Eq. (19.47) and verifying the initial conditions, $P_n(0) = 0$, $n > 0$. This is an example of the Poisson distribution.

The Poisson distribution is defined with the probabilities

$$p(n) = \frac{\mu^n}{n!} e^{-\mu}, \qquad X = n = 0, 1, 2, \ldots \tag{19.51}$$

and is exhibited in Fig. 19.6. The random variable X is discrete. The probabilities are properly normalized because $e^{-\mu}\sum_{n=0}^{\infty} \frac{\mu^n}{n!} = 1$. The mean value and variance,

$$\begin{aligned} \langle X \rangle &= e^{-\mu} \sum_{n=1}^{\infty} n \frac{\mu^n}{n!} = \mu e^{-\mu} \sum_{n=0}^{\infty} \frac{\mu^n}{n!} = \mu, \\ \sigma^2 &= \left\langle X^2 \right\rangle - \langle X \rangle^2 = \mu(\mu+1) - \mu^2 = \mu, \end{aligned} \tag{19.52}$$

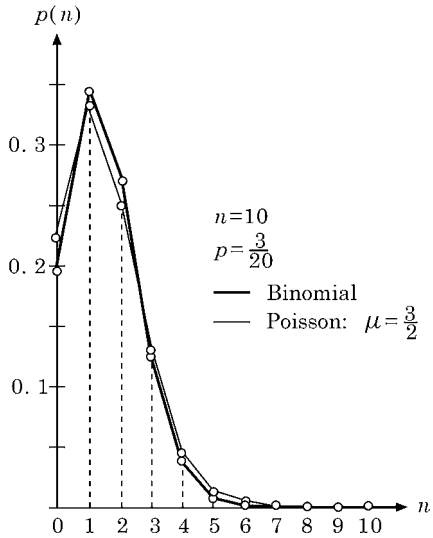

FIGURE 19.6 Poisson distribution compared with binomial distribution.

follow from the characteristic function

$$\left\langle e^{itX}\right\rangle = \sum_{n=0}^{\infty} e^{itn-\mu}\frac{\mu^n}{n!} = e^{-\mu}\sum_{n=0}^{\infty}\frac{(\mu e^{it})^n}{n!} = e^{\mu(e^{it}-1)}$$

by differentiation and setting $t = 0$, using Eq. (19.17).

A Poisson distribution becomes a good approximation of the binomial distribution for a large number n of trials and small probability $p \sim \mu/n$, μ a constant.

THEOREM: *In the limit $n \to \infty$ and $p \to 0$ so that the mean value $np \to \mu$ stays finite, the binomial distribution becomes a Poisson distribution.*

To prove this theorem, we apply Stirling's formula (Chapter 8) $n! \sim \sqrt{2\pi n}(n/e)^n$ for large n to the factorials in Eq. (19.42), keeping x finite while $n \to \infty$. This yields for $n \to \infty$:

$$\frac{n!}{(n-x)!} \sim \left(\frac{n}{e}\right)^n \left(\frac{e}{n-x}\right)^{n-x} \sim \left(\frac{n}{e}\right)^x \left(\frac{n}{n-x}\right)^{n-x}$$
$$\sim \left(\frac{n}{e}\right)^x \left(1+\frac{x}{n-x}\right)^{n-x} \sim \left(\frac{n}{e}\right)^x e^x \sim n^x,$$

and for $n \to \infty$, $p \to 0$, with $np \to \mu$:

$$(1-p)^{n-x} \sim \left(1-\frac{pn}{n}\right)^n \sim \left(1-\frac{\mu}{n}\right)^n \sim e^{-\mu}.$$

Table 19.1

$i \rightarrow$	0	1	2	3	4	5	6	7	8	9	10
$n_i \rightarrow$	57	203	383	525	532	408	273	139	45	27	16

Finally, $p^x n^x \rightarrow \mu^x$, so altogether

$$\frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \rightarrow \frac{\mu^x}{x!} e^{-\mu}, \qquad n \rightarrow \infty, \tag{19.53}$$

which is a Poisson distribution for the random variable $X = x$ with $0 \leq x < \infty$. This limit theorem is a particular example of the **laws of large numbers**.

Exercises

19.4.1 Radioactive decays are governed by the Poisson distribution. In a Rutherford–Geiger experiment the number n_i of emitted α particles is counted in $n = 2608$ time intervals of 7.5 seconds each. In Table 19.1 n_i is the number of time intervals in which i particles were emitted. Determine the average number λ of emitted particles, and compare the n_i of Table 19.1 with np_i computed from the Poisson distribution with mean value λ.

19.4.2 Derive the standard deviation of a Poisson distribution of mean value μ.

19.4.3 The number of α decay particles of a radium sample is counted per minute for 40 hours. The total number is 5000. How many 1-minute intervals are there expected to be with (a) 2, (b) 5 α particles?

19.4.4 For a radioactive sample, 10 decays are counted on average in 100 seconds. Use the Poisson distribution to estimate the probability of counting 3 decays in 10 seconds.

19.4.5 ^{238}U has a half-life of 4.51×10^9 years. Its decay series ends with the stable lead isotope ^{206}Pb. The ratio of the number of ^{206}Pb to ^{238}U atoms in a rock sample is measured as 0.0058. Estimate the age of the rock assuming that all the lead in the rock is from the initial decay of the ^{238}U, which determines the rate of the entire decay process, because the subsequent steps take place far more rapidly.
Hint. The decay constant λ in the decay law $N(t) = Ne^{-\lambda t}$ is related to the half-life T by $T = \ln 2/\lambda$.

ANS. 3.8×10^7 years.

19.4.6 The probability of hitting a target in one shot is known to be 20%. If five shots are fired independently, what is the probability of striking the target at least once?

19.4.7 A piece of uranium is known to contain the isotopes $^{235}_{92}$U and $^{238}_{92}$U as well as from 0.80 g of $^{206}_{82}$Pb per gram of uranium. Estimate the age of the piece (and thus Earth) in years.
Hint. Assume the lead comes only from the $^{238}_{92}$U. Use the decay constant from Exercise 19.4.5.

19.5 GAUSS' NORMAL DISTRIBUTION

The bell-shaped Gauss distribution is defined by the probability density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{[x-\mu]^2}{2\sigma^2}\right), \qquad -\infty < x < \infty, \tag{19.54}$$

with mean value μ and variance σ^2. It is by far the most important continuous probability distribution and is displayed in Fig. 19.7.

It is properly normalized because, substituting $y = \frac{x-\mu}{\sigma\sqrt{2}}$, we obtain

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-y^2}\, dy = \frac{2}{\sqrt{\pi}} \int_{0}^{\infty} e^{-y^2}\, dy = 1.$$

Similarly, substituting $y = x - \mu$, we see that

$$\langle X \rangle - \mu = \int_{-\infty}^{\infty} \frac{x-\mu}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \int_{-\infty}^{\infty} \frac{y}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}}\, dy = 0,$$

the integrand being odd in y, so the integral over $y > 0$ cancels that over $y < 0$. Similarly we check that the standard deviation is σ.

From the normal distribution (by the substitution $y = \frac{x-\langle X \rangle}{\sigma}$)

$$P\big(|X - \langle X \rangle| > k\sigma\big) = P\left(\frac{|X - \langle X \rangle|}{\sigma} > k\right) = P\big(|Y| > k\big)$$

$$= \sqrt{\frac{2}{\pi}} \int_{k}^{\infty} e^{-y^2/2}\, dy = \sqrt{\frac{4}{\pi}} \int_{k/\sqrt{2}}^{\infty} e^{-z^2}\, dz = \operatorname{erfc} \frac{k}{\sqrt{2}},$$

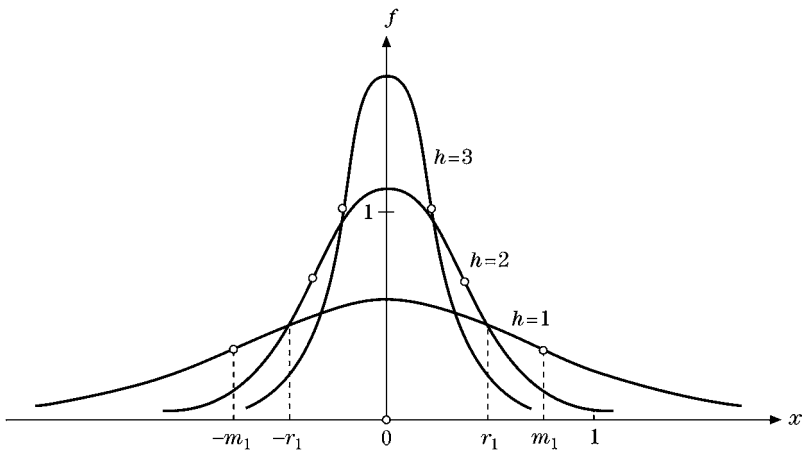

FIGURE 19.7 Normal Gauss distribution for mean value zero and various standard deviations $h = 1/\sigma\sqrt{2}$.

$$P\left(|X-\langle X\rangle| \geq 3\sigma\right) \sim 0.0027, \qquad P\left(|X-\langle X\rangle| \geq 2\sigma\right) \sim 0.0455, \tag{19.55}$$

we can evaluate the integral for $k = 1, 2, 3$ and thus extract the following numerical relations for a normally distributed random variable X with $\langle X\rangle$ of which the last one is interesting to compare with Chebychev's inequality (see Eq. (19.21).) giving $\leq 1/9$ for an **arbitrary** probability distribution instead of ~ 0.0027 for the 3σ-rule of the **normal** distribution.

ADDITION THEOREM: *If the random variables X, Y have the same normal distributions, that is, the same mean value and variance, then $Z = X + Y$ has normal distribution with twice the mean value and twice the variance of X and Y.*

To prove this theorem, we take the Gauss density as

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad \text{with } \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2}\, dx = 1,$$

without loss of generality. Then the probability density of (X, Y) is the product

$$f(x, y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{2\pi} e^{-(x^2+y^2)/2}.$$

Also, Eq. (17.36) gives the density for $Z = X + Y$ as

$$g(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-(x-z)^2/2}\, dx.$$

Completing the square in the exponent,

$$2x^2 - 2xz + z^2 = \left(x\sqrt{2} - \frac{z}{\sqrt{2}}\right)^2 + \frac{z^2}{2},$$

we obtain

$$g(z) = \frac{1}{2\pi} e^{-z^2/4} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(x\sqrt{2} - \frac{z}{\sqrt{2}}\right)^2\right) dx.$$

Using the substitution $u = x - \frac{z}{2}$, we find that the integral transforms into

$$\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(x\sqrt{2} - \frac{z}{\sqrt{2}}\right)^2\right) dx = \int_{-\infty}^{\infty} e^{-u^2}\, du = \sqrt{\pi},$$

so the density for $Z = X + Y$ is

$$g(z) = \frac{1}{2\sqrt{\pi}} e^{-z^2/4}, \tag{19.56}$$

which means it has mean value zero and variance 2, twice that of X and Y.

In a special limit the discrete Poisson probability distribution is closely related to the continuous Gauss distribution. This limit theorem is another example of the **laws of large numbers**, which are often dominated by the bell-shaped normal distribution.

THEOREM: *For large n and mean value μ, the Poisson distribution approaches a Gauss distribution.*

To prove this theorem for $n \to \infty$, we approximate the factorial in the Poisson's probability $p(n)$ of Eq. (19.51) by Stirling's asymptotic formula (see Chapter 8),

$$n! \sim \sqrt{2n\pi}\left(\frac{n}{e}\right)^n, \qquad n \to \infty,$$

and choose the deviation $v = n - \mu$ from the mean value as the new variable. We let the mean value $\mu \to \infty$ and treat v/μ as small but v^2/μ as finite. Substituting $n = \mu + v$ and expanding the logarithm in a MacLaurin series, keeping two terms, we obtain

$$\begin{aligned}
\ln p(n) &= -\mu + n\ln\mu - n\ln n + n - \ln\sqrt{2n\pi} \\
&= (\mu+v)\ln\mu - (\mu+v)\ln(\mu+v) + v - \ln\sqrt{2\pi(\mu+v)} \\
&= (\mu+v)\ln\left(1 - \frac{v}{\mu+v}\right) + v - \ln\sqrt{2\pi\mu} \\
&= (\mu+v)\left(-\frac{v}{\mu+v} - \frac{v^2}{2(\mu+v)^2}\right) + v - \ln\sqrt{2\pi\mu} \\
&\sim -\frac{v^2}{2\mu} - \ln\sqrt{2\pi\mu},
\end{aligned}$$

replacing $\mu + v \to \mu$ because $|v| \ll \mu$. Exponentiating this result we find that for large n and μ

$$p(n) \to \frac{1}{\sqrt{2\pi\mu}} e^{-v^2/2\mu}, \tag{19.57}$$

which is a Gauss distribution of the continuous variable v with mean value 0 and standard deviation $\sigma = \sqrt{\mu}$.

In a special limit the discrete binomial probability distribution is also closely related to the continuous Gauss distribution. This limit theorem is another example of the **laws of large numbers**.

THEOREM: *In the limit $n \to \infty$, so that the mean value $np \to \infty$, the binomial distribution becomes Gauss' normal distribution. Recall from Section* 19.4 *that, when $np \to \mu < \infty$, the binomial distribution becomes a Poisson distribution.*

Instead of the large number x of successes in n trials, we use the deviation $v = x - pn$ from the (large) mean value pn as our new continuous random variable, under the condition that $|v| \ll pn$ but v^2/n is finite as $n \to \infty$. Thus, we replace x by $v + pn$ and $n - x$ by $qn - v$ in the factorials of Eq. (19.42), $f(x) \to W(v)$ as $n \to \infty$, and then apply Stirling's formula. This yields

$$W(v) = \frac{p^x q^{n-x} n^{n+1/2} e^{-n+x+(n-x)}}{\sqrt{2\pi}(v+pn)^{x+1/2}(qn-v)^{n-x+1/2}}.$$

Here we factor out the dominant powers of n and cancel powers of p and q to find

$$W(v) = \frac{1}{\sqrt{2\pi pqn}} \left(1 + \frac{v}{pn}\right)^{-(v+pn+1/2)} \left(1 - \frac{v}{qn}\right)^{-(qn-v+1/2)}.$$

In terms of the logarithm we have

$$\begin{aligned}
\ln W(v) &= \ln \frac{1}{\sqrt{2\pi pqn}} - (v + pn + 1/2) \ln\left(1 + \frac{v}{pn}\right) \\
&\quad - (qn - v + 1/2) \ln\left(1 - \frac{v}{qn}\right) \\
&= \ln \frac{1}{\sqrt{2\pi pqn}} - (v + pn + 1/2)\left(\frac{v}{pn} - \frac{v^2}{2p^2n^2} + \cdots\right) \\
&\quad - (qn - v + 1/2)\left(-\frac{v}{qn} - \frac{v^2}{2q^2n^2} + \cdots\right) \\
&= \ln \frac{1}{\sqrt{2\pi pqn}} - \left[\frac{v}{n}\left(\frac{1}{2p} - \frac{1}{2q}\right) + \frac{v^2}{n}\left(\frac{1}{2p} + \frac{1}{2q}\right) + \cdots\right],
\end{aligned}$$

where

$$\frac{v}{n} \to 0, \qquad \frac{v^2}{n}$$

is finite and

$$\left(\frac{1}{2p} + \frac{1}{2q}\right) = \frac{p+q}{2pq} = \frac{1}{2pq}.$$

Neglecting higher orders in v/n, such as v^2/p^2n^2 and v^2/q^2n^2, we find the large n limit

$$W(v) = \frac{1}{\sqrt{2\pi pqn}} e^{-v^2/2pqn}, \tag{19.58}$$

which is a Gaussian distribution in the deviations $x - pn$, with mean value 0 and standard deviation $\sigma = \sqrt{npq}$. The large mean value pn (and the discarded terms) restricts the validity of the theorem to the central part of the Gaussian bell shape, excluding the tails.

Exercises

19.5.1 What is the probability for a normally distributed random variable to differ by more than 4σ from its mean value? Compare your result with the corresponding one from Chebychev's inequality. Explain the difference in your own words.

19.5.2 Let $X_1, X_2, \ldots, X_n$ be independent normal random variables with the same mean $\bar{x}$ and variance σ^2. Show that $\frac{\sum_i X_i/n - \bar{x}}{\sqrt{n}\sigma}$ is normal with mean zero and variance 1.

19.5.3 An instructor grades a final exam of a large undergraduate class, obtaining the mean value of points M and the variance σ^2. Assuming a normal distribution for the number M of points, he defines a grade F when $M < m - 3\sigma/2$, D when $m - 3\sigma/2 < M < m - \sigma/2$, C when $m - \sigma/2 < M < m + \sigma/2$, B when $m + \sigma/2 < M < m + 3\sigma/2$, A when $M > m + 3\sigma/2$. What is the percentage of As, Fs; Bs, Ds; Cs? Redesign the cutoffs so that there are equal percentages of As and Fs (5%), 25% Bs and Ds, and 40% Cs.

19.5.4 If the random variable X is normal with mean value 29 and standard deviation 3, what are the distributions of $2X - 1$ and $3X + 2$?

19.5.5 For a normal distribution of mean value m and variance σ^2, find the distance r such that half the area under the bell shape is between $m - r$ and $m + r$.

19.6 STATISTICS

In statistics, probability theory is applied to the evaluation of data from random experiments or to samples to test some hypothesis because the data have random fluctuations due to lack of complete control over the experimental conditions. Typically one attempts to estimate the mean value and variance of the distributions, from which the samples derive, and to generalize properties valid for a sample to the rest of the events at a prescised confidence level. Any assumption about an unknown probability distribution is called a **statistical hypothesis**. The concepts of tests and confidence intervals are among the most important developments of statistics.

Error Propagation

When we measure a quantity x repeatedly, obtaining the values x_j at random, or select a sample for testing, we determine the mean value (see Eq. (19.10)) and the variance,

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j, \qquad \sigma^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2,$$

as a measure for the error, or spread from the mean value $\bar{x}$. We can write $x_j = \bar{x} + e_j$, where the error e_j is the deviation from the mean value, and we know that $\sum_j e_j = 0$. (See the discussion after Eq. (19.15).)

Now suppose we want to determine a known function $f(x)$ from these measurements; that is, we have a set $f_j = f(x_j)$ from the measurements of x. Substituting $x_j = \bar{x} + e_j$ and forming the mean value from

$$\begin{aligned}\bar{f} &= \frac{1}{n}\sum_j f(x_j) = \frac{1}{n}\sum_j f(\bar{x} + e_j)\\ &= f(\bar{x}) + \frac{1}{n}f'(\bar{x})\sum_j e_j + \frac{1}{2n}f''(\bar{x})\sum_j e_j^2 + \cdots\\ &= f(\bar{x}) + \frac{1}{2}\sigma^2 f''(\bar{x}) + \cdots,\end{aligned} \tag{19.59}$$

we obtain the average value $\bar{f}$ as $f(\bar{x})$ in lowest order, as expected. But in second order there is a correction given by half the variance with a scale factor $f''(\bar{x})$. It is interesting to compare this correction of the mean value with the average spread of individual f_j from the mean value $\bar{f}$, the variance of f. To lowest order, this is given by the average of the sum of squares of the deviations, in which we approximate $f_j \approx \bar{f} + f'(\bar{x})e_j$, yielding

$$\sigma^2(f) \equiv \frac{1}{n}\sum_j (f_j - \bar{f})^2 = \big(f'(\bar{x})\big)^2 \frac{1}{n}\sum_j e_j^2 = \big(f'(\bar{x})\big)^2 \sigma^2. \tag{19.60}$$

In summary we may formulate somewhat symbolically

$$f(\bar{x} \pm \sigma) = f(\bar{x}) \pm f'(\bar{x})\sigma$$

as the simplest form of error propagation by a function of one measured variable.

For a function $f(x_j, y_k)$ of two measured quantities $x_j = \bar{x} + u_j$, $y_k = \bar{y} + v_k$, we obtain similarly

$$\begin{aligned}\bar{f} &= \frac{1}{rs}\sum_{j=1}^{r}\sum_{k=1}^{s} f_{jk} = \frac{1}{rs}\sum_{j=1}^{r}\sum_{k=1}^{s} f(\bar{x}+u_j, \bar{y}+v_k) \\ &= f(\bar{x},\bar{y}) + \frac{1}{r}f_x \sum_j u_j + \frac{1}{s}f_x \sum_k v_k + \cdots,\end{aligned}$$

where $\sum_j u_j = 0 = \sum_k v_k$, so again $\bar{f} = f(\bar{x}, \bar{y})$ in lowest order. Here

$$f_x = \frac{\partial f}{\partial x}(\bar{x},\bar{y}), \qquad f_y = \frac{\partial f}{\partial y}(\bar{x},\bar{y}) \tag{19.61}$$

denote partial derivatives. The sum of squares of the deviations from the mean value is given by

$$\sum_{j=1}^{r}\sum_{k+1}^{s}(f_{jk} - \bar{f})^2 = \sum_{j,k}(u_j f_x + v_k f_y)^2 = s f_x^2 \sum_j u_j^2 + r f_y^2 \sum_k v_k^2,$$

because $\sum_{j,k} u_j v_k = \sum_j u_j \sum_k v_k = 0$. Therefore the variance is

$$\sigma^2(f) = \frac{1}{rs}\sum_{j,k}(f_{jk} - \bar{f})^2 = f_x^2\sigma_x^2 + f_y^2\sigma_y^2, \tag{19.62}$$

with f_x, f_y from Eq. (19.61); and

$$\sigma_x^2 = \frac{1}{r}\sum_j u_j^2, \qquad \sigma_y^2 = \frac{1}{s}\sum_k v_k^2$$

are the variances of the x and y data points. Symbolically the error propagation for a function of two measured variables may be summarized as

$$f(\bar{x} \pm \sigma_x, \bar{y} \pm \sigma_y) = f(\bar{x},\bar{y}) \pm \sqrt{f_x^2\sigma_x^2 + f_y^2\sigma_y^2}.$$

As an application and generalization of the last result, we now calculate the error of the mean value $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$ of a sample of n individual measurements x_j, each with

spread σ. In this case the partial derivatives are given by $f_x = \frac{1}{n} = f_y = \cdots$ and $\sigma_x = \sigma = \sigma_y = \cdots$. Thus, our last error propagation rule tells us that errors of a sum of variables add quadratically, so the uncertainty of the arithmetic mean is given by

$$\bar{\sigma} = \frac{1}{n}\sqrt{n\sigma^2} = \frac{\sigma}{\sqrt{n}}, \tag{19.63}$$

decreasing with the number of measurements n.

As the number n of measurements increases, we expect the arithmetic mean $\bar{x}$ to converge to some true value x. Let $\bar{x}$ differ from x by α and $v_j = x_j - x$ be the true deviations; then

$$\sum_j (x_j - \bar{x})^2 = \sum_j e_j^2 = \sum_j v_j^2 + n\alpha^2.$$

Taking into account the error of the arithmetic mean, we determine the spread of the individual points about the unknown true mean value to be

$$\sigma^2 = \frac{1}{n}\sum_j e_j^2 = \frac{1}{n}\sum_j v_j^2 + \alpha^2.$$

According to our earlier discussion leading to Eq. (19.63), $\alpha^2 = \frac{1}{n}\sigma^2$. As a result

$$\sigma^2 = \frac{1}{n}\sum_j v_j^2 + \frac{\sigma^2}{n},$$

from which the **standard deviation of a sample in statistics** follows:

$$\sigma = \sqrt{\frac{\sum_j v_j^2}{n-1}} = \sqrt{\frac{\sum_j (x_j - x)^2}{n-1}}, \tag{19.64}$$

with $n-1$ being the number of control measurements of the sample. This modified mean error includes the expected error in the arithmetic mean.

Because the spread is not well defined when there is no comparison measurement, that is, when $n = 1$, the variance is sometimes defined by Eq. (19.64), in which we replace the number n of measurements by the number $n-1$ of control measurements in statistics.

Fitting Curves to Data

Suppose we have a sample of measurements y_j (for example, a particle moving freely, that is, no force) taken at known times t_j (which are taken to be practically free of errors; that is, the time t is an ordinary independent variable) that we expect to be linearly related as $y = at$, our hypothesis. We want to fit this line to the data.

First we minimize the sum of deviations $\sum_j (at_j - y_j)^2$ to determine the slope parameter a, also called the **regression coefficient**, using the method of least squares. Differentiating with respect to a we obtain

$$2\sum_j (at_j - y_j)t_j = 0,$$

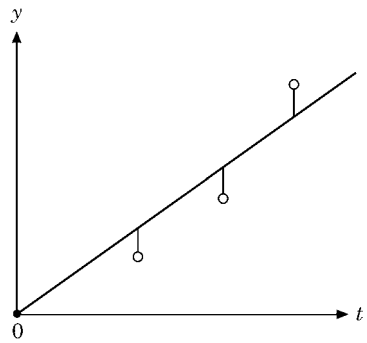

FIGURE 19.8 Straight line fit to data points (t_j, y_j) with t_j known, y_j measured.

from which

$$a = \frac{\sum_j t_j y_j}{\sum_j t_j^2} \tag{19.65}$$

follows. Note that the numerator is built like a sample covariance, the scalar product of the variables t, y of the sample. As shown in Fig. 19.8, the measured values y_j do not lie on the line as a rule. They have the spread (or root mean square deviation from the fitted line)

$$\sigma = \sqrt{\frac{\sum_j (y_j - at_j)^2}{n-1}}.$$

Alternatively, let the y_j values be known (without error) while t_j are measurements. As suggested by Fig. 19.9, in this case we need to interchange the role of t and y and to fit the line $t = by$ to the data points. We minimize $\sum_j (by_j - t_j)^2$, set the derivative with respect

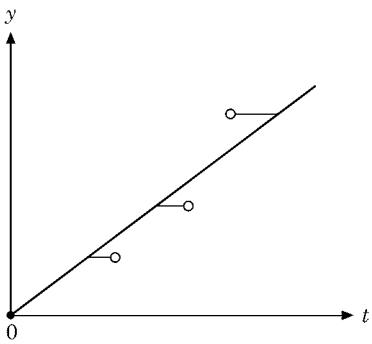

FIGURE 19.9 Straight line fit to data points (t_j, y_j) with y_j known, t_j measured.

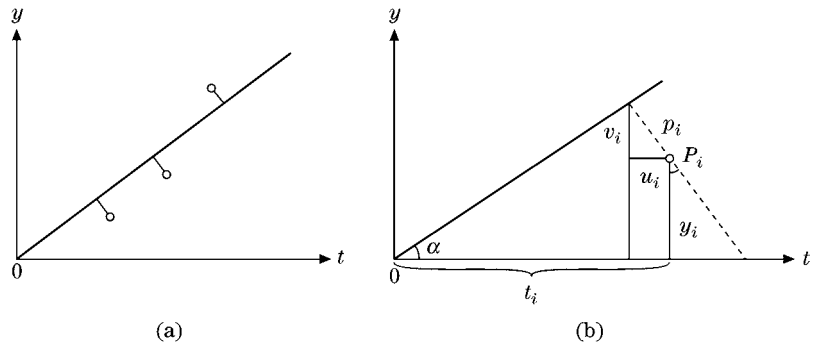

FIGURE 19.10 (a) Straight line fit to data points (t_j, y_j). (b) Geometry of deviations u_j, v_j, d_j.

to b equal to zero, and find similarly the slope parameter

$$b = \frac{\sum_j t_j y_j}{\sum_j y_j^2}. \tag{19.66}$$

In case both t_j and y_j have errors (we take t and y to have the same units), we have to minimize the sum of squares of the deviations of both variables and fit to a parameterization $t \sin\alpha - y \cos\alpha = 0$, where t and y occur on an equal footing. As displayed in Fig. (19.10a) this means geometrically that the line has to be drawn so that the sum of the squares of the distances d_j of the points (t_j, y_j) from the line becomes a minimum. (See Fig. 19.10b and Chapter 1.) Here $d_j = t_j \sin\alpha - y_j \cos\alpha$, so $\sum_j d_j^2 =$ minimum must be solved for the angle α. Setting the derivative with respect to the angle equal to zero,

$$\sum_j (t_j \sin\alpha - y_j \cos\alpha)(t_j \cos\alpha + y_j \sin\alpha) = 0,$$

yields

$$\sin\alpha \cos\alpha \sum_j \left(t_j^2 - y_j^2\right) - \left(\cos^2\alpha - \sin^2\alpha\right) \sum_j t_j y_j = 0.$$

Therefore the angle of the straight-line fit is given by

$$\tan 2\alpha = \frac{2\sum_j t_j y_j}{\sum_j (t_j^2 - y_j^2)}. \tag{19.67}$$

This least-squares fitting applies when the measurement errors are unknown. It allows assigning at least some kind of error bar to the measured points. Recall that we did not use errors for the points. Our parameter a (or α) is most likely to reproduce the data under these circumstances. More precisely, the least-squares method is a maximum-likelihood estimate of the fitted parameters when it is reasonable to assume that the errors are independent and normally distributed with **the same deviation for all points**. This fairly

strong assumption can be relaxed in "weighted" least-squares fits called **chi square fits**.[4] (See also Example 19.6.1.)

The χ^2 Distribution

This distribution is typically applied to fits of a curve $y(t, a, \ldots)$ with parameters $a, \ldots$ to data t_j using the method of least squares involving the **weighted** sum of squares of deviations; that is,

$$\chi^2 = \sum_{j=1}^{N} \left(\frac{y_j - y(t_j, a, \ldots)}{\Delta y_j} \right)^2$$

is minimized, where N is the number of points and r is the number of adjusted parameters $a, \ldots$. This quadratic merit function gives more weight to points with small measurement uncertainties Δy_j.

We represent each point by a normally distributed random variable X with zero mean value and variance $\sigma^2 = 1$, the latter in view of the weights in the χ^2 function. In a first step, we determine the probability density for the random variable $Y = X^2$ of a single point that takes only positive values. Assuming a zero mean value is no loss of generality because, if $\langle X \rangle = m \neq 0$, we would consider the shifted variable $Y = X - m$, whose mean value is zero. We show that if X has a Gauss normal density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}, \qquad -\infty < x < \infty,$$

then the probability of the random variable Y is zero if $y \leq 0$, and

$$P(Y < y) = P(X^2 < y) = P(-\sqrt{y} < X < \sqrt{y}) \qquad \text{if } y > 0.$$

From the continuous normal distribution $P(y) = \int_{-\infty}^{y} f(x)\, dx$, we obtain the probability density $g(y)$ by differentiation:

$$\begin{aligned} g(y) &= \frac{d}{dy}\big[P(\sqrt{y}) - P(-\sqrt{y})\big] = \frac{1}{2\sqrt{y}}\big(f(\sqrt{y}) + f(-\sqrt{y})\big) \\ &= \frac{1}{\sigma\sqrt{2\pi y}} e^{-y/2\sigma^2}, \qquad y > 0. \end{aligned} \tag{19.68}$$

This density, $\sim e^{-y/2\sigma^2}/\sqrt{y}$, corresponds to the integrand of the Euler integral of the gamma function. Such a probability distribution

$$g(y) = \frac{y^{p-1}}{\Gamma(p)(2\sigma^2)^p} e^{-y/2\sigma^2}$$

is called a **gamma distribution** with parameters p and σ. Its characteristic function for our case, $p = 1/2$, is proportional to the Fourier transform

$$\begin{aligned} \left\langle e^{itY} \right\rangle &= \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty e^{-y(1/2\sigma^2 - it)} \frac{dy}{\sqrt{y}} = \frac{1}{\sigma\sqrt{2\pi}\,(\frac{1}{2\sigma^2} - it)^{1/2}} \int_0^\infty e^{-x} \frac{dx}{\sqrt{x}} \\ &= \left(1 - 2it\sigma^2\right)^{-1/2}. \end{aligned}$$

[4] For more details, see Chapter 14 of Press et al. in the Additional Readings of Chapter 9.

Since the χ^2 sample function contains a sum of squares, we need the following theorem.

ADDITION THEOREM: *for the gamma distributions*: *If the independent random variables* Y_1 *and* Y_2 *have a gamma distribution with* $p = 1/2$, *and the same* σ *then* $Y_1 + Y_2$ *has a gamma distribution with* $p = 1$.

Since Y_1 and Y_2 are independent, the product of their densities (Eq. (19.36)) generates the characteristic function

$$\left\langle e^{it(Y_1+Y_2)}\right\rangle = \left\langle e^{itY_1}e^{itY_2}\right\rangle = \left\langle e^{itY_1}\right\rangle\left\langle e^{itY_2}\right\rangle = \left(1-2it\sigma^2\right)^{-1}. \tag{19.69}$$

Now we come to the second step. We assess the quality of the fit by the random variable $Y = \sum_{j=1}^{n} X_j^2$, where $n = N - r$ is the number of degrees of freedom for N data points and r fitted parameters. The independent random variables X_j are taken to be normally distributed with the (sample) variance σ^2. (In our case $r = 1$ and $\sigma = 1$.) The χ^2 analysis does not really test the assumptions of normality and independence, but if these are not approximately valid, there will be many outlying points in the fit. The addition theorem gives the probability density (Fig. 19.11) for Y,

$$g_n(y) = \frac{y^{\frac{n}{2}-1}}{2^{n/2}\sigma^n\Gamma(\frac{n}{2})}e^{-y/2\sigma^2}, \qquad y > 0,$$

and $g_n(y) = 0$ if $y < 0$, which is the χ^2 distribution corresponding to n degrees of freedom. Its characteristic function is

$$\left\langle e^{itY}\right\rangle = \left(1-2it\sigma^2\right)^{-n/2}.$$

Differentiating and setting $t = 0$ we obtain its mean value and variance

$$\langle Y\rangle = n\sigma^2, \qquad \sigma^2(Y) = 2n\sigma^4. \tag{19.70}$$

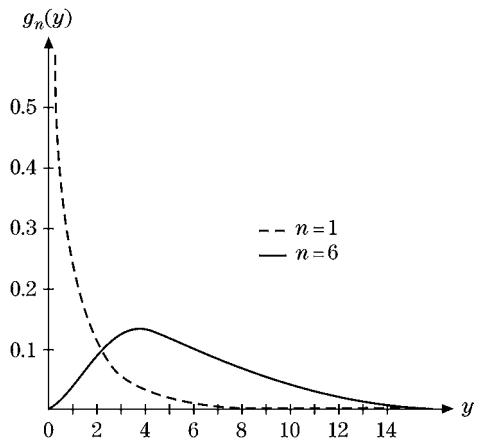

FIGURE 19.11 χ^2 probability density $g_n(y)$.

Table 19.2 χ^2 Distribution

n	$v = 0.8$	$v = 0.7$	$v = 0.5$	$v = 0.3$	$v = 0.2$	$v = 0.1$
1	0.064	0.148	0.455	1.074	1.642	2.706
2	0.446	0.713	1.386	2.408	3.219	4.605
3	1.005	1.424	2.366	3.665	4.642	6.251
4	1.649	2.195	3.357	4.878	5.989	7.779
5	2.343	3.000	4.351	6.064	7.289	9.236
6	3.070	3.828	5.348	7.231	8.558	10.645

Entries are χ_v for the probabilities $v = P(\chi^2 \geq \chi_v^2) = \frac{1}{2^{n/2}\Gamma(n/2)} \int_{\chi_v^2}^{\infty} e^{-y/2} y^{(n/2)-1}\, dy$ for $\sigma = 1$.

Tables give values for the χ^2 probability for n degrees of freedom,

$$P(\chi^2 \geq y_0) = \frac{1}{2^{n/2}\sigma^n \Gamma(\frac{n}{2})} \int_{y_0}^{\infty} y^{n/2-1} e^{-y/2\sigma^2}\, dy$$

for $\sigma = 1$ and $y_0 > 0$. To use Table 19.2 for $\sigma \neq 1$, rescale $y_0 = v_0\sigma^2$ so that $P(\chi^2 \geq v_0\sigma^2)$ corresponds to $P(\chi^2 \geq v_0)$ of Table 19.2. The following example will illustrate the whole process.

Example 19.6.1

Let us apply the χ^2 function to the fit in Fig. 19.8. The measured points $(t_j, y_j \pm \Delta y_j)$ with errors Δy_j are

$$(1, 0.8 \pm 0.1), \qquad (2, 1.5 \pm 0.05), \qquad (3, 3 \pm 0.2).$$

For comparison, the maximum-likelihood fit, Eq. (19.65), gives

$$a = \frac{1 \cdot 0.8 + 2 \cdot 1.5 + 3 \cdot 3}{1 + 4 + 9} = \frac{12.8}{14} = 0.914.$$

Minimizing instead,

$$\chi^2 = \sum_j \left(\frac{y_j - a t_j}{\Delta y_j} \right)^2,$$

gives

$$0 = \frac{\partial \chi^2}{\partial a} = -2 \sum_j \frac{t_j (y_j - a t_j)}{(\Delta y_j)^2},$$

or

$$a = \frac{\sum_j \frac{t_j y_j}{(\Delta y_j)^2}}{\sum_j \frac{t_j^2}{(\Delta y_j)^2}}.$$

In our case

$$a = \frac{\frac{1\cdot 0.8}{0.1^2} + \frac{2\cdot 1.5}{0.05^2} + \frac{3\cdot 3}{0.2^2}}{\frac{1^2}{0.1^2} + \frac{2^2}{0.05^2} + \frac{3^2}{0.2^2}} = \frac{1505}{1925} = 0.782$$

is dominated by the middle point with the smallest error, $\Delta y_2 = 0.05$. The error propagation formula (Eq. (19.62)) gives us the variance σ_a^2 of the estimate of a,

$$\sigma_a^2 = \sum_j (\Delta y_j)^2 \left(\frac{\partial a}{\partial y_j}\right)^2 = \sum_j \frac{\frac{t_j^2}{(\Delta y_j)^2}}{\left(\sum_k \frac{t_k^2}{(\Delta y_k)^2}\right)^2} = \frac{1}{\sum_j \frac{t_j^2}{(\Delta y_j)^2}}$$

using

$$\frac{\partial a}{\partial y_j} = \frac{\frac{t_j}{(\Delta y_j)^2}}{\sum_k \frac{t_k^2}{(\Delta y_k)^2}}.$$

For our case, $\sigma_a = 1/\sqrt{1925} = 0.023$; that is, our slope parameter is $a = 0.782 \pm 0.023$.

To estimate the quality of this fit of a, we compute the χ^2 probability that the two independent (control) points miss the fit by two standard deviations; that is, on average each point misses by one standard deviation. We apply the χ^2 distribution to the fit involving $N = 3$ data points and $r = 1$ parameter, that is, for $n = 3 - 1 = 2$ degrees of freedom. From Eq. (19.70) the χ^2 distribution has a mean value 2 and a variance 4. A rule of thumb is that $\chi^2 \approx n$ for a reasonably good fit. Then $P(\chi^2 \geq 2) \sim 0.496$ is read off Table 19.2, where we interpolate between $P(\chi^2 \geq 1.386^2) = 0.50$ and $P(\chi^2 \geq 2.408^2) = 0.30$ as follows:

$$\begin{aligned} P\left(\chi^2 \geq 2\right) &= P\left(\chi^2 \geq 1.386^2\right) - \frac{2 - 1.386^2}{2.408^2 - 1.386^2}\left[P\left(\chi^2 \geq 1.386^2\right) - P\left(\chi^2 \geq 2.408^2\right)\right] \\ &= 0.5 - 0.02 \cdot 0.2 = 0.496. \end{aligned}$$

Thus the χ^2 probability that, on average, each point misses by one standard deviation is nearly 50% and fairly large. ■

Our next goal is to compute a **confidence interval** for the slope parameter of our fit. A confidence interval for an a priori unknown parameter of some distribution (for example, a determined by our fit) is an interval that contains a not with certainty but with a high probability p, the confidence level, which we can choose. Such an interval is computed for a given sample. Such an analysis involves the Student t distribution.

The Student t Distribution

Because we always compute the arithmetic mean of measured points, we now consider the sample function

$$\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j,$$

where the random variables X_j are assumed independent with a normal distribution of the same mean value m and variance σ^2. The addition theorem for the Gauss distribution tells us that $X_1 + \cdots + X_n$ has the mean value nm and variance $n\sigma^2$. Therefore $(X_1 + \cdots + X_n)/n$ is normal with mean value m and variance $n\sigma^2/n^2 = \sigma^2/n$. The probability density of the variable $\bar{X} - m$ is the Gauss distribution

$$\bar{f}(\bar{x} - m) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left(-\frac{n(\bar{x} - m)^2}{2\sigma^2}\right). \tag{19.71}$$

The key problem solved by the Student t distribution is to provide estimates for the mean value m, when σ **is not known**, in terms of a sample function whose distribution is **independent** of σ. To this end, we define a rescaled sample function (traditionally called) t:

$$t = \frac{\bar{X} - m}{S}\sqrt{n-1}, \qquad S^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X})^2. \tag{19.72}$$

It can be shown that t and S are independent random variables. Following the arguments leading to the χ^2 distribution, the density of the denominator variable S is given by the gamma distribution

$$d(s) = \frac{n^{(n-1)/2} s^{n-2} e^{-ns^2/2\sigma^2}}{2^{\frac{n-3}{2}}\Gamma(\frac{n-1}{2})\sigma^{n-1}}. \tag{19.73}$$

The probability for the ratio $Z = X/Y$ of two independent random variables X, Y with normal density for $\bar{f}$ and d as given by Eqs. (19.71) and (19.73) is (Eq. (19.40))

$$R(z) = \int_{-\infty}^{z}\int_{-\infty}^{\infty} f(yz)d(y)|y|\,dy\,dz, \tag{19.74}$$

so the variable $V = (\bar{X} - m)/S$ has the density

$$\begin{aligned} r(v) &= \int_0^\infty \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left(-\frac{nv^2s^2}{2\sigma^2}\right) \frac{n^{(n-1)/2} s^{n-2} e^{-ns^2/2\sigma^2}}{2^{\frac{n-3}{2}}\Gamma(\frac{n-1}{2})\sigma^{n-1}} s\,ds \\ &= \frac{n^{n/2}}{\sigma^n\sqrt{\pi}2^{(n-2)/2}\Gamma(\frac{n-1}{2})} \int_0^\infty e^{-ns^2(v^2+1)/2\sigma^2} s^{n-1}\,ds. \end{aligned}$$

Here we substitute $z = s^2$ and obtain

$$r(v) = \frac{n^{n/2}}{\sigma^n\sqrt{\pi}2^{n/2}\Gamma(\frac{n-1}{2})} \int_0^\infty e^{-nz(v^2+1)/2\sigma^2} z^{(n-2)/2}\,dz.$$

Now we substitute $\Gamma(1/2) = \sqrt{\pi}$, define the parameter

$$a = \frac{n(v^2+1)}{2\sigma^2},$$

and transform the integral into $\Gamma(n/2)/a^{n/2}$ to find

$$r(v) = \frac{\Gamma(n/2)}{\sqrt{\pi}\Gamma(\frac{n-1}{2})(v^2+1)^{n/2}}, \qquad -\infty < v < \infty.$$

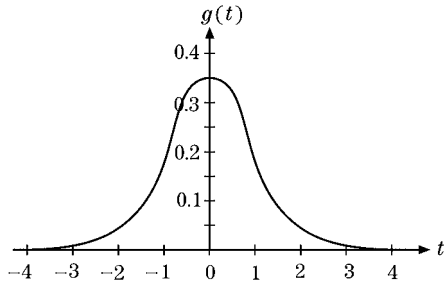

FIGURE 19.12 Student t probability density $g_n(y)$ for $n = 3$.

Table 19.3 Student t Distribution

p	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$
0.8	1.38	1.06	0.98	0.94	0.92
0.9	3.08	1.89	1.64	1.53	1.48
0.95	6.31	2.92	2.35	2.13	2.02
0.975	12.7	4.30	3.18	2.78	2.57
0.99	31.8	6.96	4.54	3.75	3.36
0.999	318.3	22.3	10.2	7.17	5.89

Entries are the values C in $P(C) = K_n \int_{-\infty}^{C} (1 + \frac{t^2}{n})^{-(n+1)/2}\, dt = p$, n is the number of degrees of freedom.

Finally we rescale this expression to the variable t in Eq. (19.72) with the density (Fig. 19.12)

$$g(t) = \frac{\Gamma(n/2)}{\sqrt{\pi(n-1)}\Gamma(\frac{n-1}{2})(1+\frac{t^2}{n-1})^{n/2}}, \qquad -\infty < t < \infty, \tag{19.75}$$

for the Student t distribution, which manifestly does not depend on m or σ. The probability for $t_1 < t < t_2$ is given by the integral

$$P(t_1, t_2) = \frac{\Gamma(n/2)}{\sqrt{\pi(n-1)}\Gamma(\frac{n-1}{2})} \int_{t_1}^{t_2} \frac{dt}{(1+\frac{t^2}{n-1})^{n/2}}, \tag{19.76}$$

and $P(z) \equiv P(-\infty, z)$ is tabulated. (See Table 19.3 for example.) Also, $P(\infty, -\infty) = 1$ and $P(-z) = 1 - P(z)$, because the integrand in Eq (19.76) is even in t, so

$$\int_{-\infty}^{-z} \frac{dt}{(1+\frac{t^2}{n-1})^{n/2}} = \int_{z}^{\infty} \frac{dt}{(1+\frac{t^2}{n-1})^{n/2}}$$

and

$$\int_{z}^{\infty} \frac{dt}{(1+\frac{t^2}{n-1})^{n/2}} = \int_{-\infty}^{\infty} \frac{dt}{(1+\frac{t^2}{n-1})^{n/2}} - \int_{-\infty}^{z} \frac{dt}{(1+\frac{t^2}{n-1})^{n/2}}.$$

Multiplying this by the factor preceding the integral in Eq. (19.76) yields $P(-z) = 1 - P(z)$. In the following example we show how to apply the Student t distribution to our fit of Example 19.6.1.

Example 19.6.2 CONFIDENCE INTERVAL

Here we want to determine a confidence interval for the slope a in the linear $y = at$ fit of Fig. 19.8. We assume

- first that the sample points (t_j, y_j) are random and independent, and
- second that, for each fixed value t, the random variable Y is normal with mean $\mu(t) = at$ and variance σ^2 independent of t.

These values y_j are measurements of the random variable Y, but we will regard them as single measurements of the independent random variables Y_j with the same normal distribution as Y (whose variance we do not know).

We choose a confidence level, $p = 95\%$, say. Then the Student probability is

$$P(-C, C) = P(C) - P(-C) = p = -1 + 2P(C),$$

hence

$$P(C) = \frac{1}{2}(1 + p),$$

using $P(-C) = 1 - P(C)$, and

$$P(C) = \frac{1}{2}(1 + p) = 0.975 = K_n \int_{-\infty}^{C} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} dt,$$

where K_{n-1} is the factor preceding the integral in Eq. (19.76). Now we determine a solution $C = 4.3$ from Table 19.3 of Student's t distribution, with $n = N - r = 3 - 1 = 2$ the number of degrees of freedom, noting that $(1 + p)/2$ corresponds to p in Table 19.3.

Then we compute $A = C\sigma_a/\sqrt{N}$ for sample size $N = 3$. The confidence interval is given by

$$a - A \le a \le a + A, \qquad \text{at } p = 95\% \text{ confidence level.}$$

From the χ^2 analysis of Example 17.6.1 we use the slope $a = 0.782$ and variance $\sigma_a^2 = 0.023^2$, so $A = 4.3\frac{0.023}{\sqrt{3}} = 0.057$, and the confidence interval is determined by $a - A = 0.782 - 0.057 = 0.725$, $a + A = 0.839$, or

$$0.725 < a < 0.839 \qquad \text{at 95\% confidence level.}$$

Compared to σ_a, the uncertainty of a has increased due to the high confidence level. A look at Table 19.3 shows that a decrease in confidence level, p, reduces the uncertainty interval, and increasing the number of degrees of freedom, n, would also lower the range of uncertainty. ■

Exercises

19.6.1 Let ΔA be the error of a measurement of A, etc. Use error propagation to show that

$$\left(\frac{\sigma(C)}{C}\right)^2 = \left(\frac{\sigma(A)}{A}\right)^2 + \left(\frac{\sigma(B)}{B}\right)^2$$

holds for the product $C = AB$ and the ratio $C = A/B$.

19.6.2 Find the mean value and standard deviation of the sample of measurements $x_1 = 6.0, x_2 = 6.5, x_3 = 5.9, x_5 = 6.2$. If the point $x_6 = 6.1$ is added to the sample, how does the change affect the mean value and standard deviation?

19.6.3 (a) Carry out a χ^2 analysis of the fit of case b in Fig. 19.9 assuming the same errors for the t_i, $\Delta t_i = \Delta y_i$, as for the y_i used in the χ^2 analysis of the fit in Fig. 19.8. (b) Determine the confidence interval at 95% confidence level.

19.6.4 If $x_1, x_2, \ldots, x_n$ are a sample of measurements with mean value given by the arithmetic mean $\bar{x}$ and the corresponding random variables X_j that take the values x_j with the same probability are independent and have mean value μ and variance σ^2, then show that $\langle \bar{x} \rangle = \mu$ and $\sigma^2(\bar{x}) = \sigma^2/n$. If $\bar{\sigma}^2 = \frac{1}{n}\sum_j (x_j - \bar{x})^2$ is the sample variance, show that $\langle \bar{\sigma}^2 \rangle = \frac{n-1}{n}\sigma^2$.

Additional Readings

Kreyszig, E., *Introductory Mathematical Statistics: Principles and Methods.* New York: Wiley (1970).

Suhir, E., *Applied Probability for Engineers and Scientists.* New York: McGraw-Hill (1997).

Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill (1991).

Ross, S. M., *First Course in Probability*, 5th ed., Vol. A. New York: Prentice-Hall (1997).

Ross, S. M., *Introduction to Probability Models*, 7th ed. New York: Academic Press (2000).

Ross, S. M., *Introduction to Probability and Statistics for Engineers and Scientists*, 2nd ed. New York: Academic Press (1999).

Chung, K. L., *A Course in Probability Theory Revised*, 3rd ed. New York: Academic Press (2000).

Devore, J. L., *Probability and Statistics for Engineering and the Sciences*, 5th ed. New York: Duxbury Pr. (1999).

Montgomery, D. C., and G. C. Runger, *Applied Statistics and Probability for Engineers*, 2nd ed. New York: Wiley (1998).

Degroot, M. H., *Probability and Statistics*, 2nd ed. New York: Addison-Wesley (1986).

Bevington, P. R., and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, 3rd. ed. New York: McGraw-Hill (2003).

General References

Additional, more specialized references are listed at the end of each chapter.

1. E. T. Whittaker and G. N. Watson, *A Course of modern Analysis*, 4th ed. Cambridge, UK: Cambridge University Press (1962), paperback. Although this is the oldest of the references (original edition 1902), it still is the classic reference. It leans strongly toward pure mathematics, as of 1902, with full mathematical rigor.
2. P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, 2 vols. New York: McGraw-Hill (1953). This work presents the mathematics of much of theoretical physics in detail but at a rather advanced level. It is recommended as the outstanding source of information for supplementary reading and advanced study.

3. H. S. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics*, 3rd ed. Cambridge, UK: Cambridge University Press (1972). This is a scholarly treatment of a wide range of mathematical analysis, in which considerable attention is paid to mathematical rigor. Applications are to classical physics and to geophysics.
4. R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Vol. 1 (1st English ed.). New York: Wiley (Interscience) (1953). As a reference book for mathematical physics, it is particularly valuable for existence theorems and discussions of areas such as eigenvalue problems, integral equations, and calculus of variations.
5. F. W. Byron Jr., and R. W. Fuller, *Mathematics of Classical and Quantum Physics*, Reading, MA: Addison-Wesley, reprinted, Dover (1992). This is an advanced text that presupposes a moderate knowledge of mathematical physics.
6. C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*. New York: McGraw-Hill (1978).
7. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Applied Mathematics Series-55 (AMS-55). Washington, DC: National Bureau of Standards, U.S. Department of Commerce; reprinted, Dover (1974). As a tremendous compilation of just what the title says, this is an extremely useful reference.

This page intentionally left blank

INDEX

Numbers

1-forms, 304–5
2-forms, 305–6
3-forms, 306–7

A

Abelian group, 242
Abel's equation, 1016
Abel's test, 351, 665, 882
Abel's theorem, 882
absolute convergence, 340–42, 350, 363
addition
  of matrices, 178–79
  of series, 324–25
  of tensors, 136
addition rule, 1111
addition theorem for spherical harmonics, 797–802
  Bessel functions, 636
  derivation of addition theorem, 798–800
  Legendre polynomials, 798
  trigonometric identity, 797–98
adjoint operator property, 208
algebraic form, 405
aliasing, 916–17
alternating series, 339–42
  absolute convergence, 340–42
  exercises, 342
  Leibniz criterion, 339–40
  overview, 339
analytic continuation, 432–34
analytic functions, 415–18
  z^*, 416
  z^2, 415
analytic landscape, 489–90
angular Mathieu equation, 872
angular momentum, 18, 215, 267, *see also* orbital angular momentum
  coupling, 266–78
    Clebsch–Gordan coefficients $\mathsf{SU}(2)$ and $\mathsf{SO}(3)$, 267–70
    exercises, 277–78
    overview, 266–67
    spherical tensors, 271–74
    young tableaux for $\mathsf{SU}(n)$, 274–77
angular momentum operators, 261
  Clebsch–Gordan coefficients, 803
  orbital, 793–96
  spherical harmonics, 793
    vector spherical harmonics, 813–16
annihilation operator, 824
anomalous dispersion, 998
anticommutation relation, 18
anti-Hermitian matrices, 221–23
  eigenvalues
    degenerate, 223
    and eigenvectors of real symmetric matrices, 221–22
  overview, 221
antisymmetry
  antisymmetric matrices, 204
  and determinants, 168–72
    Gauss elimination, 170–72
    overview, 168–70
  of tensors, 137, 147
area law for planetary motion, 116–19
Argand diagram, 405
associated Legendre equation functions, *see* Legendre equation, functions polynomials
associative, 2
asymptotic expansions, 719–25
asymptotic forms
  of factorial function $\Gamma(1+s)$, 494–95
  of Hankel function, 493–94
asymptotic series, 389–96
  Bessel functions, 722
  confluent hypergeometric functions, 393
  cosine and sine integrals, 392–93
  definition of, 393–94
  exercises, 394–96
  incomplete gamma function, 389–92
  overview, 389
  overview: integral representation expansion, 719–23
  steepest descent, 489–95
  Stokes' method, 719, 724
asymptotic values, Bessel functions, 722, 729
attractor, 1081
autonomous differential equations, 1091–93
average value, 1117–18
axes, *see* rotations; symmetry

axial vector, 143–44
axis, coordinate, 4–5, 7–11, 195–99
azimuthal dependence — orthogonality, 787

B

Baker–Hausdorff formula, 225
basin of attraction, 1081
Bernoulli and Riccati equations, 1089–90
Bernoulli numbers, 376–89, 473–74
  Euler–Maclaurin integration formula, 380–82
  exercises, 385–89
  improvement of convergence, 385
  overview, 376–79
  polynomials, 379–80
  Riemann Zeta function, 382–84
Bessel functions, 675–739, 865
  asymptotic expansions, 719–25
    exercises, 723–25
    expansion of an integral representation, 720–23
  asymptotic values, 722
  closure equation, 696
  of first kind, 675–93
    alternate approaches, 685–86
    Bessel functions of nonintegral order, 686
    Bessel's differential equation, 678–79
    Bessel's differential equation: self-adjoint form, 694
    confluent hypergeometric representation, 865
    cylindrical resonant cavity, 682–85
    cylindrical wave guide, 705
    exercises, 686–93
    Fourier transform, 933
    Fraunhofer diffraction, circular aperture, 680–82
    generating function for integral order, 675–77
    integral representation, 679–80
    Laplace transform solution, 983–84
    orthogonality, 694
    recurrence relations, 677–78
    second kinds, 699–707
    series solution, 570–72, 676–79
    singularities, 564
    spherical, 725–39
    Wronskian, 702–5
  Fraunhofer diffraction, 680–82
  Hankel functions, 707–13
    contour integral representation of, 709–11
    cylindrical traveling waves, 708
    definitions, 707–8
    exercises, 711–13
  Helmholtz equation, 683–84, 725
  Laplace's equation, 695
  modified, 713–19
    asymptotic expansion, 711, 719
    exercises, 716–19
    Fourier transform, 716
    generating function, 709
    integral representation, 720–23
    Laplace transform, 933
    recurrence relations, 714–16
    series form, 714
  Neumann functions, Bessel functions of second kind, 699–707
    coaxial wave guides, 703–4
    definition and series form, 699–700
    exercises, 704–7
    other forms, 701
    recurrence relations, 702
    Wronskian formulas, 702–3
  of nonintegral order, 686
  orthogonality, 694–99
    Bessel series, 695
    continuum form, 696
    electrostatic potential in a hollow cylinder, 695–96
    exercises, 697–99
    normalization, 695
  recurrence relations, 677
  spherical, 725–39
    asymptotic values, 729
    definitions, 726–29
    exercises, 732–39
    limiting values, 729–30
    orthogonality, 731
    particle in a sphere, 731–32
    recurrence relations, 730
  spherical waves, 730
  in wave guides, 703–4
  zeros, 682
Bessel's differential equation, 678–79, 684
  self-adjoint form, 694
Bessel's equation, 983–84
Bessel series, 695
Bessel's inequality, 651–52
beta function, 520–26
  definite integrals, alternate forms, 521–22
  derivation of Legendre duplication formula, 522–23
  incomplete, 523
  Laplace convolution, 993
  verification of $\pi\alpha/\sin\pi\alpha$ relation, 522
bifurcates, 1083
bifurcations in dynamical systems, 1103–4
  Hopf, 1101, 1103–4, 1107
  pitchfork, 1083, 1086, 1101, 1107
binomial coefficient, 356
binomial distribution, 1128–30
binomial expansion, 1129
binomial probability distribution, 1129

binomial theorem, 356–57
Biot and Savart law, 780–82
black hole, optical path near event horizon of, 1041–42
Bohr radius, 844
Born approximation, 603
Bose–Einstein statistics, 1064, 1115
boundary conditions, 542–43
  Cauchy, 542
  Dirichlet, 543
  hollow cylinder, 695
  magnetic field of current loop, 778–82
  Neumann, 543
  ring of charge, 761
  sphere in uniform electric field, 759
  Sturm–Liouville theory, 627–29
  waveguide, coaxial cable, 703
bound state, 627
box counting dimension, 1086
branch cut (cut line), 409
branch points, 440–42
  and multivalent functions, 447–50
  of order 2, 440–42
Bromwich integral, 994–95
Butterfly effect, 1079

C

calculus of residues, 455–82, *see also* definite integrals
  Cauchy principal value, 457–60
  exercises, 474–82
  Jordan's lemma, 466–68
  overview, 455
  pole expansion of meromorphic functions, 461
  product expansion of entire functions, 462–63
  residue theorem, 455–56
calculus of variations, 1037–77
  applications of the Euler equation, 1044–52
    exercises, 1049–52
    soap film, 1045–46
    soap film — minimum area, 1046–49
    straight line, 1044–45
  dependent and an independent variable, 1038–44
    alternate forms of Euler equations, 1042
    concept of variation, 1038–41
    exercises, 1043–44
    missing dependent variables, 1042–43
    optical path near event horizon of a black hole, 1041–42
  Lagrangian multipliers, 1060–65
    constraints, 1060–72
    cylindrical nuclear reactor, 1062
    exercises, 1063–65
    particle in a box, 1061–62
  Rayleigh–Ritz variational technique, 1072–76
    exercises, 1074–76
    ground state eigenfunction, 1073
    Sturm–Liouville equation, 1072
    vibrating string, 1074
  several dependent and independent variables, 1058–59
    exercises, 1059
    relation to physics, 1059
  several dependent variables, 1052–58
    exercises, 1055–56
    Hamilton's principle, 1053–54
    Laplace's equation, 1057–58
    moving particle — Cartesian coordinates, 1054
    moving particle — circular cylindrical coordinates, 1054–55
  several independent variables, exercises, 1058
  surface of revolution, 1046
  uses of, 1037
  variation with constraints, 1065–72
    exercises, 1070–72
    Lagrangian equations, 1066–67
    Schrödinger wave equation, 1069–70
    simple pendulum, 1067–68
    sliding off a log, 1068–69
Cartesian components, 4
Cartesian coordinates, 554–55
  unit vectors, 5
Casimir operators, 265
Catalan's constant, 384, 513
catenoid, catenary of revolution, 1046
Cauchy (Maclaurin) integral test, 327–30, 418–30
  contour integrals, 418–20
  derivatives, 426–27
  exercises, 424–25, 429–30
  Goursat proof, 421–23
  Morera's theorem, 427–28
  multiply connected regions, 423–24
  overview, 418, 425–26
  Stokes' theorem proof, 420–21
Cauchy boundary conditions, 542
Cauchy criterion, 322
Cauchy inequality, 428
Cauchy principal value, 457–60, 471
Cauchy–Riemann conditions, 413–18
  analytic functions, 415–18
    z^*, 416
    z^2, 415
  exercises, 416–18
  overview, 413–15
causality, 486–87
cavities, cylindrical, 682–85
Cayley–Klein parameters, 252
center or cycle, 1100–1101

central force, 117
central force field, 39–40, 44–46
centrifugal potentials, 72
chain rule, 34
chaos in dynamical systems, 1105–6
chaotic attractor, 1085
character, 184, 293
characteristics, 538–41
Chebyshev differential equation, 559
Chebyshev polynomials, 848–59
  generating functions, 848
  Gram–Schmidt construction, 646
  hypergeometric representations, 862
  orthogonality, 854–55
  recurrence relation, 850
  recurrence relations — derivatives, 852–53
  shifted, 646, 850
  trigonometric form, 853–54
  type I, 849–52
  type II, 849
chi-squared (χ^2) distribution, 1143–45
Christoffel symbols, 154–56, 314
circular cylinder coordinates, 115–23
  area law for planetary motion, 116–19
  exercises, 120–23
  Navier–Stokes term, 119
  overview, 115–16
circular cylindrical coordinates, 555–56
  expansion, 601–2
circular membrane, Bessel functions, 693, 708
classes and character, 293
Clausen functions, 909
Clebsch–Gordan coefficients, 267–70, 803
Clifford algebra, 211–12
closed-form solutions, 810–12
closure, of Bessel function, 89, 248, 696
closure, of spherical harmonics, 790, 792
coaxial wave guides, 703–4
commutative, 1
commutator, 180, 225, 231, 249, 253, 262, 264
comparison tests, 325–26
completeness of eigenfunctions
  of Fourier series: of Sturm–Liouville eigenfunctions, 649–51
  of Hilbert–Schmidt: of integral equations, 1031–33
complex variables, 403–54, 455–97, *see also* calculus of residues; Cauchy–Riemann conditions; functions; mapping; saddle points (steepest descent method); singularities
  algebra using, 404–13
    calculus of residues, 455–82
    complex conjugation, 407–8
    exercises, 409–13
    overview, 404–5
    permanence of algebraic form, 405–7
  Cauchy's integral formula, 425–30
    derivatives, 426–27
    exercises, 429–30
    Morera's theorem, 427–28
    overview, 425–26
  Cauchy's integral theorem, 418–25
    Cauchy–Goursat proof, 421–23
    contour integrals, 418–20
    exercises, 424–25
    multiply connected regions, 423–24
    overview, 418
    Stokes' theorem proof, 420–21
  dispersion relations, 482–89
    causality, 486–87
    exercises, 487–89
    optical dispersion, 484–85
    overview, 482–83
    Parseval relation, 485–86
    symmetry relations, 484
  Laurent expansion, 430–38
    analytic continuation, 432–34
    exercises, 437–38
    Schwarz reflection principle, 431–32
    Taylor expansion, 430–31
  overview, 403–4, 455
conditional convergence, 340
conditional probability, 1112
condition number, of ill-conditioned systems, 234
Condon–Shortley phase conventions, 270
confidence interval, 1146, 1149
confluent hypergeometric functions, 863–69
  asymptotic expansions, 866
  Bessel and modified Bessel functions, 865
  Hermite functions, 866
  integral representations, 865
  Laguerre functions, 837–48
  miscellaneous cases, 866
  Whittaker functions, 866
  Wronskian, 868
conformal mapping, 451–54
conjugation, complex, 407–8
connected, simply or multiply, 60, 95, 420, 423, 426, 435
conservation theorem, 309
conservative force, 34, 69
constant 1-forms, 305
constant **B** field, vector potentials of, 44
contiguous function relations, 861
continuation, analytic, 432–34
continuity equation, 40–42
continuity of power series, 364
continuous random variable, 1117–19
continuum form, 696

contour integral representation, Hankel functions, 709–11
contour integrals, 418–20
contour of integration, simple pole on, 468–69
contraction, 139
contravariant tensor, 135, 156–58, 160, 162
contravariant vector, 134
convergence, rate of, 334, 345
convergence of infinite series, 321, 903
  absolute, 340–42, 350
  improvement of, and Bernoulli numbers, 385
  of infinite product, 397–98
  of power series, 363
  rate, 334, 345
  and rational approximations, 345
  tests, 325–39, *see also* Cauchy (Maclaurin) integral test
    comparison, 325–26
    exercises, 335–39
    Gauss', 332–33, 357
    improvement of, 334–39
    Kummer's, 330–32
    overview, 325
    partial sum approximation, 390
    Raabe's, 332
  uniform and nonuniform, 348–49
convolution (Faltungs) theorem, 951–55, 990–94
  driven oscillator with damping, 991–93
  Fourier transform, 931–32, 936–45
  Laplace transform, 965–1003
  Parseval's relation, 952–53
coordinates, *see also* circular cylinder coordinates; curved coordinates and vectors; orthogonal coordinates; spherical polar coordinates
  axes, rotation of, *see* rotations
  curvilinear, 104, 105, 110, 111, 112
  divergence of coordinate vector, 39
  Laplacian in orthogonal, 316
  rotation of, 199
correlation, 1122
cosets and subgroups, 293–94
cosines
  asymptotic expansion, 392–93
  confluent hypergeometric representation, 867
  cosine transform, 939
  direction, 196–97
  direction cosines (orthogonal matrices), 4, 196–97, 201
  functions of infinite products, 398–99
  infinite product, 398, 462
  integral, 392
  integral of in denominator, 464–65
  integrals in asymptotic series, 392–93
  law of, 16–17, 118, 745
  law of, theorem, 16
  theorem, 16
coupling, angular momentum, *see* angular momentum
covariance, 1122
covariance of Maxwell's equations, Lorentz, *see* Lorentz covariance of Maxwell's equations
covariant derivative, 151, 156
covariant vectors, 134, 152–53
  tensor, 135, 156–58, 160, 162
Cramer's rule, 166
creation operator, 824
criterion, Leibniz, 339–40
critical point, 1091–93
critical strip, 897
critical temperature, 1086
crossing conditions, 484
cross product, 18–22, *see also* triple vector products
  exercises, 22–25
  overview, 18–22
  of vectors, 315
crystallographic point and space groups, 299–300
curl, $\nabla\times$, 43–49
  central force field, 44–46
  as differential vector operator, 112–13
  exercises, 47–49
  gradient of dot product, 46
  integral definitions of gradient, divergence and, 58–59
  integration by parts of, 47
  overview, 43
  as tensor derivative operator, 162–63
  vector potential of constant **B** field, 44
curl, $\nabla\times$
  central force field
    in circular cylindrical coordinates, 118
    in curvilinear coordinates, 112–13
    in spherical polar coordinates, 126
  irrotational, 45
curved coordinates and vectors, 103–33, *see also* circular cylinder coordinates; orthogonal coordinates; spherical polar coordinates
  differential operators, 110–14
    curl, 112–13
    divergence, 111–12
    exercises, 113–14
    gradient, 110
    overview, 110
  overview, 103
  special coordinate systems, 114–33
curves, fitting to data, 1140–43
curvilinear coordinates, 104, 105, 110, 111, 112
cut line (branch cut), 409
cylinder coordinates, circular, *see* circular cylinder coordinates

cylindrical coordinates, 104
cylindrical symmetry, 617
cylindrical traveling waves, 708

D

d'Alembertian, 141
d'Alembert ratio test, 326–27
damped oscillator, 979–80
damped simple harmonic oscillation, 980
decay, kaon, 282–83
definite integral (Euler), 500–501
definite integrals
  evaluation of, 463
  exponential forms, 471–82
    Bernoulli numbers, 473–74
    factorial function, 472–73
  $\int_{-\infty}^{\infty} f(x)\,dx$, 465–66
  $\int_{-\infty}^{\infty} f(x)e^{iax}\,dx$, 466–71
    quantum mechanical scattering, 469–71
    simple pole on contour of integration, 468–69
  $\int_0^{2\pi} f(\sin\theta, \cos\theta)\,d\theta$, 464–65
degeneracy, of Schrödinger's wave equation, 638
degenerate eigenfunctions, 638
degenerate eigenvalues, 223, 638
Del (∇), 42, 43
  for central force, 127
  successive applications of, 49–54
    electromagnetic wave equation, 51–53
    exercises, 53–54
    Laplacian of potential, 50–51
    overview, 49–50
delta function, Dirac, 83–85, 669–70, 975
  Bessel representation, 935
  in circular cylindrical coordinates, 601
  derivation, 937–38
  eigenfunction expansion, 89, 650
  exercises, 91–95
  Fourier integral, 90
  Fourier representation, 90
  Green's function and, 592–610
  impulse force, 975
  integral representations for, 90
  Laplace transform, 975
  overview, 83–87
  phase space, 88
  point source, 88, 593
  quantum theory, 955–61
  representation by orthogonal functions, 88–89
  sequences, 83, 86
  sine, cosine representations, 943
  in spherical polar coordinates, 82, 599
  theory of distributions, 86
  total charge inside sphere, 88
De Moivre's formula, 408–13
denominator, integral of cos in, 464–65
dependent and independent variables, 1038–44
  alternate forms of Euler equations, 1042
  concept of variation, 1038–41
  missing dependent variables, 1042–43
  optical path near event horizon of a black hole, 1041–42
derivative operators, tensor
  curl, 162–63
  divergence, 160–61
  exercises, 162–63
  Laplacian, 161–62
  overview, 160
derivatives, 426–27, *see also* exterior derivative
  covariant, 156
  gauge covariant, 76
descending power series solutions, 369, 781
descent, steepest, *see* saddle points (steepest descent method)
determinants, 165–239
  antisymmetry, 168–72
    Gauss elimination, 170–72
    Gauss–Jordan elimination, inversion, 185–86
    overview, 168–70
  exercises, 174–76
  Gram–Schmidt procedure, 173–74
    overview, 173–74
    vectors by orthogonalization, 174
  homogeneous linear equations, 165–66
  inhomogeneous linear equations, 166–67
  Laplacian development by minors, 167–68
  linear dependence of vectors, 172–73
  overview, 165
  product theorem, 181
    representation of a vector product, 20
    secular equation, 218
    solution of a set of homogenous equations, 165
    solution of a set of nonhomogenous equations, 166
deuteron, 626–27
diagonal matrices, 182–83, 215–31, *see also* anti-Hermitian matrices
  eigenvectors and eigenvalues, 216–19
  exercises, 226–31
  functions of, 224–26
  Hermitian, 219–21
  moment of inertia, 215–16
differential equations, 535–619, 751–52
  first-order differential equations, 543–53
    exact differential equations, 545–47
    exercises, 550–53
    linear first-order ODEs, 547–50
    nonlinear, 1088–1102
    parachutist, 544–49

RL circuit, 549–50
separable variables, 544–45
Fuchs' theorem, 573
heat flow, or diffusion, PDE, 611–18
alternate solutions, 614–15
special boundary condition again, 615–16
specific boundary condition, 612–13
spherically symmetric heat flow, 616–17
homogeneous, 536, 548–50, 565
linear independence of solutions
second solution, 581–83
series form of the second solution, 583–85
nonhomogeneous equation — Green's function, 592–610
circular cylindrical coordinate expansion, 27, 601–2
exercises, 607–10
form of Green's functions, 596–98
Legendre polynomial addition theorem, 600–601
quantum mechanical scattering — Green's function, 603–6
quantum mechanical scattering — Neumann series solution, 602–3
spherical polar coordinate expansion, 598–600
symmetry of Green's function, 595–96
partial differential equations, 535–43
boundary conditions, 542–43
classes of PDEs and characteristics, 538–41
examples of, 536–38
introduction, 535–36
nonlinear PDEs, 541–42
particular solution, 548, 565
second solution, 573–78
exercises, 588–92
linear dependence, 580–81
linear independence, 580
linear independence of solutions, 579–80
second solution for the linear oscillator equation, 583
second solution of Bessel's equation, 586–87
second solution logarithmic term, 585, 700
separation of variables, 554–62
Cartesian coordinates, 554–55
circular cylindrical coordinates, 555–56
exercises, 560–62
spherical polar coordinates, 557–60
series solutions — Frobenius' method, 565–78
exercises, 574–78
expansion about x_0, 569
Fuchs' theorem, 573
limitations of series approach — Bessel's equation, 570–71
regular and irregular singularities, 572–73
symmetry of solutions, 569
singular points, 562–65
differential forms, 304–19, *see also* pullbacks
1-forms, 304–5
2-forms, 305–6
3-forms, 306–7
exercises, 318–20
exterior derivative, 307–9
Hodge operator *, 314–19
cross product of vectors, 315
Laplacian in orthogonal coordinates, 316
Maxwell's equations, 316–20
overview, 314–15
Legendre's equation, 752
overview, 304
Stokes' theorem on, 313–14
differentials, exact, *see* thermodynamics
differential vector operators, 110–14
adjoint, 621, 623, 634–35
curl, 112–13
del, 33
divergence, 111–12
exercises, 113–14
gradient, 110
overview, 110
differentiation, 904–5
differentiation of power series, 364
diffraction, 680–82, 1017
diffusion equation, *see* differential equations, heat flow PDE
digamma and polygamma functions, 510–16
digamma functions, 510–11
Maclaurin expansion, computation, 512
polygamma function, 511–12
series summation, 512
dihedral groups, D_n, 299
dimension
box-counting, 1086
Hausdorff, 1086
Kolmogorov, 1086
dimensionality theorem, 298
dipoles, interaction energy, magnetic dipole, radiation fields, *see* electric dipole
Dirac bra-ket notation, 177
Dirac delta function, *see* delta function, Dirac
Dirac matrices, 209–12
direction cosines (orthogonal matrices), 4, 196–97, 201
direct product, 139–41
exercises, 140–41
and matrix multiplication, 181–82
overview, 139–40
of tensors, 140
direct tensor, 181
direct tensor product, 139

Dirichlet conditions, 543, 882
kernel, 910
Dirichlet problem, 617
Dirichlet series, 326
discontinuities, behavior of, 886
discontinuous functions, 888
discrete Fourier transform, 914–19
aliasing, 916–17
fast Fourier transform, 917
limitations, 916
orthogonality over discrete points, 914–15
discrete groups, 291–304
classes and character, 293
crystallographic point and space groups, 299–300
dihedral groups, D_n, 299
exercises, 300–304
subgroups and cosets, 293–94
threefold symmetry axis, 296–99
twofold symmetry axis, 294–96
discrete random variable, 1116–17
dispersion relations, 482–89
causality, 486–87
crossing relations, 484
exercises, 487–89
Hilbert transform, 485
optical dispersion, 484–85
overview, 482–83
Parseval relation, 485–86
sum rules, 484
symmetry relations, 484
displacement, 158
dissipation, 1101–3
distributive, 13
divergence, ∇, 38–43
of central force field, 39–40
circular cylindrical, 115–23
coordinates, Cartesian, 4–7
of coordinate vector, 39
curvilinear coordinates, 111
as differential vector operator, 111–12
exercises, 42–43
integral definitions of gradient, curl, and, 58–59
integration by parts of, 40
overview, 38–39
physical interpretation, 40–42
solenoidal, 42
spherical polar, 123–33
as tensor derivative operator, 160–61
divergent series, 325–26
Doppler shift, 360
dot products, 12–17
exercises, 17
gradient of, 46
invariance of scalar product under rotations, 15–17
overview, 12–15
double series, rearrangement of, 345–48
driven oscillator with damping, 991–93
dual tensors, 147–48
duplication formula for factorial functions, see Legendre duplication formula
dynamical systems, dissipation in, 1102–3

E

E, Lorentz transformation of the electric field, 287–88
Earth's gravitational field, 758–59
Earth's nutation, 973–74
eigenfunctions, 624, 626–27
Bessel's inequality, 651–52
completeness of, 649–61
of Fourier series: of Sturm–Liouville eigenfunctions, 649–51
of Hilbert–Schmidt: of integral equations, 1031–33
eigenvalue equation, 667–68
expansion, Green's function, 662–74
of Dirac delta function, 88–90
of Hermitian differential operators, 635, 649–51
of square wave, 637
expansion coefficients, 658
orthogonal, 636, 637, 1030–32
Schwarz inequality, 652–54
summary — vector spaces, completeness, 654–58
variational calculation, 1072–76
eigenvalues, 217–18, 223, 624–27, 634–35
of Hermitial differential operator, 634
of Hermitial matrices, 219–23
of Hilbert–Schmidt integral equation, 1030–36
of normal matrices, 231–32
of real symmetric matrices, 215–19
variational principle for, 1072
eigenvectors, 216–19, 221–22
eightfold way (weight diagram), 257
Einstein's energy relation, 281
Einstein's summation convention, 136
velocity addition law, 283, 290
electrical charge inside spheres, 88
electric dipole potential, Legendre polynomial expansion, 745
electromagnetic invariants, 288
electromagnetic wave equation, 51–53
electromagnetic waves, 981–82
electrostatic multipole expansion, 599
electrostatic potential, 593
in hollow cylinder, 695–96

of ring of charge, 761–62
electrostatics, physical basis, 741–42
elimination, *see* determinants
elliptical drum, 872–73
elliptic integrals, 370–76
definitions of, 372
exercises, 374–76
of first kind, 372
hypergeometric representations, 373, 860
limiting values, 374
overview, 370
period of simple pendulum, 370–71
of second kind, 372
series expansion, 372–73
elliptic PDEs, 538
empty set, 1111
energy
potentials, 309
relativistic, 356–57
equality of matrices, 178
equations, *see also* linear equations; Maxwell's equations; Poisson's equation
of motion and field, 142
error integrals, 530
asymptotic expansion, 393
confluent hypergeometric representation, 864
error propagation, 1138–40
essential (irregular) singular point, 563
Euler angles, 202–3
Euler equation, 1040
alternate forms of, 1042
applications of, 1044–52
soap film, 1045–46
soap film — minimum area, 1046–49
straight line, 1044–45
Euler identity, 224
product formula, 382–84
Euler integrals, 502
Euler–Maclaurin integration formula, 380–82, 517
Euler–Mascheroni constant, 330
Euler product for Riemann Zeta function, 382–84
event horizon, 1041
exact differential equations, 545–47
exact differentials, *see* thermodynamics
expansion, *see also* Laurent expansion; Taylor's expansion
pole, of meromorphic functions, 461
product, of entire function, 462–63
of series, 372–73
expansion coefficients, 658
expansion of functions, Legendre series, 757–58
expectation value, 630, 955–56, 1117–18
exponential forms, 471–82
Bernoulli numbers, 473–74
factorial function, 472–73
exponential function, of Maclaurin theorem, 354–55
exponential integral, 527–30
exponential of diagonal matrix, 225–26
exponential transform, 938
exterior derivative, 307–9
extreme or stationary value, 36, 881, 1038–40

F

Factorial function $\Gamma(1+s)$
asymptotic form of, 494–95
complex argument, 499–506
contour integrals, 505
digamma function, 510
double factorial notation, 505
Gamma functional relation, 503
infinite product, 501
integral (Euler) representation, 500
Legendre duplication formula, 503
Maclaurin expansion, 512
polygamma functions, 511
steepest descent asymptotic formula, 495
Stirling's (formula) series, 516–18
factorial notation, 503–5
faithful group, 243
Faraday's law, 66–68
fast Fourier transform, 917
Feigenbaum number, 1083–84
Fermage equation, 950
Fermat's principle, 157, 1041, 1050
Fermi–Dirac statistics, 1064, 1115
field equations, 142
finite wave train, 940–41
first-order differential equations, 543–53
exact differential equations, 545–47
linear first-order ODEs, 547–50
separable variables, 544–45
fixed and movable singularities, special solutions, 1090
Floquet's theorem, 877
force as gradient of potentials, 36
forced classical oscillators, 457–60
force field, central, *see* central force field
Fourier–Bessel series, 695
Fourier expansions of Mathieu functions, 919–29
integral equations and Fourier series for Mathieu functions, 919–23
leading coefficients for ce_0, 926–29
leading coefficients of se_1, 923–26
Fourier integral theorem, 937
development of, 936–38
exponential form, 937
Fourier–Mellin integral, 995
Fourier representation, of Dirac delta function, 90
Fourier series, 881–930

advantages, uses of, 888–92
change of interval, 890–91
completeness, *see* general properties of Fourier series
convergence, 893, 903–10
differentiation, *see* general properties of Fourier series
discontinuous functions, 888
exercises, 891–92
integration, *see* general properties of Fourier series
periodic functions, 888–90
applications of, 892–903
exercises, 898–903
full-wave rectifier, 893–94
infinite series, Riemann Zeta function, 894–98
square wave — high frequencies, 892–93
discrete Fourier transform, 914–19
discrete fourier transform, 915–16
discrete Fourier transform — aliasing, 916–17
exercises, 918–19
fast Fourier transform, 917
limitations, 916
orthogonality over discrete points, 914–15
Fourier expansions of Mathieu functions, 919–29
exercises, 929
integral equations and Fourier series for Mathieu functions, 919–23
leading coefficients for ce_0, 926–29
leading coefficients of se_1, 923–26
general properties, 881–88
behavior of discontinuities, 886
completeness, 883–84
complex variables — Abel's theorem, 882
exercises, 886–88
sawtooth wave, 885–86
square wave, 892
Sturm–Liouville theory, 885
summation of a Fourier series, 882–83
Gibbs phenomenon, 910–14
calculation of overshoot, 912–13
exercises, 913–14
square wave, 911–12
summation of series, 910
orthogonality, 636–37
properties of, 903–10
convergence, 903
differentiation, 904–5
exercises, 905–10
integration, 904
summation of, 882–83
Fourier transform, of Gaussian, 932
Fourier transform of derivatives, 946–51
heat flow PDE, 948–49
inversion of PDE, 949–50
wave equation, 947–48
Fourier transforms, 486–87, 931–32
aliasing, 916
convolution (Faltungs) theorem, 951
delta function derivation, 937
transfer functions, 961
Fourier transforms — inversion theorem, 938–46
cosine transform, 939
exponential transform, 938
fast Fourier transform, 917
finite wave train, 940–41
Fourier integral, 931, 936–39
momentum space representation, 955
sine transform, 939–40
uncertainty principle, 941
Fourier transform solution, 1013
fractals, 1086–88
fractional order, 427
Fraunhofer diffraction, Bessel function, 680–82
Fredholm equation, 1005–6
Frobenius' method, series solutions, 565–78
Fuchs' theorem, 573
full-wave rectifier, 893–94
functional equation, Riemann Zeta function, 896
Gamma function, 500
functions, 817–80, *see also* analytic functions
Chebyshev polynomials, 848–59
exercises, 855–59
generating functions, 848
orthogonality, 854–55
recurrence relations — derivatives, 852–53
trigonometric form, 853–54
type I, 849–52
type II, 849
of complex variable, 408–13
confluent hypergeometric functions, 863–69
Bessel and modified Bessel functions, 865
exercises, 867–69
Hermite functions, 866
integral representations, 865
miscellaneous cases, 866
entire, 415, 451, 462–63
exponential, of Maclaurin theorem, 354–55
factorial, 472–73
Hermite functions, 817–36
alternate representations, 819
applications of the product formulas, 831–32
direct expansion of products of Hermite polynomials, 828–30
exercises, 832–36
generating functions — Hermite polynomials, 817–18

orthogonality, 821–22
quantum mechanical simple harmonic oscillator, 822–27
recurrence relations, 818–19
Rodrigues' representation, 820–21
threefold Hermite formula, 827–28
hypergeometric functions, 859–63
contiguous function relations, 861
exercises, 862–63
hypergeometric representations, 861–62
Laguerre functions, 837–48
associated Laguerre polynomials, 841–43
differential equation — Laguerre polynomials, 837–41
exercises, 845–48
hydrogen atom, 843–45
Mathieu functions, 869–79
elliptical drum, 872–73
exercises, 879
general properties of Mathieu functions, 874
quantum pendulum, 873
radial Mathieu functions, 874–79
separation of variables in elliptical coordinates, 870–71
of matrices, 224–26
meromorphic, 451, 461–62, 478
multivalent, and branch points, 447–50
rotation of, 251
series of, 348–52
Abel's test, 351
exercises, 352
overview, 348
uniform and nonuniform convergence, 348–49
Weierstrass *M* test, 349–50

G

gamma distribution, 1143
gamma function, *see also* factorial function, 499–533
beta function, 520–26
definite integrals, alternate forms, 521–22
derivation of Legendre duplication formula, 522–23
exercises, 523–26
incomplete beta function, 523
verification of $\pi\alpha/\sin\pi\alpha$ relation, 522
definitions, simple properties, 499–510
definite integral (Euler), 500–501
double factorial notation, 505
exercises, 506–10
factorial notation, 503–5
infinite limit (Euler), 499–500
infinite product (Weierstrass), 501–3
integral representation, 505–6
digamma and polygamma functions, 510–16
Catalan's constant, 513
digamma functions, 510–11
exercises, 513–16
Maclaurin expansion, computation, 512
polygamma function, 511–12
series summation, 512
incomplete gamma functions and related functions, 527–33
error integrals, 530
exercises, 530–33
exponential integral, 527–30
of infinite product, 398–99
Stirling's series, 516–20
derivation from Euler–Maclaurin integration formula, 517
exercises, 518–20
Stirling's series, 518
gauge covariant derivative, 76
theory, 259
transformation, 76
gauge theory, 76, 241
Gauss elimination, 170–72
Gauss' differential equation, 312–13, 318, 614–15, 617–18
hypergeometric differential equation, 576, 859–62, 873
Gauss error integral, 500
asymptotic expansion, 530
Gauss' fundamental theorem of algebra, 428, 463
Gauss–Jordan matrix inversion, 185–87
Gauss' law, 52, 79–83, 594
Gauss' normal distribution, 1134–38
Gauss' notation, 503
Gauss–Seidel iteration technique, 172
Gauss–Seidel method, 226
Gauss' test, 332–33, 357
Legendre series, 333
Gauss' theorem, 60–64
alternate forms of, 62–64
exercises, 62–64
overview, 62
overview, 60–61
pullbacks, 312–13
Gegenbauer polynomials, *see* ultraspherical polynomials
general parabolic solution, 540
general properties, 881–88
behavior of discontinuities, 886
completeness, 883–84
complex variables — Abel's theorem, 882
of Mathieu functions, 874
sawtooth wave, 885–86
Sturm–Liouville theory, 885
summation of a Fourier series, 882–83

general tensors, 151–60, *see also* Christoffel symbols
covariant derivative, 156
exercises, 158–60
geodesics and parallel transport, 157–60
metric tensor, 151–54
overview, 151
generating function, 741–49, 848, 1014–15
associated Laguerre polynomials, 624–25, 841–45
associated Legendre functions, 771–82, 788
associated Legendre polynomials, 773
Bernoulli numbers, 376–89, 473–74, 517
Bernoulli polynomials, 379–80
Bessel functions, modified, 711, 713–19, 723–24, 865
Chebyshev polynomials, 848–59
extension to ultraspherical polynomials, 747
Hermite polynomials, 817–30
for integral order, 675–77
Laguerre polynomials, 647, 837–41, 843–45
Legendre polynomials, 742–44
linear electric multipoles, 744–45
physical basis — electrostatics, 741–42
ultraspherical polynomials, 747
vector expansion, 745–47
generators of continuous groups, 246–61, *see also* rotations; SU(2)
exercises, 260–61
overview, 246–50
geodesic equation, 157
geodesics, 157–60
geometrical interpretation of gradient, 35–38
integration by parts of, 36–37
of potential, force as, 36
geometric series, 322–23
Gibbs phenomenon, 886, 910–14
calculation of overshoot, 912–13
square wave, 911–12
summation of series, 910
global behavior, 1093–1101
Goursat proof of Cauchy's integral theorem, 421–23
gradient
in Cartesian coordinates, 37, 51, 113, 134
in circular cylindrical coordinates, 118
in curvilinear coordinates, 110
in spherical polar coordinates, 126
gradient, curvilinear coordinates, 110
gradient, ∇, 32–38
as differential vector operator, 110
of dot product, 46
exercises, 37–38
geometrical interpretation, 35–38
integration by parts of, 36–37
of potential, force as, 36
integral definitions of divergence, curl, and, 58–59
overview, 32–34
of potential, 34
Gram–Schmidt orthogonalization, 642–49
Gram–Schmidt procedure, 173–74
overview, 173–74
vectors by orthogonalization, 174
gravitational potentials, 72
great circle, 1040
Green's function, 662–74
construction of, one dimension, 598, 599, 663–65, 670
construction of, two dimension, 597–98
construction of, three dimension, 597–98
and Dirac delta function, 669–70
eigenfunction, eigenvalue equation, 667–68
eigenfunction expansion, 662–82
electrostatic analog, 592, 665
form of, 596–98
Helmholtz, 598
Helmholtz equation, 662
integral — differential equation, 665–67
Laplace operator, 598
circular cylindrical expansion, 601–6
spherical polar expansion, 598–600
linear oscillator, 668–69
modified Helmholtz, 598
nonhomogeneous equation, 592–610
one-dimensional, 663–65
Poisson's equation, 669–70
symmetry of, 595–96
Green's theorem, 61–62, 593
Gregory series, 362, 368
ground state eigenfunction, 1073
group theory, 241–320, *see also* angular momentum; differential forms; generators of continuous groups; homogeneous Lorentz group
character, 184
definition of, 242–43
discrete, 291–93
classes and character, 293
crystallographic point and space, 299–300
dihedral, D_n, 299
exercises, 300–304
subgroups and cosets, 293–94
threefold symmetry axis, 296–99
twofold symmetry axis, 294–96
faithfulness, 243
homomorphic, 243
homomorphism and isomorphism, 243–45
homomorphism SU(2)–SO(3), 252–6
isomorphic, 243

Lorentz covariance of Maxwell's equations, 283–91
electromagnetic invariants, 288
exercises, 289–91
overview, 283–86
transformation of **E** and **B**, 287–88
overview, 241–42
permutation groups, 301–3
quantum chromodynamics (QCD), 258
reducible and irreducible representations, 245–46
special unitary group SU(2), 251
vierergruppe, 292–93, 296
Gutzwiller's trace formula, 898

H

Hadamard product, 208
Hamilton–Jacobi equation, 539
Hamilton's principle, 1053–54
Hankel functions and Lagrange equations of motion, 493–94, 707–13
asymptotic forms, 493–94, 723
contour integral representation of the Hankel functions, 709–11
cylindrical traveling waves, 708
definition, by Neumann function, 707–8
definitions, 707–8
series expansion, 707, 715
spherical, 728–29
Wronskian formulas, 708
Hankel transforms, 933
harmonic functions, 539
harmonic oscillator, 822–27, 958–59
harmonics, *see also* spherical harmonics, vector spherical harmonics
harmonic series, 323–24
Hausdorff, 225, 249, 1086
heat flow PDE, 948–49
Heaviside expansion theorem, 442, 1003
Heaviside shifting theorem, 981
Heaviside unit step function, 93, 996
Heisenberg uncertainty principle, 732, 941
Helmholtz diffusion equation, 536, 537
Helmholtz equation, 556, 557, 613
Bessel function, 683–84, 725
Green's function, 662
spherical coordinates, 725
Helmholtz operators, 598
Helmholtz's theorem, 95–101
exercises, 100–101
overview, 95–96
Hermite functions, 817–36, 866
alternate representations, 819
applications of the product formulas, 831–32
confluent hypergeometric representation, 866
direct expansion of products of Hermite polynomials, 828–30
generating functions — Hermite polynomials, 817–18
Gram–Schmidt construction, 642
orthogonality, 821–22
quantum mechanical simple harmonic oscillator, 822–27
recurrence relations, 818–19
Rodrigues representation, 820–21
threefold Hermite formula, 827–28
Hermite polynomials
direct expansion of products of, 828–30
generating functions, 817–18
orthogonality integral, 821
recurrence relations, 818–19
Rodrigues representation, 820–21
Hermitian matrices, 184, 209
and matrix diagonalization, 219–21
unitary and, 208–15
exercises, 212–15
overview, 208–9
Pauli and Dirac, 209–12
Hermitian matrices, anti-, 221–23
eigenvalues
degenerate, 223
and eigenvectors of real symmetric matrices, 221–22
overview, 221
Hermitian operators, 629–30, 634–42
completeness of eigenfunctions, 649–58
degeneracy, 638
expansion in orthogonal eigenfunctions — square wave, 637
Fourier series — orthogonality, 636–37
integration interval, 628–29
orthogonal eigenfunctions, 636
properties of, 634–38
in quantum mechanics, 630
real eigenvalues, 634–35
Hilbert matrix, determinant, 235
Hilbert–Schmidt theory, 1029–36
nonhomogeneous integral equation, 1032–34
orthogonal eigenfunctions, 1030–32
symmetrization of kernels, 1029–30
Hilbert space, 7, 535, 629, 638, 658, 885
Hilbert transforms, 483
Hodge * operator, 314–19
cross product of vectors, 315
Laplacian in orthogonal coordinates, 316
Maxwell's equations, 316–20
overview, 314–15
holomorphic functions (analytic or regular functions), 415
homogeneous equations, 165–66, 565

homogeneous Lorentz group, 278–83
exercises, 283
kinematics and dynamics in Minkowski space–time, 280–83
overview, 278–80
homomorphic group, 243
homomorphism, 243–45
overview, 243
rotations, 244–45
SU(2) and SO(3), 252–56
hooks, 275–76
Hubble's law, 7
hydrogen atom, 843–45, 957–58
Schrödinger's wave equation, 843
hyperbolic PDEs, 538
hypercharge, 257
hypergeometric equation
alternate forms, 860, 864
second independent solution, 860
singularities, 564, 859, 864, 865, 873
hypergeometric functions, 859–63
contiguous function relations, 861
hypergeometric representations, 861–62

I

ill-conditioned matrices, 234–35
imaginary part, 407
impulsive force, 975–76
incomplete gamma function, 389–92
confluent hypergeometric representation, 500, 864
recurrence relations, 399, 512
independence, linear, 173, 579–81, 643, 665, 703
of solutions of ordinary differential equations, 579–81
of vectors, 173, 579
indicial equation, 566
inertia matrix, moment of, 215–16
infinite limit (Euler), 499–500
infinite product (Weierstrass), 499–500, 501–3
infinite products, 378, 383, 396–99, 499, 501–3
convergence, 397–98
cosine, 398–99
entire functions, 462
gamma function, 398–99
sine, 398–99
infinite series, 321–401, *see also* alternating series; power series; Taylor's expansion
algebra of, 342–48
alternating series, 342–43
convergence, 342–45
convergence: absolute, 342–44
convergence: Cauchy integral, 327–29
convergence: Cauchy root, 326
convergence: comparison, 325–26
convergence: conditional, Leibniz criterion, 344
convergence: D'Alembert ratio, 326–27
convergence: Gauss', 332–33
convergence: improvement of, 345
convergence: Kummer's, 330–32
convergence: Maclaurin integral, 327–30
convergence: Raabe's, 332
convergence: tests of, 325–35
convergence: uniform, 348–51, 363–64
divergence of squares, 344
double series, 345–47
exercises, 347–48
overview, 342–44
rearrangement of double, 345–48
asymptotic series, 389–96
cosine and sine integrals, 392–93
definition of, 393–94
exercises, 394–96
incomplete gamma function, 389–92
overview, 389
Bernoulli numbers, 376–89
Euler–Maclaurin integration formula, 380–82
exercises, 385–89
improvement of convergence, 385
overview, 376–79
polynomials, 379–80
Riemann Zeta function, 382–84
elliptic integrals, 370–76
definitions of, 372
exercises, 374–76
limiting values, 374
overview, 370
period of simple pendulum, 370–71
series expansion, 372–73
of functions, 348–52
Abel's test, 351
exercises, 352
overview, 348
uniform and nonuniform convergence, 348–49
Weierstrass *M* test, 349–50
fundamental concepts, 321–25
addition and subtraction of, 324–25
exercises, 325
geometric, 322–23
harmonic, 323–24
overview, 321–22
power series, 363–66, 578–79
products of, 396–401
convergence of, 397–98
exercises, 399–401
overview, 396–97
sine, cosine, and gamma functions, 398–99
Riemann's theorem, 894–98

infinity, *see* singularity, pole, essential singularity
inhomogeneous linear equations, 166–67
inhomogeneous ordinary differential equation (ODE), Green's function solutions, 663–64
inner product and matrix multiplication, 179–81
integral definitions of gradient, divergence, and curl, 58–59
integral — differential equation, 665–67
integral equations, 1005–36
  and Fourier series for Mathieu functions, 919–23
  Fredholm equations, 1005, 1007, 1010, 1013, 1018, 1021–24, 1030–32
  Hilbert–Schmidt theory, 1029–36
    exercises, 1034–36
    nonhomogeneous integral equation, 1032–34
    orthogonal eigenfunctions, 1030–32
    symmetrization of kernels, 1029–30
  integral transforms, generating functions, 1012–18
    exercises, 1015–18
    Fourier transform solution, 1013
    generalized Abel equation, convolution theorem, 1014
    generating functions, 1014–15
  introduction, 1005–12
    definitions, 1005–6
    exercises, 1011
    linear oscillator equation, 1009–11
    momentum representation in quantum mechanics, 1006–7
    transformation of a differential equation into an integral equation, 1008–9
  Neumann series, separable (degenerate) kernels, 1018–29
    exercises, 1025–29
    Neumann series, 1018–19
    Neumann series solution, 1020–21
    numerical solution, 1023–25
    separable kernel, 1021–22
  Volterra equations, 991, 1005–6, 1009–11, 1021
integral form, Neumann functions, 701
integral representations, 505–6, 679–80
  for Dirac delta function, 90
  expansion of, 720–23
integrals, *see also* Cauchy (Maclaurin) integral test; definite integrals; elliptic integrals
  contour integration, 463, 471, 503, 522, 603
  differentiation of, 590
  evaluation of, 810
  Lebesgue, 649, 657
  line, 55–56, 65–67, 440
  of products of three spherical harmonics, 803–6
  Riemann, 55, 60–61, 605, 636
  Stieltjes, 86, 873
  surface, 56–57
  volume, 57–58
integral test, Cauchy, *see* Cauchy (Maclaurin) integral test
integral transforms, 931–1004
  convolution (Faltungs) theorem, 990–94
    driven oscillator with damping, 991–93
    exercises, 993
  convolution theorem, 951–55
    exercises, 953–55
    Parseval's relation, 952–53
  development of the Fourier integral, 936–38
    Dirac delta function derivation, 937–38
    Fourier integral — exponential form, 937
  Fourier transform, 931–32
  Fourier transform of derivatives, 946–51
    heat flow PDE, 948–49
    inversion of PDE, 949–50
    wave equation, 947–48
  Fourier transform of Gaussian, 932
  Fourier transforms — inversion theorem, 938–46
    cosine transform, 939
    exercises, 942–46
    exponential transform, 938
    finite wave train, 940–41
    sine transform, 939–40
    uncertainty principle, 941
  Fourier transforms of derivatives, 950–51
  generating functions, 1012–18
    Fourier transform solution, 1013
    generalized Abel equation, convolution theorem, 1014
    generating functions, 1014–15
  integral transforms, 931–35
    exercises, 934–35
    Fourier transform, 931–32
    Fourier transform of Gaussian, 932
    Laplace, Mellin, and Hankel transforms, 933
    linearity, 933–34
  inverse Laplace transform, 994–1003
    Bromwich integral, 994–95
    exercises, 1000–1003
    inversion via calculus of residues, 996
    summary — inversion of Laplace transform, 999–1000
    velocity of electromagnetic waves in a dispersive medium, 997–99
  Laplace, Mellin, and Hankel transforms, 933
  Laplace transform of derivatives, 971–78
    Dirac delta function, 975
    Earth's nutation, 973–74
    exercises, 976–78
    impulsive force, 975–76
    simple harmonic oscillator, 973

Laplace transforms, 965–71
definition, 965
elementary functions, 965–66
exercises, 970–71
inverse transform, 967–68
partial fraction expansion, 968–69
step function, 969–70
linearity, 933–34
momentum representation, 955–61
exercises, 959–61
harmonic oscillator, 958–59
hydrogen atom, 957–58
other properties, 979–89
Bessel's equation, 983–84
damped oscillator, 979–80
derivative of a transform, 982–83
electromagnetic waves, 981–82
exercises, 985–89
integration of transforms, 984
limits of integration — unit step function, 985
RLC analog, 980–81
substitution, 979
translation, 981
transfer functions, 961–64
exercises, 964
significance of $\Phi(t)$, 963–64
integration, 904, *see also* path-dependent work
Euler–Maclaurin formula, 380–82
by parts of curl, 47
by parts of divergence, 40
by parts of gradient, 36
of power series, 364
simple pole on contour of, 468–69
of transforms, 984
of vectors, 54–60
exercises, 59–60
overview, 54–56
integration interval $[a, b]$, 628–29
interpolating polynomials, 194
interpretation, *see* geometrical interpretation of gradient; physical interpretation of divergence
interitem, 1111
invariance of scalar product under rotations, 15–17
invariants, electromagnetic, 288
inverse Laplace transform, 994–1003
Bromwich integral, 994–95
inversion via calculus of residues, 996
summary — inversion of Laplace transform, 999–1000
velocity of electromagnetic waves in a dispersive medium, 997–99
inverse matrix, 200
inverse operator, 934, 1019
inverse transform, 967–68
inversion, 445–46
matrix, 184–87
Gauss–Jordan, 185–87
overview, 184–85
of PDE, 949–50
of power series, 366
via calculus of residues, 996
irreducible representations, 245–46
irreducible tensors, 149–51
irregular (essential) singular point, 563
irregular sign changes, series with, 341–42
irregular singularities, 572–73
irrotational, 45
isomorphic group, 243
isomorphism, 243–45
overview, 243
rotations, 244–45
isospin, SU(2), 256–60
isospin **I**, 257

J

Jacobian, 107–8
parity transformation, 146
Jacobians for polar coordinates, 108–10
Jacobi–Anger expansion, 687
Jacobi identity, 248
Jacobi technique, 226
Jordan's lemma, 468
Julia set, 1087

K

kaon decay, 282–83
Kepler's laws of planetary motion, 116–17
kinematics and dynamics in Minkowski space–time, 280–83
Kirchhoff diffraction theory, 426
Klein–Gordon equation, 537
Korteweg–deVries equation, 542
Kronecker delta, 10, 136–37
Kronecker product, 181
Kronig–Kramers optical dispersion relations, 484, 485
Kummer's equation, *see* confluent hypergeometric equation
Kummer's test, 330–32

L

ladder operators, approach to orbital angular momentum, 262–64
Lagrange's equations, 1066
Lagrangian, 1053–54
Lagrangian equations, 1066–67
Lagrangian multipliers, 1060–65
cylindrical nuclear reactor, 1062
particle in a box, 1061–62

Laguerre functions, 837–48
associated Laguerre polynomials, 841–43
differential equation — Laguerre polynomials, 837–41
hydrogen atom, 843–45
Laguerre polynomials, 624
associated, 624–25
confluent hypergeometric representation, 866
generating function, 841–42
integral representation, 843
orthogonality, 843
recurrence relations, 842
Rodrigues' representation, 767, 842
confluent hypergeometric representation, 866
differential equation, 837–41
generating function, 837–39
Gram–Schmidt construction, 647
orthogonality, 624, 647
recurrence relations, 840, 842
Rodrigues' formula, 839
Schrödinger's wave equation, 843
self-adjoint form, 624, 840
singularities, 564
Laplace, Mellin, and Hankel transforms, 933
Laplace function, 598
Laplace's equation, 536, 1057–58
Bessel functions, 695
Legendre polynomials, 760, 761
solutions, 50–51, 96, 443, 452, 539, 559–60
Laplace series
expansion theorem, 790–91
gravity fields, 791
Laplace transforms, 965–71
convolution theorem, 521, 1000, 1003, 1014
definition, 965
of derivatives, 971–78
Dirac delta function, 975
Earth's nutation, 973–74
impulsive force, 975–76
simple harmonic oscillator, 973
elementary functions, 965–66
inverse transform, 967–68
partial fraction expansion, 968–69
step function, 969–70
table of transforms, 967–68, 979
translation, 1000
Laplacian
in Cartesian coordinates, 51–52, 554–55
in circular cylindrical coordinates, 119
development by minors, 167–68
in orthogonal coordinates, 316
of potentials, 50–51
spherical polar coordinates, 126
as tensor derivative operator, 161–62
Laurent expansion, 430–38, 466, 472
analytic continuation, 432–34
exercises, 437–38
Schwarz reflection principle, 431–32
Taylor expansion, 430–31
law of cosines, 16–17, 118, 745
leading coefficients for ce_0, 926–29
leading coefficients of se_1, 923–26
least squares, method of, 1119
Legendre duplication formula, derivation of, 522–23
Legendre equation, Maxwell's equation, 779
self-adjoint form, 623, 625
Legendre functions, 741–816
addition theorem, 798
addition theorem for spherical harmonics, 797–802
derivation of addition theorem, 798–800
exercises, 800–802
trigonometric identity, 797–98
alternate definitions of Legendre polynomials, 767–70
exercises, 769–70
Rodrigues' formula, 767
Schlaefli integral, 768–69
associated, 772
associated Legendre functions, 771–86
associated Legendre polynomials, 772–74
equation, 558, 771–72, 778–82, 788
Fourier transform, 770
Gram–Schmidt construction, 789
hypergeometric representation, 861
lowest associated Legendre polynomials, 774
magnetic induction field of a current loop, 778–82
orthogonality, 776–78
parity, 776
poles, 760, 782
recurrence relations, 775
Rodrigues' formula, 772–73
Schlaefli integral, 768
second kind, 806–12
self-adjoint form, 771–72
special values, 774–75
generating function, 741–49
exercises, 747–49
extension to ultraspherical polynomials, 747
Legendre polynomials, 742–44
linear electric multipoles, 744–45
physical basis — electrostatics, 741–42
vector expansion, 745–47
integrals of products of three spherical harmonics, 803–6
application of recurrence relations, 804–5
exercises, 805–6
orbital angularmomentum operators, 793–97

orbital angular momentum operators, 796–97
orthogonality, 756–67
Earth's gravitational field, 758–59
electrostatic potential of a ring of charge, 761–62
exercises, 762–66
expansion of functions, Legendre series, 757–58
polarization of dielectric, 764
ring of electric charge, 761–62
sphere in a uniform field, 759–61
recurrence relations and special properties, 749–56
differential equations, 751–52
exercises, 754–56
parity, 753
recurrence relations, 749–50
special values, 752
sphere in uniform electric field, 759–61
upper and lower bounds for $P_n(\cos\theta)$, 753–4
of second kind, 806–13
closed-form solutions, 810–12
exercises, 812–13
$Q_n(x)$ functions of the second kind, 809–10
series solutions of Legendre's equation, 807–9
spherical harmonics, 786–93
azimuthal dependence — orthogonality, 787
exercises, 791–93
Laplace series, expansion theorem, 790–91
Laplace series — gravity fields, 791
polar angle dependence, 788
spherical harmonics, 788–90
vector spherical harmonics, 813–16
Legendre polynomials, 644–46, 741, 742–44
associated
generating function, 773
recurrence relations, 775
generating function, 743
by Gram–Schmidt orthogonalization, 644–46
Laplace's equation, 760, 761
orthogonality integral, 777
recurrence relations, 749–50
Rodrigues' formula, 761
Schlaefli integral, 768
Legendre's duplication formula, 503
Legendre's equation, 625
differential form, 752
Legendre's equation
self-adjoint form, 625, 771
Legendre's equation
series solutions of, 807–9
Legendre series, 333, 807–9
recurrence relations, 807–8
Leibniz criterion, 339–40
formula for differentiating an integral, 590, 776
formula for differentiating a product, 771
Lerch's theorem, 967
Levi-Civita symbol, 146–47
L'Hôpital's rule, 365
Lie groups and algebras, 243, 248, 264–66
limits of integration — unit step function, 985
limits to values of elliptic integrals, 374
linear electric multipoles, 744–45
linear equations
homogeneous, 165–66
inhomogeneous, 165–66
linear independence of solutions, 581–83
linearity, 933–34
linearly dependent solutions, 549
linearly independent solutions, 549
linear operator, 87, 176, 208, 535, 622, 650
differential operator, 42–43, 249, 261, 285, 304, 307, 554, 569, 629, 634, 664
integral operator, 768, 1019
linear oscillator
Green's function, 668–69
linear oscillator equation, 1009–11
linear transformation law, 152
line integrals, 55
Liouville's theorem, 428
liquid drop model, 785
logistic map, 1080–84
Lommel integrals, 697
Lorentz covariance of Maxwell's equations, 283–91
electromagnetic invariants, 288
exercises, 289–91
overview, 283–86
transformation of E and B, 287–88
Lorentz–Fitzgerald contraction, 148
Lorentz gauge, 52
Lorentz group, *see* homogeneous Lorentz group
lowering operator, 263
lowest associated Legendre polynomials, 774
Lyapunov exponents, 1085–86

M

Maclaurin expansion, series computation, 512
Maclaurin integral test, 327–30
Riemann Zeta function, 329–30
Maclaurin theorem, 354–55
exponential function, 354–55
logarithm, 355
overview, 354
Madelung constant, 347
magnetic, 19, 46, 51, 66–67, 69, 74–76, 96, 100, 127–28, 144–45, 283–84, 288, 306, 311, 317, 447, 537, 638, 685, 703–5, 746, 778–82, 974

magnetic field constant (**B** field), 44
magnetic flux across an oriented surface, 306
magnetic induction field of current loop, 778–82
magnetic moments, 145
magnetic vector potentials, 44, 74–76, 127–28
Mandelbrot set, 1087–88
mapping
  complex variables, 443–51
    branch points and multivalent functions, 447–50
    exercises, 450–51
    inversion, 445–46
    overview, 443
    rotation, 444
    translation, 443–44
  conformal, 451–54
    exercises, 453–54
Mathieu equation
  angular, 872
  modified, 872
  radial, 872
Mathieu functions, 869–79, 921
  elliptical drum, 872–73
  Fourier expansions of, 919–29
    integral equations and Fourier series for Mathieu functions, 919–23
    leading coefficients for ce_0, 926–29
    leading coefficients of se_1, 923–26
  general properties of Mathieu functions, 874
  quantum pendulum, 873
  radial Mathieu functions, 874–79
  separation of variables in elliptical coordinates, 870–71
matrices, 165–239, *see also* determinants; diagonal matrices; orthogonal matrices
  addition and subtraction, 178–79
  adjoint, 208
  angular momentum matrices, 253, 273
  anticommuting sets, 236
  antihermitian, 221–23, 231
  antisymmetric, 204
  definition, 176
  diagonalization, 215–26
  direct product, 181–82
  equality, 178
  Euler angle rotation, 202–3
  exercises, 187–95
  Gauss–Jordan matrix inversion technique, 185
  Hermitian and unitary, 208–15
    exercises, 212–15
    overview, 208–9
    Pauli and Dirac, 209–12
  inversion of, 184–87
    Gauss–Jordan, 185–87
    overview, 184–85
  ladder operators, 263
  matrix multiplication, 176, 179–81
  moment of intertia, 215–16, 220
  multiplication, 179–80
    direct product, 181–82
    inner product, 179–81
    by scalar, 179
  normal, 231–39
    exercises, 236–39
    ill-conditioned systems, 234–35
    normal modes of vibration, 233–34
    overview, 231–32
  null matrix, 178
  orthogonal matrix, 201, 206, 209
  overview, 176–78
  product theorem, 181
  quaternions, 204, 212
  rank, 178
  relation to tensor, 206
  representation, 177, 184, 205, 208, 212
  self-adjoint, 209
  similarity transformation, 205
  skewsymmetric, 204
  symmetric, 204
  traces, 139, 183–84
  transposition, 177
  unitary, 209
  vector transformation law, 198
Maxwell's equations, 51, 284, 316–20
  derivation of wave equations, 52
  dual transformation, 290
  Gauss' law, 52–53
  Legendre equation, 779
  Lorentz covariance of, 283–91
    electromagnetic invariants, 288
    exercises, 289–91
    overview, 283–86
    transformation of **E** and **B**, 287–88
  Oersted's law, 52–53
mean value theorem, 353
Mellin transforms, 897, 933
meromorphic functions, 439, 461
  integral of, 466
  pole expansion of, 461
metric, curvilinear coordinates, 105
metric tensor, 151–54
  Christoffel symbols as derivatives of, 155–56
minimal substitution, 76
Minkowski space, 136, 278–79
Minkowski space–time, *see* kinematics and dynamics in Minkowski space–time
minor, 168
minors, Laplacian development by, 167–68
missing dependent variables, 1042–43
Mittag-Leffler theorem, 461

mixed tensor, 136, 139, 154
modes of vibration, normal, 233–34
modified Bessel functions, 713–19
  asymptotic expansion, 711, 719
  Fourier transform, 716
  generating function, 709
  integral representation, 720–23
  Laplace transform, 933
  recurrence relations, 714–16
  series form, 714
modified Helmholtz operator, 598
modified Mathieu equation, 872
modulus, 406
moment of inertia matrix, 215–16
momentum, *see* angular momentum; orbital angular momentum
momentum representation, 955–61
  harmonic oscillator, 958–59
  hydrogen atom, 957–58
  Schrödinger wave equation, 957–58
momentum representation in quantum mechanics, 1006–7
monopole, 745–46
Morera's theorem, 427–28
motion
  area law for planetary, 116–19
  equations of, 142
moving particle
  Cartesian coordinates, 1054
  circular cylindrical coordinates, 1054–55
multiplet, 245
multiplication
  of matrices
    direct product, 181–82
      of vectors, 182
    inner product, 179–81
    by scalar, 179
multiply connected regions, 423–24
multipole expansion, electrostatic, 599
multivalent functions, 447–50
multivalued function, 409
mutually exclusive, 1109–10

N

Navier–Stokes equations, 119
Neumann boundary conditions, 543
Neumann functions, 511
  asymptotic form, 602–3
  Bessel functions of second kind, 699–707
    coaxial wave guides, 703–4
    definition and series form, 699–700
    other forms, 701
    recurrence relations, 702
    Wronskian formulas, 702–3
  Fourier transform, 602
  Hankel function definition, 707–8
  integral form, 701
  recurrence relations, 702
  spherical, 507, 727
  Wronskian formulas, 702–3
Neumann functions, integral form, 701
Neumann problem, 617
Neumann series, 1018–19
Neumann series, separable (degenerate) kernels, 1018–29
  numerical solution, 1023–25
  separable kernel, 1021–22
Neumann series solution, 1020–21
neutron diffusion theory, 369, 951
Newton's second laws, 130, 233, 370, 973, 1054
node, spiral, 1097, 1103–4
non-Cartesian tensors, 140
nonessential (regular) singular point, 563
nonhomogeneous equation — Green's function, 592–610
  circular cylindrical coordinate expansion, 601–2
  form of Green's functions, 596–98
  Legendre polynomial addition theorem, 600–601
  spherical polar coordinate expansion, 598–600
  symmetry of Green's function, 595–96
nonhomogeneous integral equation, 1032–34
nonlinear differential equations (NDEs), 1088–89
  autonomous differential equations, 1091–93
  Bernoulli and Riccati equations, 1089–90
  bifurcations in dynamical systems, 1103–4
  center or cycle, 1100–1101
  chaos in dynamical systems, 1105–6
  dissipation in dynamical systems, 1102–3
  fixed and movable singularities, special solutions, 1090
  local and global behavior in higher dimensions, 1093–94
  routes to chaos in dynamical systems, 1106–7
  saddle point, 1095–97
  spiral fixed point, 1098–1100
  stable sink, 1095
nonlinear methods and chaos, 1079–1108
  introduction, 1079–80
  logistic map, 1080–84
    exercises, 1084
  nonlinear differential equations (NDEs), 1088–89
    autonomous differential equations, 1091–93
    Bernoulli and Riccati equations, 1089–90
    bifurcations in dynamical systems, 1103–4
    center or cycle, 1100–1101
    chaos in dynamical systems, 1105–6
    dissipation in dynamical systems, 1102–3
    exercises, 1089, 1102, 1106, 1107

fixed and movable singularities, special solutions, 1090
local and global behavior in higher dimensions, 1093–94
routes to chaos in dynamical systems, 1106–7
saddle point, 1095–97
spiral fixed point, 1098–1100
stable sink, 1095
sensitivity to initial conditions and parameters, 1085–88
exercises, 1088
fractals, 1086–88
Lyapunov exponents, 1085–86
nonuniform convergence, 348–49
normalization, 695
normal matrices, 231–39
exercises, 236–39
ill-conditioned, 234–35
normal modes of vibration, 233–34
overview, 231–32
null matrix, 178
number operator, 823
numbers, Bernoulli, *see* Bernoulli numbers
numerical solution, 1023–25
nutation, 973–74

O

Oersted's law, 52, 66–68
Olbers' paradox, 337
operators, differential vector, *see* differential vector operators
optical dispersion, 484–85
optical path near event horizon of black hole, 1041–42
orbital angular momentum, 251, 261–66
exercises, 266
ladder operator approach, 262–64
Lie groups and algebras, 264–66
Lie groups and operators, order of, 264–65
operators, 793–97
overview, 261
rotation of, 251
order 2 branch points, 440–42
order parameter, 1086
ordinary differential equations (ODEs), linear first-order, 547–50
oriented surface, magnetic flux across, 306
orthogonal coordinates
Laplacian in, 316
in $\mathbb{R}^3$, 103–10
exercises, 109–10
Jacobians for polar coordinates, 108
overview, 103–8
orthogonal eigenfunctions, 636, 637, 1030–32
orthogonal functions, representation of Dirac delta function by, 88–89
orthogonal groups, 243
orthogonal group $\mathsf{SO}(3)$, 254, 256
orthogonality, 694–99, 731, 776–78, 821–22, 854–55
Bessel series, 695
continuum form, 696
curvilinear coordinates, 104
Earth's gravitational field, 758–59
electrostatic potential in hollow cylinder, 695–96
electrostatic potential of ring of charge, 761–62
expansion of functions, Legendre series, 757–58
Fourier series, 636–37
Fourier series: Hilbert–Schmidt integral equations, 919–29
normalization, 695
over discrete points, 914–15
sphere in a uniform field, 759–61
Sturm–Liouville differential equations, 1031
of vectors, 14
orthogonality condition, 10, 198
orthogonality integral
Hermite polynomials, 821
Legendre polynomials, 777
spherical harmonics, 788
orthogonality relations, spherical harmonics, 814
orthogonalization, Gram–Schmidt, 173, 642–7
orthogonal matrices, 195–208, 209
applications to vectors, 197–98
direction cosines, 196–97
Euler angles, 202–3
exercises, 206–8
inverse, 200
overview, 195–96
relation to tensors, 206
symmetry properties, 203–5
transpose matrix, $\tilde{\mathsf{A}}$, 200–202
two-dimensional conditions for, 199–200
orthonormal functions, 642–44
polynomials, 646–47
vectors, 174
oscillator
damping, 979–80, 991–93, 1101
driven, 457, 565, 991–93
forced classical, 457–60
harmonic, 822–27
integral equation for, 991–93
Laplace transform solution, 991–93
linear, 668–69
momentum space wave function, 569
self-adjoint equation, 679
series solution of differential equation, 568–69

singularities in harmonic oscillator differential equation, 564
oscillatory series, 322
overshoot, calculation of, 912–13

P

parabolic PDEs, 538
parallelogram addition law, 2–3
parallel transport, 157–60
parity, 753, 776, 939
Bessel functions, 687, 735
Chebyshev functions, 569
differential operator, 569
Fourier cosine, sine transforms, 939
Hermite functions, 569
Legendre functions, 569
Legendre functions, associated, 776
Legendre functions, second kind, 812
spherical harmonics, 776, 791
vector spherical harmonics, 814
parity transformation, 143
Parseval relation, 485–86, 952–53
partial differential equations (PDEs), 535–43
bicharacteristics of, 538
boundary conditions, 542–43
characteristics of, 538–41
classes of, 538–41
elliptic, 538
examples of, 536–38
harmonic functions, 539
hyperbolic, 538
introduction, 535–36
inversion of, 949–50
nonlinear, 541–42
parabolic, 538
partial fraction expansion, 968–69
partial sum approximation, 390
particle
in a box, 1061–62
in a sphere, 731–32
particle motion
Cartesian coordinates, 1054
circular cylindrical coordinates, 1054–55
quantum mechanical, 827
in rectangular box, 1061–62
in right circular cylinder, 1062
in sphere, 731–32
path-dependent work, 56–60
integral definitions of gradient, divergence, and curl, 58–59
overview, 56
surface integrals, 56–57
volume integrals, 57–58
Pauli matrices, 209–12
pendulums, period of simple, 370–71
periodic functions, 888–90
boundary conditions, 636, 883–5
permutations and combinations, counting of, 1114–15
phase of a complex function, 409
phase of a complex number, 409
phase space, 88, 1080, 1091
physical interpretation of divergence, 40–42
pi, π, 219, 379, 396–400, 586
Leibniz formula, 590, 776, 886
Wallis formula, 399
pion photoproduction threshold, 282–83
pi-sine relation, verification of, 522
planetary motion, area law for, 116–19
Pochhammer symbol, 859–60, 864
Poincaré section, 1088–89, 1104–6
point and space groups, crystallographic, 299–300
point source equation, 662
Poisson distribution, 1130–33
Poisson's equation, 536
and Gauss' law, 81–83
Green's function, 669–70
polar angle dependence, 788
polar coordinates, Jacobians for, 108–10, *see also* spherical polar coordinates
polar vectors, 143–44
pole expansion of meromorphic functions, 461
poles, 439
simple, on contour of integration, 468–69
polygamma functions, 511–12
Catalan's constant, 513
polynomials, Bernoulli, 379–80
potential energy, 309
potentials, 68–79, *see also* thermodynamics
gradient of, 34
force as, 36
Laplacian of, 50–51
overview, 68
scalar, 68–72
centrifugal, 72
gravitational, 72
overview, 68–71
vector, 73–79
of constant **B** field, 44
exercises, 77–79
magnetic, 74–76
potential theory
conservative force, 69
electrostatic potential, 593
scalar potential, 70
vector potential, 73, 311
power series, 363–70
continuity, 364
convergence, 363
uniform and absolute, 363

differentiation and integration, 364
exercises, 366–70
inversion of, 366
overview, 363
uniqueness theorem, 364–65
L'Hôpital's rule, 365
prime numbers, 379, 382–83, 897
prime number theorem, asymptotic, 897–8
primitive, 898
principal axis, 216
principal value, 409
probability, 1109–51
binomial distribution, 1128–30
exercises, 1129
repeated tosses of dice, 1128–30
definitions, simple properties, 1109–15
conditional probability, 1112
counting of permutations and combinations, 1114–15
exercises, 1115
probability for A or B, 1110–11
scholastic aptitude tests, 1112–14
Gauss' normal distribution, 1134–38
exercises, 1137–38
Poisson distribution, 1130–33
exercises, 1133
random variables, 1116–28
continuous random variable: hydrogen atom, 1117–19
discrete random variable, 1116–17
exercises, 1127–28
repeated draws of cards, 1123–26
standard deviation of measurements, 1119–23
sum, product, and ratio of random variables, 1126–27
statistics, 1138
χ^2 distribution, 1143–45
confidence interval, 1149
error propagation, 1138–40
exercises, 1150
fitting curves to data, 1140–43
student t distribution, 1146–49
probability for A or B, 1110–11
product convergence theorem, 344
products, *see also* cross product; direct product; dot products; scalars
expansion of entire functions, 462–63
of infinite series, 396–401
convergence of infinite, 397–98
exercises, 399–401
overview, 396–97
sine, cosine, and gamma functions, 398–99
product theorem, 181
projection operators, 644
projections, of vectors, 4, 12
pseudoscalars, 146
pseudotensors, 142–51
dual tensors, 147–48
exercises, 149–51
irreducible tensors, 149–51
Levi-Civita symbol, 146–47
overview, 142–46
pseudovectors, 146
pullbacks, 309–13
Gauss' theorem, differential form, 312–13
overview, 309–10
Stokes' theorem, 310–11
differential form, 311

Q

QCD (quantum chromodynamics), 259
$Q_n(x)$ functions of the second kind, 809–10
quadrupole, 149, 745–47, 805
quantization, 624–25, 1054
quantum chromodynamics (QCD), 259
quantum mechanical scattering, 469–71
quantum mechanical simple harmonic oscillator, 822–27
quantum mechanics, sum rules, 484
angular momentum, 189–92, 251–6, 261–4, 266–70
configuration space representation, 957–58
expectation values, 263
hydrogen atom, 245–46
momentum representation, 1006–7
Schrödinger representation, 1006–7
quantum pendulum, 873
quasiperiodic, 1100–1101, 1107
quaternions, 189, 204, 212
quotient rule, 141–42
equations of motion and field equations, 142
exercises, 142
overview, 141–42

R

$\mathbb{R}^3$, orthogonal coordinates in, *see* orthogonal coordinates
Raabe's test, 332
radial Mathieu equation, 872
radial Mathieu functions, 874–79
radioactive decay, 553, 977–78, 1129, 1130–31, 1133
raising operator, 263
random variables, 1116–28
continuous random variable: hydrogen atom, 1117–19
discrete random variable, 1116–17
standard deviation of measurements, 1119–23
sum, product, and ratio of random variables, 1126–27

rank, 264
of matrices, 178
of tensor, 133
rapidity, 280
rational approximations, 345
ratio test, Cauchy, d'Alembert, 326
Rayleigh formulas, 730
Rayleigh–Ritz variational technique, 1072–76
ground state eigenfunction, 1073
vibrating string, 1074
real part, 407
rearrangement of double series, 345–48
reciprocal, 916
reciprocal lattice, 27
reciprocity principle, 665
rectifier, full-wave, 893–94
recurrence relations, 567, 714–16, 749–50, 775, 818–19
application of, 804–5
associated Legendre polynomials, 775
Bernoulli numbers, 377
Bessel functions, 677
Bessel functions, spherical, 730
Chebyshev polynomials, 850
confluent hypergeometric functions, 866
derivatives, 852–53
exponential integral, 529
factorial function, gamma, 504, 506, 529, 995
Hankel functions, 708
Hermite polynomials, 818–19
hypergeometric functions, 861
Laguerre functions, associated, 842
Legendre polynomials, 749–50
Legendre series, 807–8
modified Bessel functions, 714–16
Neumann functions, 702
polygamma functions, 512
and special properties, 749–56
differential equations, 751–52
parity, 753
recurrence relations, 749–50
special values, 752
upper and lower bounds for $P_n(\cos\theta)$, 753–54
spherical Bessel functions, 730
reducible representations, 245–46
reflection principle, Schwarz, 431–32
regression coefficient, 1140–41
regular (nonessential) singular point, 563
regular functions (holomorphic or analytic functions), 415
regular singularities, 572–73
relations, *see* dispersion relations
relativistic energy, 356–57
repeated draws of cards, 1123–26
repellant, 1082
repellor, 1092
representation
fundamental, 259–60, 274–75
irreducible, 246, 263, 265, 268, 274, 276, 293, 298
reducible, 246
residue theorem, 455–56, 472; *see also* calculus of residues
resonant cavity, 682–85
Riccati equation, 1089–90
Riemann–Christoffel curvature tensor, 138
Riemann integral, 60–61
Riemann manifold, 313–14
Riemann's theorem, 344
Riemann surface, 448–49
Riemann Zeta function, 329–30, 334–39
and Bernoulli numbers, 382–84
Fourier series evaluation, 329
table of values, 382
infinite series, 894–98
Riesz' theorem, 177
RLC analog, 980–81
RL circuit, 549–50
Rodrigues' formula, 767
Laguerre polynomials, 839
associated, 767, 842
Rodrigues representation, 820–21
Hermite polynomials, 820–21
root diagram, 264
root test, Cauchy, 326
rotations, 176, 444
of coordinate axes, 7–12
exercises, 12
vectors and vector space, 11–12
of coordinates, 199
of functions and orbital angular momentum, 251
groups SO(2) and SO(3), 250
invariance of scalar product under, 15–17
isomorphic and homomorphic, 244–45
Rouché's theorem, 463
routes to chaos in dynamical systems, 1106–7

S

saddle points (steepest descent method), 489–97, 1092, 1095–97, 1103
analytic landscape, 489–90
asymptotic forms
of factorial function Γ, 494–95
of Hankel function $H_\nu^{(1)}(s)$, 493–94
exercises, 496–97
factorial function $\Gamma(z)$, 494–95
Hankel function $H_\nu^{(1)}$, 493–94
overview, 489
sample space, 1109–10

sawtooth wave, 885–86
scalar potential, 33–34, 70, 536
scalar quantities, 1
scalars, 7, 133
  multiplication of matrices by, 179
  potentials, 68–72
    centrifugal, 72
    gravitational, 72
    overview, 68–71
  products, 12–17
    exercises, 17
    invariance of under rotations, 15–17
    overview, 12–15
    triple, 25–27
scattering, quantum mechanical, 469–71
Schlaefli integral, 709, 768–69
  Legendre polynomials, 768
Schmidt orthogonalonization, *see* Gram–Schmidt
scholastic aptitude tests, 1112–14
Schrödinger's wave equation, 624
  degeneracy of, 638
  hydrogen atom, 843
Schrödinger wave equation, 76, 537, 1069–70
  momentum space representation, 957–58
  variational derivations, 1069–70
Schur's lemma, 265
Schwarz inequality, 652–54
  generalized, 661
Schwarz reflection principle, 431–32
second-rank tensors, 135–36
section, *see* Poincaré section
secular equation, 218
selection rules, 265
self-adjoint eigenvalue equations, 595
self-adjoint matrices, 209
self-adjoint ODEs, 622–34
  boundary conditions, 627–28
  deuteron, 626–27
  eigenfunctions, eigenvalues, 624–25
  Hermitian operators, 629
  Hermitian operators in quantum mechanics, 630
  integration interval $[a, b]$, 628–29
  Legendre's equation, 625
self-adjoint operator, 623, 630
  Sturm–Liouville theory, differential equations, 622–30
semiconvergent series, 391
sensitivity to initial conditions and parameters, 1085–88
  fractals, 1086–88
  Lyapunov exponents, 1085–86
separable kernel, 1021–22
separable variables, 544–45
separation of variables in elliptical coordinates, 870–71
series, *see* infinite series
series approach, 570–71
  Bessel's equation, limitations of, 570–71
  Chebyshev, 338, 569
  Hermite, 567
  hypergeometric, 569, 859, 862
  incomplete beta, 523
  Laguerre, 569
  Legendre, 333, 507, 569, 757–62
  shifted polynomials, 836
    Chebyshev, 850
    Legendre, 629
  ultraspherical, 338
series form, 714
  of second solution, 583–85
series solutions — Frobenius' method, 565–78
  expansion about x_0, 569
  Fuchs' theorem, 573
  limitations of series approach — Bessel's equation, 570–71
  regular and irregular singularities, 572–73
  symmetry of solutions, 569
sign changes, series with alternating, irregular, 341–42
similarity transformation, 205
simple harmonic oscillator, 973
simple pendulum, 1067–68
simple pole on contour of integration, 468–69
sine
  confluent hypergeometric representation, 393
  functions of infinite products, 398–99
  integrals in asymptotic series, 392–93
sine transform, 939–40
singularities, 438–43
  branch points, 440–42
    of order 2, 440–42
  exercises, 442–43
  fixed, 1090
  Laurent series, 439
  movable, 1090
  overview, 438
  poles, 439
singular points, 562–65
  essential (irregular), 563
  irregular (essential), 563
  nonessential (regular), 563
  regular (nonessential), 563
sink, 1092–96, 1101
$\sin x$
  infinite product representation, 378, 392–93, 398, 439, 468–69, 583, 679–80, 728–29, 730, 885, 889, 892, 911, 950, 1021, 1097–98
  power series, 358, 406, 410
skewsymmetric matrices, 204

Slater determinant, 274
small oscillations, 370–71
SO(2) rotation groups, 250
SO(3)
  Clebsch–Gordan coefficients, 267–70
  homomorphism, 252–56
  rotation groups, 250
soap film, 1045–46
soap film — minimum area, 1046–49
solenoidal, 42
soliton solutions, 542
space, vector, *see* vectors
space and point groups, crystallographic, 299–300
space–time, Minkowski, *see* kinematics and dynamics in Minkowski space–time
special unitary groups
  SU(2), Pauli spin matrices, 189, 203–4, 250–66, 267–70, 274–76
  SU(3), Gell-Mann matrices, 212, 256–60, 265–66, 274–76
  SU(n), Young tableaux, 274–76
special unitary group SU(2), 252
special values, 774–75
spectral decomposition, 219, 225, 635
sphere in a uniform field, 759–61
spheres, total charge inside, 88
spherical Bessel functions, 725–39
  asymptotic values, 729
  definitions, 726–29
  limiting values, 729–30
  recurrence relations, 730
spherical components, 271
spherical coordinates, Helmholtz equation, 725
spherical harmonics, 264, 786–93
  addition theorem for, 797–802
    derivation of addition theorem, 798–800
    trigonometric identity, 797–98
  angular momentum operators, 793
  azimuthal dependence — orthogonality, 787
  Condon–Shortley phase conventions, 270
    integrals of, 804
    ladder operators, 796–97
  Laplace series, expansion theorem, 790–91
  orthogonality integral, 788
  orthogonality relations, 814
  polar angle dependence, 788
  spherical harmonics, 788–90
  vector spherical harmonics, 813–16
spherical polar coordinates, 123–33, 557–60
  exercises, 128–33
  expansion, 598–600
  magnetic vector potential, 127–28
  ∇, $\nabla\cdot$, $\nabla\times$ for central force, 127
  overview, 123–26
  unit vectors, 123
spherical symmetry, 616
spherical tensor operator, 271
spherical tensors, 271–74
spherical waves, Bessel functions, 730
spinors, 138–39
  exercises, 138–39
  overview, 138
spinor wave functions, 212
spiral fixed point, 1098–1100
spiral node, 1097, 1103
spiral repellor, 1097, 1103
square integrable, 485, 487, 653, 658, 882
squares of series, divergent, 344
square wave, 911–12
square wave — high frequencies, 892–93
stable sink, 1095
standard deviation, 1119, 1140
standard deviation of measurements, 1119–23
stark effect, 576, 847
statistical hypothesis, 1138
statistics, 1138
  χ^2 distribution, 1143–45
  confidence interval, 1149
  error propagation, 1138–40
  fitting curves to data, 1140–43
  student t distribution, 1146–49
steepest descent, method of, 489–96
  factorial function, 494–95
  Hankel functions, 493–94
  modified Bessel functions, 720
step function, 969–70
Stirling's expansion, 495
Stirling's series, 516–20
  derivation from Euler–Maclaurin integration formula, 517
Stokes' theorem, 64–68
  alternate forms of, 66–68
    Oersted's and Faraday's laws, 66–68
    overview, 66
  on differential forms, 313–14
    Riemann manifold, 313–14
  exercises, 67–68
  overview, 64–65
  proof, 420–21
  pullbacks, 310–11
straight line, 1044–45
strange attractor, 1085, 1086
string, Lagrangian of a vibrational, 1058, 1071
structure constants, 248, 251, 259
student t distribution, 1146–49
Sturm–Liouville theory, 885
Sturm–Liouville theory — orthogonal functions, 621–74
  completeness of egenfunctions, 649–61
    Bessel's inequality, 651–52

expansion coefficients, 658
Schwarz inequality, 652–54
summary — vector spaces, completeness, 654–58
Sturm–Liouville theory — orthogonal functions
completeness of Eigenfunction, 659–61
Gram–Schmidt orthogonalization, 642–49
exercises, 647–49
Legendre polynomials by Gram–Schmidt orthogonalization, 644–46
Green's function — eigenfunction expansion, 662–74
eigenfunction, eigenvalue equation, 667–68
exercises, 670–74
Green's function and the Dirac delta function, 669–70
Green's function integral — differential equation, 665–67
Green's functions — one-dimensional, 663–65
linear ocillator, 668–69
Hermitian operators, 634–42
degeneracy, 638
exercises, 639–42
expansion in orthogonal eigenfunctions — square wave, 637
Fourier series — orthogonality, 636–37
orthogonal eigenfunctions, 636
real eigenvalues, 634–35
self-adjoint ODEs, 622–34
boundary conditions, 627–28
deuteron, 626–27
eigenfunctions, eigenvalues, 624–25
exercises, 631–34
Hermitian operators, 629
Hermitian operators in quantum mechanics, 630
integration interval $[a, b]$, 628–29
Legendre's equation, 625
SU(2)
Clebsch–Gordan coefficients, 267–70
isospin and SU(3) flavor symmetry, 256–60
and SO(3) homomorphism, 252–56
SU(3) flavor symmetry, 256–60
subgroups and cosets, 293–94
substitution, 979
subtraction
of matrices, 178–79
of series, 324–25
of sets, 1111
of tensors, 136
sum, product, and ratio of random variables, 1126–27
summation convention, 136–37, 139
summation of series, 910
sum rules, 484
SU(n), young tableaux for, 274–78
superposition principle for homogenous ODEs, PDEs, 536
surface integrals, 56–57
symmetric matrices, 204
symmetric tensor, 137
symmetrization of kernels, 1029–30
symmetry, 889
axes
threefold, 296–99
twofold, 294–96
cylindrical, 617
properties of orthogonal matrices, 203–5
relations, 484
of solutions, 569
spherical, 616
SU(3) flavor, 256–60
of tensors, 137

T

tableaux for SU(n), Young, 274–78
Taylor's expansion, 352–63, 430–31
binomial theorem, 356–57
relativistic energy, 356–57
exercises, 358–63
Maclaurin theorem, 354–55
exponential function, 354–55
logarithm, 355
overview, 354
multiple variables, 358
overview, 352–54
tensor analysis
contravariant tensor, 135–36, 153
contravariant vector, 134–35, 139, 152–54, 158
covariant tensor, 135–36, 158
covariant vector, 134–35, 139–40, 152–54, 156, 158
definition, 133–36
displacement, 158
isotropic tensor, 137
non-Cartesian tensors, 140
parallel transport, 158
scalar quantity, 8, 15, 57, 134, 149, 179
spherical components, 271
spherical tensor operator, 271
symmetry–asymmetry, 137
tensor density, *see* pseudotensors
tensor derivative operators
curl, 162–63
divergence, 160–61
exercises, 162–63
Laplacian, 161–62
overview, 160

tensors, *see also* derivative operators, tensor; direct product; general tensors; pseudotensors; quotient rule; spinors
general tensors, 151–60, *see also* Christoffel symbols
covariant derivative, 156
exercises, 158–60
geodesics and parallel transport, 157–60
metric tensor, 151–54
overview, 151
relation to orthogonal matrices, 206
spherical, 271–74
vector analysis in, 133–63
addition and subtraction of, 136
contraction, 139
overview, 133–35
second-rank, 135–36
summation convention, 136–37
symmetry–antisymmetry, 137
thermodynamics, 72–79
exact differentials, 72–76
overview, 72–73
vector potential, 73–79
exercises, 77–79
magnetic, 74–76
Thomas precession, 280
threefold Hermite formula, 827–28
threefold symmetry axis, 296–99
time-dependent diffusion equation, 536
time-independent diffusion equation, 536
Titchmarsh theorem, 487
trace, 139
trace formula, 224–25
Gutzwiller's, 898
traces of matrices, 183–84
trajectory, 38, 1088, 1091–92, 1097, 1099, 1100, 1103, 1105
transfer function, 962
transfer functions, 961–64
significance of $\Phi(t)$, 963–64
transform, derivative of, 982–83; *see also* cosines; exponential; Fourier; Fourier–Bessel; Hankel; Laplace; Mellin; sine
transformation law, 134
transformation of differential equation into integral equation, 1008–9
transformation of **E** and **B**, Lorentz, 287–88
translation, 443–44, 981
transport, parallel, 157–60
transpose matrix, $\tilde{A}$, 200–202
transposition, 177
triangle inequalities, 406
triangle rule, 268–69
trigonometric form, 853–54
trigonometric identity, 797–98
triple scalar products, 25–27, 165–66
triple vector products, 27–29, 46
BAC–CAB rule, 28, 46, 51
exercises, 27–32
overview, 27
Tschebyscheff, *see* Chebyshev
two-dimensional conditions for orthogonal matrices, 199–200
twofold symmetry axis, 294–96

U

ultraspherical polynomials, extension to, 747
equation, 853, 861
self-adjoint form, 854
uncertainty principle in quantum theory, 941
uniform convergence, 348–49, 363
union of sets, 1111
uniqueness theorem, 364–65
descending power series, 781
inverse operator, 181
Laurent expansion, 433
of power series, 364–65
uniqueness theorem, L'Hôpital's rule, 365
unitary groups, 243
unitary matrices, *see also* Hermitian matrices
algebra, 210–12
Heaviside, 93, 981, 985, 996
ring, 180, 187
unit element of group, 293
unit step function, 191
vector space, 208
unit vectors, 123
Cartesian coordinates, 5
circular cylindrical, 116
orthogonality relation, 651
spherical polar, 201–2
spherical polar coordinates, 126
upper and lower bounds for $P_n(\cos\theta)$, 753–4

V

values, limiting, of elliptic integrals, 374
variables, *see also* complex variables
dependent, 1052–58
Hamilton's Principle, 1053–54
Laplace's equation, 1057–58
moving particle — Cartesian coordinates, 1054
moving particle — circular cylindrical coordinates, 1054–55
dependent and independent, 1038–44, 1058–59
alternate forms of Euler equations, 1042
concept of variation, 1038–41
missing dependent variables, 1042–43
optical path near event horizon of a black hole, 1041–42

multiple, of Taylor's expansion, 358
separation of, 554–62
variance, 1120
variation, concept of, 1038–41
variation of the constant, 548
variations, *see also* calculus of variations
variation with constraints, 1065–72
Lagrangian equations, 1066–67
Schrödinger wave equation, 1069–70
simple pendulum, 1067–68
sliding off a log, 1068–69
vector analysis
parallelogram addition law, 2–3
reciprocal lattice, 27
rotation of coordinates, 205
transformation law, 134
vector definition, 8–9
representation of, 9
vector expansion, 745–47
vector field, 7
vector integrals
line integrals, 55
surface integrals, 56–57
volume integrals, 57
vector potential, 73, 311
vector product, 5, 11, 18–22, 25–29, 43–44, 46, 147, 272
vector quantities, 1
vectors, 1–101, *see also* curved coordinates and vectors; divergence, ∇; gradient, ∇; integration; potentials; rotations; Stokes' theorem; tensors
applications of orthogonal matrices to, 197–98
components, 8, 11, 153, 654
contravariant, 134–35, 139, 152–54, 158
covariant, 134–35, 139, 152–54, 156, 158
cross product of, 315
curl, $\nabla\times$, 43–49
of central force field, 44–46
exercises, 47–49
gradient of dot product, 46
integration by parts of, 47
overview, 43
potential of constant **B** field, 44
definitions and elementary approach, 1–7
exercises, 6–7
differential vector operators, 110–14
curl, 112–13
divergence, 111–12
exercises, 113–14
gradient, 110
overview, 110
Dirac delta function, 83–95
exercises, 91–95
integral representations for, 90–95
overview, 83–87
phase space, 88
representation by orthogonal functions, 88–89
total charge inside sphere, 88
direction, 5
direct product of, 182
elementary approach to, 1–7
elementary approach to, exercises, 6–7
Gauss' law, 79–83
exercises, 82–83
Poisson's equation, 81–83
Gauss' theorem, 60–64
alternate forms of, 62
exercises, 62–64
Green's theorem, 61–62
overview, 60–61
by Gram–Schmidt orthogonalization, 174–76
Helmholtz's theorem, 95–101
exercises, 100–101
overview, 95–96
irrotational, 45–46
linear dependence of, 172–73
normal, 15, 56
or cross product, 18–22
exercises, 22–25
orthogonal, 108, 173
overview, 1
potentials, magnetic, 127–28
scalar or dot product, 12–17
exercises, 17
invariance of under rotations, 15–16
serve as a basis, 5
space, 11–12
exercises, 12
successive applications of ∇, 49–54
electromagnetic wave equation, 51–53
exercises, 53–54
Laplacian of potential, 50–51
overview, 49–50
triangle law of addition, 1–2
triple product of, 27–29
exercises, 27–32
triple scalar product, 25–27
vector space, linear space, 7, 11–12, 173, 177, 208, 219, 247–48, 314, 638, 644, 650, 654–55, 657–58, 884
vector spaces, completeness, 654–58
vector spherical harmonics, 813–16
vector transformation law, 21
velocity of electromagnetic waves in a dispersive medium, 997–99
vibrating string, 1074
vibration, normal modes of, 233–34
vierergruppe, 292–93, 296

Volterra equation, 1005–6
volume integrals, 57–58
von Staudt–Clausen theorem, 379

W

Wallis' formula, 399
wave diffusion equation (Helmholtz diffusion equation), 536, 537
wave equation, 947–48
  anomalous dispersion, 999
  derivation from Maxwell's equation, 52
  Fourier transform solution, 947–48
  Laplace transform solution, 948
wave equation, electromagnetic, 51–53
wave functions, 624
  spinor, 212
wave guides, coaxial, Bessel functions, 703–4
Weierstrass infinite-product form of $\Gamma(z)$, 499–500
Weierstrass M test, 349–50
weight diagram, 257, 258, 260, 265
weight vectors, 265
Whittaker functions, 866, 919
Wigner–Eckart theorem, 273
WKB expansion, 394
work, potential, 34, 70–72
Wronskian, 550
Wronskian determinant, 579–80
Wronskian formulas, 702–3
  absence of third solution, 587, 589
  Bessel functions, 702–4
  Bessel functions, spherical, 601, 723
  Chebyshev functions, 582–83, 591
  confluent hypergeometric functions, 868
  Green's function, construction of, 592, 703
  linear independence of functions, 579–80, 665, 703
  second solution of differential equations, 583–87
  solutions of self-adjoint differential equation, 702

Y

Young tableaux for $\mathsf{SU}(n)$, 274–77

Z

zero-point energy, 732, 826
zeros, Bessel function, 682
zeta function, *see* Riemann Zeta function