Classical Mechanics **Herbert Goldstein** **1950**

Table of contents: (numbers in parens indicate book page numbers)

Comments on the author and book. 3
Chapter 1: Survey of the Basics (1) (mainly, the Lagrangian formulation of mechanics) 5
1.1 Mechanics of a particle (1) 5
1.2 Mechanics of a System of Particles (4) 5
1.3 Constraints (10). 6
1.4 D'Alembert's principle and Lagrange's equations (14) 6
1.5 Velocity-dependent potentials and the dissipative function (18). 17
1.6 Simple applications of the Lagrangian Formulation (22) 19
Chapter 2: Variational Principles and Lagrange's Equations (30) 20
2.1 Hamilton's Principle (30). 20
2.2 Calculus of Variations (31) 20
2.3 Derivation of Lagrange's Equations from Hamilton's Principle (36) 22
2.4 Extension of Hamilton's Principle to more general systems. (38) 22
2.5 Advantages of a variational principle formulation (44). 25
2.6 Conservation theorems and symmetry properties (47) 26
Chapter 3: The 2-body problem with a central force (58) 28
3.1 Reduction to the equivalent 1-body problem (58) 28
3.2 Central Force Problems (59). 28
3.3 The equivalent one-dim problem, and the classification of orbits (63) 29
3.4 The virial theorem (69). 31
3.5 Differential equation for the orbit (71) 32
3.6 The Kepler problem: inverse square law force. (76) 33
3.7 Central Force Scattering (81) 37
3.8 The Lab Frame (85). 41
Chapter 4: Rigid Body Kinematics (93) 49
4.1 The independent coordinates of a rigid body (93) 50
4.2 Real orthogonal transformations (97). 50
4.3 Formal properties of the transformation matrix (101). 50
4.4 The Euler angles (107). 50
4.5 The Cayley-Klein parameters (109). 50
4.6 Euler's Theorem on the motion of a rigid body (118). 51
4.7 Infinitesimal rotations (124) 52
4.8 Rate of change of a vector (132). 54
Clarification of Notational Ambiguity. 54
4.9 The Coriolis Force (135) 56
Chapter 5: Rigid Body Equations of Motion (143) 59
5.1 Angular Momentum and Kinetic Energy of motion about a point (143). 59
5.2 Tensors, Dyads and Dyadics (146) 60
5.3 The inertial tensor and the moment of inertia (149) 62
5.4 The eigenvalues of the inertia tensor and the principle axis transformation (151). 62
5.5. Methods of solving rigid body problems and the Euler equations of motion (156). 63
5.6 Force-free motion of a rigid body (159). 64

5.7 The heavy symmetrical top with one point fixed (164). 68
The Heavy Top Started with a Drop. 70
Heavy top started vertically (p 173) 71
Some Applications (p174). 72
5.8 Precession of charged bodies in a magnetic field (176) 72
The Classical Motion of a Rigid Charged Body in a Magnetic Field 72
Chapter 6. Special Relativity in Classical Mechanics (p 185) 73
6.1 The basic program(185). 73
6.2 The Lorentz Transformation (187). 73
6.3 The covariance of equations (194). 73
6.4 What happens to F = ma ? (199) 74
6.5 Relativistic Lagrangian situation (205) 74
6.6 Covariant Lagrangian situation (207) 75
Chapter 7. The Hamilton Equations of Motion (p 215) 75
7.1 Legendre Transformations and the Hamilton Equations of Motion (215) 75
7.2 Cyclic coordinates and Routh's procedure (218) 76
7.3 Conservation Theorems and Physical Significance of H (220) 77
7.4 Hamilton's Equations from a variational principle (225). 77
7.5 Principle of Least Action (228) 77
Comment on the word "action". 82
Chapter 8. Canonical Transformations (p 237) 83
8.1 The Equations of Canonical Transformation (237). 83
8.2. Examples of canonical transformations (244) 84
8.3 The integral invariants of Poincare (247) 84
8.4 Lagrange and Poisson Brackets as canonical invariants (250). 85
8.5 Equations of motion in Poisson bracket notation (255). 87
8.6 Infinitesimal contact transformations, constants of the motion, symmetry (258) 88
8.7 The angular momentum Poisson bracket relations (263). 88
8.8 Liouville's Theorem (206) 89
Chapter 9. Hamilton-Jacobi Theory (p 273) 89
9.1 The Hamilton-Jacobi Identity for the Hamilton's Principle Function (273) 89
9.2 A HJE example: the harmonic oscillator (277). 90
9.3 Hamilton's Characteristic Function W (279) 90
Review of the Table: 92
9.4 Separation of Variables in the HJE (284) 92
9.5 Action-angle variables (288) 93
9.6 Further properties of action-angle variables (294) 95
9.7 The Kepler problem in action-angle variables (299) 98
Part I: The basics 98
Part II: Computing the J_i action variables 101
9.8 Hamilton-Jacobi theory, geometric optics, and wave mechanics (307) 104
History Added: 114
Chapter 10. Small Oscillations (p 318) 115
10.1 Formulation of the Problem (318). 116
10.2 The eigenvalue equation and the principle axis transformation(321) 116
10.3 Frequencies of free vibration and normal coordinates (329). 121

Comparison with the inertia tensor situation. 123
10.4 Free vibrations of a linear triatomic molecule (333). 123
10.5 Forced vibrations and dissipative forces (338). 126
Chapter 11. Continuous Systems and Fields (p 347)∂ 130
11.1. Transition from discrete to continuous system (347). 130
11.2 The Lagrangian formulation for continuous systems (350). 130
11.3 Sound vibrations in a gas (355). 131
11.4 The Hamiltonian Formulation (355) 133
11.5 Descriptions of Fields (364). 134

Comments on the author and book.

This fellow died in 2005, basically wrote this book post war while he had a temp job as a Harvard instructor. His main gig was at Columbia where he specialized in nuclear radiation shielding. A second edition of this book appeared in 1980, but I am reading my original copy.

A *third* edition appeared in **2001** with coauthors **Poole and Safko**. This would be used in a current day course I suppose. Here is the contents of that book:

1. Survey of the Elementary Principles.
2. Variational Principles and Lagrange's Equations.
UPDATED! 3. The Central Force Problem.
4. The Kinematics of Rigid Body Motion.
5. The Rigid Body Equations of Motion.
UPDATED! 6. Oscillations.
REVISED! 7. The Classical Mechanics of the Special Theory of Relativity.
8. The Hamiltonian Equations of Motion.
9. Canonical Transformations.
10. Hamilton-Jacobi Theory and Action Angle Variables.
NEW! 11. Classical Chaos.
REVISED! 12. Canonical Perturbation Theory.
13. Introduction to Lagrangian and Hamiltonian Formulations for Continuous Systems and Fields.
Appendixes.
Select Bibliography.
Index.

This sight maintains errata for the 3rd edition: http://astro.physics.sc.edu/goldstein/

You can see how this compares to my 1950 edition which has 11 chapters. My chapter 10 on oscillations appears earlier as chapter 6 in the new book. Chapters on chaos and perturbation theory have been added. I suspect much of the book is unchanged! All 11 chapters have the exact same names as in the 1950 book.

Note added 10.31.16. I found errata on 2nd and 3rd edition but not first. For 3rd it is a website supposedly maintained by John L. Safko at South Carolina. http://astro.physics.sc.edu/goldstein/ . *My* PDF has no copyright page, so I don't know its printing number. But it seems to have corrections for 1,2,3 printings, and it seems *not* to have corretions for 4 and 5; for example in my edition the item for page 13 is not

corrected. Thus it seems my GPS copy is either printing 4 or 5. My PDF date is Aug 2006. There are many errata for all the editions! I don't know how to find printing dates online. There is now a 2013 "international edition" with a different cover, from Pearson. My integral error page page 93 does not appear in any of these errata lists.

It turns out that all three authors are dead, and all in the not-too-distant past:

Herbert Goldstein: Died 1.12.05 http://www.columbia.edu/cu/news/05/02/herbertGoldstein.html

Charles P. Poole Jr: Died 11.1.15 at age 88,

http://obits.dignitymemorial.com/dignity-memorial/obituary.aspx?n=Charles-Poole&lc=9817&pid=176301825&uuid=0d732362-339a-469c-aae6-65be8d0d08e5

John L. Safko: Died 6.15.16 at age 77, did 38 years at South Carolina, a good teacher it says

http://www.legacy.com/obituaries/thestate/obituary.aspx?n=john-safko&pid=180380299&fhid=32171

Preface: Various arguments are given why a solid basis in *classical* mechanics is good for people about to study quantum mechanics and field theory. He throws in special relativity and velocity-dependent potentials.

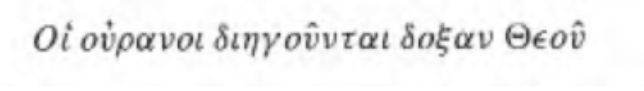
The preface ends with a Hebrew word that I don't know how to translate or even type. I tried some Yiddish dictionaries, no luck. But then I found it on the web with a fancy search: "I had first encountered this *insider's code* in a secular environment many years ago when reading herbert goldstein's majesterial text on classical mechanics (published 1950, addison wesley) as he closed his preface with the mysterious hebrew lettering."

The word is really an **abbreviation**, not a word, and you do the letters right to left. So if I ignore the thing that looks like a ", I can get these 6 letters: O - B L Sh N T which you can reverse to say T N Sh L B O which roughly agrees with the following phrase:

TVSLBO Tom V'nishlom Shevoch L'Eil Boreih Olom (Hebrew acronym to mark the conclusion of a published work) -- means: "finished and complete is (this?) praise to god, creator of the world" In other places I see it written as " Tam V'nishlam Shevach L-eil Boreih Olam."

Web: There is a touching tribute to the individual that can be derived from the traditional custom of writing in conclusion of a Jewish book, the Hebrew phrase "Tam V'nishlam Shevach L'El Borei Olam" - "The book is complete, praise be to God"

The 3rd edition has a similar but different ending to the Preface, this time in Greek instead of Hebrew,



which is this from Psalm 19,1:

	3588	3772	1334	1391	2316
19:1	οι	ουρανοί	διήγουνται	δόξαν	θεού
	The	heavens	describe	*the* glory	of God,

****

Chapter 1: Survey of the Basics (1) (mainly, the Lagrangian formulation of mechanics)

****

My Lagrange doc paper constains a lot of this stuff.

1.1 Mechanics of a particle (1). We assume **F** = d**p**/dt and just work from there. **p** is conserved if no force. **L** is defined and shown to be conserved if no torque **N**. Work as Fdx is integrated to show that work causes a change in kinetic energy T. A conservative force is one whose path integral is zero, which in turn means it can be represented as the gradient of a potential V. We soon end up with T+V being conserved. He has not mentioned that many of his statements are only true non-relativistically.

Question: If the force in a problem can be written as the gradient of a potential V which is velocity-dependent, is that still a **conservative force**? Gold seems to avoid asking this question. The work done in moving in a closed loop must be 0. Obviously if $V = kx|v_x|$ so that $F = k|v_x|$ describes a friction situation, we know that work is done against friction so positive work is done going around the loop, so such a force cannot be conservative (which really means non-heat energy is conserved I think). So here at least is an example of a "velocity dependent potential" which is non-conservative. On the other hand, EM theory gives an effective potential U which I think IS conservative. Probably this is because the Lorentz force **v** x **B** is always perpendicular to **v** so no work gets done. So if a force can be written as the gradient of a position-only-dependent potential, the force is conservative. If the potential is velocity dependent, then the force may or may not be conservative.

1.2 Mechanics of a System of Particles (4). Assumption here is that $F_{ij} = - F_{ji}$ and lies along the line between the particles i and j (either polarity). With this assumption, many conclusions are reached. The center of mass R is defined and it's momentum is conserved if no external force. Total angular momentum is conserved if no external torque. On page 7, L is written as the sum $L_{cms} + L_{internal}$ relative to the CMS. The only dependence on the origin location here is through R, the CMS location. On page 9 he has a potential $V_{ij}(r_{ij})$ which gives rise to the internal forces (between particles i and j), and represents the internal potential energy. The total potential is then given top of page 10 where V_i is from external forces, and the 1/2 just from double counting. We conclude that T+V is again conserved for the system as a whole, no great surprise. Beware E&M fields where forces don't meet the assumptions here, and cause violations since they carry momentum and angular momentum and energy. "

Comments: In this fairly long 6 page section, forces on a particle are partitioned into "internal" and "external" forces. Internal means forces of particles acting on each other. This is *different* from a later partition of the total force on a particle as "constraint force" and "other forces; constraints are not mentioned in this section. It is only if you make the assumption $\mathbf{F}_{ij} = - \mathbf{F}_{ji}$ for the internal forces that you can show the famous "rigid body" results quoted in my paragraph above such as the existence of a center of mass vector **R** and the fact that $M\ddot{\mathbf{R}} = \Sigma \mathbf{F}_{ext}$ and the partitioning of angular momentum and energy into CMS and local components, and so on. So $\mathbf{F}_{ij} = - \mathbf{F}_{ji}$ is a very major assumption used in this section 1.2,

1.3 Constraints (10). To dispel the notion that you just "turn the crank", this little topic is brought up. Sometimes constraints can be written as equations like top of page 11, in which case they are **holonomic**. [I think holo- means whole, suggesting there are no differentials appearing in the constraints.] Otherwise they are non-holonomic which means they are harder to deal with. Example of the latter is any inequality. Holonomic really means you have a constraint equation which allows you to remove a variable from the problem (an "integrable constraint"). Then on page 12 if we have k constraints, we can come up with a number of "generalized coordinates" q_i which equals the true number of degrees of freedom in your problem (here 3N-k), then the 3N true coordinates $\mathbf{r}_i$ which are not independent can be written in terms of the generalized q_i ones (note that in general the q_i won't come in vectors like the $\mathbf{r}_i$ do, often the q_i are angles). For example, a bead constrained to a fixed 1D string in 3space, would nominally have 3 degrees of freedom x,y and z, but in reality has only one degree of freedom q_1 = distance along the string. Rolling without slipping is *non*holonomic because you cannot express it as an equation on the $\mathbf{r}_i$. It is a velocity condition of the point of contact (thus involving $d\mathbf{r}_i$). He considers a rolling vertical disk on a flat surface, comes up with nice coordinates. He can write down the constraints as differentials, but this does not count as holonomic. A good problem, I commented long ago. These problems require "custom solutions".

A main point here: If you start with 3N coordinates for a system of **N particles**, and if you have k *holonomic* constraints, you can then define a set of 3N-k generalized coordinates q_i which are independent variables! This independence is crucial for doing Lagrangian mechanics as we shall see. If your constraints are non-holonomic, you have to modify Lagrangian mechanics with those Lagrange multipliers (as we shall also see).

Comment: Nowhere in this 4 page section is there an explicit partitioning of forces acting on particle k as a sum of constraint forces plus other forces. No symbol is used for any force really.

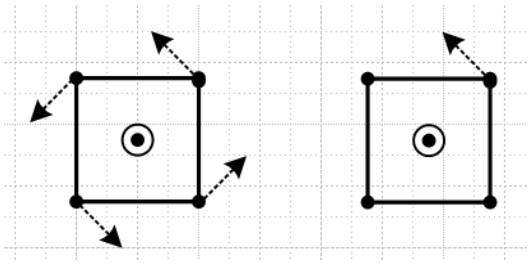
Comment: Gold uses the symbol N as the number of particles in a system.

1.4 D'Alembert's principle and Lagrange's equations (14).

This section now has 11 pages of raw notes below mainly because I just did not understand the approach which is centered on D'Alembert's Principle. Now I think it all makes sense, but it took a long time.
[1.10.12 while reviewing M&F was led to review Goldstein a bit.] [By 9.29.16 I have written Lagrange App D which helps on a lot of this stuff below, and have read the R&S paper 2006 on $\delta\mathbf{r}$.]

Here we have something new to me: the notion of a virtual displacement $\delta\mathbf{r}_i$ of the coordinates of a system of particles each having a label i. In my first reading, I was totally confused, but then in Jan 2012 I read a little Indian PDF on the subject and will rewrite my notes here based on what I learned there.

That PDF discusses both an "allowed displacement $d\mathbf{r}_i$" and a "virtual displacement $\delta\mathbf{r}_i$". In either case, you think of a SET of displacements, one for each particle in the system, so write as $\{ \delta\mathbf{r}_i \}$ as a reminder of this fact. As an example, consider a 4 particle rigid-body system:



The set of four arrows on the left shows an example of a legal "allowed displacement" set of 4 arrows $\{d\mathbf{r}_i\}$. The four particles can move differentially as shown because this comprises an allowed rotation of the rigid body about some origin. The only allowed system displacement sets $\{d\mathbf{r}_i\}$ in this 2D world would be sets which correspond to either a rotation about some origin, or to a translation (or a combination). The picture on the right shows a non-allowed set $\{d\mathbf{r}_i\}$ where three of the four arrows are zero displacement. This set would require deformation of the "sticks" holding the 4 particles together and this violates the constraints.

The Indian PDF clarifies the "time" aspect of things by saying that a "virtual displacement" is just the difference of two "allowed displacements" (which difference is not allowed to be zero). Their idea is this: from the chain rule acting on the constraint set $f_i(\mathbf{r}_j\ , t) = 0$ (k equations for G, s for the Indians) we have this fact, which I write in four different notations

$\Sigma_{j=1}^{N} \Sigma_{a=1}^{3}\ \partial f_i/\partial(\mathbf{r}_j)_a\ \partial(\mathbf{r}_j)_a/\partial t\ + \partial f_i/\partial t\ = 0$ // N = number of particles
$\Sigma_{j=1}^{N} \Sigma_{a=1}^{3}\ \partial f_i/\partial(\mathbf{r}_j)_a\ (\mathbf{v}_j)_a\ + \partial f_i/\partial t\ = 0$ // a = index like x,y,z
$\Sigma_{j=1}^{N} \nabla^{(j)} f_i \bullet \mathbf{v}_j\ + \partial f_i/\partial t\ = 0$ // i = constraint number i
$\Sigma_{j=1}^{N}\ \partial f_i/\partial \mathbf{r}_j \bullet \mathbf{v}_j\ + \partial f_i/\partial t\ = 0$ // this is the Indian notation

For an "allowed" displacement set, one has $d\mathbf{r}_j = \mathbf{v}_j dt$ in the {..} sense described above for the system. Thus the last equation above can be written

$\Sigma_{j=1}^{N}\ \partial f_i/\partial \mathbf{r}_j \bullet \mathbf{v}_j dt\ + (\partial f_i/\partial t)\ dt\ = 0$
$\Sigma_{j=1}^{N}\ \partial f_i/\partial \mathbf{r}_j \bullet d\mathbf{r}_j\ + (\partial f_i/\partial t)\ dt\ = 0$

By defining a $\delta\mathbf{r}_j$ virtual displacement as any difference of allowed $d\mathbf{r}_j$, the Indians conclude that

$\Sigma_{j=1}^{N}\ (\partial f_i/\partial \mathbf{r}_j) \bullet \delta\mathbf{r}_j\ = 0$ // time cancels in difference i = 1...k

and it is in this "sense" that $\delta\mathbf{r}_j$ does not involve a dt movement. Here we can interpret $(\partial f_i/\partial \mathbf{r}_j)$ as some kind of "force" and then the above is some kind of "principle of zero virtual work". We know that we cannot set the 3n scalar members of the set $\{\delta\mathbf{r}_i\}$ to arbitrary values due to the constraints -- we can only have sets $\{\delta\mathbf{r}_i\}$ which are "allowed" by the constraints. In other words, the 3n variations δr_i are not "independent". Maybe later I can tie the above in with Gold's discussion.

Fact: The above equation says that the $\delta\mathbf{r}_i$ are linearly dependent! That is the key point.

So now let's move back to the Goldstein discussion on page 14. The plan here is to start with a system in static equilibrium, and then later make it be dynamic. We start this long section with (1-38). Why is this true in equilibrium? It says $\Sigma_i \delta W_i = 0$. If the total force $\mathbf{F}_i$ on particle i is zero, then sure, $\mathbf{F}_i = 0$ and so we cannot deny that $\mathbf{F}_i \bullet \delta \mathbf{r}_i = 0$ and then sum to get (1-38). This equation seems to say nothing interesting. He assumes equilibrium here (clearly stated). Again, the individual terms in this sum are 0 because the total force on each particle is 0 because we are at rest.

In (1-39) we recognize that some forces actually come from the constraints, called $\mathbf{f}_i$. Then $\mathbf{F}^a_i$ refers to the "applied force", meaning any non-constraint force. This separation seems reasonable, and then (1-40) is a restatement of (1-38).

Question: For the first time now he has partitioned forces into constraint forces and "applied" forces. If there are internal inter-particle forces F_{ij}, into which category do they go? He uses the same $F_i^{(a)}$ symbol that he used in the previous partition of internal + external forces. Gold does not answer this question. R&S refer to the $F_i^{(a)}$ portion as "external forces". How could Gold ignore this question? A simple example would be a set of N planets moving around. Are the gravitational attractions "applied" or are they "constraint" ?

Let's look right now at some other sources:

Gold 3rd edition (w Poole, Safko). I just found this as PDF for the first time. The text in this area is unchanged, so nothing said about the partition. But good now to have this book for the first time. I can give refs to multiple editions now.

Greenwood Classical Dynamics (in Dover): I cannot find this as PDF, but good search views are provided in google books

https://books.google.com/books?id=x7rj83I98yMC&q=alembert#v=snippet&q=alembert&f=false

Section 1.4 is on virtual work. Refers to some constraints as "workless constraints". Uses same R_i and F_i used by R&S. Same work=0 for static. In the partition F_i is "applied" force. Good text, lots of examples, but no mention of "internal forces".

L&L: Have a djvu, but no good hits on "internal forces"

Web search: I found someone asking this same question, but not useful.

Forces PDF: Found and saved a pdf that classifies forces into external + internal + constraints. So at least someone is talking about this. But too elementary to mention virtual work.

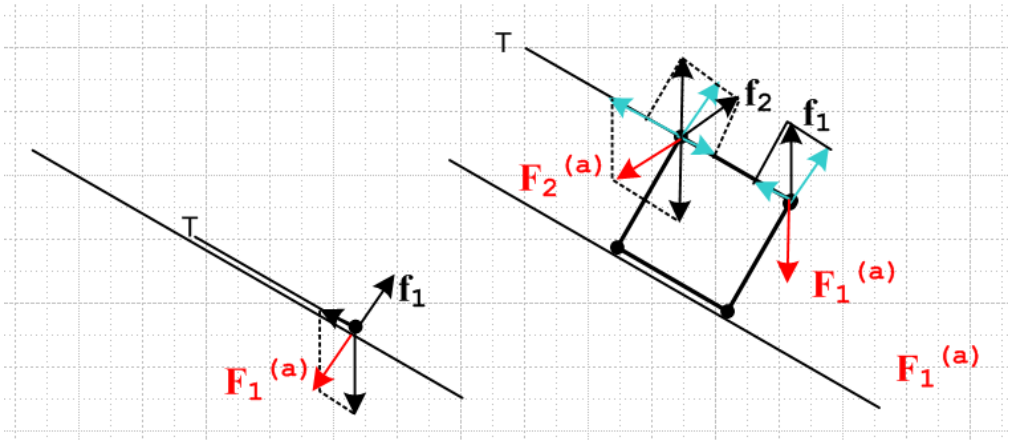
wiki on virtual work: States that the internal forces of a rigid body or those of an ideal joint are classified as constraint forces, mainly because there is no virtual work for them! But no d'Alembert so cannot see my answer directly. Searching now on: constraint forces "internal forces" "virtual work"

my answer: If you by fiat make the division R + F as being "constraint plus other", then the division depends on what you define to be the constraint forces. For a rigid body of masses and sticks, you want to

classify the internal forces as constraint forces because these constraints are associated with zero work. but for a semi-rigid body with springs instead of sticks between the masses (planets eg), I think you would classify the internal force on a particle (force due to all the other particles) as an "other" force and not as a constraint force, because if you classify as constraint, they you have contraints which do work. In the limit that the springs become infinitely stiff, you would reclassify the internal forces as constraint forces. So I guess I am happy now with this question and its answer.

I have added my answer to Lagrange Appendix D.

Two Examples:



The left example is a particle on an inclined plane with a string tension **T** holding things at rest. Here, the total force $\mathbf{F}_1$ acting on particle 1 is zero, as in 1-38. We write this as $\mathbf{F}_1 = \mathbf{F}_1^{(a)} + \mathbf{f}_1$. The constraint force $\mathbf{f}_1$ is normal to the surface as shown. Then $\mathbf{F}_1^{(a)}$ is the sum of the gravity down force and the string tension force, so $\mathbf{F}_1^{(a)}$ is then equal and opposite to $\mathbf{f}_1$. The allowed or virtual displacement would be along the inclined plane, so one would have $\mathbf{f}_1 \bullet \delta\mathbf{r}_1 = 0$ due to perpendicularity. Then 1-41 says that $\mathbf{F}_1^{(a)} \bullet \delta\mathbf{r}_1 = 0$ and we see this is true also by perpendicularity.

Now let's switch to the example shown on the right of a 4-particle rigid object at rest on an inclined plane still held by a string. In this case, $\mathbf{f}_1$ is the sum of the two green arrows forces exerted by the sticks, and $\mathbf{F}_1^{(a)}$ is the gravity down force all by itself. The allowed $\delta\mathbf{r}_i$ set for our object would be four arrows down the plane (not shown). So notice right off the bat that now $\mathbf{f}_1 \bullet \delta\mathbf{r}_1 \neq 0$ and also $\mathbf{F}_1^{(a)} \bullet \delta\mathbf{r}_1 \neq 0$ for this individual particle! So suddenly all of G's discussion "goes hazy". How do we know that this is true for our 4-particle example:

$$\Sigma_{i=1}^{4}\ \mathbf{f}_i \bullet \delta\mathbf{r}_i = 0 \quad ?$$

I have drawn in all the arrows for particle 2 where things are radically different in terms of force directions. But here is the key point. The two green arrows along the top stick are equal and opposite "internal" forces. Thus we will have $\mathbf{f}_2^{(top)} \bullet \delta\mathbf{r}_2 + \mathbf{f}_1^{(top)} \bullet \delta\mathbf{r}_1 = 0$ and we can regard these two terms as 2 of 8 terms in the sum above. Top means due only to the top stick. A similar pair would be for the bottom stick. The side sticks all have things like $\mathbf{f}_1^{(side)} \bullet \delta\mathbf{r}_1 = 0$ due to perpendicularity and this accounts for the other four terms. Thus, the above sum will in fact be valid! Gold just does not give the detail that is needed here, so I will now provide it. What is really true for our example is this

$\Sigma_{i=1}^{4}\ \mathbf{f}_i = 0$

where $\mathbf{f}_i$ is the total "internal force" on particle i, formerly written as $F^{(i)}$. Write [(i) means internal]

$\mathbf{f}_i = \Sigma_{j\neq i}\ \mathbf{F}^{(i)}{}_{ij}$

where the total internal force on i comes from all the other particles in the rigid object. Then consider

$\Sigma_{i=1}^{N}\ \mathbf{f}_i = \Sigma_{i=1}^{N}\ \Sigma_{j\neq i}\ \mathbf{F}^{(i)}{}_{ij} = \Sigma_{pairs,\ j\neq i}\ \mathbf{F}^{(i)}{}_{ij}$

This is the sum over all pairs ij where i≠j, so there are (N,2) terms in the sum. now write

$= \Sigma_{pairs,\ i<j}\ \mathbf{F}^{(i)}{}_{ij} + \Sigma_{pairs,\ i>j}\ \mathbf{F}^{(i)}{}_{ij} = \Sigma_{pairs,\ i<j}\ [\ \mathbf{F}_{ij} + \mathbf{F}_{ji}]$

Now Gold's assumption p 4 is that these internal forces are equal and opposite within each pair, as would be the case if we had "sticks" connecting each pair (or gravity or electrostatic force etc). Then each term in the above sum over pairs i<j is 0, and thus we have proven that $\Sigma_{i=1}^{4}\ \mathbf{f}_i = 0$ for "this kind of system". Then since the $\delta\mathbf{r}_i$ virtual displacement set of arrows consists of equal arrows pointing down the plane, we then arrive at $\Sigma_{i=1}^{n}\ \mathbf{f}_i \bullet \delta\mathbf{r}_i = 0$ which is the right part of 1-40.

In conclusion, I agree that $\Sigma_{i=1}^{n}\ \mathbf{f}_i \bullet \delta\mathbf{r}_i = 0$ as a sum, whereas the individual terms labelled by i are in fact not zero. We saw this in our second example above. So here we are

$$\Sigma_{i=1}^{N}\ \mathbf{F}^{(a)}{}_i \bullet \delta\mathbf{r}_i = 0 \qquad (1\text{-}41)$$

The truth of this equation is not trivially obvious, since the individual terms are not 0. For each particle, the symbol $\mathbf{F}^{(a)}{}_i$ means all forces applied to that particle OTHER than those of the contraints. In our example 2, particle 1 had only gravity, whereas particle 2 had gravity + string tension. Since the above is not trivially true but is only true for systems in which $\Sigma_{i=1}^{n}\ \mathbf{f}_i \bullet \delta\mathbf{r}_i = 0$, and since the terms in the sum have dimensions of work $\mathbf{F}\bullet d\mathbf{r}$, and since the set $\{\delta\mathbf{r}_i\}$ must be an allowed "virtual" displacement (not an arbitrary displacement set for the system}, this result (1-41) earns itself a formal name which is **"the principle of virtual work"**. One comment about (1-41) is that we seem to have simplified the problem in some sense by removing constraint forces from the equation. This is static equilibrium only !!

The history is unclear here but http://books.google.com/books?id=jmKxffjQtasC suggests that maybe the Flemish statics guy Stevin (1586, 1605) came up with this idea. Newton Principia was 1687.

Having read my notes above, I have to say that Gold is just totally unclear in this section p 14-15. No meaning for the symbol $\delta\mathbf{r}_i$ is given, no definition. I agree with the Indians, Goldstein has just dropped the ball here. Going back to the right-side figure above, forget the string, and the thing can slide on the frictionless surface. Any allowed displacement is 4 downhill arrows, so any Indian virtual displacement would be the same, so call each $\delta\mathbf{r}$! Then if f_i are the four constraint forces, you get

$\Sigma_{i=1}^{4}\ \mathbf{f}_i \bullet \delta\mathbf{r}_i = (\Sigma_{i=1}^{4}\ \mathbf{f}_i) \bullet \delta\mathbf{r} = 0 \bullet \delta\mathbf{r} = 0$ // as noted in detail above

where the individual terms are NOT zero but the sum is zero. So this is an example where the inclined plane is not moving, but the individual terms are not zero. So App D needs more work. // I have now upgraded App D with an example of this type.

Continue now in Goldstein on page 15:

Up to this point, Gold has been talking virtual work = 0 in statics situation. Now comes dynamics:

So, now we are going to just start over and replace $\mathbf{F}_i$ by ($\mathbf{F}_i$- d$\mathbf{p}_i$/dt). Then (1-38) becomes (1-41'), true because $\mathbf{F}_i$- d$\mathbf{p}_i$/dt = 0 by Newton's Law. If we break out the constraint forces, we get A at top of page 16. And again if we restrict to our special systems as above in which $\Sigma_i \mathbf{f}_i \bullet \delta\mathbf{r}_i = 0$, we get (1-42). Notice that the quantity ($\mathbf{F}_i^a$- d$\mathbf{p}_i$/dt) might not be zero, but the dot product sum is zero. Again, this gets the fancy name of **"D'Alembert's Principle":**

$$\Sigma_{i=1}^{n} \left(\mathbf{F}^{(a)}_i - \dot{\mathbf{p}}_i \right) \bullet \delta\mathbf{r}_i = 0 \qquad (1\text{-}42)$$

This principle is again non-trivial, just as was its static version. Very importantly, each term in the sum is not separately zero. We do not have $\mathbf{F}^{(a)}_i = m_i\mathbf{a}_i$ because that is true is $\mathbf{F}_i = m_i\mathbf{a}_i$ and if you exclude the contraint forces the result is in general not true -- F = ma is true if you include ALL the forces. Nevertheless, the sum above is true. Jean le Rond d'Alembert did this in 1743 which you see is 50 years after Principia.

On page 16 after stating (1-42) Gold simplifies his notation so that after this $\mathbf{F}_i$ means $\mathbf{F}_i^{(a)}$. It is true that in (1-42) the constraint forces $\mathbf{f}_i$ do not appear at all.

Independent "generalized" coordinates (p 16 middle). He now imagines that n is the number of "independent" coordinates so that p 16 B is true: $\mathbf{r}_i = \mathbf{r}_i(q_j, t)$. He differentiates this to get (1-43) which is just fine.

$$\mathbf{r}_i = \mathbf{r}_i(q_j, t)$$

$$d\mathbf{r}_i/dt = \mathbf{v}_i = \Sigma_{j=1}^{n} (\partial\mathbf{r}_i/\partial q_j)(\partial q_j/\partial t) + \partial\mathbf{r}_i/\partial t \qquad (1\text{-}43)$$

$$d\mathbf{r}_i = \Sigma_{j=1}^{n} (\partial\mathbf{r}_i/\partial q_j)(\partial q_j/\partial t)dt + (\partial\mathbf{r}_i/\partial t)dt$$

$$d\mathbf{r}_i = \Sigma_{j=1}^{n} (\partial\mathbf{r}_i/\partial q_j)\,\dot{q}_j dt + (\partial\mathbf{r}_i/\partial t)dt$$
$$d\mathbf{r}'_i = \Sigma_{j=1}^{n} (\partial\mathbf{r}_i/\partial q_j)\,\dot{q}'_j dt + (\partial\mathbf{r}_i/\partial t)dt$$

Here my idea is that these are two different allowed displacements and they differ because the generalized velocity sets are set differently, one with prime the other not. Things like ($\partial\mathbf{r}_i/\partial q_j$) and ($\partial\mathbf{r}_i/\partial t$) are more or less constants, determined by $\mathbf{r}_i = \mathbf{r}_i(q_j, t)$, these cannot be different for different displacements. Now subtract to get

$$\delta\mathbf{r}_i = \Sigma_{j=1}^{n} (\partial\mathbf{r}_i/\partial q_j)\,[\,\dot{q}'_j - \dot{q}_j\,]dt$$

Then write

$$[\,\dot{q}'_j - \dot{q}_j\,]dt \;=\; \partial_t\,[\,q'_j - q_j\,]\,dt \;=\; q'_j - q_j \;=\; \delta q_j$$

Then we arrive at

$$\delta \mathbf{r}_i = \Sigma_{j=1}^{n}\,(\partial \mathbf{r}_i/\partial q_j)\,\delta q_j \qquad (1\text{-}44)$$

I think this is how the Indians would derive (1.44)

Note added 9.29. Could write the above as

$$d\mathbf{r}_i = \Sigma_{j=1}^{n}\,(\partial \mathbf{r}_i/\partial q_j)\,d\mathbf{q}_j + (\partial \mathbf{r}_i/\partial t)dt$$
$$d\mathbf{r}'_i = \Sigma_{j=1}^{n}\,(\partial \mathbf{r}_i/\partial q_j)d\mathbf{q}'_j + (\partial \mathbf{r}_i/\partial t)dt$$

Then

$$\delta \mathbf{r}_i \;= d\mathbf{r}_i - d\mathbf{r}'_i \;=\; \Sigma_{j=1}^{n}\,(\partial \mathbf{r}_i/\partial q_j)\,[d\mathbf{q}_j - d\mathbf{q}'_j] \;=\; \Sigma_{j=1}^{n}\,(\partial \mathbf{r}_i/\partial q_j)\,\delta \mathbf{q}_j$$

This is really a trivial point. The notion of $\delta\mathbf{q}_j = d\mathbf{q}_j - d\mathbf{q}'_j$ is exactly the same as the R&S notion.

Both the virtual $\delta\mathbf{r}_i$ displacements and the virtual δq_i displacements must be taken as differences of the corresponding allowed displacements. Gold comments in parens that this then supports time-dependent constraints, and the Indians would agree.

So (1-44) is a very important ingredient and Gold sort of fudges it, but I now accept it. Then (1-45) is a no-brainer and he has then defined his generalized force Qj.

So Gold now wants to process D'Alembert 1-42 into something involving only these q_j coordinates. He comments that if we could cast this into the form $\Sigma_j G_j \delta q_j = 0$ where δq_j are virtual displacements in the *generalized* coordinates, then since these are independent, we could conclude that this result tells us that $G_j = 0$.

If we skip the algebra for the moment and cut to the chase, we see that we can in fact make such a formula and it is shown in (1-49) and we have $G_j \equiv [\ \ldots\ ]$ as shown there. This object G_j is seen to involve various derivatives of kinetic energy T, and the *generalized force* Q_j that corresponds to the generalized coordinate q_j. Accepting this for the moment, we then have (1-50) on page 18 top. The operator acting on T is the same as that we know appears in "Lagrange's equation", which will appear in a moment. In some sense, (1-50) says that T is driven by the generalized force Q_j. At this point, we have not assumed that forces come from a potential V, or that such a V even exists.

Now we go an extra step and assume that our regular force $\mathbf{F}_i$ (meaning the applied force) is conservative and so can be obtained as $\mathbf{F}_i = -\nabla_i V(..r_i...)$. Again skipping algebra steps, we then arrive at the famous **Lagrange Equation** of motion (1-53) where we have defined a new interesting object known as **"The Lagrangian L"** and **L = T - V**, not to be confused with the Hamiltonian which is H = T

+ V = total energy. The L thing is some weird energy difference thing. The derivatives in (1-53) are wrt to generalized coordinate q_j and the generalized velocity $\dot{q}_j$, as well as wrt time.

$$d/dt(\partial L/\partial \dot{q}_j) - \partial L/\partial q_j = 0$$

where each term has dimensions of energy over dim(q_j). The generalized forces Q_j no longer appear! We get one equation for each of our independent q_j coordinates. Assumes conservative forces (which means those derivable from a potential energy gradient. This V in turn could still have velocities in it.)

Comment: If you know the transform $\mathbf{r}_i = \mathbf{r}_i(q_j, t)$ and you know the real forces F_i. you can eaily compute the generalized forces Q_i using (1-46), so the Q_i are totally well-defined.

Now let's go back and look at the details of the algebra.

Note: I have done this I think in a simpler and clearer way in my Appendix D, but I leave these details here

********** start algebra **********

Start with p 16 B and apply d/dt:

$$(d\mathbf{r}_i/dt) = (\partial \mathbf{r}_i/\partial q_j)\,(dq_j/dt) + (\partial \mathbf{r}_i/\partial t) \qquad \text{or} \qquad \dot{r}_j = (\partial \mathbf{r}_i/\partial q_j)\,\dot{q}_j + (\partial \mathbf{r}_i/\partial t) \qquad (1\text{-}43)$$

Next, assume (1-44) based on discussion earlier so that

$$\delta \mathbf{r}_i = \Sigma_{j=1}^{n} (\partial \mathbf{r}_i/\partial q_j)\,\delta q_j \qquad (1\text{-}44)$$

Now the first term in D'Alambert is this, where we can take F^i as either total or applies (noted above)

$$\Sigma_{i=1}^{N}\ \mathbf{F}_i \bullet \delta \mathbf{r}_i \qquad N = \text{number of particles in system}$$

Now insert 1-44 to get

$$\Sigma_{i=1}^{N}\ \mathbf{F}_i \bullet \delta \mathbf{r}_i = \Sigma_{i=1}^{N}\ \mathbf{F}_i \bullet [\Sigma_{j=1}^{n} (\partial \mathbf{r}_i/\partial q_j)\,\delta q_j]$$

$$= \Sigma_{j=1}^{n} [\Sigma_{i=1}^{N}\ \mathbf{F}_i \bullet (\partial \mathbf{r}_i/\partial q_j)]\,\delta q_j$$

$$= \Sigma_{j=1}^{n} [Q_j]\,\delta q_j = 0 \qquad (1\text{-}45) \qquad Q_j \equiv \Sigma_{i=1}^{N}\ \mathbf{F}_i \bullet (\partial \mathbf{r}_i/\partial q_j) \qquad (1\text{-}46)$$

The second term in D'Alembert is (ignoring minus sign for the moment)

$$\Sigma_{i=1}^{N}\ \dot{\mathbf{p}}_i \bullet \delta \mathbf{r}_i = \Sigma_{i=1}^{N}\ m_i\, \ddot{\mathbf{r}}_i \bullet \delta \mathbf{r}_i \qquad (\text{p } 17\ A)$$

$$= \Sigma_{i=1}^{N} \; m_i \ddot{\mathbf{r}}_i \bullet [\Sigma_{j=1}^{n} (\partial \mathbf{r}_i/\partial q_j) \, \delta q_j]$$

$$= \Sigma_{i=1}^{N} \Sigma_{j=1}^{n} \; m_i \ddot{\mathbf{r}}_i \bullet (\partial \mathbf{r}_i/\partial q_j) \, \delta q_j \qquad \text{(p 17 B)}$$

Now write

$$d/dt[\, m_i \dot{\mathbf{r}}_i \bullet (\partial \mathbf{r}_i/\partial q_j)] = m_i \ddot{\mathbf{r}}_i \bullet (\partial \mathbf{r}_i/\partial q_j) + m_i \dot{\mathbf{r}}_i \bullet d/dt \, (\partial \mathbf{r}_i/\partial q_j)$$

or

$$m_i \ddot{\mathbf{r}}_i \bullet (\partial \mathbf{r}_i/\partial q_j) = d/dt[\, m_i \dot{\mathbf{r}}_i \bullet (\partial \mathbf{r}_i/\partial q_j)] - m_i \dot{\mathbf{r}}_i \bullet d/dt \, (\partial \mathbf{r}_i/\partial q_j)$$

Then sum this on i to get

$$\Sigma_i \, m_i \ddot{\mathbf{r}}_i \bullet (\partial \mathbf{r}_i/\partial q_j) = \Sigma_i \, \{ \, d/dt[\, m_i \dot{\mathbf{r}}_i \bullet (\partial \mathbf{r}_i/\partial q_j)] - m_i \dot{\mathbf{r}}_i \bullet d/dt \, (\partial \mathbf{r}_i/\partial q_j) \, \} \qquad \text{(1-47)}$$

Now take (1.43) and apply $\partial/\partial q_k$ to both sides to get

$$\partial/\partial q_k (d\mathbf{r}_i/dt) = \partial/\partial q_k \, \Sigma_j [(\partial \mathbf{r}_i/\partial q_j) \, (dq_j/dt)] + \partial/\partial q_k \, (\partial \mathbf{r}_i/\partial t)$$

Now one does not think of (dq_j/dt) as being a function of q_j so pull that from the bracket

$$\partial/\partial q_k (d\mathbf{r}_i/dt) = \Sigma_j \, \partial/\partial q_k \, [(\partial \mathbf{r}_i/\partial q_j)] \, (dq_j/dt) + \partial/\partial q_k \, (\partial \mathbf{r}_i/\partial t)$$

Or just think

$$\partial/\partial q_k \; (dq_j/dt) = d/dt \, (dq_j/\partial q_k) = d/dt \, (\delta_{ij}) = 0$$

and then claim is justified. Now swap k↔ j to get

$$\partial/\partial q_j (d\mathbf{r}_i/dt) = \Sigma_k \, \partial/\partial q_j \, [(\partial \mathbf{r}_i/\partial q_k)] \, (dq_k/dt) + \partial/\partial q_j \, (\partial \mathbf{r}_i/\partial t)$$

Interchange on left, rewrite on right

$$d/dt(\partial \mathbf{r}_i/\partial q_j) = \Sigma_k \, (\partial^2 \mathbf{r}_i/\partial q_j \partial q_k) \, \dot{q}_k + \partial^2 \mathbf{r}_i/\partial q_j \partial t \qquad \text{p 17 C}$$

or

$$(\partial \mathbf{v}_i/\partial q_j) = \Sigma_k \, (\partial^2 \mathbf{r}_i/\partial q_j \partial q_k) \, \dot{q}_k + \partial^2 \mathbf{r}_i/\partial q_j \partial t$$

where the LHS's are trivially the same: $d/dt(\partial \mathbf{r}_i/\partial q_j) = (\partial \mathbf{v}_i/\partial q_j)$ (*)

Now start again with (1-43) and apply $\partial/\partial \dot{q}_k$ to both sides to get

$$(d\mathbf{r}_i/dt) = \Sigma_j (\partial \mathbf{r}_i/\partial q_j) \, (dq_j/dt) + (\partial \mathbf{r}_i/\partial t)$$

$$\partial/\partial\dot{q}_k\,(d\mathbf{r}_i/dt) \;=\; \partial/\partial\dot{q}_k\,\Sigma_j\,(\partial\mathbf{r}_i/\partial q_j)\,(dq_j/dt) \;+\; \partial/\partial\dot{q}_k\,(\partial\mathbf{r}_i/\partial t)$$

Now r_i is not a function of $\dot{q}_k$ (only of q_k) and so we get

$$\partial/\partial\dot{q}_k\,(d\mathbf{r}_i/dt) \;= \Sigma_j\,(\partial\mathbf{r}_i/\partial q_j)\,\delta_{kj} \;=\; (\partial\mathbf{r}_i/\partial q_k)$$

$$\partial\mathbf{v}_i/\partial\dot{q}_k \;= \partial\mathbf{r}_i/\partial q_k$$

and change k to j

$$\partial\mathbf{v}_i/\partial\dot{q}_j \;= \partial\mathbf{r}_i/\partial q_j \qquad (1\text{-}48)$$

Now I will add my own line. Rewrite

Now start with 1-47 again

$$\Sigma_i\, m_i\, \ddot{\mathbf{r}}_i \bullet (\partial\mathbf{r}_i/\partial q_j) \;= \Sigma_i\,\{\, d/dt[\, m_i\, \dot{\mathbf{r}}_i \bullet \underset{1\text{-}48}{(\partial\mathbf{r}_i/\partial q_j)}] - m_i\, \dot{\mathbf{r}}_i \bullet \underset{(*)}{d/dt\,(\partial\mathbf{r}_i/\partial q_j)}\,\} \qquad (1\text{-}47)$$

Replace as shown to get

$$\Sigma_i\, m_i\, \ddot{\mathbf{r}}_i \bullet (\partial\mathbf{r}_i/\partial q_j) \;= \Sigma_i\,\{\, d/dt[\, m_i\mathbf{v}_i \bullet \partial\mathbf{v}_i/\partial\dot{q}_k] - m_i\, \mathbf{v}_i \bullet (\partial\mathbf{v}_i/\partial q_j)\,\} \qquad \text{p 17 E}$$

Now look at {...} in p 17F. It can be written as

$$\{...\}_{17F} = d/dt\,[\, m\,\mathbf{v}_i \bullet \partial\mathbf{v}_i/\partial\dot{q}_k] \;- m_i\,\mathbf{v}_i \bullet (\partial\mathbf{v}_i/\partial q_j) = \{...\}_{17E}$$

Therefore we have from p 17E

$$\Sigma_i\, m_i\, \ddot{\mathbf{r}}_i \bullet (\partial\mathbf{r}_i/\partial q_j) \;= \Sigma_i\;\{...\}_{17F}$$
$$\Sigma_i\;\dot{\mathbf{p}}_i \bullet (\partial\mathbf{r}_i/\partial q_j) \;= \Sigma_i\;\{...\}_{17F}$$
$$\Sigma_i\;\dot{\mathbf{p}}_i \bullet \Sigma_j(\partial\mathbf{r}_i/\partial q_j)\delta q_j \;= \Sigma_j\,\Sigma_i\;\{...\}_{17F}\,\delta q_j$$
$$\Sigma_i\;\dot{\mathbf{p}}_i \bullet \delta\mathbf{r}_i \;= \Sigma_j\,\Sigma_i\;\{...\}_{17F}\,\delta q_j$$

The LHS is the desired piece of (1-42) , and this RHS is

$$\Sigma_j\,\Sigma_i\;\{...\}_{17F}\,\delta q_j \;= \Sigma_j\,[d/dt\,(\partial T/\partial\dot{q}_j) - \;(\partial T/\partial q_j)]\;\}\,\delta q_j$$

which is to say

$$\Sigma_i\;\dot{\mathbf{p}}_i \bullet \delta\mathbf{r}_i \;= \Sigma_j\,[d/dt\,(\partial T/\partial\dot{q}_j) - \;(\partial T/\partial q_j)]\;\}\,\delta q_j \qquad (**)$$

Now recall from above that the first term in (1-42) was

$\Sigma_{i=1}^{N}\ \mathbf{F}_i \bullet \delta\mathbf{r}_i = \Sigma_{j=1}^{n} [Q_j]\ \delta q_j$

$\Sigma_i \mathbf{F}_i \bullet \delta\mathbf{r}_i = \Sigma_j Q_j\ \delta q_j$

Therefore we can write (1-42) this way

$\Sigma_j Q_j\ \delta q_j - \Sigma_j [d/dt\ (\partial T/\partial \dot{q}_j) - (\partial T/\partial q_j)]\ \}\ \delta q_j = 0$

$\Sigma_j \{ Q_j - [d/dt\ (\partial T/\partial \dot{q}_j) - (\partial T/\partial q_j)]\ \}\ \delta q_j = 0$

$\Sigma_j \{ [d/dt\ (\partial T/\partial \dot{q}_j) - (\partial T/\partial q_j)] - Q_j\ \}\ \delta q_j = 0$ (1-49)

Now recall that we could write p 16B because we could eliminate some of our r_i type variables using our holonomic constraints, thus givein us the n independent variables q_j and this from above we can say

$[d/dt\ (\partial T/\partial \dot{q}_j) - (\partial T/\partial q_j)] = Q_j$ $\qquad T = \Sigma_i (1/2\ m_i v_i^2)$ (1-50)
"Lagrange's Equations"

********** end of long algebra section **********

The above equation involves the generalized coordinates q_j and generalized velocities $\dot{q}_j$ and generalized forces Q_j as determined by 1-46. The equation is true in the presence of our constraints! If we had no constraints and $q_i \sim r_i$ (vector issue of course), then the first term on the left is just $m_i \dot{v}_i$ and Q = F and we just have ma = F. But the whole point is that 1-50 is valid with constraints where the q_i are some wierd but independent coordinates and T may somehow depend on the q_i as well as the $\dot{q}_i$.

Now assume that the real forces $\mathbf{F}_i = -\nabla_i V(\mathbf{r}_i)$, for example

$\mathbf{F}_2 = -\nabla_{r2} V(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, ...)$

Then we know from 1-46 that

$Q_j \equiv \Sigma_i \mathbf{F}_i \bullet (\partial \mathbf{r}_i/\partial q_j) = \Sigma_i -\nabla_i V \bullet (\partial \mathbf{r}_i/\partial q_j) = -\Sigma_i (\partial V/\partial \mathbf{r}_i) \bullet (\partial \mathbf{r}_i/\partial q_j)$

But the chain rule then says (this assumes that $V = V(\mathbf{r}_i$; but *not* t and not $\mathbf{v}_i)$)

$Q_j = -\partial V/\partial q_j$ (1-51)

Comment: Conservative force means all closed line integral paths give 0. If force comes from potential $V(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, ...)$ then this is true. If there is velocity dependence in V, then whether or not the force is conservative probably depends on the details. The force is likely NOT to be conservative. This would be

the case if there are any frictional forces. But I could imagine that somehow the velocity dependence could be such that the force somehow remains conservative. A quick web scan did not clarify this point.

So our Lagrange equations (1-51) are now

$$[d/dt\,(\partial T/\partial \dot{q}_j) - (\partial T/\partial q_j)] = Q_j = -\partial V/\partial q_j$$

$$[d/dt\,(\partial T/\partial \dot{q}_j) - (\partial T/\partial q_j) + \partial V/\partial q_j\,] = 0$$

$$[d/dt\,(\partial T/\partial \dot{q}_j) - (\partial [T-V]/\partial q_j)] = 0$$

$$[d/dt\,(\partial [T-V]/\partial \dot{q}_j) - (\partial [T-V]/\partial q_j)] = 0$$

$$[d/dt\,(\partial L/\partial \dot{q}_j) - (\partial L/\partial q_j)] = 0 \qquad (1\text{-}53)$$

and these are what most people call "Lagrange's equations", written in terms of the Lagrangian L.
So we have finally battled our way through section 1-4 !

Assumptions: holonomic constraints, and contraint forces respect $\Sigma_i \mathbf{f}_i \bullet \delta \mathbf{r}_i = 0$, and in this form we must have all forces from a conservative potential V.

You can be sure that our very dense Section 1.4 here is the subject of much of other mechanics books! Lagrangian dynamics. OK 1.8.12 OK 9.29.16

1.5 Velocity-dependent potentials and the dissipative function (18). Two separate topics are treated here: (1) particle in EM field, (2) particle with linear friction. Both involve "velocity dependent forces" and in fact both cases have forces *linear* in velocity. These situations cannot be handled by "velocity independent potentials" V because, were that so, we would compute $\mathbf{F} = -\nabla V$ and then **F** would be velocity independent.

We have to treat the first situation because this case will apply to very standard problems involving EM fields acting on charged particles, which comprise the subject of a huge portion of physics!

Let's ignore the opening comments for the moment and look at the E&M algebra. Everything is quoted from Jackson, I will say. We define **A** in (1-57) and define ϕ in equation above (1-58). Then (1-58) gives the **E** field as a function of the potentials **A** and ϕ, complementing (1-57) which gives the **B** field in terms of just **A**. The Lorentz Force can be written in terms of potentials as (1-59) which we got from the quoted result (1-56). Goldstein is not deriving all these E&M results, he is just quoting many and then stating equations he is interested in. We see in any event that the Lorentz force has a term which is a linear function of velocity in (1-59).

To avoid using a fancy vector identity, Goldstein does lots of details by hand, all of which I verified, and he is then able to write the Lorentz force as in equation β on page 21. Then if he defines U as shown in (1-60), he can write the Lorentz force as in equation γ. Now let's pause on this as just a pure algebra result, and let's go back to the section's opening comments.

The comments say: *Suppose* you could express the generalized force Q_j as the Lagrange-like set of derivatives of some "generalized potential" U as shown in (1-54). If this were possible to do, then you could rewrite (1-50) as the Euler-Lagrange equation (1-53) provided you used U in place of the potential V. Now we can think of the real force and real coordinate as examples of the generalized versions of the

same, if there are no constraints. In equation γ on page 21, we have in fact written the real force for a certain problem (charged particle in an EM field) as exactly that combination of derivatives acting on a generalized potential U which in turn is given in (1-60). Therefore, we can apply the normal Euler Lagrange equation to problems like this in E&M *provided* we use L = T - U. This final L Lagrangian then has the form L = T - qϕ + q/c **A**• **v** . So this is NOT an L which you write as L = T - V.

Note added 9.29.16. Here is the math to prove the claim of the opening section. We have (1-50)

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_j}\right) - \frac{\partial T}{\partial q_j} = Q_j$$

and then (1-54) says

$$\frac{d}{dt}\left(\frac{\partial U}{\partial \dot{q}_j}\right) - \frac{\partial U}{\partial q_j} = Q_j$$

So if these equations are both true we could subtract them to get

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_j}\right) - \frac{\partial L}{\partial q_j} = 0 \qquad \text{provided we take } L = T - U.$$

In this section the generalized and real coordinates are the same, so equation γ on page 21 matches (1-54).

Comments: This is of course a purely classical result. However, we know for a point charge in motion at position $\mathbf{r}_1$ and velocity **v** that $\rho = q\,\delta(\mathbf{r}-\mathbf{r}_1)$ and $\mathbf{J} = q\mathbf{v}\delta(\mathbf{r}-\mathbf{r}_1)$ and we know that we can think of the combination $(\rho,\mathbf{J}) = J^\mu$ as a conserved 4-current in that $\partial_\mu J^\mu = 0$. Suppose we define $(\phi,\mathbf{A}/c) = A^\mu$ as a 4-potential of some kind, so that $(-\phi,\mathbf{A}/c) = A_\mu$. Then we can say,

$$L = T - q\phi + q/c\ \mathbf{A}\bullet\mathbf{v} = T + A_\mu J^\mu\ \delta(\mathbf{r}-\mathbf{r}_1)$$

Although a little strange, this shows that we have the Lorentz covariant combination of A and J. This last term is of course more or less the interaction Lagrangian of QED field theory, where we would have something like $J^\mu = e\bar{\psi}\,\gamma^\mu\psi$ for the electron current and A^μ being the photon field. We are getting a little preview or "omen" of field theory here.

This expression L = T - qϕ + q/c **A**• **v** is not easy to find in my books, so Goldstein has done us all a big favor by writing it down and showing its significance. Jackson has it on page 407 (how could he NOT write this somewhere). He talks about p_μ instead of J_μ and that fixes up my slight mess above, so we then have

$$L = T - q\phi + q/c\ \mathbf{A}\bullet\mathbf{v} = T + (q/m)A_\mu p^\mu \qquad \text{where } p^\mu = m(1,\mathbf{v})$$

where I have fudged some relativistic details for now because I cannot remember them.

We now have a complete subject change in mid section p 21. We know that if we have non-conservative forces, we can use the Lagrange form given earlier in (1-50). We want to show how we can incorporate friction into our analysis, provided frictional force is proportional to velocity as in $\mathbf{F}_i = -k\mathbf{v}_i$. In this case, we first derive the **Rayleigh Dissipative Function** $\mathcal{F}$ as bottom of page 21. The algebra on page 22 top shows that we can think of $2\mathcal{F}$ as the rate at which the frictional process consumes energy. The main result, however, is that we can write this frictional effect as a generalized force $Q_j = -(\partial\mathcal{F}/\partial\dot{q}_j)$ where we don't at this point really have to specify details of our generalized coordinates. So we can use this form for Q_j in our result (1-50) and we end up with equation α on page 22. We cannot incorporate $\mathcal{F}$ into the Lagrangian as in our previous situation, so we have to carry the extra term shown. But now we can use the Lagrangian method to solve problems which include linear friction! And earlier we learned how to incorporate EM fields as well. So this has been a fairly heavy-duty Goldstein section! We will now finally consider some "problems". [But we don't get a problem illustrating the friction thing until page 45 in the next chapter.] OK 1/9/12 OK 9/29/16

Comment: Neither of these examples fits into a simple mold where you have force $\mathbf{Q} = -\nabla V(\mathbf{q}_i, \dot{\mathbf{q}}_i, t)$. In the EM case there are potentials φ and **A**, but **Q** isn't just a gradient of these things as (1-59) shows. Maybe the friction example does sort of fit this mold because you end up with $Q_i = -\partial\mathcal{F}/\partial\dot{q}_i$ where yes, this is a "velocity gradient" and not a normal position gradient, and $\mathcal{F} = (1/2)\, k\, \dot{q}_i^2$ plays the role of V.

1.6 Simple applications of the Lagrangian Formulation (22) Claim is made that since T and V are scalars, calculations are simpler than those dealing with vector forces. Observation is made that you can always write T in the form given in (1-62) and following, for any choice of generalized coordinates. And if there is no explicit time activity in the **r**(**q**) equations (**scleronomous** = time independent constraints) , only the a_{jk} term survives, quite obvious. Thus, in this case, T is a "homogeneous function of degree 2" in the various $\dot{q}_j$, meaning like $x^2 - 3xy + 2y^2$. So don't forget that this quadratic form of T is related to the idea of time-independent constraints! He now looks at some examples:

1(a) particle in space, arbitrary force **F** (no potential given), so T = as shown, and everything boils down to **F** = m**a**. In this situation, the generalized coordinates are selected to be the same as the Cartesian ones.

1(b) same but in planar polar coords. Now take r and θ as the coordinates, compute T in terms of them. The **r**(**q**) equations are for example x = rcosθ. But we maintain two generalized coords here, no constraints to reduce the number. We compute the generalized forces and use the (1-50) version of Lagrange. That is, we have no potential V here, but we have some generalized forces. One is the radial force, the other is the tangential force which causes a torque which drives the ang mom.

2. Atwood's machine -- two weights and a pulley. Only one coordinate x, write T and V, turn the crank to get top page 26. There was a constraint in this problem -- the positions of the two weights are correlated so only one variable x is needed, not two. We never worry about this constraint in the L formulation, and correspondingly, the tension in the rope is not revealed in the solution.

3. Bead sliding on a uniformly rotating wire. When treated in r,θ coordinates, the one holonomic constraint is that θ = ωt which write as θ - ωt = 0 = f(θ,t), This is then a time-dependent constraint. So we have two variables and one constraint, this leaves r as the only degree of freedom. Our L formalism tells

us that $\ddot{r} = r\omega^2$ and we see the bead accelerating outward as expected. The details of the constraint forces on the bead never enter our realm of concern!

Author then reviews a series of mechanics texts, with some interesting comments. Whittaker's famous book has no drawings! Milne's makes a mess of vectors. These are mostly very famous names. I at least read through all the problems, they are not bad, nine of them. [I have the Whittaker 1917 2nd Ed now, and it has a whole chapter on the 3-body problem. I read a little and found that this system consists of 9 2nd order PDE's. That would be a great chapter to read some day.]

Summary: In this Chapter 1, we explained the interpretation of d'Alembert's Principle involving applied forces (and ignoring constraint forces) and virtual displacements. Then we used this to derive two versions of the Lagrange equation of motion. The first involves T and Q_i and does not require conservative forces and is (1-50) on page 18. The second (1-53) *does* require conservative forces derived from a potential V, and this contains L = T-V, a quantity known as the Lagrangian, and no Q_i appears. We saw how one can use a "generalized" potential U in place of V in the Lagrangian to handle a charged particle in an EM field, and we saw also how one can add a term to Lagrange's equation to handle linear frictional forces. Then a few simple problems were considered after a bibliography with discussion.

Comments 10.28.16. My Lagrange doc Appendix D has clarified nearly all of the mysteries present in Goldtstein's Chapter 1. But I really should prepare a better summary maybe with pasted stuff from G.

As of 10.29.16 I am pretty happy with Goldstein Chapter 1. Wrote much of it into Lag doc.

****

Chapter 2: Variational Principles and Lagrange's Equations (30)

****

2.1 Hamilton's Principle (30). This principle says that $I = \int L\, dt$ is stationary along the actual path of motion, where L is of course the Lagrangian T - V. In other words, $\delta I = 0$. In more detail, you think of t as a parameter and $L = L(q_i(t), \dot{q}_i(t), t)$. The endpoints are some t_1 and t_2. We know that "physics" will result in some definite solution for the path of motion which would be the various $q_i(t)$. It is certainly not obvious by just staring that the physical path $q_i(t)$ minimizes or maximizes this integral I.

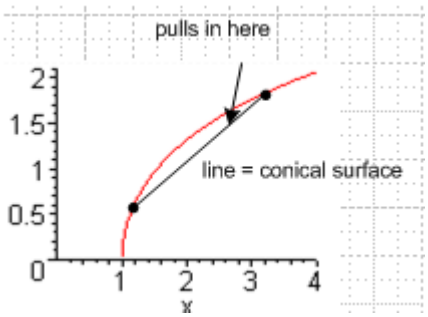
2.2 Calculus of Variations (31). Here we work in one dimension and replace the parameter t with parameter x, and the overdot • means d/dx. We start with the very general form (2-3). We imagine some parameterization of the alternate paths y(t) as in (2-4) where we add parameter α. Then by "doing the math", we find that $\delta J = 0$ is the same as saying $[\, \partial f/\partial y - d/dx(df/d\dot{y}) \,] = 0$ which is (2-11) and which has our familiar Lagrange form if x were time t and if f were L. A key step is the parts integration in 2-8 where the "parts" vanish because we assume the endpoints are fixed so the derivative $dy/d\alpha = 0$ at the endpoints. The picture on page 32 shows two possible paths in some sense separated by "α" but this separation goes to 0 at the end points.

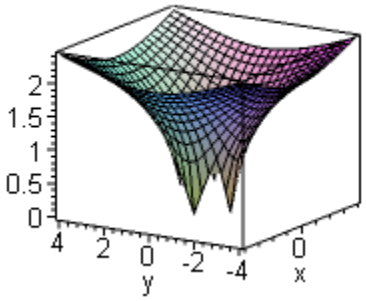
At this point we pause to do three simple sample problems.

In the first p34, he shows that y = mx+b is the form of the path that minimizes our integral J which is set up to be the distance along an *arbitrary* path connecting two points. The solution method is extremely

obscure but it works! We just turn the little crank. The path length is ds = dx $\sqrt{1 + \dot{y}^2}$ and this is what we integrate to get integral J.

The second problem asks what curve you would use to minimize a surface of revolution's area. The curve which defines the area has two endpoints. You might think that a conical surface would be the answer, since this would at least minimize the distance between the two endpoints, but that is of course not the answer because the more distant points generate more area. We write the area as an integral with the single parameter x, and then it is in the form (2-3) for a J integral, and we turn the crank. The solution curve is an inverse hyperbolic cosine, the solution has the form y = a $ch^{-1}(x/a)$ + b. We have the two conditions that $y(x_1) = y_1$ and $y(x_2) = y_2$ and these determine a and b. The result looks a little like what he has plotted, but is more like this: (solid sym axis is vertical graph axis)



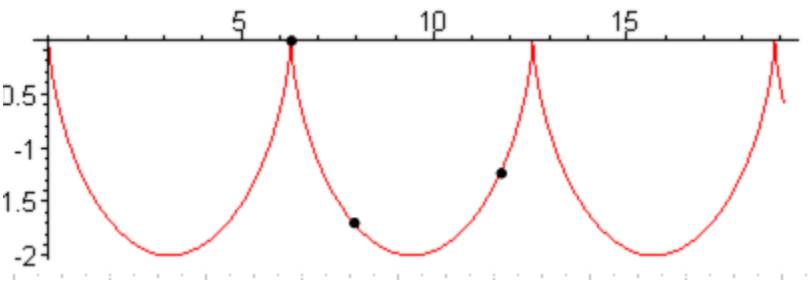


You can see that the shape pulls the surface in more quickly than a conical surface, to start cutting down on area! But too much pulling would increase the horizontal component of the surface.

This problem is interesting because I don't know how else you would solve it! There are an infinite number of ways to parameterize a curve, you would not likely hit upon the right one. You might do a polynomial and then recognize the series for the ch function in your result from regular calculus where you just try to minimize area. Again, if we write the area as integral J, we know how to make $\delta J = 0$ with our Lagrange-like equation. It seems clear that this must be a minimum from the problem's situation.

The third problem is the **brachistochrone** problem, the word from Gk meaning "shortest time". [10.30.16: My description here is **wrong**, I have now lots of stuff on this problem in a separate folder in Mechanics. Bead starts at origin and slides to some ending point on the leftmost down-hump. Point A is always the origin, and point B is some point on the first down-hump.] The problem is this: first pick two points A and B in 2-space like the two dots below. Then imagine a frictionless rigid wire which constrains a bead sliding on it and describe this wire by a red curve. The problem is then to find the shape of the curve that minimizes the *time* for the bead to slide from A to B. The integral is of dt = ds/v and our author leaves the solution to the problems, but we know the answer is that the curve is a cycloid which starts at a cusp and passes through the second point:

```
plot([t-sin(t), -1 + cos(t), t=0..20]);          // How you use Maple to do a parametric curve!
```



We start at the upper dot shown at (x,y) = (6.3) on the horizontal axis, and the solution curve heads straight down, then bends off toward the destination point. Depending on where the destination is, we might have a minimum in the curve or we might not. Again, how would you find this solution by some other method? [sorry about errors here]

Summary of this section: we defined $J = \int dx\, f(y, \dot{y}; x)$ and we showed that a Lagrange-like equation causes $\delta J = 0$. We did three examples in which the parameter was a geometric "x", never time. In these examples, we found the curve y(x) that caused $\delta J = 0$, and in all three cases this represented a minimum of J. We minimized a distance, then an area, and then a time.

2.3 Derivation of Lagrange's Equations from Hamilton's Principle (36). We first extend f to multiple variables as in (2-12), and we imagine the role of our single parameter α as shown in (2-13). We do the exact same math and we arrive at the generalized multi-variable result (2-16) which is again "Lagrange like". Again, we assume endpoints are fixed in the nD space and the solution is a set of $y_i(x)$. In the function f we start with, we assume that the y_i are independent, otherwise the result does not follow. The Lagrange-like equations are called **the Euler-Lagrange equations**.

Now we just make the associations shown below (2-2') [p 38] where parameter x becomes t, and we end up with exactly the "Lagrange equations" of mechanics. So this is just a special case of the E-L equations.

Conclusion: We have derived the Lagrange equations of motion from Hamilton's Principle of having $\delta I = 0$ where I is the integral of the Lagrangian over an arbitrary path. You could of course have done this derivation in the other direction. The two are equivalent. But in our examples, we learned some things outside this mechanics bailiwick. [See Lag doc Appendix D for detail of this section.]

2.4 Extension of Hamilton's Principle to more general systems. (38) Above we got our result (Hamilton's Principle $\Leftrightarrow$ Lagrange equations) in terms of the Lagrangian L = T-V, but this assumes the existence of a potential V.

In the first logic here, we consider using T+W instead of L = T-V inside the integral $\delta I = 0$. W has the unusual definition 2-18 which has dimensions of work, but is not the usual differential work. When we think about δW, we are supposed to think of the forces as constant, and we are varying the path δr or δq. So δW is the differential "virtual" work associated with varying the path at some point on the path. Then $\delta W \sim F\, \delta r$ and we get to (2-19). Hamilton now says integral of $\delta T + \delta W$ is zero.

Hold on this for a second, and consider this side alley: If there exists a potential V, we can write Q = dV/dq as earlier, add a zero term, and we are at the bottom of page 39. In this case, we end up with our previous $\delta I(L)=0$ version. So here we have just shown that if you put T+W inside, or L inside, you get the same thing, for the case of a conservative force. You would put the T+W inside in the case that you had a

non-potential-derived force. In both cases we assume holonomic constraints so we can say the δq_i are independent.

But, if there is no V, you can probably still use the T+W form! So again starting from the T+W form, he derives (2-21) which is the same as the non-potential result (1-50) we got back on page 18. So, the T+W gives this result as well. Holonomic means the q_i are independent. We are now done thinking about non-potential-derived forces and the T+W method.

Next, suppose the n q_i are not independent because we have m **differential constraints** of the form (2-22). On page 32 top he shows that a holonomic constraint implies a differential constraint, but not vice versa. So the case of m differential constraints is a broader set of problems which includes holonomic constraints.

Comment Added: In the previous chapter, we had N particles so 3N variables. There were N $\mathbf{r}_i$ vectors. In that chapter we managed to cut this down to n = 3N-k independent generlized variables q_i because we had k holonomic constraints of the form 1-35. In the current chapter section, there is no mention of the $\mathbf{r}_i$ anymore. We start off here assuming we have n coordinates q_i. Perhaps we got there from the same starting system and a set of holonomic constraints. However we got to the n q_i, we NOW want to add to the soup a new set of m non-holonoic constraints as in 2-23. This means that only n-m of the q_i are independent, and we are going to pick the *first* n-m of the q_i as the independent q_i.

If we are doing our virtual differentials, we then get m constraints (2-23). The next 4 equations are obvious, and note that we have introduced m "parameters" called λ_ℓ, one for each of our m differential constraints. They are arbitrary so far. But then in (2-27) we write out "the last m expressions" which appear inside (2-26), and we force them to be zero. But then (2-27) is m equations in the m unknowns λ_ℓ, so we know there will be a set of values for the m λ_ℓ which will allow these m expressions to be zero. Notice that the k sum in (2-26) goes all the way up to n, but because of the 0's in (2-27) we can now run this sum up to n-m, which gives us (2-28). This equation involves only n-m of the q_i and we can regard these as independent since we had m constraints, so then we set the inner guts of (2-28) to 0 which gives (2-29), for k = 1,2...n-m. But we already had these being 0 for the last m in (2-27), so we can say now that all n of the factors are 0 and this is what (2-30) says. *So this is the thing that replaces the simpler result (1-53) on page 18 if we have differential non-holonomic constraints.* We have this extra stuff in the equation involving the **Lagrange multipliers** λ_ℓ and the constants $a_{\ell k}$ from the differential constraints.

We can think then of (2-30) as being n equations with n+m unknowns, which are the n q_i and the m λ_ℓ. But the constraint equations written as (2-31) gives us m more equations, so we then have n+m equations in n+m unknowns.

The next paragraph (last on page 42) gives a way to interpret the term on the RHS of (2-30). Remember what we did. We usually think of our set of $\{q_i\}$ as being independent [and maybe we had a *larger* set of N $\{\mathbf{r}_i\}$ with a set of 3N-n holonomic constraints which brought is down to the set of n $\{q_i\}$]. But *now* we are saying that *in* our set $\{q_i\}$, only n-m of them are independent due to our m differential constraints. This then gives rise to the RHS of (2-30). If all the $\{q_i\}$ were independent, the RHS of (2-30) would be 0 for k = 1..n. So we can think of the RHS as somehow "arising from the non-holo constraints".

Now before continuing on this line, we need to go back to the logic leading up to equation (1-50) on page 18. In deriving this equation, we assumed on page 15 (1-40) that the forces of constraint "did no virtual work", and this limited us to a set of rigid body problems with no friction. Suppose we had decided to maintain the constraint forces in our development there (these forces were there called f_i). Then on the RHS of (1-50) we would have ended up with $Q_j + Q_j'$ where the $Q_j' = \Sigma_i\, f_i\, \partial r_i/\partial q_i$. Then in going from (1-50) to (1-53), we would have "removed" Q_j from the picture, but the Q_j' would have remained on the

RHS of (1.53). But looking at the RHS of (2-30), we see that we can then identify $Q_j' = \Sigma_\ell \lambda_\ell a_{\ell k}$. So we can interpret this term as the generalized constraint forces! [A+]

So in this kind of solution, unlike our earlier situation, here we will find our solutions q_i and we will also find the constraint forces of the problem. That is the n+m unknowns!

So when does one use this Lagrange multiplier / differential constraint method? (1) Maybe it is "inconvenient" to reduce your set of $\{q_i\}$ to a smaller set of those that are just independent ; (2) you want to know what the forces of constraint are; I would then add: (3) you have a differential constraint.

[Note added 10.30.16: Most of the above exists in Lag doc App D.]

Now we come to the **example** of the **hoop rolling down a plane without slipping**! There is a lot that we need to say about this seemingly simple example.

(1) This is a problem that involves friction! The constraint here does work on the hoop as it rolls, unlike the case where a hoop slides down a ramp with no friction. The friction causes the hoop to rotate and build up angular energy. Thus, we cannot use (1-53) to solve this equation if we think of the set $\{q_i\}$ containing two variables x and θ. We cannot use (1-53) because this current problem violates the assumptions about the constraint forces that went into the derivation of (1-53).

(2) The hoop has moment of inertia Mr^2, it is not a disk, hence T is as shown on page 43.

(3) We could use $rd\theta = dx$ to say that $r\theta = x$, and then we could reduce the problem to just one q_i and then we *could* use (1-53) for just one variable. But here we pretend this is "inconvenient to do" and we treat $rd\theta = dx$ as the differential constraint (r) dθ + (-1) dx = 0 so we get the two a's top of page 44. So here we have n = 2 variables $q_i = \{x,\theta\}$ and we have m=1 differential constraints. We will thus have only one λ. Our E-L equations (2-30) then become (2-34) and (2-35). This leads to a complete solution of the problem as shown for our n+m = 3 unknowns which are θ, x and λ. The result is fascinating. If we look at the x equation, we can regard (-1) λ as a constraint force acting on this variable x, and we see that the constraint pushes back against the hoop's center of mass, causing it to roll down the plane at only half the frictionless rate we would have gotten. From the θ point of view, the constraint force is (r)λ and this force makes a torque which rotates the hoop, whereas in the frictionless problem the hoop would not have an angular acceleration.

Note added 9.30.16. I am now confused about the origin of equation (2-26), more specifically, about the use of (2-20') in its derivation. As he says, (2-20') applies to a conservative system with a potential included in L = T+V. But with non-holonomic constraints, are things still conservative? The equation (2-20) differs from (2-20') and I don't follow. For a conservative system, I agree that the (...) in (2-20') is zero since this is just Lagrange's equations. Maybe he has mixed holo and non-holo constraints here? I'm sorry, the logic flow is lacking here, so I would like to see another source on this subject. [Lag App D]

Poole: The presentation is different here, and they quote a 1996 paper on the subject and bring in semi-holonomic constraints and it is all very ugly. But at the start they do make the claim that the Lagrange equations under the integral are valid with any constraints and that the holo constraint assumption is only needed to claim that the δq_i are independent. So with nonholo they won't be independent, but the equation (2-20') is still valid since its same derivation works (starting with Hamilton's principle for L). This derivation unfortunately is presented with other variable names on page 37. At the top of that page it is stated that the y_i (the q_i) are independent, but as of (2-15) this fact has not yet been used. So (2-15) is valid even if the y_i are NOT independent, and changing names you then get (2-20') where the δq_i now are NOT independent. Then we do the λ_i "trick": You can add the λ_i term in (2-26) "for free" since that

term is 0. Then you can select the λ_i to kill off m of the brackets in Σ_k for m dependent δq_i terms, and then the rest of the terms are independent, and THEN you get the desired result (2-27) if you have the right λ_i. and you have now derived equation (2-30) which is my goal here. Then the claim that you can interpret $\Sigma_i \lambda_i a_{ik} = Q'_k$ to be forces which would simulate the non-holo constraints!

On page 47 of Poole it is said "for the simplifying assuption that $\lambda_i = \lambda_i(t)$, so they add least mention that detail. I tried to find the 2nd Ed but could not.

I have now written this up in Lagrange Appendix D, I think it is now OK.

2.5 Advantages of a variational principle formulation (44). Claim is that the variational method works best when you do have independent coordinates. It is "elegant" because a one-liner contains all the physics. Also, L is a Lorentz scalar, so method is fully frame independent.

The next topic here is not as dramatic to me as it is to the author. I know that the same differential equations govern mechanics and E&M and other things, so I am used to applying the math methods of one field in another. Here he makes this comment for the Lagrangian approach in general. His example is a very good one, however, of a set of 3 (in this case) fully interlinked RLC circuits each with a voltage driver. It is not at all obvious to me where this Lagrangian comes from! The energy stored in an inductor is $\frac{1}{2}LI^2$ and $\frac{1}{2}CV^2$ for a cap, so that is not where these terms are coming from. There is no real connection between L = T+V here. Rather, we just have a candidate proposed L which gives the right equations so it must be an acceptable L.

Let's write Kirchhoff for the top loop (#1):

$$-E_1 + I_1R_1 + Q_1/C + \dot{I}_1L_1 + M_{12}\dot{I}_2 + M_{13}\dot{I}_3 = 0$$

Now apply d_t to this to get

$$\dot{I}_1R_1 + I_1/C + \ddot{I}_1L_1 + M_{12}\ddot{I}_2 + M_{13}\ddot{I}_3 = \dot{E}_1$$

Rearrange terms

$$\ddot{I}_1L_1 + (M_{12}\ddot{I}_2 + M_{13}\ddot{I}_3) + \dot{I}_1R_1 + I_1/C = \dot{E}_1$$

$$\ddot{I}_1L_1 + \Sigma_{k\neq 1} M_{1k}\ddot{I}_k + \dot{I}_1R_1 + I_1/C = \dot{E}_1$$

and this then is where Gold's 2-39 comes from. The equations for loops 2 and 3 have the same form, and then if we treat these as E-L equations, we can work backwards somehow and get his L. The model for the Lagrange equatinos is p 22α which includes the dissipative $\mathcal{F}$ term which we know is 2-38, the power burned in the resistors.

I have never seen this example anywhere! It is not in any of my E&M books, and it is hard to find on the web. So it makes his point quite well. The current I_i is the generalized coordinate q_i and the velocity $\dot{q}_i$ is $\dot{I}_i$. In his Lagrangian, the constants L and M are like "masses", while 1/C is a "spring constant" and the $\dot{E}_i$ are like F = qE = I $\dot{E}_i$, so interpret $\dot{E}_i$ as your "force" .

He then says there is a more subtle point that the variational idea when transferred to some different type of system implies a structural analogy with your original system. Thus it is that you might "carry over" the idea of quantizing a particle to the idea of quantizing a field. QED was quite new when Goldstein wrote this book [1950], at least the Feynman version which I think was 1949.

2.6 Conservation theorems and symmetry properties (47) You can learn a lot about your problem even if you cannot do the full double integration (F = ma) to solve it exactly. In general, if we have n variables q_i, we will have n 2nd order differential equations, each having two constants for 2n required constants, and these are provided for example by the n q_i and the n $\dot{q}_i$ initial conditions. Gold does not mention that these in general are *coupled* 2nd order ODE's.

If we define the n **generalized momenta** as $\partial L/\partial \dot{q}_i \equiv p_i$, then the L equation says $\dot{p}_i = \partial L/\partial q_i$. For a problem with no q_i dependence in V and the usual T, we find that $p_i = m\dot{q}_i = mv_i$ which is the familiar result. But for other Lagrangians like our E&M one, we get an extra term eA_i/c as shown in (2-42), and this can be interpreted as the field momentum. Since our generalized force is $Q_i = -\partial V/dq_i$, we see that we have a **"generalized Newton's law"** which is $\dot{p}_i = Q_i$ which involves the generalized force and the generalized momentum. [this is in Lag App D]

If a certain coordinate q_i does not appear in L, it is called **cyclic**. In this case, $\dot{p}_i = \partial L/\partial q_i = 0$ and the corresponding momentum is seen to be conserved. G comments on Routh's Method, but we are going to do that later in the book.

Our author then does two very well presented sections, taken right from The Book.

Conservation of linear momentum: In the first, we set up dq_i as apropo for a coordinate translation in some direction $\hat{\mathbf{n}}$ as shown in the page 50 picture. He shows first that the generalized momentum p_i is the component of $m\mathbf{v}_i$ in the $\hat{\mathbf{n}}$ direction (top p 51). Notice that he is considering a system of many particles, and he assumes that $d\mathbf{r}_i = dq_j\hat{\mathbf{n}}$ for ALL of the particles i, where dq_j is a change in only one of the generalized coordinates which we assume somehow causes this uniform translation of all the Cartesian coordinates. He then says that, if this q_i is cyclic and so does not appear in L, then we know $\dot{p}_i = 0$. So we then associate in this case conservation of linear momentum in the n direction with this cyclic q_i. Because L is cyclic in q_i , p_i is conserved. This seems pretty clear even without his pictures.

Conservation of angular momentum: He then repeats this discussion for angular momentum. Everything is the same but now the picture is shown on page 52 and **n** is now the axis of rotation. In this situation, our selected q_j causes the change in all the $\mathbf{r}_i$ as shown (they rotate about **n**). In this case, he first shows that $Q_j = \mathbf{n}\bullet\mathbf{N}$ where **N** is the total torque on the system. So now the generalized force is the component of torque along the **n** direction. He goes on to show that $p_j = \mathbf{n}\bullet\mathbf{L}$ and we then have $\dot{p}_j = Q_j$ saying that torque drives ang mom in the usual way. And if L is cyclic in q_j , then $\dot{p}_j = 0$ and our projection of ang mom is conserved.

Everything in this section is done with very simple math and leans on results from page 17 having to do with the relationship between the $\mathbf{r}_i$ and the q_j. Gold notes that conservation of p and L are associated with translational and rotational symmetries.

New topic starts on page 53. If we assume that L has no *explicit* time dependence, then he claims this is the same as time-independent constraints, scleronomous p 23. As a counter example, on page 55 we have a constraint which is moving in time, so perhaps you would have some V(t) and hence L(t).

Assuming no explicit time dependence, *and* assuming no velocities in V (in order to use the Lagrange equation in L), author shows that a certain quantity $H \equiv \Sigma_i \dot{q}_i p_i - L$ is conserved, $dH/dt = 0$.

So far we don't know much about what H is. But, if T is homogeneous of degree 2 in the $\dot{q}_i$ [which author identifies with no explicit t dependence, see p 23], then we can use Euler's theorem on T as shown in (2-51), and we quickly end up showing that in this case, H = T+V = total energy.

The section ends with a subtle issue of whether or not the work done by forces of constraint is included in this H. The answer is that it is not, but with time-independent constraints no such work is done anyway. The V that appears in the Lagrangian formulation only involves the "applied" forces. A situation where the constraint is time dependent is shown page 55 where the path of constraint is itself moving. So our actions in this last section would not apply to this situation.

He has not yet referred to this H thing as a "Hamiltonian".

Chapter ends as usual with some annotated references and some problems, all of which I have at least read through, but I did none of the problems. [Have now down the brach thing.]

Note Added: In this section, H is defined as $\Sigma p\dot{q} - L$ and is given no name. On page 54 if we assume conservative systems (meaning V independent of velocity), we have $p_i = \partial T/\partial \dot{q}_i$ [a key fact! If V is function of velocity, all you know is $p_i = \partial L/\partial \dot{q}_i$] . We then have

$\Sigma p_i \dot{q}_i = \Sigma \dot{q}_i \partial T/\partial \dot{q}_i = 2T$ // By a tricky use of "Euler's Theorem"

Let's pause to look at this theorem. Suppose $f(x_i)$ is such that $f(\lambda x_i) = \lambda^n f(x_i)$. For example, suppose you had a function $f(x_i) = f(x_1, x_2) = A_{11}x_1^2 + A_{12}x_1x_2 + A_{13}x_2^2$. Then $f(\lambda x_1, \lambda x_2) = \lambda^2 f(x_1, x_2)$ and this example is then "homogeneous of degree 2". Now, on page 23 G shows that if you express T of a system in any generalized coordinates q_i, it will be "quadratic" in this n=2 sense as long as the equations of transformations $\mathbf{r} = \mathbf{r}(q_i,t)$ are not functions of t (scleronomous). Therefore, the above "tricky use" equation is true for scleronomous generalized coordinates.

Now with the above in place, we have $H = \Sigma p\dot{q} - L = (2T) - (T-V) = T + V$ = total energy!! So, the association of H with total energy only applies when (1) V is conservative; (2) transformation between physical and generalized coordinates is time-independent; (3) there is an implicit third condition to worry about here, namely, that any possible constraints be holonomic. Only then does the set of generalized coordinates q_i contain 3N-k *independent* coordinates q_i. This is needed to make Lagrange equations 1-50 be valid without Lagrange multiplier modification. Presumably, if we have some kind of dissipative constraints, they are blocked by condition 1 above where the dissipation is modeled inside V.

So, exactly when can we say that H = T+V is the conserved total energy of a system? I think if the three conditions above are met, this will be true.

The example on page 55 violates our second condition because $\mathbf{r} = \mathbf{r}(q_i,t)$ is a function of time. A rotating frame of reference would be another example I think.

If we are given a Hamiltonian which is obviously in the form T+V with a conservative potential V and expressed in coordinates which are scleronomously related to physical Cartesian coordinates and we have no non-holonomic constraints, I think we can conclude that H = E = conserved total energy of the system. // review to here on 10.2.16, and again 10.30.16

Comment: There is a ton of advanced Lagrange mechanics stuffed into the first 50 pages of this book! And there are many tons lying ahead, book is 372 pages. I read this book part in 2009 then rest in 2011.

****

Chapter 3: The 2-body problem with a central force (58)

****

3.1 Reduction to the equivalent 1-body problem (58) We start off assuming a quite general Lagrangian of the form 3-1 where the potential might contain higher order derivatives of r, V(**r**,$\dot{\mathbf{r}}$, ...). **R** is the $\mathbf{r}_{cms}$. The potential is assumed to be a function only of the *relative* **r** and its derivatives. Moving to the CMS rest frame and defining the famous **reduced mass** μ, we find that this limited 2-body problem **reduces** to a simple equivalent one-body problem for particle of mass μ and we define an origin at **r** = 0.

Note added 10.31.16. (1) Why does spherical symmetry of V(r) in the equivalent single-particle problem (particle goes around origin somehow) imply **L** conserved? Answer: since **F** = -∇V, the force acting on the particle is in the $\hat{\mathbf{r}}$ direction. Torque on the particle is **r** x **F** which is then 0, hence **L** conserved in full 3D. (2) Why is motion planar? We know **L** = **r** x **p**. Then **L** • **r** = **r** • **r** x **p** = **r x r** • **p** = **0** • **p** = 0. Since this is true for any position **r**, **L** is normal to the plane of motion. Thus the plane of motion is the plane that is perp to **L**. Since **L** = constant (conserved, no torque), this plane never changes.

3.2 Central Force Problems (59). We now *further assume* that V has the simple form V(r) which means the force is along $\hat{\mathrm{r}}$. Since the problem is now spherically symmetric, any angle you come up with about an axis through the origin will be a cyclic coordinate. We know that **L** is conserved as a 3-vector. A simple argument shows that any motion must lie in a plane: for each **r** on the orbit, we know that **r•L** = 0, which means all such **r** lie in a plane perp to **L**. You might as well select a z-axis to line up with **L**, and then your problem has only two variables we can call **r and θ** which described the planar motion.

We then write the Lagrangian as in 3-6 (T now expressed in polar coordinates r,θ), see pencil notes bottom of page for v^2. Since θ does not appear in L (as predicted, θ is cyclic), we know that p_θ is conserved. But $p_\theta = mr^2\dot{\theta}$, and of course this "canonical momentum" is really ℓ about the **L** axis, so we write $\ell = mr^2\dot{\theta}$ which we regard as a "first integral" of one of the two Lagrange equations (that is to say, 1-53 for $q_i = \theta$). Goldstein then shows that $r^2\dot{\theta}/2 = dA/dt$, so conservation of ℓ is really the same as the notion of **constant areal velocity** -- equal area is swept out in equal time. This **Kepler Law #2** is true for an *arbitrary* central potential V(r), not just inverse square.

G then writes down the second Lagrange equation (the one in r) in 3-10, defines f(r) in the obvious way. He replaces $\dot{\theta}$ using the previous "first integral" to get 3-12 which involves $\ddot{r}$, $1/r^3$ and f(r).

Meanwhile, we know the T+V is the conserved energy so we write 3-13 and define this energy as E. Again we replace $\dot{\theta}$ and get the form 3-15 for E which now involves $\dot{r}$, $1/r^2$ and V(r). Web says that an archaic term for integration is **quadrature**, derived from an old idea of finding a square whose area was equal to the area under a curve. So these first and second integrations are called "quadratures" by our archaic Goldstein. Equation 3-15 can be formally integrated to find t(r) as shown in 3-18, from which you could presumably obtain the inverse r(t). Notice how the "first integral" got rid of $\ddot{r}$ which was sitting in Lagrange equation 3-10. In similar fashion, we integrate 3-8 to get 3-20 to give θ(t).

So, embodied in 3-18 and 3-20 we have formal solutions to the arbitrary central force V(r) two-body problem! Given any V(r), select a value for E and $\boldsymbol{\ell}$ and do a numerical integration to find your solution.

Nice comment is made here about the four integration constants. It is good to have E and $\boldsymbol{\ell}$ be among these constants, since E and $\boldsymbol{\ell}$ survive the transition to quantum mechanics from classical mechanics. Nothing is yet said about the signs of these two numbers. Other int constants are r_0 and θ_0.

Note added 10.31.16. I think this means the following. If you were to specify E, $\boldsymbol{\ell}$, r_0 and θ_0, then r(t) and θ(t) are fully determined. So pick a point r_0,θ_0 in the plane, specify E and $\boldsymbol{\ell}$, and there is only one orbit that works, only one possible **r**(t) for that particle, but there must be a ± hiding somewhere. I guess it would be the sign of the square root in (3-16) which determines the sign of $\dot{r}$.

Note added 10.31.16. The EL equations are 2nd order PDEs, so to get a solution you in effect have to do two levels of integration. Doing the first level ("first integrals") is easy for this type of problem, no work really needed. The first level integrals yielded equations which involve first derivatives of $\dot{r}$ and $\dot{\theta}$. It is the second level of integration to get r and θ that requires more work, and these are the formal integrals shown on page 63.

3.3 The equivalent one-dim problem, and the classification of orbits (63). What do we know without doing a lot of work? From 3-16 we know v_r as function of r, and from 3-21 we know v, so we can then compute $v_\theta = \sqrt{v^2 - v_r^2}$. Thus, even without solving the problem, we know all about the velocity vector **v** as a function of r (but we don't know r as a function of t without integrating.) Remember that **v** is a "first integral" and we have done that level of the problem.

Next: Looking at 3-12, we seem to have an F = ma problem in variable r, but there is an extra force term $\boldsymbol{\ell}^2/mr^3$, so we can think of this as some kind of extra fictitious force term in the one-dim problem and incorporate it into an effective total force f '(r) arising from an effective potential V'(r) as shown in 3-22 and 3-22'. This extra term is in fact the familiar centrifugal force mv^2/r where $v = v_\theta$. This reminds me of the "radial equation" for the hydrogen atom which is a "one dimensional problem" for a fixed $\boldsymbol{\ell}$.

First example: consider attractive inverse square force as shown. Plot the two terms which make up V' (dashed lines in Fig 3-3) and plot their sum (done for some $\boldsymbol{\ell} \neq 0$). For small r, the fictitious term wins and V' goes positive, then we have a nadir as shown and then V=0 at infinity. For $E \geq 0$ have unbound orbits (hyperbolas and parabola), for E<0 can have bound orbits (ellipses). Notion of point of closest approach r_1 for unbound orbits as shown page 66. Bound orbits (which become bound states in QM) might not be "closed", but move somehow between two circles in general case. So this example shows how you can "classify" the orbits for a known problem, and you get similar results for a similar V(r).

Second example: attractive inverse quartic force has plots for V and V' as top page 68. Now the negative $1/r^4$ term "wins" at small r and you get V' as shown. In this interesting case, for a given energy E you can have both a bound and a separate unbound orbit type.

Third example: attractive linear force (HO), and you get V and V' as shown bottom page 68 for this case. In both cases there are only bound orbits. These are ellipses since each direction has same HO frequency.

So the main point: even if you are unable or unwilling to solve a central force problem, you can probably make some clear statements about the nature of its orbits by plotting the effective potential V' of the equivalent one-dim problem.

Comments added 10.31.16. Goldstein is paying a lot of preparatory attention to quantum mechanics. In QM you have the same notion of potentials and turning points. Classically these turning points are "hard", but in QM they are soft, allowing penetration ("tunneling") of ψ through the walls of the potential, because as long as the walls are not infinitely steep, ψ has to be smooth at the wall and that means it can go through it a little bit. This has implications for semiconductors and many things. The classical r_0 and θ_0 become lost in QM since orbits are just ψ clouds, but E and ℓ remain. I can see how the case on p 68 with a central "well" implies possible "bound states" for elementary particles, and particle theory has been littered over decades with "potential models" for things, such as the Yukawa potential. Lots of history.

"Controversy" regarding the cover of the 3rd edition (11.12.16) Note page 66 Fig 3.7 which shows some generic classical orbit having two turning points. This Fig 3.7 is drawn differently in GP&S, so wit

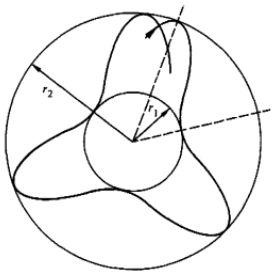


FIGURE 3.7 The nature of the orbits for bounded motion.

and this is the picture that appears on the cover. Here is a complaint from a book Galactic Dynamics

3.13 [2] Let $\mathbf{r}(s)$ be a plane curve depending on the parameter s. Then the **curvature** is

$$K = \frac{|\mathbf{r}' \times \mathbf{r}''|}{|\mathbf{r}'|^3}, \tag{3.331}$$

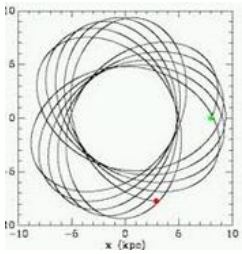
where $\mathbf{r}' \equiv \mathrm{d}\mathbf{r}/\mathrm{d}s$. The local radius of curvature is K^{-1}. Prove that the curvature of an orbit with energy E and angular momentum L in the spherical potential $\Phi(r)$ is

$$K = \frac{L\,\mathrm{d}\Phi/\mathrm{d}r}{2^{3/2} r[E - \Phi(r)]^{3/2}}. \tag{3.332}$$

Hence prove that no orbit in any spherical mass distribution can have an inflection point (in contrast to the cover illustration of Goldstein, Safko, & Poole 2002).

A 2003 letter (I have in PDF) comments on errors that have survived all 3 editions of Goldstein and which talk about this picture. The writer Martin Tiersten claims that you cannot have out-facing curvature in any "attractive" central force orbit. The corrected picture for the above appears on the cover of Marion and Thornton, competitors in this market, showing that you basically have a rotating ellipse situation for such an orbit,





His easy derivation is that $F_n = mv^2/R$ as the normal force to the center at any point on the orbit, where R is the instantaneous radius or curvature. Thus since F_n never changes sign, R can never change sign and must always be concave toward the origin! So Goldstein's original on p 66 is wrong as are the follow-on two pictures in 3rd Ed. Very good. The authors Safko and Poole respond apologetically.

BUT, I don't see why F_n must always point to the center. For a single-power potential like $V = k/r^n$ you would have $F_n = -nk/r^{n+1}$ so in that case Tiersten's comment applies. But if you have two terms with different powers such as $V = k/r - k'/r^3$ then $F_n = -kr + 3k'/r^4$ with $k,k' > 0$. In this case F_n can have either sign. This then is how Safko ended up stating things in the errata, from which I quote:

These corrections will make the 1st, 2th, and 3rd printings of this text the same as the 6th printing which is the last revised printing.

Corrections to the 6th printing are listed separately and will be revised as discovered.

page #	where	correction
Cover	Comment on the cover figure	The cover figure shows the motion of a particle around a central force source. There are two central forces acting; an attractive force proportional to $1/r^3$ and a repulsive force proportional to $1/r^4$. The figure was probably inspired by the hard core repulsive nuclear force. Another possibility is that the attractive force is proportional to $1/r^3$ and the repulsive force is proportional to $1/r^3$. The second possibility can be symbolicly integrated.

Here is another version of the comment with the above typo corrected,

page#	Correction
Cover	The cover picture shows the motion of a particle under central forces. There is a $1/r^3$ attractive and a $1/r^4$ repulsive force. or another possibility is an $1/r^2$ attractive and a $1/r^3$ repulsive force. As discussed in Section 3.6 the orbit does not close. Special thanks to Thomas Fischaleck and Ian Gatland for pointing this out to us.

So I think Tiersten's letter only applies to a purely-attractive central force where the sign cannot change. It is true though that Goldstein on page 66 *implies* that his picture applies to the simple $V = -k/r$ force shown on page 65. He is in a discussion of inverse square force when he draws his picture. So the authors clarify their covers with the above comments (bailing them out).

3.4 The virial theorem (69). This theorem applies to any set of particles having any total forces $\mathbf{F}_i$. It is not necessary even to have a potential. The only requirement is that the motion be bounded so that in effect after enough time the orbit repeats so that $G \equiv \Sigma_i \mathbf{p}_i \bullet \mathbf{r}_i$ at t = 0 repeats at some huge t = T. The

theorem states that <T> = - (1/2) < Σ_i $\mathbf{F}_i \bullet \mathbf{r}_i$> for time averages, as shown on p 70. The RHS here is called the "virial". Wiki: The word virial for the right-hand side of the equation derives from vis, the Latin word for "force" or "energy", and was given its technical definition by Rudolf Clausius in 1870.[1]

Nominative	vis	vires
Genitive	----	virium
Dative	----	viribus
Accusative	vim	vires/viris
Ablative	vi	viribus

force in singular, strength in plural

We start with some basic algebra and we quickly arrive at 3-26 which is the virial theorem for an arbitrary system of particles, where $\mathbf{F}_i$ are applied *and* constraint forces. The theorem gives a way to compute the time average (periodic or long term) of the kinetic energy T as a time average of the sum of dot products shown, the **virial of Clausius**. G says this is useful in both ideal and non-ideal gas physics where you do in fact have constraints, like walls of a container, but only gives us a reference.

Specialize to forces from a **potential V**, get 3-27. Of course here you cannot have constraint forces unless they are coming out of the potential somehow, like the U business earlier.

Specialize to **single particle**, get 3-28.

Specialize to **power-law** V(r), get 3-29.

Specialize to n = -2 for **inverse square force**, get 3-30, which is a famous result <T> = -1/2 <V>.

3.5 Differential equation for the orbit (71). Back to central force f(r) again. Can always replace time derivative with θ derivative as in 3-31. Take our earlier result 3-12 (from the r Euler-Lagrange equation, sub'd in), define u = 1/r, and end up with 3-34 which is an ODE for the orbit. It is second order in θ, t does not appear but u does, so you can solve this thing for θ(u) for example. You need $\boldsymbol{\ell}$, m and f(1/u) as inputs to this equation. People are usually more interested in the orbit than say r(t). If you arrange for a turning point to be at θ = 0, then orbit is symmetrical in θ.

If we rewrite 3-17 in terms of u, we can write θ(u) as an integral as in 3-37. This is in effect the solution of the little ODE, but we reuse earlier work to get this result. The book then gives us a list of values of "n" for which the integral can be done in terms of "simple functions", assuming a power law central potential. The simplest are arcsin and arcsinh integrals which he calls "**circular functions**". These arise when the radical R is a quadratic. But this only gives you cases n = -2 and -3. For n = -1, you are talking a log potential and nothing really works right, we are "out of the power bailiwick".

The next more complex class of cases that can be integrated are the **elliptical functions** which involve R as a quadratic radical as shown page 74. This brings in the cases n = 0 as the only new one. By doing little changes of variable, one can get n = -5 and -7, and then another change gives +5 and +3. So the bottom line is that if you insist on closed form circular or elliptic functions, you have to have n = 1,3,5 or you need -n = 0,2,3,4,5,7. Some fractional n lead also to elliptics.

Goldstein harps on this a bit because in 1950 you could not just throw an integral into "Maple" on your home PC, or do a simple numerical integration to find an answer. And as always, we like closed form results where possible.

Summary: Central potential context. We already had t = integral of some F(E,$\boldsymbol{\ell}$,r) in (3.18) to find t(r) Due to the simple first integral fact $\boldsymbol{\ell}$dt = mr^{2}dθ it is easy to replace "t" by "θ" in any ODE or integral. If you replace in this way t by θ, you end up with ODE's or integrals allowing you to compute θ(r) instead of

t(r), and θ(r) or r(θ) is the equation of an **orbit**. If you specialize to power law central potentials, you can ask for which powers n do the integrals generate known functions, and which do not. For some powers including inverse square n = -2 you get "circular functions" like sin or sh or their arcs, while for other powers you get elliptic functions, and yet others you get functions having no name. It is often convenient to use u ≡ 1/r in place of r when doing work in this area.

3.6 The Kepler problem: inverse square law force. (76) If only Kepler had a copy of this book in 1600. However, Leibnitz and Newton did not discover calculus until the 1670's, so Kepler might have been unable to read the book the way I can read it now. In this section are derived the remaining two of Kepler's laws, they just fall out from the simple theory like rain from the sky.

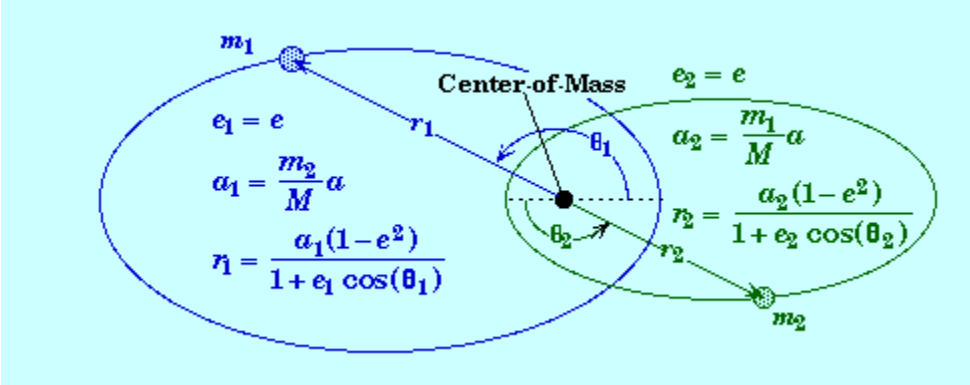
We assume $f = -k/r^2$ and solve the little orbit DE to find result 3-43 which says right off the bat: all orbits are conic sections, with some unknown integration constants. But then we use the integral we had earlier for θ and "do it again", this time obtaining 3-46. The added feature here is that we now have a formula for the orbit eccentricity ε as a function of E, ℓ, m and k (remember that m is the reduced mass). I think G has made a **sign error** in this last calculation but the error does not propagate into the answer. It probably involves a different branch of the square root or something like that. I could not find errata for this 1950 edition on the web, but learned that there is a third 2001 edition.

Note added 11.12.16. There *is* a sign error in (3-45) then then also in equation β. But this has no effect on the main result (3.46) because cos(θ-θ') is an even function of its argumenty. I wrote this up in Euler doc. I have corrected both sign errors in red in Goldstein copy.

Now that we have a formula for r(θ), namely 3-46, we could integrate 3-19 (top p78) to get θ(t), but this is not done in this section of the book. Here we are content with knowledge of the orbits.

So, we take 3-46 = 3-47 and discuss the nature of the orbits relative to the size of ε. This of course agrees with the *general* orbit discussion earlier for the central force problem. For a circular orbit, time averages equal quantities so we use the virial theorem to get 3-50, not that exciting, but an application of our little virial theorem of a previous section.

One point of confusion that caught me. On page 66 we see some orbits bounded by r_1 and r_2 circles. One could imagine having an inscribed ellipse and one would be tempted to say a = r_2 and b = r_1. One must remember, however, that the "r" in this picture is the "relative r" which is $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$. In the heavy sun with planet picture, this would be distance between planet and sun, and sun is at a focus, so here you see that in fact $r_1 + r_2 = 2a$. For a binary star where masses might be similar, the picture in the CMS frame of the resulting orbits is like this:



Again, we have $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$. We know the solution for r is "an ellipse", but when this is translated back into the orbits for r_1 and r_2 using page 58, we obtain a picture like that above for the CMS frame (non-reduced) problem. Each star does an ellipse of the same eccentricity, and the two stars are always on a line through the CMS as shown, and the two reach closest approach to the CMS at the same time. You can arrange for one ellipse to be entirely within the other.

Perhaps in the later editions this situation is examined in more detail. As you increase the mass of one of the binaries, it's ellipse gets smaller and closes down around the focus of the ellipse of the lighter mass. A nice simulation is given at http://ltc.cit.cornell.edu/courses/astro101/java/binary/binary.htm which shows the possible locations of the two ellipses and what happens as you increase mass. For equal masses, the two ellipses always overlap and are identical in shape and size.

Now back to the reduced mass or one-mass-is-huge view. We get a list of conics versus ε on page 78. then on page 79 we relate a, C and ε and we find that "a" is a function of E and not of $\boldsymbol{\ell}$. Finally, using that fact that the area sweep rate is constant and area of ellipse is πab, we derive an expression for the period of the orbit τ in 3-54. We find that $\tau^2 \propto a^3$ which is the main idea of Kepler Law #3. Again, m = μ and k = force constant. If we apply this to gravitational force and planet/sun, we get 3-56 which would be the actual Kepler result which says $\tau^2 = \text{const} * a^3$ with the *same* constant for all planets. Only the mass of the sun and G and a enter in this formula.

Update 11.14.16. A loose-bolt detail here. If you look at the ellipse equation (3.46) you see a leading constant $C = mk/\boldsymbol{\ell}^2$ which G fails to identify as "C". The usual ellipse equation is for r, not 1/r, and there we know that the leading constant is $a(1-\varepsilon^2)$. Thus we can identify $1/C = a(1-\varepsilon^2)$ which says $a = [C(1-\varepsilon^2)]^{-1}$. G eventually shows this fact in p 79 (3-51). Now since $b = \sqrt{a^2-c^2}$ and $c = \varepsilon a$ we know that $b = a\sqrt{1-\varepsilon^2}$. But we just showed that $\sqrt{1-\varepsilon^2} = \sqrt{1/(aC)} = a^{-1/2}\sqrt{\boldsymbol{\ell}^2/mk}$. This then leads to our desired result that $b = a^{1/2}\sqrt{\boldsymbol{\ell}^2/mk}$. The ellipse area is then $A = \pi ab = \pi a^{3/2}\sqrt{\boldsymbol{\ell}^2/mk}$ which later figures into the 3rd law. In fact the difference between a and a_1 is very small except for Jupiter. In any event, Kepler #3 as he saw it is (3-56).

Update 11.14.16. G starts with the simple 3-54 version of Kepler #3 for the 1-body problem. For the 2-body problem you replace mass with μ and assume sun dominates so you get (3-56). But in this formula, G does not point out that a still refers to the 1-body problem orbit, which reduces to a_1 and a_2 for the real orbits of the two objects as I show in binary doc. However, the key point is that Kepler is observing things basically from the frame of reference of the sun, so a is then the correct thing to use, not a_1.

Again, it is interesting to contemplate the history of mankind during which the motion of the planets was gradually understood, and moved from religious to scientific reality. Wiki notes that Kepler tried to fit his work in with biblical quotations and existing astrology, his time being 1610.

In the Bohr atom, we have roughly that $\mu = m_e$ and $k = e^2$ so in this classical atom, you ought to have elliptical motion for electrons in bound states. We always have $\varepsilon = 1$ if $\boldsymbol{\ell} = 0$ so S states are "circular" and higher $\boldsymbol{\ell}$ states are "elliptical". In classical mechanics, orbit is always in a plane, but in the quantum atom, the plane must be statistically distributed at all angles, giving the spherical S orbital. Would be interesting to read more on this subject -- connection between classical and quantum atoms.

Another comment on the sign error. GR7 has it right!

Note added 10.31.16. Sign Confusion page 77. In 1st edition the integral (3.45) has a plus sign, and in 3rd edition we also have a + sign,

$$\int \frac{dx}{\sqrt{\alpha+\beta x+\gamma x^2}} = \frac{1}{\sqrt{-\gamma}} \arccos -\frac{\beta+2\gamma x}{\sqrt{q}}, \qquad (3.51)$$

// wrong

What do GR7 have to say? From their page page 94 I quote:

Notation: $R = a + bx + cx^2, \quad \Delta = 4ac - b^2$

2.261[11] For $n = -1$

$$\int \frac{dx}{\sqrt{R}} = \frac{1}{\sqrt{c}} \ln\left(\frac{2\sqrt{cR}+2cx+b}{\sqrt{\Delta}}\right) \quad [c>0] \qquad \text{TI (127)}$$

$$= \frac{1}{\sqrt{c}} \operatorname{arcsinh}\left(\frac{2cx+b}{\sqrt{\Delta}}\right) \quad [c>0, \quad \Delta>0] \qquad \text{DW}$$

$$= \frac{1}{\sqrt{c}} \ln(2cx+b) \quad [c>0, \quad \Delta=0] \qquad \text{DW}$$

$$= \frac{-1}{\sqrt{-c}} \arcsin\left(\frac{2cx+b}{\sqrt{-\Delta}}\right) \quad [c<0, \quad \Delta<0] \qquad \text{TI (128)}$$

In our application, what GR7 call Δ is negative, so the last entry is relevant. Goldstein has q = -Δ. Now start with the above and use this rule

arcsin(α) + arccos(α) = π/2

Thus the above integral can be written,

$$I = \frac{-1}{\sqrt{-c}} \left[- \arccos\left(\frac{b+2cx}{\sqrt{q}}\right) + \pi/2 \right]$$

But you can drop the π/2 since this is an indefinite integral. We then have

$$I = \frac{1}{\sqrt{-c}} \arccos\left(\frac{b+2cx}{\sqrt{q}}\right)$$

Now use another rule which says

arccos(α) = π - arccos(-α)

Again drop the π and you end up with

$$I = -\frac{1}{\sqrt{-c}} \arccos\left(-\frac{b+2cx}{\sqrt{q}}\right)$$

right

If GR7 is correct, then both 1st and 3rd Ed are missing the leading overall minus sign. [correct] GR7 agrees with Schuam p 72.

Now let's assume for the moment that Goldstein is correct and that (3-44) is correct, which in 3rd and 1st is this

$$\theta = \theta' - \int \frac{du}{\sqrt{\frac{2mE}{l^2} + \frac{2mku}{l^2} - u^2}}, \qquad (3.50)$$

Now write

$q = (2mk/\ell^2)^2 \left(1 + 2E\ell^2/mk^2 \right)$ which I have verified on scratch.

Then the $\int$.... object in (3-44) is equal to this

$$I_G = + \frac{1}{\sqrt{-c}} \arccos\left(- \frac{b+2cx}{\sqrt{q}} \right) = \arccos\left(- \frac{b+2cx}{\sqrt{q}} \right) \qquad c = -1$$

$$\frac{b+2cx}{\sqrt{q}} = \left[(2mk/\ell^2) - 2u \right] / \left\{ (2mk/\ell^2) \sqrt{1+ 2E\ell^2/mk^2} \right\}$$

$$= \left[1 - 2u(\ell^2/2mk) \right] / \left\{ \sqrt{1+ 2E\ell^2/mk^2} \right\}$$

$$= \left[1 - u\ell^2/mk \right] / \left\{ \sqrt{1+ 2E\ell^2/mk^2} \right\}$$

and then

$$- \frac{b+2cx}{\sqrt{q}} = \left[u\ell^2/mk - 1 \right] / \left\{ \sqrt{1+ 2E\ell^2/mk^2} \right\}$$

This implies that

$\theta = \theta' - \left[\arccos\left(\left[1 - u\ell^2/mk \right] / \left\{ \sqrt{1+ 2E\ell^2/mk^2} \right\} \right) \right)$ wrong sign after θ'

This matches both editions, and I quote it from 3rd ed

$$\theta = \theta' - \arccos \frac{\frac{l^2 u}{mk} - 1}{\sqrt{1 + \frac{2El^2}{mk^2}}}. \qquad (3.54)$$

So everybody is in agreement except GR7 !! All cleared up in Euler doc!

3.7 Central Force Scattering (81). History puts the Kepler application first (1620). The Bohr atom came later in 1913-15 (Livesey p 158), for inverse square central forces. Here we do the classical inverse square force *scattering* problem, and this is the Rutherford discovery of 1911-13.

We shall capture previous work using 3-64 for the "orbit". After all, the trajectory of a scattered particle is in fact going to be a piece of a hyperbolic orbit, and we have ε taken from 3-46 and restated in 3-65, where we replace angular momentum $\boldsymbol{\ell}$ by impact parameter s according to 3-59. Here are details.

[this section was rewritten 11.15.16]

[81] Lots of words about scattering that I am OK with. A ring of solid angle dΩ is defined, and the cross section concept is defined in 3-57.

[82] A decent scattering picture and definition of the scattering angle Θ. Simple connection between impact parameter s and angular momentum $\boldsymbol{\ell}$ in 3.59, based on L = rxp and E as K.E. Then eq 3.60 describes what I wrote in as dN with ds > 0. It happens that in this case, dΘ < 0 just from the way things work (see picture), so we have a strange-looking minus sign. Solve for σ to get 3-62. Notice that the beam intensity I cancels out. Cross section has dimensions of area and we have the same minus sign.

[83] Projectile has Z', target has Z, so we get a value for force constant k as shown (3.36). The force is assumed repulsive here!! Look at 3-41 to see that this means k > 0 so V = |k|/r repulsive. He then directly copies 3-46 with this value of k and it has the standard polar variables r and θ as always. He can replace $\boldsymbol{\ell}$ with s from 3-59 so ε is as shown in 3-65. To keep r positive, we must have 3-66.

Break time. Here I went off and wrote notes in hyperbola.doc. I wanted to know exactly what a hyperbola's two branches look like in polar coordinates. It is a rather long story actually. I only worked with horizontal hyperbolas. You have to take a different path depending on whether you want the origin at the left or at the right focus. Here are the results:

Origin at left focus:

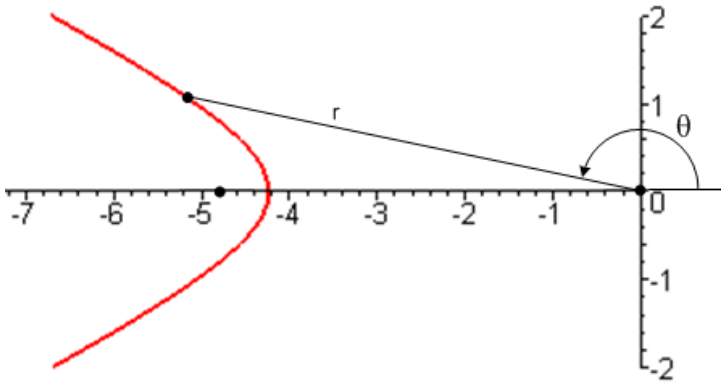
r_1	$r_+ = \dfrac{a(\varepsilon^2-1)}{\varepsilon\cos(\theta)+1}$	$\cos(\theta) > -1/\varepsilon$	makes the left branch
r_2	$r_- = \dfrac{a(\varepsilon^2-1)}{\varepsilon\cos(\theta)-1}$	$\cos(\theta) > +1/\varepsilon$	makes the right branch

Origin at right focus:

r_3	$r_+ = -\dfrac{a(\varepsilon^2-1)}{1+\varepsilon\cos(\theta)}$	$\cos(\theta) < -1/\varepsilon$	makes the left branch
r_4	$r_- = -\dfrac{a(\varepsilon^2-1)}{-1+\varepsilon\cos(\theta)}$	$\cos(\theta) < +1/\varepsilon$	makes the right branch

Notice that $r_3 = -r_1$ and $r_4 = -r_2$. Numerator always positive for hyperbola.

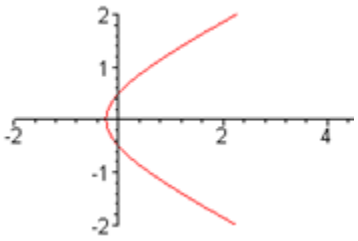
Goldstein's equation 3-64 fits the mold of r_3 above. Here is a sample of such a left branch.



(1)

You think of the hyperbola as the scattering path of a particle which is repelled by a fixed particle at the origin (the right side focus). Polar coordinate r is the distance from that origin to a point on the path, and that is the distance over which your repulsive potential operates.

Goldstein does not want to do this, but if you somehow had an attractive potential, the path would be better modelled by r_4 which has this curve



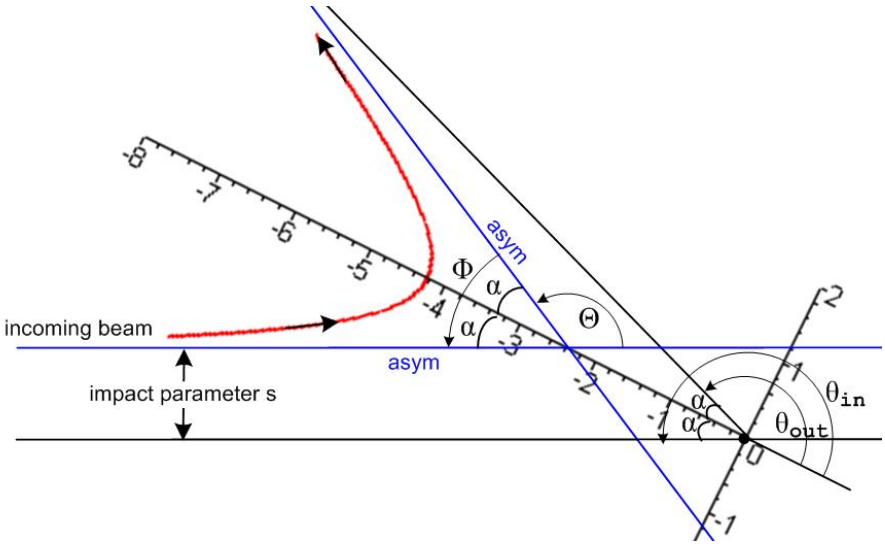
(2)

Here the attractive potential source is located at the origin which is still the "right focal point", and we just ignore the left branch here which is off the page to the left.

Goldstein Fig 3-15. This is a bit confusing for several reasons:
(1) his equation 3-64 directly implies my Fig 1 above where particle comes in from the left. But in this figure, he shows the particle path instead being on the right side.
(2) he does not show the location of the attractive potential, I added it. This is my (2) above.
(3) for some reason there are ink blobs on both his asymptotes which just adds confusion and Jim's book has this as well.

The third edition completely avoids this picture!

So I now take Fig (1) and rotate slightly it to get a scattering picture as follows:



I note that the slanted black line and the slanted asymptote are really parallel for a point on the hyerbola which is very far away. This picture shows many useful things.

(1) the scattering angle Θ is the same as in G's Fig 3-13, it is the standard scattering angle away from straight ahead. .

(2) For far-away incoming and outgoing hyberbola points I show θ_{in} and θ_{out} in official polar coordinates, same as in my earlier pix above.

(3) Angle Φ is then $\Phi \equiv \pi - \Theta = 2\alpha = \theta_{in} - \theta_{out}$

(4) It is true that

$$\cos(\theta_{out}) = \cos(\theta_{in}) = -1/\varepsilon$$

since in general $\cos(\theta) < -1/\varepsilon$ for r_3 but these are both extremal points on the hyperbola. I have marked θ_{in} and θ_{out} on Gold Fig 3-14.

(5) Also it is true that (just look at the above drawing),

$$\theta_{out} = \pi - \alpha = \pi - \Phi/2$$
$$\theta_{in} = \pi + \alpha = \pi - \Phi/2$$

(6) this math then follows at once (these two equations appear as page 83 A)

$$-1/\varepsilon = \cos\theta_{out} = \cos(\pi - \Phi/2) = -\cos(\Phi/2) \quad \Rightarrow \quad \cos(\Phi/2) = 1/\varepsilon$$

$$\sin(\Theta/2) = \sin([\pi - \Phi]/2) = \sin(\pi/2 - \Phi/2) = \cos(\Phi/2) = 1/\varepsilon$$

(7) then it is trivial that (this is equation B on page 83)

$\cot^2(\Theta/2) = \varepsilon^2 - 1$ // based on $\sin(\Theta/2) = 1/\varepsilon$ and thus $\csc(\Theta/2) = \varepsilon$

(8) But $\varepsilon^2 - 1$ is clearly the second term in the radical in 3-65, and so 3-67 follows. This gives the scattering angle Θ in terms of the impact parameter s. Note that $E = (1/2)\, mv_0^2 > 0$ when the particle is far away, and energy conservation means that E = T+V maintains this value on the entire path. When it gets close to the repulsive center, T slows down and V increaes so E is maintained. This is elastic scattering, energy in = energy out for scattered particle.

Finally we reach the end of page 83 !!

[84] Now rewrite 3-67 as page 84 A. Then compute $ds/d\Theta$ using 3-62 and this gives B and then finally simplifying 3-68. This is as he says the very famous Rutherford scattering formula with that inverse 4th power of the sine of the scattering angle. This peaks strongly when $\Theta \approx 0$ which corresponds to large s.

Note the cross section naming conventions,

	Goldstein		particle physics sources
cross section	$\sigma(\Theta)$	=	$d\sigma/d\Omega$

This G's equation C really says, in particle notation,

$$\sigma_{tot} = \int d\sigma = \int (d\sigma/d\Omega) d\Omega$$

G notes that the Rutherford cross section he gets agrees with the QM version non-rel, and I have done the relativstic calculation many times in Bjorken Drell (I think). For this kind of scattering, $\sigma_{tot} = \infty$ because there is scattering at any s you want. This is also a well-known fact.

This is the Rutherford classical result as derived in my Livesey book page 154 where $E = 1/2mv^2$. Rutherford did the theory in 1911, Geiger and Muller did the experiments in 1913, and this showed the existence of a very small nucleus and the planetary model of the atom (Bohr model) was born. Most beam particles go straight through, only those that come very close to a nucleus get deflected a lot, hence the $1/\sin^4(\Theta/2)$ factor. It showed that the $1/r^2$ force law was accurate down to something like 10^{-14} m. The beam was alpha particles created by a radon source.

The Livesey derivation assumes a massive target like a gold nucleus, so the CMS kinematics are very close to the lab kinematics. But Goldstein is now going to do this correctly in the next section.

Note retained from 10.12.08: On page 83 top G sets up the problem for a positively charged projectile (like an α particle) scattering off a positively charged center (like a nucleus). On that same page, he comments just in passing that if we had one object positive and the other negative -- the attractive case -- the picture on page 84 still applies, the only difference being that the scattering center is on the right side at F instead of on the left side at F'. So the orbit is unchanged. So we find the interesting result that the cross section is the same in either the attractive or repulsive case, something that is not obvious to me at first glance.

The third edition adds more stuff to this section discussing other potential shapes. In particular, if you have a combination of attractive and repulsive potentials you can get orbits which circle the target one or more times before leaving. The authors talk about rainbow and glory scattering in passing. I guess they felt this was the right place to add more content about more general scattering situations. There is always the issue of classical versus quantum analysis. I am happy without this extra stuff. I would like at some point to do the gravity assist analysis when an object takes energy off a moving planet. I don't think I have ever done that. Maybe it is trivial.

3.8 The Lab Frame (85).

[p 85]

New notes 11.16.16. I have to say that this is NOT a very clear Goldstein section. His notation is not very good, he assumes things that are not obvious at least to me. He just gives waypoints instead of a clean derivation of his results. I derived all these equations "at one time" but have no record of these derivations. Some equations don't really need a record, but others do.

On first reading, I am confused all three figures. I agree that in the CMS frame where $\mathbf{R} = 0$ all the time we can use results obtained earlier to relate the effective 1-body problem to the physical 2-body problem. In the 1-body problem the orbit of that body is a hyperbola (branch) and we then have these equations (taken from binary star doc). I follow G by priming the CMS vectors.

$$\mathbf{r} = \mathbf{r'}_1 - \mathbf{r'}_2 \qquad \mathbf{R} = (m_1\mathbf{r'}_1 + m_2\mathbf{r'}_2)/(m_1 + m_2) = 0$$
$$\mathbf{r'}_1 = (\mu/m_1)\,\mathbf{r} \qquad \mathbf{r'}_2 = -(\mu/m_2)\,\mathbf{r}$$

In the binary star situation, the ellipse shape was translated into the ellipse shape of the orbit of each particle because in r'_1,θ'_1 for example $\theta'_1 = \theta$ and $r'_1 = (\mu/m_1)r$. Similarly $\theta'_2 = \theta+\pi$ and $r'_2 = (\mu/m_2)r$ and this caused the orbits to have opposite orientations. The same thing happens here in the CMS frame. The 1-body hyperbolic orbit translates into a pair of "opposite" physical hyperbolic orbits and that is what is shown in Fig 3-17. Goldstein assumes the reader will figure out what I have just stated.

[p 86]

<u>Proof that CMS scattering angle is the same as the 1-body scattering angle</u>

The first puzzler is why the CMS scattering angle should be the same as the 1-body scattering angle. He draws pictures on top of p 86 and simply states that this is so, as if it were totally obvious. How can I prove this claim? In the CMS consider first for the 1-body position (think of ∞ as a very large but finite #)

$$\mathbf{r}_{in} \equiv \mathbf{r}(t=-\infty) \qquad \mathbf{r}_{out} \equiv \mathbf{r}(t=+\infty) \qquad \mathbf{r}_{in} \bullet \mathbf{r}_{out} = r_{in}r_{out}\cos\Theta$$

and this Θ would be the 1-body scattering angle. Now define a bunch more CMS objects,

$$\mathbf{r'}_{1in} \equiv \mathbf{r'}_1(t=-\infty) \qquad \mathbf{r'}_{1out} \equiv \mathbf{r'}_1(t=+\infty)$$

$\mathbf{r'_{2in}} \equiv \mathbf{r'_2}(t=-\infty)$ $\qquad$ $\mathbf{r'_{2out}} \equiv \mathbf{r'_2}(t=+\infty)$

$\mathbf{r'_{1in}} \bullet \mathbf{r'_{1out}} = r'_{1in} r'_{1out} \cos\theta'_{1s}$
$\mathbf{r'_{2in}} \bullet \mathbf{r'_{2out}} = r'_{2in} r'_{2out} \cos\theta'_{2s}$

In this last line, θ'_{1s} is the deflection or scattering angle for particle 1. I agree that since we are in the CMS frame, the angles θ'_{1s} and θ'_{2s} have a simple relationship. I have draw these angles on G's Fig 3-17 and it is clear that these angles must be the same. Since the CMS is at rest, if one particle goes off at some angle, the other has to go off at the same angle (each angle measured from particle's initial direction), otherwise the CMS would no longer be at rest after collision. So let's call this angle just θ'_s. Then we have

$\mathbf{r'_{1in}} \bullet \mathbf{r'_{1out}} = r'_{1in} r'_{1out} \cos\theta'_s$
$\mathbf{r'_{2in}} \bullet \mathbf{r'_{2out}} = r'_{2in} r'_{2out} \cos\theta'_s$ $\qquad$ (1)

But now use

$\mathbf{r'_{1in}} = (\mu/m_1)\, \mathbf{r_{in}}$ $\qquad$ $\mathbf{r'_{1out}} = (\mu/m_1)\, \mathbf{r_{out}}$
$\mathbf{r'_{2in}} = (\mu/m_2)\, \mathbf{r_{in}}$ $\qquad$ $\mathbf{r'_{2out}} = (\mu/m_2)\, \mathbf{r_{out}}$

and then

$$\mathbf{r'_{1in}} \bullet \mathbf{r'_{1out}} = (\mu/m_1)^2\, \mathbf{r_{in}} \bullet \mathbf{r_{out}} = (\mu/m_1)^2 r_{in} r_{out} \cos\Theta = [(\mu/m_1) r_{in}]\, [(\mu/m_1) r_{out}] \cos\Theta$$

$$= [r'_{1in}]\, [r'_{1out}] \cos\Theta \qquad (2)$$

Comparing (1) and (2) we conclude that $\theta'_s = \Theta$. The same result obtains using 2 instead of 1 subscript.

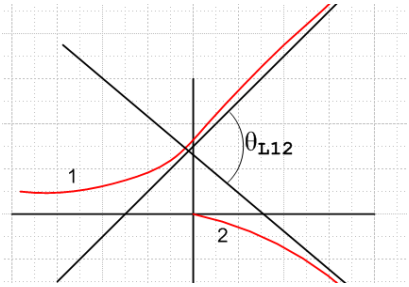
$$\mathbf{r'_{2in}} \bullet \mathbf{r'_{2out}} = (\mu/m_2)^2\, \mathbf{r_{in}} \bullet \mathbf{r_{out}} = (\mu/m_2)^2 r_{in} r_{out} \cos\Theta = [(\mu/m_2) r_{in}]\, [(\mu/m_2) r_{out}] \cos\Theta$$

$$= [r'_{2in}]\, [r'_{2out}] \cos\Theta \qquad (3)$$

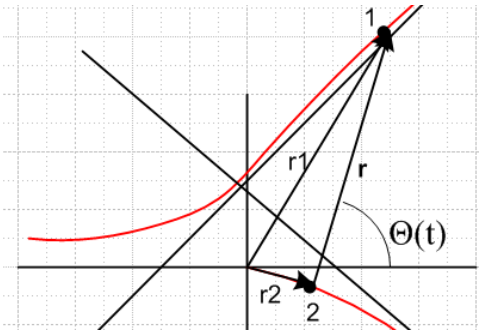
and then comparing (1) and (3) also shows $\theta'_s = \Theta$. So QED on this proof. You have to make use of equations which relate the 1-body problem to the 2-body problem as I have just done. As noted, G assumes this proof is obvious. His Fig 3-16 seems very strange. I understand the label θ there. The line segment seems to refer to $\mathbf{r_1}(t) - \mathbf{r_2}(t) = \mathbf{r}(t)$ at some instant of time, but we need things at $t = +\infty$. The figure shows the angle

What is Fig 3-16 all about?

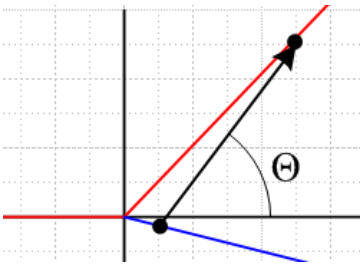
Thus purports to show the 2-body problem activity in the lab frame. Here is one interpretation of this figure



The particle tracks are shown in red, I assume each has an asymptote which I draw, and then there would be an angle between the two asymptotes which I call θ_{L12} which is the relative angle of the two particles at t = +∞. Is he claiming that this angle is the same as the 1-body problem scattering angle Θ? No, I think that is the wrong interpretation of the figure. Let's try another interpretation,



close view



far away view

Here I show the two particles in the lab frame at some finite time t, and **r**(t) has angle Θ(t). Maybe the implication is that as t→∞, Θ(t) → Θ which is the 1-body angle of **r** after the scattering. But such a picture seems unhelpful because we don't off-hand now $\mathbf{r}_1(t)$ and $\mathbf{r}_2(t)$ so hard to draw this thing at various times t. On the right I back up the camera 10 miles and draw the paths. But again, I have no idea what the asymptotic value of Θ is. This picture just seems absolutely worthless!!! It just says that there is a difference vector **r** and it ends up having the angle Θ at t = + very large time.

Conclusion: I have no idea what the purpose of Fig 3-16 is in the mind of Goldstein.

[p 87]

In the lab frame, the CMS point is located at some moving position **R**(t). The CMS coordinates of the two particles are defined by

$$\mathbf{r}'_1 = \mathbf{r}_1 - \mathbf{R} \qquad \text{// position relative to the CMS point} \qquad \text{// (3-68)}$$
$$\mathbf{r}'_2 = \mathbf{r}_2 - \mathbf{R}$$

and differentiating

$$\mathbf{v}'_1 = \mathbf{v}_1 - \dot{\mathbf{R}} \qquad \text{// position relative to the CMS point} \qquad \text{// (3-68)}$$

$$\mathbf{v'}_2 = \mathbf{v}_2 - \dot{\mathbf{R}}$$

Since there is no external force on our 2-particle system, we know that

$$\dot{\mathbf{R}}(t) = (m_1\mathbf{v}_1(t)+m_2\mathbf{v}_2(t))/(m_1+ m_2) = (\mu/m_2)\,\mathbf{v}_1(t) + (\mu/m_1)\,\mathbf{v}_2(t) = \text{constant}$$

This is true because total momentum is conserved,

$$\mathbf{P} = m_1\mathbf{v}_1(t)+m_2\mathbf{v}_2(t) = \text{constant}$$

What is Fig 3-18 showing?

Now lets define some new objects

$$\mathbf{v}_{1in} = \dot{\mathbf{r}}_{1in} = \dot{\mathbf{r}}_1(t = -\infty)$$
$$\mathbf{v'}_{1in} = \dot{\mathbf{r}}'_{1in} = \dot{\mathbf{r}}'_1(t = -\infty)$$

$$\mathbf{v}_{1out} = \dot{\mathbf{r}}_{1out} = \dot{\mathbf{r}}_1(t = +\infty)$$
$$\mathbf{v'}_{1out} = \dot{\mathbf{r}}'_{1out} = \dot{\mathbf{r}}'_1(t = +\infty)$$

In Figure 3-10 he is representing the equation

$$\mathbf{v}_{1out} = \mathbf{v'}_{1out} + \dot{\mathbf{R}}$$

By definition θ is the angle of $\mathbf{v}_{1out}$ and I agree with the two locations in Fig 3-18 where this is shown. I also agree that the angle of $\mathbf{v'}_{1out}$ is Θ because I showed that earlier, so I agree with the one Θ label in that picture. Now from the figure I can write these two equations:

$$v'_{1out}\cos\Theta + \dot{R} = \mathbf{v}_{1out}\cos\theta$$

$$v'_{1out}\sin\Theta = \mathbf{v}_{1out}\sin\theta$$

Divide them to bget

$$\tan\theta = \frac{v'_{1out}\sin\Theta}{v'_{1out}\cos\Theta + \dot{R}} \qquad \text{// agrees with (3-69)}$$

and this then relates the lab particle 1 scattering angle θ to the CMS scattering angle Θ. Now recall from above that $\mathbf{r}'_1 = (\mu/m_1)\,\mathbf{r}$ so that $\dot{\mathbf{r}}'_1 = (\mu/m_1)\,\dot{\mathbf{r}}$ or $\mathbf{v}'_1 = (\mu/m_1)\,\mathbf{v}$ and of course

$$\mathbf{v'}_{1out} = (\mu/m_1)\ \dot{\mathbf{r}}_{out}. \qquad \text{// agrees with p 87 A}$$

Now I think G is assuming this is an elastic collosion, so the speeds are the same before and after. Then $v'_{1out} = v'_{1in}$. But $\mathbf{v}'_{1in} = (\mu/m_1)\ \dot{\mathbf{r}}_{in}$. I think he uses $\mathbf{v}_0$ as a name for $\dot{\mathbf{r}}_{in}$ in the 1-body problem, so then $\mathbf{v}'_{1in} = (\mu/m_1)\ \mathbf{v}_0$. Therefore

$v'_{1out} = (\mu/m_1)\ v_0$ // agrees with (3-70)

Now at $t = -\infty$ we can identify the lab frame with the 1-body problem. A particle is incident on a fixed scattering center, so $\mathbf{v}_0$ is also $\mathbf{v}_{1in}$, the main lab parameter.

At any time in the lab frame we have some center of mass velocity $\dot{\mathbf{R}}$ given by

$\dot{\mathbf{R}} = (m_1\mathbf{v}_1+m_2\mathbf{v}_2)/(m_1+ m_2) = 0$.

At $t = -\infty$ this is just

$\dot{\mathbf{R}}_{in} = (m_1\mathbf{v}_{1in})/(m_1+ m_2)$

At $t = +\infty$ it is the same but given by

$\dot{\mathbf{R}}_{out} = (m_1\mathbf{v}_{1out} + m_2\mathbf{v}_{2out})/(m_1+ m_2)$

Setting these equal gives

$m_1\mathbf{v}_{1in} = m_1\mathbf{v}_{1out} + m_2\mathbf{v}_{2out} = (m_1+ m_2)\dot{\mathbf{R}}$

The left = sign shows momentum conservation, the second just comes from above. Equating left and right

$m_1\mathbf{v}_0 \equiv m_1\mathbf{v}_{1in} = (m_1+ m_2)\dot{\mathbf{R}}$ // agrees with page 87 B

$\Rightarrow \quad \dot{\mathbf{R}} = [\, m_1/ (m_1+ m_2)]\ \mathbf{v}_0 = [\, m_2m_1/ (m_1+ m_2)]\ (1/m_2)\ \mathbf{v}_0 = (\mu/m_2)\ \mathbf{v}_0$ // (3-71)

Now go back to (3-69) which we already proved,

$$\tan\theta = \frac{v'_{1out}\sin\Theta}{v'_{1out}\cos\Theta + \dot{R}}$$

(3-71) (3-70)

and replace $\dot{R} = (\mu/m_2)\ v_0$ and $v'_{1out} = (\mu/m_1)\ v_0$, so then

$$\tan\theta = \frac{(\mu/m_1)\, v_0 \sin\Theta}{(\mu/m_1)\, v_0 \cos\Theta + (\mu/m_2)\, v_0} = \frac{(\mu/m_1)\ \sin\Theta}{(\mu/m_1)\cos\Theta + (\mu/m_2)} = \frac{\mu\ \sin\Theta}{\mu\cos\Theta + \mu\,(m_1/m_2)}$$

$$= \frac{\sin\Theta}{\cos\Theta + (m_1/m_2)} \qquad \text{// agrees with (3-72)}$$

and now we have a relatively simple relationship between the cms angle Θ and the lab angle θ. The angles are thus the same if $m_1 << m_2$.

[p 88]

In p 88A, why is I the same on both sides of the equation? I is particles/area/sec and it seems that if frames are boosted, you would have I' ≠ I. This issue is not clarified in 3rd edition.

I think the remedy is to go back to 3-57 and rewrite it this way

$$\sigma(\Omega)d\Omega = \frac{\text{total particles scattered into } d\Omega \text{ during experiment}}{\text{total particles in experiment}}$$

Now there is no mention of intensity or time. The beam is turned on at t = 0 and off at some t = T and during the experiment the total incident particles was 1,000,000 and the number scattered into dΩ was perhaps 100. If two observers in frame S and S' watch this experiment, they both agree that the experiment involve 1,000,000 particles total. The guy in frame S sees 100 into dΩ, and the guy in frame S' sees those same 100 but into his dΩ' and dΩ ≠ dΩ' in general. So then

$$\sigma(\Omega)d\Omega = \frac{\text{total particles scattered into } d\Omega \text{ during experiment}}{\text{total particles in experiment}}$$

$$= \frac{\text{total particles scattered into } d\Omega' \text{ during experiment}}{\text{total particles in experiment}} = \sigma'(\Omega)d\Omega' = \frac{100}{1\text{ million}}$$

I might write

$\frac{d\sigma}{d\Omega}$ = cross section in the lab frame = σ(Ω) of Goldstein

$\frac{d\sigma'}{d\Omega'}$ = cross section in the CMS frame = σ'(Ω) of Goldstein

and in this notation we have

$$\frac{d\sigma}{d\Omega}\, d\Omega = \frac{d\sigma'}{d\Omega'}\, d\Omega' = \frac{100}{1\text{ million}}$$

So we then get, in G language,

$$\sigma'(\Omega') = \sigma(\Omega)\ \frac{d\Omega}{d\Omega'}$$

Now let S = lab and S' = CMS frame. Then we have using the rings for dΩ

$$\sigma_{CMS}(\Theta) = \sigma_{LAB}(\theta)\ \frac{d\Omega}{d\Omega'} \ = \ \sigma_{LAB}(\theta)\ \frac{d(\cos\theta)}{d(\cos\Theta)}$$

But let's write this the other way

$$\sigma_{LAB}(\theta) \ = \sigma_{CMS}(\Theta)\ \frac{d(\cos\Theta)}{d(\cos\theta)}$$

Whereas on the previous page G used primes to indicate the CMS frame and no primes for the LAB frame, here he suddenly exacdtly reverses that notation (horrible!) and writes the above as

LAB CMS

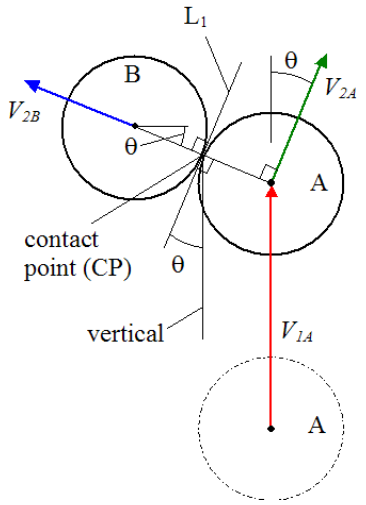
$$\sigma'(\theta) \ = \sigma(\Theta)\ \frac{d(\cos\Theta)}{d(\cos\theta)} \qquad \text{// agrees with (3-73)}$$

OK, fine. If you really wanted to know the fraction you have to do a lot of work based on 3-72 which relates the angles Θ and θ , and G does not carry out this task for general m_1 and m_2. But for the special case $m_1 = m_2$ which applies to say neutron-proton scattering, the relation is simply $\theta = \Theta/2$. Why?

$$\tan\theta = \frac{\sin\Theta}{\cos\Theta \ + (m_1/m_2)} = \frac{\sin\Theta}{\cos\Theta \ + 1} = \frac{2\sin(\Theta/2)\cos(\Theta/2)}{2\cos^2(\Theta/2)} = \frac{\sin(\Theta/2)}{\cos(\Theta/2)} = \tan(\Theta/2)$$

$$\Rightarrow\ \theta = \Theta/2 \qquad \text{// agrees with (3-74)}$$

Example: if you backscatter 180° in the CMS, $\theta = 90°$. How can that be? Well in fact the incoming ball has stopped completely so the 90° does not mean anything. But suppose you are just a little off center so in the CMS you have 179°. Then the initial particle goes off very slowly sideways, it is true. It is also true in billiards which is a much different kind of interaction but I guess is still a central force:



It happens that in billiards you always get as 90° angle in the lab frame as shown above from the web.

Now, for cross section relation you then get,

$$\frac{d(\cos\Theta)}{d(\cos\theta)} = (\sin\Theta/\sin\theta)\, d\Theta/d\theta = 2\,(\sin\Theta/\sin\theta) = 2\,(\sin 2\theta/\sin\theta) = 4\sin\theta\cos\theta/\sin\theta = 4\cos\theta$$

and then you get

$\sigma'(\theta) = 4\cos(\theta)\,\sigma(2\theta)$ // agrees with (3-75)
LAB CMS

[p 89]

G goes on to compute how much KE is transferred from the incident particle to the target. He gives a general result (3-76) which I won't derive here (looks easy) and then notes if $m_1 = m_2$ the result is very simple

$v_1/v_0 = \cos\theta$

Example: for a full backscatter, $\theta = 90°$ in effect, as noted above, and then v_1 lab after colliosion = 0. Goldstein's point is that ALL the energy of the incoming particle is transferred to the target. G notes that in a nuclear pile moderator the hot neutrons collide with H atoms in the moderator and slow down and this increases their fission probability. H does not work well since it absorbs neutrons, but deuterium and carbon are pretty good if not optimal.

G concludes stating that you can consider scattering as a black-box operation and our work here involving cross sections applies regardless of what happens in the box, quantum or not.

G got his PhD in 1943, at the peak of the nuclear weapon era (Hiro 1945). He was at the MIT Rad Lab doing radar, but later he actually worked on reactor design and moderators and shielding. A prof at Colombia 61-05 where he died emeritus. So you can see why that subject shows up in this book of 1950.

My old notes: We define velocities in both frames, do some geometry, use E and p conservation (not very clearly, however) and derive formula 3-72 which relates the lab scattering angle script-θ to the CMS scattering angle Θ. A simple counting rule relates the lab and CMS cross sections as shown in 3-73. One can write this out in detail, but I guess it is a bit messy so G does not do that. He looks at the $m_1 = m_2$ case and finds the simple result 3-74 which says the lab angle is half the CMS angle, and in the lab you can get at most 90 degree scattering, as in billiards. This is all a consequence of conservation of energy and momentum in an elastic collision where no heat is generated and there is no final change in potential energy either. He comments at the end that this means that all this math applies to "whatever happens" to actually make the collision, be it billiard balls, Coulomb force, or something else.

At the time of this book 1950, there was no "particle physics" as there was later, so this simple application of E and p conservation was not found in mechanics books the way it is now found in every particle physics books in a kinematics section. Of course we now have relativistic kinematics in these books of which this non-rel stuff is a special case.

If you want to slow down some neutrons, you put in a moderator of light nuclei. Heavy nuclei don't recoil so there is no KE transfer and the neutrons don't slow down. C^{12} works OK, H^1 in paraffin works better. This was Goldstein's area of expertise, so he threw in this comment. Slowed neutrons react better to maintain nuclear reactions in a pile, but I am not wandering off on that subject right now. I did read about the Wigner heating effect and how it caused the Windgate Fire in Britain in 1957, a scary event for all involved.

Note: Maple says that, if we use Θ = y and script- θ = x, the conversion factor for the cross section is given by dcosy/dcosx = below, where b = m1/m2.

$$\frac{(\cos(y)+b)\sqrt{\dfrac{2\cos(y)\,b+b^2+1}{(\cos(y)+b)^2}\,(2\cos(y)\,b+b^2+1)}}{\cos(y)\,b+1}$$

Obviously the whole thing will a big mess in the general case. For b=1, above becomes
$2\sqrt{2}\,(\cos(y)+1)^{1/2}$ and everything simplifies, but I don't care much about these details right now.

Chapter ends with the usual comments on some books, Whittaker and Sommerfeld and Slater, names from olden days of text books. He fails to put publication dates on books for some reason! I then read through all the problems, many of which are very interesting. For example, mixing $1/r^2$ with $1/r^3$ gives a precessing elliptical orbit. And so ends this long "two-body central force" chapter.

****

Chapter 4: Rigid Body Kinematics (93)

****

4.1 The independent coordinates of a rigid body (93). G gives several arguments why there are 6 independent coordinates for a rigid body, but this is old hat for me. You position a reference point somewhere, then you align a set of axes with three Euler rotations. He talks all about the direction cosines of unit vectors in two systems rotated relative to each other, primed and unprimed. There are 9 such cosines, but of course only three are independent ones. // I reread this section 11.17.16 and it includes the definition of a "rigid" body, meaning $r_{ij} = c_{ij}$ for all particles which is N(N-1)/2 non-independent conditions on the 3N coordinates. Yes, there are 9 direction cosines which connect two sets of unit vector axes, but only three are independent. Here just for fun, suppose I write

$x'_i = R_{ij}x_j$ R = matrix of 9 direction cosines like $R_{ij} = \hat{\mathbf{i}} \bullet \hat{\mathbf{j}}' = \cos(\theta_{ij})$

I know that to preserve unit vector lengths, R must be real orthogonal so $RR^T = 1$ or $R_{ij}R_{kj} = \delta_{ik}$ which can be written

$R_{\ell 1}R_{k1} + R_{\ell 2}R_{k2} + R_{\ell 3}R_{k3} = 0$ $\ell \neq k$

$R_{\ell 1}R_{\ell 1} + R_{\ell 2}R_{\ell 2} + R_{\ell 3}R_{\ell 3} = 1$ $\ell = 1,2,3$ // this is (4-8)

A rotation matrix R has only 3 independent parameters. There is no need ever to reread this section, I grok it.

4.2 Real orthogonal transformations (97). So here he does what I just said above. He does not use the word real. Here we put the nine direction cosines into a 3x3 matrix and we show that $AA^T = 1$ so $A^T = A^{-1}$ and all that stuff. He then talks about the active and passive meanings of a rotation, but does not use those words. Done forever with this section.

4.3 Formal properties of the transformation matrix (101). Then get AB ≠ BA for sequential rotations, Finally on page 105 he defines the similarity transformation which changes a matrix A to A' but preserves the determinant. He mentions $A^\dagger = A^{T*}$ and talks about a complex unitary matrix, but does not use this idea yet. I guess linear algebra was not as familiar then in 1950 as it is now to technical students of many disciplines, so G felt the need to go through all this stuff. He refers us to M&M Chap 10 which is where I went to get the basics when I was doing my matrix review. All fine, do not reread.

4.4 The Euler angles (107). He calls them "Eulerian angles". He does things in terms of passive rotations, which for me means the axes are moved, not the vectors in 3space. Thus, all sinθ signs are changed relative to my usual active matrix notes. And he uses "R_x" in the middle, and comments on this a lot in a huge footnote, variations in conventions. He ends up with

$R(\psi,\theta,\phi) = B(\psi)C(\theta)A(\phi) = R_z(\psi)R_x(\theta)R_z(\phi)$ with passive rotation forms // (4-43,44,45)

He does us the favor of multiplying these things out on page 109 and also writing the inverse. No need to read this again. These last results are (4-46,47). // I had Maple verify both his big matrices, all OK.

4.5 The Cayley-Klein parameters (109). I know already how to write the above Euler rotation in the 2D representation of the rotation group, and the results are as in my notes except again for minus signs from passive versus active. Again we can think of R(ψ,θ,ϕ) as above with the exact same three parameters and

the result is as shown in 4-67 p 116 which no doubt agrees with something in Levitt modulo minus signs. We know there is a 1:2 homomorphism between SO(3) and SU(2) so that Q and -Q both correspond to the same rotation. If you write your 2x2 matrix as Q = (α,β,δ,γ) then we know there are not really 8 independent real parameters (only three as he carefully shows, since unitary and unimodular). But written in this general form, the parameters α,β,δ,γ are called the Cayley-Klein parameters and they were found to be useful in fancy gyroscope problems, probably we shall see some in the next chapter. Felix Klein (1849-1925) was the guy who used them. The 2π rotation causing -1 is noted. The 2D space here is called "the u,v space" and G comments that it relates to spin 1/2 things like electrons. I know that in fact the u,v space is the J = 1/2 representation of the rotation group and yes, we call such things spinors. Finally, he shows that if you write P = (z, x–,x+,-z) as a 2x2 matrix, then P' = QPQ† gives you the 3D coordinates in the rotated frame. Also, P = P† and trP = 0 obviously. You can write P = **r**•**σ** where σ are the Pauli matrices. In (4-68) G expresses the Cayley-Kleins in terms of the Euler angles for example α = α(ψ,θ,ϕ). I am happy to say that by now I have all this stuff completely nailed down in various notes, but it would have seemed strange to me back in 1970 when I probably had this class with Sidney Coleman.

Just adding a few notes 11.17.16. Q is his SU(2) matrix (4-49). It must be unitary so it has the simpler form (4-56). You can rotate a 3-vector **r** into **r'** by doing P' = QPQ† where $P = \begin{pmatrix} z & x+iy \\ y-iy & -z \end{pmatrix}$ though I have never found this very useful. G shows that using $Q = \begin{pmatrix} \alpha & \beta \\ \delta & \gamma \end{pmatrix}$ you can produce the 3x3 corresponding rotation matrix which is stated in (4-63). He then writes the SU(2) specific rotation matrices and gets the SU(2) version of the Euler angle rotation in (4-67), so he does the mapping both ways SU(2) ↔ SO(3). This is another section I never need to read again, just read these notes!

4.6 Euler's Theorem on the motion of a rigid body (118). Now we get back to some physics. First, the word "displacement" for a rigid body means *any* change in its "position/orientation". A rotation is a displacement, so is a translation, so is a combination of the two. I had the idea that it only meant translation d**r**, but that is wrong. **Euler's Theorem** says this: any displacement of a rigid body in which some point in the body stays fixed can be represented as a rotation about some axis. It seems obvious to me that any such displacement must be a rotation about that fixed point. But an arbitrary rotation I know can be written R(ψ,θ,ϕ) = exp(+iγ**J**•**n**) which is then a rotation of γ about the **n** axis. So to me, with what I know, this is not a very exciting theorem. And then **Chasle's theorem** says that the most general displacement is a combination of a translation and a rotation.

In our group-savvy modern era, we know that the product of the three Euler rotations is a single rotation about some axis which I would write as R(ψ,θ,ϕ) = exp(+iγ**J**•**n**) and I am sure I could find the direction **n** and the angle γ by consulting my notes for a very short while. [Note added 11.19.16: this connection was not in my notes, but it is now, see this directory "Euler Angles compared with exp(-iu n.J) parameters"]. At one time, this must have seemed interesting enough to call it Euler's Theorem. Similarly, Chastle's Theorem is just saying that the position of a rigid body can be identified with an element of the Galilean Group. Goldstein never mentions SU(2) or groups (so far), but I am sure the latest edition does this more systematically.

The details of this section are as follows. We want to show that there is some vector **R** which is not changed by a selected general rotation matrix A, since that would mean that rotation was about the axis of that vector. So that would mean A**R** = λ**R** (**R** = **n,** the rotation axis) meaning that perhaps it could change in length by matrix A. We know this cannot happen, so we know λ = 1.

G now uses this problem A**R** = 1**R** to present a general discussion of eigenvalues problems for a real-orthogonal matrix. He writes the equation as (A-λI)**R** = 0 in 4-75 and 4-76. I would write this differently as (R-λI)**x** = 0. To have a solution, one muast have det(A-λI) = 0 which is the secular equation (secular because first used for long term planetary motions), and then we have a cubic in λ which implies 3 roots. He shows that there is a transformation matrix **X** which brings A to diagonal form, so then $X^{-1}AX = \Lambda$ where $\Lambda = \text{diag}(\lambda_1,\lambda_2,\lambda_3)$. He then shows these facts (his "lemmas"). I give my own proof of each:

1. The three eigenvalues have unit length.
   $R\mathbf{x} = \lambda\mathbf{x}$, $R^*\mathbf{x} = \lambda^*\mathbf{x}$, $|R\mathbf{x}|^2 = R^*\mathbf{x} \bullet R\mathbf{x} = \mathbf{x} \bullet R^{*T}R\mathbf{x} = \mathbf{x}\bullet\mathbf{x} = |\mathbf{x}|^2$, but $|\lambda\mathbf{x}|^2 = |\lambda|^2|\mathbf{x}|^2$ QED
2. At least one of the three eigenvalues is real.
   Cubic 4-83 must have a real root as graphs shows.
3. The product of the eigenvalues must be 1 (because det(A) = 1).
   $X^{-1}AX = \Lambda \Rightarrow \det(A) = \det(\Lambda) = \lambda_1\lambda_2\lambda_3$ and det(A) = 1 since "special" SO(3).
4. If λ is an eigenvalue, then so is λ*.
   Coefficients in secular equation are real, so poly(λ) = 0 ⇒ poly(λ*) = 0

Based on these lemmas, if there are three roots on the unit circle, one must be real and the other must be a complex pair. The real root can only be +1 or -1. Need detΛ = 0 which means $\lambda_1\lambda_2\lambda_2^* = 1$ so $\lambda_1 = 1$. The conclusion after all this work is that λ = 1 is an eigenvalue! Therefore, you can find R such that A**R** = **R** or I would say R**x** = **x** so you have found a vector which the general rotation R does not alter, so R must be a rotation about bout this axis, and thus you have proved Euler's Theorem.

So what exactly is the solution vector **n** such that R**n** = **n**? Assume without LOG that **n** is a unit vector. That is, what is the eigenvector here? There must be some rotation Q such that Q**n** = $\hat{\mathbf{z}}$ so write

$$R\mathbf{n} = \mathbf{n} \quad \Rightarrow \quad QR\mathbf{n} = \hat{\mathbf{z}} \quad \Rightarrow \quad QRQ^{-1}\hat{\mathbf{z}} = \hat{\mathbf{z}} \quad \Rightarrow \quad QRQ^{-1} = R_{\mathbf{z}}(\Phi)$$

But we know the form of $R_{\mathbf{z}}(\Phi)$ and we know $\text{tr}(R_{\mathbf{z}}(\Phi)) = 1 + 2\cos\Phi$. But this is also tr(R), so if you are given R, you can deduce Φ in this way

$$1 + 2\cos\Phi = \text{tr}(R) = \Sigma_i R_{ii}$$

In my doc noted above, this means that $1 + 2\cos u = R_{11} + R_{22} + R_{33}$ and that gives the result I found there.

The final paragraph says that Chasle's Theorem that the most general motion is a rotation and a translation is pretty obvious, and so the position/orientation of a rigid body can be described by 6 parameters relative to some starting position.

4.7 Infinitesimal rotations (124). We now continue this sort of backward presentation. An obvious example shows that rotations don't commute, but here he is doing active rotations of an object. I now have to change notation, and I would say A = 1 + η as the small rotation, replacing his ε with η. Then in our usual generator notation, I would say $\eta = \exp(i\theta\hat{\mathbf{n}}\bullet\mathbf{J}) - 1 \approx i\theta\hat{\mathbf{n}}\bullet\mathbf{J}$ where **J** are the usual SO(3) generator matrices. We then connect to his notation by saying $d\boldsymbol{\Omega} = \theta\hat{\mathbf{n}}$ and I also know that $i\,(J_k)_{\ell m} = \varepsilon_{k\ell m}$ and now we avoid having to use δ for ε, etc etc etc. Note that matrix J_k is antisymmetric, thus so is η. We can then say

$d\mathbf{r} = [i\theta\hat{\mathbf{n}}\bullet\mathbf{J}]\,\mathbf{r} \quad \Rightarrow$

$dr_j = [i\theta\hat{\mathbf{n}}\bullet\mathbf{J}]_{jk}\, r_k = [d\boldsymbol{\Omega}\bullet i\mathbf{J}]_{jk}\, r_k = [\, d\Omega_m\, iJ_m]_{jk}\, r_k = d\Omega_m\,(iJ_m)_{jk}\, r_k = d\Omega_m\,\varepsilon_{mjk} r_k$

$= \varepsilon_{jkm}\, r_k\, d\Omega_m = [\,\mathbf{r} \times d\boldsymbol{\Omega}]_j \Rightarrow d\mathbf{r} = \mathbf{r} \times d\boldsymbol{\Omega}$ // as in 4-94

Then in the p 132 figure we realize that $d\boldsymbol{\Omega}$ must point along the rotation axis and $d\Omega$ must be the amount of rotation, so we are back to $d\boldsymbol{\Omega} = \theta\hat{\mathbf{n}}$. Yes, the generators J_k are anti-symmetric matrices, so A and η are antisymmetric

Note added: A is antisymmetric only in its infinitesimal form. More generally A has three components

$$R_n(u) = \exp(-iu\mathbf{n}\bullet\mathbf{J}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + (\cos u - 1)\begin{pmatrix} 1-n_1^2 & -n_1n_2 & -n_1n_3 \\ -n_1n_2 & 1-n_2^2 & -n_2n_3 \\ -n_1n_3 & -n_2n_3 & 1-n_3^2 \end{pmatrix} + \sin u \begin{pmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{pmatrix}$$

where the first term is both sym and antisym, the second term is sym, and the third term is antisym. For infinitesimal you only have the first and third term so

$$R_n(u) \approx \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + u \begin{pmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{pmatrix} \qquad u = \text{small angle}$$

Notice that the result $d\mathbf{r} = \mathbf{r} \times d\boldsymbol{\Omega}$ is telling us the change in a position point's coordinates due to a small rotation. We just derived this from the passive rotation rule for a vector **r**, namely that

$\mathbf{r}' = \exp(+i\, d\boldsymbol{\Omega}\bullet\mathbf{J})\,\mathbf{r} \approx \mathbf{r} + i\, d\boldsymbol{\Omega}\bullet\mathbf{J}\,\mathbf{r} \quad \Rightarrow d\mathbf{r} = i\, d\boldsymbol{\Omega}\bullet\mathbf{J}\,\mathbf{r} = \mathbf{r} \times d\boldsymbol{\Omega}$

Reminder on the last equality

$dr_i = [i\, d\boldsymbol{\Omega}\bullet\mathbf{J}\,\mathbf{r}]_i = i\,\Omega_k (J_k)_{is} r_s = \varepsilon_{kis} r_s \Omega_k = \varepsilon_{isk}\, r_s \Omega_k = [\,\mathbf{r} \times d\boldsymbol{\Omega}]_i$

Obviously we can rotate any vector **G**, not just the position vector **r**, so the above discussion results in

$\mathbf{G}' = \exp(+i\, d\boldsymbol{\Omega}\bullet\mathbf{J})\,\mathbf{G} \approx \mathbf{G} + i\, d\boldsymbol{\Omega}\bullet\mathbf{J}\,\mathbf{G} \quad \Rightarrow d\mathbf{G} = i\, d\boldsymbol{\Omega}\bullet\mathbf{J}\,\mathbf{G} = \mathbf{G} \times d\boldsymbol{\Omega}$

The whole thing is summarized in Fig 4-12 which is my "cone picture" and $d\boldsymbol{\Omega} = \Omega\hat{\boldsymbol{\Omega}}$ in place of my usual $d\boldsymbol{\theta} = d\theta\hat{\mathbf{n}}$. I normally think of Ω as being related to solid angle, so a little confusing just for me.

Also in this section he comments on **pseudo** or **axial** vectors, such as $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ and such as **n** here.

I call this presentation backwards only because one would usually begin with infinitesimal rotations, then build up the finite rotations from exponentials, and one would mention Lie Algebras and so on.

The following section led me to write my Rotating Frames of Reference document finished Aug 2012. I have added some critical comments in red.

ok to here on my 11.20.16 review.

4.8 Rate of change of a vector (132). I am sorry to say that the presentation here is very weak [at least for me] . First, we define a rotating frame (body) and a lab frame (space). The rotating "body" frame is some set of axes fixed to a tumbling object, and this will be a non-inertial frame in general. Imagine that we observe some activity from the this tumbling body frame. We might see a vector change by an amount $d\mathbf{G}_{body}$. If $\mathbf{G} = \mathbf{r}$ = a point in the rigid body, then $d\mathbf{G}_{body} = d\mathbf{r}_{body} = 0$ because the body is "rigid" relative to the body system. But for some other vector quantity **G** (perhaps the position of a fired projectile), we can have $d\mathbf{G}_{body} \neq 0$. The claim is this:

$$d\mathbf{G}_{space} = d\mathbf{G}_{body} - \mathbf{G} \times d\boldsymbol{\Omega} \qquad \text{// this is his 4-99}$$

Here now is my own derivation and interpretation of this idea:

If we apply a small active rotation of amount $\gamma = d\Omega$ about axis $\hat{\mathbf{n}}$ to a vector, then using the usual $[J_i]_{jk} = -i\varepsilon_{ijk}$ we can show that for any vector **G** we have $d\mathbf{G} = [-i\theta\hat{\mathbf{n}}\bullet\mathbf{J}]\,\mathbf{G} = -\mathbf{G} \times d\boldsymbol{\Omega}$ where $d\boldsymbol{\Omega} = \gamma\,\hat{\mathbf{n}}$. If, for example, **G** were the vector representing a *point fixed* in a rotating body as seen from the lab frame, then we would say

$$[d\mathbf{G}]_{space} = d\boldsymbol{\Omega} \times \mathbf{G}$$

Now suppose **G** is *not* a fixed point in the object, but is a point that is *moving* relative to the rotating object. Suppose during a time interval dt we have a movement $[d\mathbf{G}]_{body}$ due to this motion within the body system. Then the overall vector change in the space system will be:

$$[d\mathbf{G}]_{space} = [d\mathbf{G}]_{body} + d\boldsymbol{\Omega} \times \mathbf{G}$$

and then divide by dt to get

$$[d\mathbf{G}/dt]_{space} = [d\mathbf{G}/dt]_{body} + \boldsymbol{\omega} \times \mathbf{G} \qquad \text{where } \boldsymbol{\omega} = d\boldsymbol{\Omega}/dt \qquad \text{4-100}$$

Notice that we can express the vector **G** itself in either space or body coordinates and we might then denote these things by $\mathbf{G}^{space}$ and $\mathbf{G}^{body}$. But we need to use the same representation in all three places that **G** appears.

[But what the above really involves is $[d\mathbf{G}]_{space}/dt$ and $[d\mathbf{G}]_{body}/dt$. One needs to explain how this is related to the notation above. In retrospect, $[d\mathbf{G}]_{space}$ is the change in G "as measured in the space frame", and that is what $[d\mathbf{G}/dt]_{space}$ somehow means.]

Clarification of Notational Ambiguity. There are two distinct meanings we could attribute to a label like "body" applied to some object X:

X^{body} = the quantity X "as expressed in body frame coordinates". [I think this is wrong.]

[The correct statement would be for *components*, so could have $X^{body}{}_i$ as component of X, but there is no object called X^{body} .] [The idea of some vector $\mathbf{X}^{body}$ led me into trouble.]

X_{body} = the quantity X "as measured in the body frame" [this *is* well defined]

For the moment, I will use superscript and subscript to distinguish these two different meanings. As an example, suppose $d\mathbf{r}_{body}$ represents a displacement of a particle as measured in the body frame. We can certainly express the components of this vector in either frame, $d\mathbf{r}_{body}{}^{body}{}_i$ and $d\mathbf{r}_{body}{}^{space}{}_i$. If one is nonzero, so is the other, because these vectors are related to each other by the Euler angle rotation which describes the relative orientation of the two frames. On the other hand, suppose $d\mathbf{r}_{body} = 0$. This does not imply that $d\mathbf{r}_{space} = 0$!

As another example, consider $[\ d\mathbf{r}/dt\ ]_{body} = \mathbf{v}_{body}$ and $[\ d\mathbf{r}/dt\ ]_{space} = \mathbf{v}_{space}$. In either of these equations, we could express our coordinates [components] in either body or space frames by putting on the right superscripts. [correct]

Goldstein likes to write this as if it were some kind of "operator equation" as in 4-102, but I find that a little confusing. This operator rule would apply only to vectors. [correct]

At this point, G derives an expression for ω^{body} in terms of the Euler angles. The last 2/3 of page 134 took me a long while to figure out, but here it is. We want to start with a rigid body at some finite Euler angles ψ,θ,ϕ relative to some standard orientation at 0,0,0, and then, starting from that location, we want to look at three infinitesimal Euler rotations such as dϕ about its Euler axis. Think of things perhaps this way, where $R = R_z(\psi)R_x(\theta)R_z(\phi) = \exp(i\ d\mathbf{n}\bullet\mathbf{J})$:

$$R + dR = \exp(i\ [\ \mathbf{n} + d\mathbf{n}]\bullet\mathbf{J}) = \exp(i\ d\mathbf{n}\bullet\mathbf{J})\ \exp(i\ \mathbf{n}\bullet\mathbf{J}) = \exp(i\ d\mathbf{n}\bullet\mathbf{J})\ R = R + R\ [id\mathbf{n}\bullet\mathbf{J}\ ]$$

where R, dR, J_i are 3x3 matrices and the first step is similar to $R_z(\phi+d\phi) = R_z(d\phi)\ R_z(\phi)$. The infinitesimal rotation we are talking about is then represented by $dR = R\ [id\mathbf{n}\bullet\mathbf{J}\ ]$.

Each of our three cases results in a certain $d\mathbf{n}$ such as $d\mathbf{n} = d\phi\hat{\mathbf{n}} = d\phi\hat{\mathbf{z}}$. We look at the drawings on page 107 to see this fact. Then we need to express this in primed body coordinates (here we have $d\mathbf{n} = d\phi(0,0,1)$ being in space coordinates). We then compute $d\mathbf{n}/dt = \boldsymbol{\omega}$ in body coordinates for each of these three cases and we give these names like $\boldsymbol{\omega}_\phi$ and *I* go on to denote the frame of reference, for example $\boldsymbol{\omega}_\phi{}^{space} = (d\phi/dt)\ (\ 0,0,1)$. Finally we think of $d\mathbf{n}$ = the sum of these three small changes and we add them all up and we get the result 134. Here now are the details.

First, look at a small dϕ rotation. In this case from the first picture on page 107 we have $d\mathbf{n} = d\phi\hat{\mathbf{n}} = d\phi\hat{\mathbf{z}}$ so we would then say that $\boldsymbol{\omega}_\phi{}^{space} = \dot{\phi}\hat{\mathbf{z}} = \dot{\phi}(\ 0,0,1)$. To get this in the body system coordinates, we apply to $\hat{\mathbf{z}} = (0,0,1)$ the matrix A shown on page 109 and we find that
$\boldsymbol{\omega}_\phi{}^{body} = \dot{\phi}\ (\ \sin\psi\ \sin\theta, \cos\psi\ \sin\theta, \cos\theta)$ and this agrees with the first equation quoted page 134.

Second, consider a small θ rotation. In this case, from the middle picture on page 107 we know that $\hat{\mathbf{n}}$ = that line of nodes = $\cos\phi\ \hat{\mathbf{x}} + \sin\phi\ \hat{\mathbf{y}}$. This is the axis about which we are doing our small θ rotation. We then know that $d\mathbf{n} = d\theta\ \hat{\mathbf{n}}$ and so $\boldsymbol{\omega}_\theta{}^{space} = \dot{\theta}\ \hat{\mathbf{n}} = \dot{\theta}\ [\cos\phi\ \hat{\mathbf{x}} + \sin\phi\ \hat{\mathbf{y}}]$. We could then transform $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ into the body frame, but this is a little messy. Instead, look at page 107. We can see this relationship from the last figure: $\hat{\mathbf{n}} = \cos\psi\ \hat{\mathbf{x}}' - \sin\psi\ \hat{\mathbf{y}}'$. Thus, we have $\boldsymbol{\omega}_{\theta,body} = \dot{\theta}\ \hat{\mathbf{n}} = \dot{\theta}\ [\cos\psi\ \hat{\mathbf{x}}' - \sin\psi\ \hat{\mathbf{y}}']$. I did

this on a piece of paper to make sure the sign is right. And this again agrees with second line on page 134.

Third, consider a small ψ rotation. In this case we know that $\hat{\mathbf{n}} = \hat{\mathbf{z}}'$ and so $d\mathbf{n} = d\psi\ \hat{\mathbf{n}} = d\psi\ \hat{\mathbf{z}}'$ and then we have $\boldsymbol{\omega}_{\psi}^{\text{body}} = \dot{\psi}\ \hat{\mathbf{z}}'$. So can summarize these three body results as follows:

$$\boldsymbol{\omega}_{\phi}^{\text{body}} = \dot{\phi}\,(\sin\psi\sin\theta,\ \cos\psi\sin\theta,\ \cos\theta) = \dot{\phi}\ \hat{\mathbf{n}}_{\phi}$$

$$\boldsymbol{\omega}_{\theta}^{\text{body}} = \dot{\theta}\,(\cos\psi,\ -\sin\psi,\ 0) = \dot{\theta}\ \hat{\mathbf{n}}_{\theta}$$

$$\boldsymbol{\omega}_{\psi}^{\text{body}} = \dot{\psi}\,(0,\ 0,\ 1) = \dot{\psi}\ \hat{\mathbf{n}}_{\psi}$$

Finally, we want to argue that we just add these three vectors up to get $\boldsymbol{\omega}_{\text{body}}$ which results from simultaneous small rotations in the three Euler angles. We then have all at once:

$$d\boldsymbol{\Omega} = d\phi\ \hat{\mathbf{n}}_{\phi} + d\theta\ \hat{\mathbf{n}}_{\theta} + d\psi\ \hat{\mathbf{n}}_{\psi} = d\gamma\ \hat{\mathbf{n}} = \text{an arbitrary differential "rotation"}$$

so that

$$\boldsymbol{\omega} = d\boldsymbol{\Omega}/dt = \dot{\phi}\ \hat{\mathbf{n}}_{\phi} + \dot{\theta}\ \hat{\mathbf{n}}_{\theta} + \dot{\psi}\ \hat{\mathbf{n}}_{\psi}$$

$$= (\dot{\phi}\sin\psi\sin\theta + \dot{\theta}\cos\psi,\ \dot{\phi}\cos\psi\sin\theta - \dot{\theta}\sin\psi,\ \dot{\phi}\cos\theta + \dot{\psi}) = \boldsymbol{\omega}^{\text{body}} = (x', y', z') \text{ components}$$

and finally we have derived 4-103. Notice that the three "Euler unit vectors" are not orthogonal.

4.9 The Coriolis Force (135). First, here is my own derivation, then I go back to my original raw notes.

Now for an application. Suppose we apply the above rule first to $\mathbf{G} = \mathbf{r}$,

$$[d\mathbf{r}/dt]_{\text{space}} = [d\mathbf{r}/dt]_{\text{body}} + \boldsymbol{\omega} \times \mathbf{r}$$

which we then write as

$$\mathbf{v}_{\text{space}} = \mathbf{v}_{\text{body}} + \boldsymbol{\omega} \times \mathbf{r}$$

and as noted above, this is true regardless of whether one expresses the coordinates of all these vectors in the body frame or in the space frame. For example, $\mathbf{v}_{\text{space}}^{\text{body}} = \mathbf{v}_{\text{body}}^{\text{body}} + \boldsymbol{\omega}^{\text{body}} \times \mathbf{r}^{\text{body}}$ would be one of the two possibilities.

Now apply the rule again to $\mathbf{G} = \mathbf{v}_{\text{space}}$

$$[d\,\mathbf{v}_{\text{space}}/dt]_{\text{space}} = [d\,\mathbf{v}_{\text{space}}/dt]_{\text{body}} + \boldsymbol{\omega} \times \mathbf{v}_{\text{space}}$$

$$= [d\,(\mathbf{v}_{\text{body}} + \boldsymbol{\omega} \times \mathbf{r})/dt]_{\text{body}} + \boldsymbol{\omega} \times (\mathbf{v}_{\text{body}} + \boldsymbol{\omega} \times \mathbf{r})$$

$$= [d\,\mathbf{v}_{\text{body}}/dt]_{\text{body}} + \boldsymbol{\omega} \times [d\mathbf{r}/dt]_{\text{body}} + \boldsymbol{\omega} \times \mathbf{v}_{\text{body}} + \boldsymbol{\omega} \times \boldsymbol{\omega} \times \mathbf{r}$$

$$\Rightarrow \qquad \mathbf{a}_{space} = \mathbf{a}_{body} + 2\,\boldsymbol{\omega} \times \mathbf{v}_{body} + \boldsymbol{\omega} \times \boldsymbol{\omega} \times \mathbf{r}$$

and, once again, we have our choice of which of the two frames, body or space, to express all the vectors appearing in this equation.

Back now to the raw notes:

Subscripts here are s = space and r = body. We apply our rule $(d\mathbf{G}/dt)_s = (d\mathbf{G}/dt)_r + \boldsymbol{\omega} \times \mathbf{G}$ to the case $\mathbf{G} = \mathbf{r}$, the position of a perhaps moving projectile. This gives $\mathbf{v}_s = \mathbf{v}_r + \boldsymbol{\omega} \times \mathbf{r}$ as in 4-104. Then we apply the same rule again, this time to $\mathbf{G} = \mathbf{v}_s$ and this gives

$$(d\mathbf{v}_s/dt)_s = (d\mathbf{v}_s/dt)_r + \boldsymbol{\omega} \times \mathbf{v}_s$$

The left hand side is $\mathbf{a}_s$ and the RHS becomes:

$$(d\mathbf{v}_s/dt)_r = (d[\mathbf{v}_r + \boldsymbol{\omega} \times \mathbf{r}]/dt)_r = (d\mathbf{v}_r/dt)_r + \boldsymbol{\omega} \times (d\mathbf{r}/dt)_r = \mathbf{a}_r + \boldsymbol{\omega} \times \mathbf{v}_r$$

$$\boldsymbol{\omega} \times \mathbf{v}_s = \omega \times [\mathbf{v}_r + \boldsymbol{\omega} \times \mathbf{r}] = \boldsymbol{\omega} \times \mathbf{v}_r + \boldsymbol{\omega} \times \boldsymbol{\omega} \times \mathbf{r}$$

and we end up with the big result

$$\mathbf{a}_s = a_r + 2\,\boldsymbol{\omega} \times \mathbf{v}_r + \boldsymbol{\omega} \times \boldsymbol{\omega} \times \mathbf{r} = \mathbf{F}/m$$

where the last comes from $\mathbf{F} = m\mathbf{a}$ in the inertial s frame. Then we have

$$\mathbf{F} = m\mathbf{a}_r + 2\,m\boldsymbol{\omega} \times \mathbf{v}_r + m\boldsymbol{\omega} \times \boldsymbol{\omega} \times \mathbf{r}$$

so in the body frame we can write

$$\mathbf{F}' = m\mathbf{a}_r \qquad \mathbf{F}' = \mathbf{F} - 2\,m\boldsymbol{\omega} \times \mathbf{v}_r - m\boldsymbol{\omega} \times \boldsymbol{\omega} \times \mathbf{r} = F + \text{Coriolis force} + \text{Centrifugal force}$$

Thus, in the rotating frame we see two fictitious forces as just labeled above, hence 4-107. The Centrifugal term points away from the earth in the $\hat{\rho}$ direction (if we had cylindrical coords).

Goldstein spends all of page 136 commenting on the **Centrifugal force**. Notice that we have "discovered" the presence of this force by our derivation, in addition to the presence of the Coriolis force. He shows that $\omega \sim 10^{-6}$ sec for the earth, a small number. He shows that Centrifugal is at most about 0.3% of gravity, and he points out that a plumb bob therefore does not point to the center of the earth, but points south of it if you are in the NH. This also affects the length of the resultant vector which one thinks of as the local **g**, making g smallest at the equator, so people are lighter there.

Then on page 137 he starts into the Coriolis force which pushes "projectiles" to the *right* in the NH regardless of whether the projectile is going north or south. If pure east or west, there is no Cor force. In the SH projectiles are deflected to the *left* if going north or south. At the equator, there is no Cor force no matter which way you shoot.

Goldstein does not attempt to interpret this force, but I will. Imagine the equator situation. When the bullet is launched, it has a horizontal space velocity that matches the earth's rotation. The distance ρ from earth axis to projectile is constant on the trajectory so nothing happens. Now imagine you are up at 45 degrees in the NH. You shoot north and the bullet starts with a certain large azimuthal velocity. As it travels north, this azimuthal velocity soon *exceeds* the local azimuthal velocity because ρ has shrunk! So the bullet drifts to the right because earth is rotating to the right (RH rule), having this extra azimuthal velocity. If bullet goes south, this effect is reversed, so the drift is the other way. But in both cases we get drift "to the right" of the trajectory.

Goldstein talks about four examples:

1. The projectile, which we have already discussed. You might get 1/2 degree deflection for a 100 second air-time of a projectile.

2. Wind and isobars. Because wind is deflected to the right (NH) no matter what direction it approaches a low from, in the NH you get a RH rule CCW cyclone around the low. In the SH get CW cyclone.

3. Falling object. This is a vertical projectile, if you will. You have some $\boldsymbol{\omega}$ x $\mathbf{v}$ Coriolis force here and you expect an azimuthal displacement. My physical explanation is very similar to that above. As object falls, its ρ decreases, so its original azimuthal velocity exceeds the "local azimuthal velocity" at a lower altitude, so it appears to drift to the east since that is the direction in which the earth turns. Drop something 100 feet and it will drift about .15 inch. Would perhaps apply to bombs.

4. The Foucault Pendulum. This problem has always mystified me. You know that if at the north pole, the plane of the pendulum will appear to precess one circle each 24 hours, just because the earth turns under the fixed pendulum. It is not obvious to me that there is no precession at the equator, because the problem is complicated. The point of attachment is travelling in a circle. If the pendulum at the equator were N-S in starting motion, it is hard to imagine what could cause that plane to rotate, so seems reasonable that the plane stays fixed. Same with an E-W starting position.

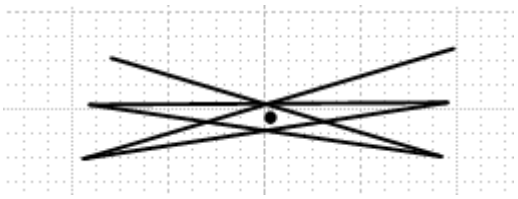
To really understand this, you have to solve the problem. Goldstein does not do this, but **Marion** does do it on pages 353-356. Let's look now at this solution! We start with three forces acting on the pendulum: gravity, string tension, and Coriolis as in (1) page 353. Marion then computes the x,y and z component of all the terms in this equation. He sets $\hat{z}$ pointing up the local vertical, and has $\hat{x}$ pointing to the south and $\hat{y}$ pointing to the east. He is using a figure on page 349 which I penciled in a bit, where λ is the latitude. All the velocity components are variables in this problem, because obviously the pendulum velocity is not constant as it swings. Using the small-angle approximation for swing angle away from vertical, we get the pair of second order ODE's shown in (8) which become (9) when α and ω_z are defined. We later see that α is and angular frequency of a free pendulum and $\omega_z = \omega\sin\lambda$, sort of a scaled down ω based on latitude, so $\omega_z = 0$ at the equator and ω at the pole.

The two differential equations are "coupled" and a clever complex notation is used to get to (12) which has the solution as described. We can use $\alpha >> \omega_z$ to simplify a bit and we get the solution shown in (19) rewritten as (20). Here, the vector $\mathbf{r}'(t)$ describes the motion of a free space pendulum (ie, our pendulum with Coriolis turned off). It could be swinging in any fixed plane. The matrix notation then shows that the pendulum appears to rotate around in the x,y plane at angular rate $\omega_z = \omega \sin\lambda$. So, for example, if it were located close to the equator, it might take 3 days to go around.

Foucault built a big one in Paris in 1851 and it rotated every 32 hours. It was a big hit! At 30 degrees latitude you would expect 2 day period.

You might think of the pendulum in this way: as it swings, it acts like a projectile, and in the NH it is deflected "to the right" on each direction of the swing, and this is what causes the circular motion. You might compute the deflection on one swing and then use that to compute the period.

A web site comments that the pendulum does not pass over its EQ position. Here I have tried to draw the "orbit" using straight lines which always deflect to the right, and this shows why that fact is true:



On each swing, the path is deflected "to the right" and fails to pass over the center point marked with a dot. In reality each of these lines is a curve bending to the right.

5. Atomic physics: rotations affect vibrations in a Coriolis manner. The rotations will create extra Coriolis forces in addition to the normal harmonic oscillator force of the potential well, this no doubt causes line blurring and various details. Our friend Herzberg has an "exhaustive discussion" of this effect, no surprise, this is the guy who wrote 6 whole books on spectra!

References: We see M&M reviewed favorably, but Jeffreys is panned for having a lousy section and a faulty picture, despite witty quotations. The usual math methods books like Courant and Hilbert are mentioned (I never had that book). Whitaker is panned for having NO diagrams at all. The last problem tells the student to go figure out the Foucault pendulum.

****

Chapter 5: Rigid Body Equations of Motion (143)

****

[Resuming here on 7.14.08, after a break of about 35 days during which we had the Lee-Rosa-Daisy visit, my visit to Cape Cod, the Chew visit, the Marilyn-Alan visit, and then a short Thomson stint.]

At this point I started my meta review, and cleaned up various items. This Chapter 5 contains another antiquated presentation presumably because "the math" was not well known in 1950 and so is here integrated into the text. // This Chapter is 41 pages long and thick with equations all of which I derived.

5.1 Angular Momentum and Kinetic Energy of motion about a point (143). We already know about the "division" idea for each [ie, X = CMS part + rotational part] . Equation 1-26 (p7) showed how you can divide up the **L**, and 1-29 (p8) how you can divide up T. It is stated that "often" V can be similarly divided. Examples: For gravity, we have only the CMS component. For **m•B** dipole, we have only the rotational component.

G then computes **L** for a set of particles in a rigid body rotating at some **ω** rate. The key fact for getting started is that $\mathbf{v}_i = \boldsymbol{\omega} \times \mathbf{r}_i$ since particles of the body are fixed in the body, this being a case of our

general theorem from Chapter 4. By brute force, we find that $\mathbf{L} = \mathbf{I\omega}$ where I is a 3x3 symmetric matrix whose elements I would write as follows:

$$I_{jk} = \Sigma_i m_i [r_i^2 \delta_{j,k} - r_{i,j} r_{i,k}] \quad \text{where } r_{i,j} = \mathbf{r_i} \bullet \hat{\mathbf{e}}_j = \mathbf{r_i} \bullet \hat{\mathbf{j}}$$

Matrix I is the inertia tensor, and we might compare to $\mathbf{p} = m\mathbf{v}$ where m ~ I. The off diagonal elements are of the form $I_{xy} = -\Sigma_i m_i x_i y_i$, diagonals are like $I_{xx} = \Sigma_i m_i (r_i^2 - x_i^2)$. The diagonal terms are the moments of inertia, the off-diagonals are the products of inertia.

[Recall that the Cartesian quadrupole moment of a charge distribution is $Q_{\alpha\beta} = \int d^3\mathbf{r}\ \rho(\mathbf{r})\ (3r_\alpha r_\beta - \delta_{\alpha\beta} r^2)$ which is traceless. The inertia tensor is similar, but without the 3, and is not traceless since it's trace will be the sum of the moments of inertia. Consider the spherical tensor moments

$$Q_{lm} \stackrel{\text{def}}{=} \int d\mathbf{r}' \rho(\mathbf{r}')\, (r')^l \sqrt{\frac{4\pi}{2l+1}} Y_{lm}^*(\theta', \phi') \quad \text{and } Q_{2,0} = \int dV\ \rho(\mathbf{r})\ (3z^2 - r^2)$$

We would express the inertia tensor as $I_{jk} = \int dV\ \rho(\mathbf{r})\ (r^2 \delta_{jk} - r_j r_k)$ for continuous mass density ρ. It then seems that we can make this claim

$$I_{jk} = \int dV\ \rho(\mathbf{r})\ (r^2 \delta_{jk} - r_j r_k) = \int dV\ \rho(\mathbf{r})\ (3\, r^2 \delta_{jk} - r_j r_k) - 2\, \delta_{jk} \int dV\ \rho(\mathbf{r})\, r^2$$
$$= Q_{jk} - 2\, \delta_{jk} \langle r^2 \rangle$$

so we can analyze I_{jk} as having a cartesian quadrupole component plus another term involving the weighted average value of r^2. Probably not too useful!]

5.2 Tensors, Dyads and Dyadics (146) A dyad [AB] is a particular class of square matrix. One writes

$$[\mathbf{AB}]_{ij} = A_i B_j$$

In quantum mechanics we would write this as [AB] = |A><B| or as AB^T in matrix notation. It is a "factorizable matrix" made from the components of two vectors, and of course not all square matrices are factorizable. Consider this example:

$$[\mathbf{r_i r_i}]_{jk} = (\mathbf{r_i})_j (\mathbf{r_i})_k$$

where $\mathbf{r_i}$ is the position of the ith particle in a rigid body. We know without thinking very hard that the identity matrix is diag(1,1,1) so the "identity dyad" would be $[1]_{jk} = \delta_{j,k}$. Thus, although the inertial tensor is not factorizable, it can be represented as a sum of factorizable terms as follows.

$$I_{jk} = \Sigma_i m_i (r_i^2 \delta_{j,k} - r_{i,j} r_{i,k})$$

$[I] = \Sigma_i m_i \ (r_i^2 \ [1] \ - [\mathbf{r}_i \mathbf{r}_i])$ // dyadic representation of inertia tensor

To clarify things, I have tried to always show the dyads with square brackets. Any linear combination of dyads is called a dyadic.

It is probably easiest to use the "matrix notation" such that $[AB] = AB^T$. In this notation with real numbers of course we know that $A^TB = A \bullet B$, the usual dot product. Here are some examples:

$[AB] \bullet C = AB^TC = A(B^TC) = A\ (B \bullet C)$ // G bottom page 147

$C \bullet [AB] = C^TA\ B^T = (C \bullet A)\ B^T$

Think of A as a column vector and A^T as a row vector, the transpose of the column vector.

Now how might you combine two dyads? Consider first "multiplication" :

$[AB]\ [CD] = (AB^T)\ (CD^T) = A\ (B^T\ C)\ D^T = (B \bullet C)\ AD^T = (B \bullet C)\ [AD]$

// noted added 2.25.17: $(A \otimes B)\ (C \otimes D) = (B \bullet C)\ (A \otimes D)$ I guess. Not sure what this means.

Another interesting combination of course is this:

$A \bullet [BC] \bullet D = A^T\ BC^T\ D = A^TB\ C^TD = (A \bullet B)(C \bullet D)$

For some strange reason, G writes this also as [BC]:[AD] which I find is pretty useless.

So now here is an example of this last combination

$n \bullet [I] \bullet n = \Sigma_i m_i \ (r_i^2\ n \bullet [1] \bullet n \ - n \bullet [\mathbf{r}_i \mathbf{r}_i] \bullet n) = \Sigma_i m_i \ (r_i^2 \ - (\mathbf{r}_i \bullet n)^2)$ // G 5-17

// noted added 2.25.17: $n \bullet [I] \bullet n = \langle \mathbf{n} \mid I \mid \mathbf{n} \rangle$

In normal matrix notation we would write

$$I_{jk} = \Sigma_i m_i \ (r_i^2\ \delta_{j,k} - r_{i,j}\ r_{i,k})$$

$$I = \Sigma_i m_i \ (r_i^2\ 1 \ - r_i\ r_i^T)$$

$$n^T\ I\ n = \Sigma_i m_i \ (r_i^2\ n^T n \ - n^T r_i\ r_i^T\ n) = \Sigma_i m_i \ (r_i^2 \ - (\mathbf{r}_i \bullet n)^2)$$

In my opinion, the dyadic notation adds nothing new whatsoever, the matrix notation is much clearer, though it requires ^{T} superscripts to keep things straight. The $\delta_{i,j}$ notation is also quite clear.

One more thing. Consider this:

$$e_1 e_2^T = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} (0\ 1\ 0) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \text{ what is meant by } \hat{1}\,\hat{2}$$

So then it is pretty clear that $[1] = 1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \hat{1}\hat{1} + \hat{2}\hat{2} + \hat{3}\hat{3}$ as in 5-13. G likes to use i,j,k for 1,2,3, this is probably some old tradition of 1950. In general, $\hat{m}\hat{n}$ is a 3x3 matrix with a single 1 in the mth row and the nth column, and of course this explains the notation such as

$[AB] = A_1B_1 \hat{1}\hat{1} + A_1B_2 \hat{1}\hat{2} + ...$ /each term is a matrix // G 5-12

So the poor student has to put up with this dyad smokescreen right in the middle of trying to learn rotational dynamics!

Question #1: Does matrix I vary in time for a rigid body? I think the answer is no. Suppose we select the principle axes to evaluate I so that I = diag(I_1, I_2, I_3). These three numbers are of course constants determined by the geometry of the rigid body. If you stupidly pick some other set of axes to compute I, you get 9 numbers, but they too are all constants. So the matrix I is like the quantity m = mass. It is a property of a rigid body and it does not vary in time. Yes, matrix I is a function of origin and axes chosen to do the evaluation, but we make that choice and we stick with it.

5.3 The inertial tensor and the moment of inertia (149). Just as we computed **L** earlier, we now compute T again using $\mathbf{v_i} = \boldsymbol{\omega} \times \mathbf{r_i}$ for particle in a rigid body. The result is $T = \omega \bullet \mathbf{I} \bullet \omega/2$ as simple algebra shows, which we compare to vmv/2 for linear action. We also find that $T = \omega \bullet L/2$ which is comparable to $T = vp/2 = (1/2)mv^2$.

Now we know that at some time instant, we can write $\boldsymbol{\omega} = \omega\hat{\mathbf{n}}$. Thus

$T = \omega \bullet \mathbf{I} \bullet \omega/2 = (\omega^2/2)\, n \bullet \mathbf{I} \bullet n = (\omega^2/2) \Sigma_i m_i (r_i^2 - (\mathbf{r_i} \bullet n)^2)$ // from above

The scalar $n \bullet \mathbf{I} \bullet n = \Sigma_i m_i (r_i^2 - (\mathbf{r_i} \bullet n)^2)$ is defined as I, called the **moment of inertia about the axis of rotation n**.

$I \equiv \mathbf{n} \bullet \mathbf{I} \bullet \mathbf{n} = \Sigma_i m_i (r_i^2 - (\mathbf{r_i} \bullet \mathbf{n})^2)$

G then shows that this "I" agrees with the conventional definition as the sum over particle masses weighted by their perp distance to the rotation axis squared.

The final act of this section is the proof of a simple theorem:

$I_a = I_b + I_{cms,a}$

where a and b refer to two parallel rotation axes, but b passes through the CMS. The last term represents the quantity Md^2 where d is the distance between the two parallel axes. So as you move your axis a further from the CMS of a rigid body, you get more inertia.

5.4 The eigenvalues of the inertia tensor and the principle axis transformation (151). This is another obfuscated section that is really quite simple. Since matrix I is symmetric and real, it is hermitian, and we know all about the eigenvalues and eigenvectors of a hermitian matrix from our general matrix studies.

First, we consider $I\psi = \lambda\psi$, just making it look like QM. We solve $\det(I-\lambda 1) = 0$ to find the three (known to be real) eigenvalues λ which he calls $I_{1,2,3}$. If they are all different, we know that the three eigenvectors are orthogonal and that they form a matrix which brings I to diagonal form by similarity.

$X^T I X = I_{diag}$ $\qquad$ $X = \{ \psi_1, \psi_2, \psi_3 \}$ = what he calls $\{ \mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3 \}$

If two eigenvalues are equal, we do GSO to still end up with three orthogonal eigenvectors. If all three are equal, then $I_{diag} = \lambda\, 1$ and then $I = X\, I_{diag}\, X^T = \lambda\, 1$ so I is diagonal to begin with, as G states.

The eigenvalues $I_{1,2,3}$ are called the **principle moments of inertia**, and the three axis R_i are called the **principle axes**. We still have our scalar $I = \mathbf{n}\bullet\mathbf{I}\bullet\mathbf{n}$ which the **moment of inertia relative to n**. If we rescale our unit vector to be $\boldsymbol{\rho} = \mathbf{n}/\sqrt{I}$, then of course we have $1 = \boldsymbol{\rho}\bullet\mathbf{I}\bullet\boldsymbol{\rho}$ which we write out as in 5-28 on page 155 and this is an equation of an **inertia ellipsoid** in ρ-space having some arbitrary orientation. The three symmetry axes of the ellipsoid are the principle axes of the problem as is not hard to show. The coefficients of the ρ terms are of course the I_{ij} matrix elements, hence the name.

Finally, one can write $I = MR_o^2$ to define an "effective" perp distance R_o of an point object of CMS mass M which would have the same I as the rigid body, and R_o is called the **radius of gyration**.

Comment added 3.12.17. Defining $\boldsymbol{\rho} \equiv \hat{\mathbf{n}}/\sqrt{I} = \hat{\boldsymbol{\omega}}/\sqrt{I}$ where I is the thing in $T = (1/2)I\omega^2$, you can then talk about "ρ-space" with coordinates ρ_x, ρ_y and ρ_z. The equation $I = \mathbf{n}\bullet\mathbf{I}\bullet\mathbf{n}$ becomes $1 = \boldsymbol{\rho}\bullet\mathbf{I}\bullet\boldsymbol{\rho}$ and for any tensor I_{ij} this equation describes the surface of an ellipsoid in ω-space. If your "r-space" basis is chosen so that I is a diagonal matrix, then the ρ-space ellipse aligns with the axis of ρ-space. Note that r-space and ρ-space are two completely different spaces.

5.5. Methods of solving rigid body problems and the Euler equations of motion (156). Everything is now in place to solve complicated top problems, it took 157 pages to get to this point! We have our Lagrangian as in 5-31, where now x,y,z without primes refer to the body axis which are chosen to align with the rigid body's principle axes -- don't miss this point! Since we have a fixed point by assumption, the CMS cannot be moving, so T is given by $T = (1/2)[\, I_1\omega_x^2 + I_2\omega_y^2 + I_3\omega_z^2]$. We now use the Lagrange equation of the form 1-50 on page 18 where T not L appears. The coordinates q_i are Euler angles which tell how to get from space to principle (body) axes. From page 52 we found that the generalized force Q_j in this situation is $\mathbf{n}\bullet\mathbf{N}$ where **n** is the rotation axis, and **N** is the torque. In our Euler angle picture page 107, we recall that the ψ rotation is about one of our body axes (z' here called z), so we set $\hat{\mathbf{n}} = \hat{\mathbf{z}}$ and this gives us our Lagrange equation as in 5-33, driven by torque N_z. Ie, we are simply applying our Lagrange equation for the coordinate ψ and doing the various derivatives. Recall that constraints have to be holonomic to be able to use 1-50. Having obtained 5-33, we appeal to symmetry to get the other two equations. We would permute the definitions of x',y',z'. I like saying it this way: we recognize that equation 5-33 is one component of this vector equation:

$I\dot{\boldsymbol{\omega}} + \boldsymbol{\omega} \times (I\boldsymbol{\omega}) = \mathbf{N}$ $\qquad$ that is $\quad I\dot{\omega}_z + \omega_x(I_2\omega_y) - \omega_y(I_1\omega_x) = N_z$

From rotation group argument, all three components of the equation must be true and we get 5-34 which are called "**Euler's equations of motion for a rigid body with one point fixed**" (such as a top).

We now get a very smart second derivation of these same equations. We known $\mathbf{N} = d\mathbf{L}/dt$ in the space inertial frame. We then say

$N_x = (dL_x/dt)_{space} = (dL_x/dt)_{body} + (\omega \times L)_x = (dL_x/dt)_{body} + \omega_y L_z - \omega_z Ly$

so we then have

$(dL_x/dt)_{body} + \omega_y L_z - \omega_z L_y = N_x$ or, since $L = I\omega$,

$I_1 (d\omega_x/dt)_{body} + \omega_y \omega_z I_3 - \omega_z \omega_y I_2 = N_x$

$I_1 (d\omega_x/dt)_{body} - \omega_y \omega_z (I_2 - I_3) = N_x$

Question 2.25.17: In $\mathbf{L} = I\boldsymbol{\omega}$, is **L** the body or space angular momentum?

Remember that we can evaluate **ω**'s components in any frame we want in this equation. The other two equations are now obvious from the same derivation. Let's restate the above on more time:

$(d\mathbf{L}/dt)_{body} + \boldsymbol{\omega} \times \mathbf{L} = \mathbf{N}$ and recall that $\mathbf{L} = I \bullet \boldsymbol{\omega}$

If there were no torques on the system, you would say

$(d\mathbf{L}/dt)_{space} = 0$
$(d\mathbf{L}/dt)_{body} = -\boldsymbol{\omega} \times \mathbf{L}$

This reminds us that the vector **L** appears to move as observed from the body frame which is non-inertial. So in the body frame of reference, you would have $\mathbf{L}_{body} = \mathbf{L}_{body}(t)$ = varies in time.

5.6 Force-free motion of a rigid body (159). We now want to think about solutions for how a rigid body moves in the absence of any torque, so 5-36 applies.

Before doing the actual solution, we are going to take a look at **Poinsot's Construction** (circa 1830). Goldstein normalizes things a little differently from Poinsot, so we have to pay attention.

First, go back to page 155 and notice this definition:

$\boldsymbol{\rho} = \mathbf{n}/\sqrt{I}$ where $\boldsymbol{\omega} = \omega\mathbf{n}$ where **n** = instantaneous rotation axis

The equation $I = \mathbf{n} \bullet I \bullet \mathbf{n}$ says that $I = I_{xx} n_x^2 + I_{xy} n_x n_y$ + etc. The three vectors **n**, **ω** and **ρ** are always along the same line, they just have different lengths. We can talk therefore about an inertia ellipsoid in any of these three spaces.

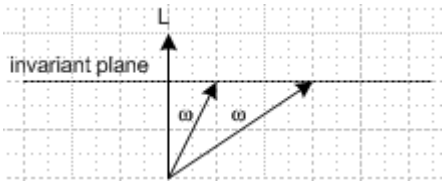
Consider these various equations:

$\mathbf{L} = I\boldsymbol{\omega}$ $2T = \boldsymbol{\omega} \bullet I\boldsymbol{\omega} = \boldsymbol{\omega} \bullet \mathbf{L} = I\omega^2$

If we measure things in the inertial space (lab) frame, we could write things like this

$\mathbf{L} = I\boldsymbol{\omega}(t)$ = constant of motion $2T = \mathbf{L} \bullet \boldsymbol{\omega}(t)$ = constant of motion

Because T and **L** are constants, the right side equation says that **ω**(t) moves in such a way that it maintains a constant dot product with **L**. That means **ω**(t) always lies on a plane known as the **invariable plane**. The distance of the vector origin below to this plane is ω cosθ = ω **L** • **ω**(t)/Lω = **L•ω** /L = 2T/L = Iω²/L. If we compute things from the space frame but using the body axes which are the principle axes, we find that $L_x = I_1\omega_x$, $L_y = I_2\omega_y$, $L_z = I_3\omega_z$, or write this as $L_i = I_i\omega_i$.



Because **L** = I**ω**(t), we are reminded of the fact that **ω** need NOT point in the same direction of **L**, something I have confused in the past. The angular velocity does not point in the same direction as the angular momentum in the general case. [Geoff Chew was talking to me about the linear version of this?]

Now we first had our ellipsoid on page 155 written in n space, to wit

$I = I_{xx}n_x^2 + I_{xy}n_xn_y + ...$ in n space

$= (1/\omega^2)\,[I_{xx}\omega_x^2 + I_{xy}\omega_x\omega_y + ...\,]$ in ω space

since **ω** = ω**n** . We can write this last as

$2T = I\omega^2 = [I_{xx}\omega_x^2 + I_{xy}\omega_x\omega_y + ...\,] \equiv G(\boldsymbol{\omega})$

which might then be called the "**kinetic energy ellipsoid**" in ω-space because its scale is set by T. If we were to take our space axes to be instantaneously aligned with the body axes and to have the origin at the center of the ellipsoid [huh?] , we would then have

$2T = I_1\omega_x^2 + I_2\omega_y^2 + I_3\omega_z^2$

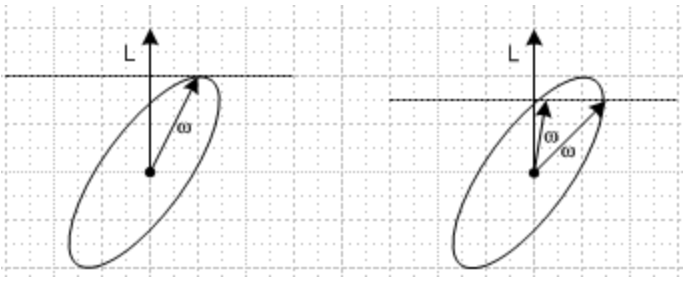
We can put this in standard form as follows

$1 = (\omega_x/a)^2 + (\omega_y/b)^2 + (\omega_z/c)^2$

where then a,b,c are the semi-axes of the ellipsoid, see picture page 51 of Schaum. Thus:

$a^2 = 2T/I_1$ $b^2 = 2T/I_2$ $c^2 = 2T/I_3$

Now I am confused about the geometric relationship of the ellipsoid to the invariable plane. All points on the ω-ellipsoid are "legal" values of **ω** from the point of view of the T ellipsoid equation. But we also have that requirement that ω must lie on the invariable plane. This gives the following possibilities:

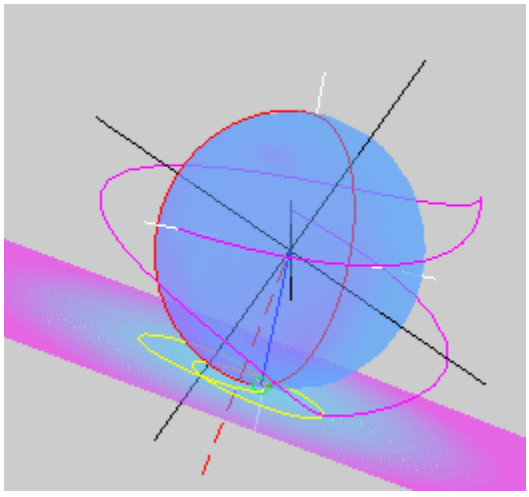


We are now going to show that the case on the right is not viable. Assuming we did our I in the principle axes, we know that $L_{\mathbf{x}} = I_1 \omega_{\mathbf{x}}$, for example. Consider the function $G(\boldsymbol{\omega})$ given above which is the equation for the ellipsoidal surface for the ellipsoid with a,b,c as stated above. It is easy to show that

$$\nabla G(\boldsymbol{\omega}) = 2\,\mathbf{L}$$

This says that the "solution" value of $\boldsymbol{\omega}$ must be such that the tangent at the solution point on the ellipse points in the direction of **L**, which is up in our pictures. That is why only the left picture applies. This means that the ellipsoid *touches* the invariant plane at a single point. We know that this point of contact is a point on the ellipsoid that has zero velocity because, after all, the ellipsoid is rotating instantaneously about the ω axis. [There is another point on the ellipsoid with this property, but it is not on the invariant plane.] This is similar to a car tire at the point of contact with the ground, velox = 0. Like the tire, the ellipse is therefore "rolling without slipping" on the invariant plane. This means the ellipse cannot roll unless it's point of contact on the plane moves (unless it rolls on a symmetry axis).

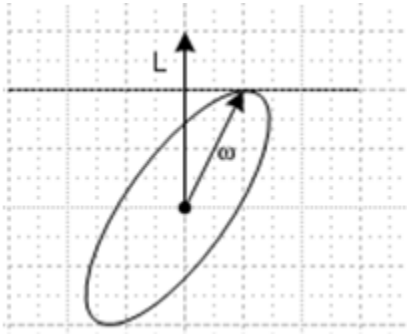
So this is the construction of Louis Poinsot. The ellipsoid rolls without slipping on the invariant plane. The origin of $\boldsymbol{\omega}$ space does not move, so the center of the ellipse stays put, and of course the invariant plane also stays put. Alternatively, the distance from the ellipsoid center to the invariable plane is constant, due to the projection idea mentioned above. Exactly *how* the ellipsoid tumbles requires a solution of the problem from the Euler equations, but we know the nature of the solution by this Poinsot construction. It is similar to doing a general study of the orbits of a central force. Here is a frame from an animation at http://www.aero.iitb.ac.in/~bhat/asymmetric.html which shows a complex example:



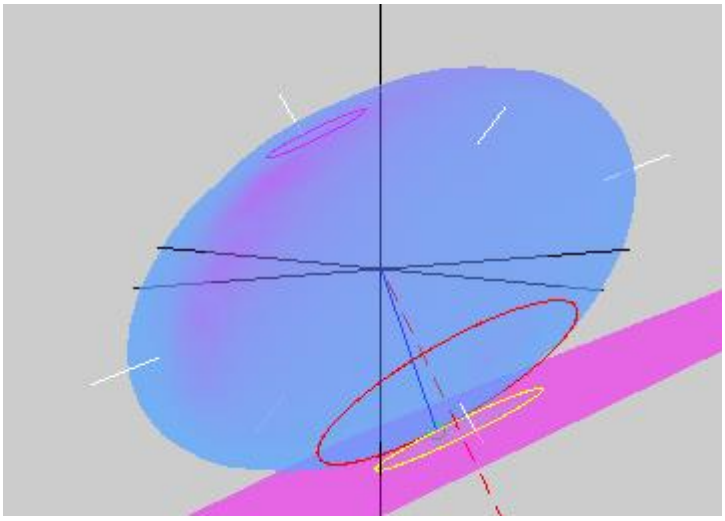
The ellipsoid has all three semi-axes different, and is started in an unstable place (rotating around the middle inertial axis), so the two curves of interest are quite complex. The yellow one on the invariant

plane (trace of the point of contact) is called the **herpolhode**. The herpolhode of course lies entirely on the plane and is a history of the point of contact. The red curve on the surface of the ellipsoid is the **polhode**, which is the history of the contact point on its ellipsoidal surface. The "pink" curve is the history of one of the extremal points of the ellipsoid and shows the amount of violent tumble action that is going on here.

Going back to our picture above



If the ellipsoid had two axes equal (were an ellipsoid of revolution b=c) you can see that the only possible motion of ω would be in a circle on the invariable plane of radius ωsinθ, and in this case **ω** travels on a cone. Note that the point of contact above is NOT at the ellipse extremum, so as the ellipse rolls, the point moves in the circle, try this with a zucchini. Here from the same animation site,



http://www.aero.iitb.ac.in/~bhat/axisymmetric.html

and you see that both the polhode and herpolhode are circular. Perhaps this is what a **near-spiral football** is doing. Remember that the short ellipse axis corresponds to the largest moment of inertia, which for the football is the long direction. So this spiral is wobbling about the spiral axis. This is an example of precession. Rigid body motion is not particularly simple, but it is solvable. The same site shows some cones rolling on cones, but I have not bothered with that detail right now.

The situation where $I_1 = I_2$, meaning object is "symmetric" in x and y. This does not mean the object is a solid of revolution, however. A square post is an example. The ellipsoid is a solid of revolution, but not the actual physical rigid body. G does not make this point, but web does. This discussion starts bottom third of page 161 where we see the simplified Euler equations. At once we find that ω_z = constant, and we find that $\boldsymbol{\omega}_T$ precesses at frequency $\Omega = (1-I_3/I_1)\omega_z$, reminds me of my Levitt math. This is exactly the situation shown in my screen clip above. Notice that Ω can have either sign. We can write $\boldsymbol{\omega}_T = \omega_T(\sin\Omega t\ \hat{\mathbf{x}} + \cos\Omega t\ \hat{\mathbf{y}})$ and the precession amplitude ω_T is a motion constant set by initial conditions. At top of page 163 we see these facts:

$$T = \tfrac{1}{2} I_1 \omega_T^2 + \tfrac{1}{2} I_3 \omega_z^2 \qquad L^2 = (I_1 \omega_T)^2 + (I_3 \omega_z)^2 \qquad // \ L_x = I_1 \omega_x \ \text{ eg}$$

which seem pretty obvious. Think T and L^2 as determining ω_T and ω_z.

G uses the earth as an obvious example. Because it is slightly oblate, it will precess while it rotates. If it really were a perfect rigid body, the prediction is that $\Omega = \omega_z/300$ which means the north pole should do a circle of some size every 300 days or 10 months. This is the torque-free effect not to be confused with the 26,000 year thing we will look at later. In reality, see a 427 day period with lots of noise, probably due to mass shifts and elastic properties of the earth.

PL likes the football as a better example, since familiar to most people. The thing wobbles unless launched with perfect spiral.

5.7 The heavy symmetrical top with one point fixed (164). This is the main case study always used in rigid body motion, we have all seen it, gyroscopes, you name it. The position of the top is described

exactly by the Euler angles, which is no doubt why they were invented in the first place. The tilt-down is θ, the fast spin axis angle is ψ, and the precession azimuth is φ. You have to look back to page 107 to fully appreciate these angles.

This problem has a torque due to gravity acting on the CMS. The direction of the torque is along the line of intersection of the initial xy plane and the tilted down equatorial plane, the "line of nodes". You see this by doing **r** x **F** with your hand. This means that components of angmom in the ψ and φ directions will be conserved, but not that on the θ direction. Correspondingly, we see that ψ and φ are cyclic and don't appear in the Lagrangian shown top of page 165 (I derived all checked equations). So we have constants of the motion L_ψ and L_φ, but he uses the generalized coordinate notation p_ψ and p_φ . Notice that both of these are functions of $\dot{\psi}$ and of $\dot{\phi}$, see mid page 165. Again, don't confused **L** with **ω**. Jumping to the middle picture on page 168, you might at first wonder now p_φ is constant when the precession goes backwards, such that $\dot{\phi}$ shows both signs during the motion (as does $\dot{\theta}$). But from 5-47, you see that $\dot{\psi}$ must compensate and go more positive when $\dot{\phi}$ goes negative to yield a constant p_φ. You might think of $\dot{\psi}$ as a constant, but it is not constant!! In fact, 5-51 shows that $\dot{\psi}$ is a function of the tilt-down angle θ which varies in time.

So we have p_ψ and p_φ as two constants of the motion, which G represents by ω dimensioned constants a and b as in 5-46 and 5-47. A third constant is total energy E as in 5-48. We then derive expressions for $\dot{\phi}(\theta)$ and $\dot{\psi}(\theta)$ in 5-50 and 5-51, and then we go on to write an ODE for θ. This ODE has the form $\dot{u}^2 = f(u)$ where $u = \cos\theta$ and f(u) is a cubic function of a,b,α,β where now α and β are constants in terms of energy E and $Mg\ell$ and indirectly a. [a and b replace constants p_ψ and p_φ]. Notice that this is not a linear ODE, but it is still easily integrable as shown in 5-56. G claims that other books go into lots of details getting exactly elliptical function solutions, but he is not going to repeat that work here. Instead, he is going to do the qualitative stuff which is really much more important.

Qualitative Analysis: So we have this cubic f(u; a,b,α,β) which we plot page 167. The places where f(u) = 0 are places where $\dot{u} = 0$ and hence where $\dot{\theta} = 0$, and hence are the θ "turning points" for the motion. Only the roots u_1 and u_2 are physical since in -1,1 range. The idea then is that u(t) moves in a way that it bounces back and forth between u_1 and u_2 in some manner, and this means that θ(t) has the same behavior and we see this in all three figures on page 168 top: θ is moving between the two circles in all the cases. By the way, the top's symmetry axis is historically called "the figure axis", and those curves are the path of the top of the top on the surface of a sphere whose angles are the θ and φ Euler angles.

We then examine three cases based on the size of u' = b/a relative to u_1 and u_2. If u' lies between the other two, then (b-au) = a(b/a - u) = a(u' - u) can have either sign as motion progresses, $u = \cos\theta$. But from 5-50, this means that $\dot{\phi}$ will have both signs, and this gives the curly middle picture on page 168. If u' is off to the right [or I suppose off to the left], then $\dot{\phi}$ has a constant sign, and this is the left picture. Finally, if u' = one of the turning points, you get the cusp case figure 3. G then shows that this is the figure that always applies if you "release" a gyroscope in the usual manner. It starts at a cusp with these initial conditions.

So we have two phenomena of interest here: precession and nutation. For the top, if you could arrange to have $u_1 = u_2$, the nutation action would go away, and you would have motion of the top tip just going around in φ, and this is the precession effect, frequency is given by 5-50 with your constant θ inserted. Note that this frequency is complete revolutions per second of the top tip about the z axis. So first, we note that this precession frequency is a function of that angle. As $\theta \rightarrow 0$ [Note: go back to other arrows in

normal.dot!] it would seen that $\dot{\phi}$ should increase, so as top tips down as energy leaks out, expect slower precession, and I think I have seen this.

In the more general case, you do get bobbing up and down which means there is $\theta(t)$ motion, and this part of the motion is called nutation. This also has a characteristic frequency which, in the examples shown, is faster than the precession frequency, about which nothing has yet been said.

The Heavy Top Started with a Drop. Up to this point, we have not assumed that the top is "heavy" or "fast", all parameters have been general, but now we specialize to wring out more useful information. Starting with 5-58 on page 169 which says the figure axis rotational KE is much larger than the gravitational PE.

But first, without approximations yet, we are going to think of that specific initial condition that we hold the thing while we spin it up, then we "drop" it to get things going. In this case, we know that our θ_0 will be the θ_2 shown in the page 168 drawings (in particular, the right one for the drop method). At t=0, we have both $\dot{\phi} = 0$ and $\dot{\theta} = 0$. The fact that $\dot{\phi}$=0 from 5-50 forces a relationship between a and b which is that $b/a = u_0 = \cos\theta_0$. Then the fact that $\dot{\theta} = 0$, which means $\dot{u}$=0, forces $f(u_0) = 0$ which then forces $\alpha/\beta = u_o$. We then just rearrange f(u) as shown in 5-59, which exposes our known zero at $u=u_0$, and our problem is now to find the other relevant root which will correspond to the lower circle at θ_1. Using 5-60, we could write an exact expression for this u_1 using the quadratic formula. This quadratic is recast in the form 5-61 where still no approximations have been made, just our starting method has been assumed.

At this point, we assume that the dimensionless ratio a^2/β is large and we are then in the heavy top approximation. Our quadratic then has the two solutions shown, but only the first is the physical one (so we know the other one is the one to the right of u=1 in the sketch on page 167). We then get the answer to our immediate question which answer is this:

$$u_0 - u_1 = \cos\theta_0 - \cos\theta_1 = x_1 = (I_1/I_3)\,(2Mg\ell/I_3\omega_z^2)\sin^2\theta_0$$

Since we know that θ_1 is a larger angle, we know that $\cos\theta_0 - \cos\theta_1$ is positive, the larger angle has the smaller cosine. We can solve to get

$$\cos\theta_1 = \cos\theta_0 - (I_1/I_3)\,(2Mg\ell/I_3\omega_z^2)\sin^2\theta_0 = \cos\theta_0 - x_1$$

This shows clearly that as $\omega_z \rightarrow \infty$, the nutation amplitude decreases to 0. A slower heavy top has more nutation amplitude that a fast one.

Next, we want to solve for $u(t) = \cos\theta(t)$. The solution is given by 5-64 which we arrange as

$$u(t) = \cos\theta(t) = \cos\theta_0 - (x_1/2)\,(1 - \cos(at))$$

Now we can interpret u in the page 164 figure as the normalized height of the top's tip, and we see that this height is nutating in a trig manner with nutation frequency $a = (I_3/I_1)\omega_z$. The actual angle $\theta(t)$ is a messier function, but we get the idea just fine. In order to get the above, we used the heavy top assumption again to approximate $\sin\theta = \sin\theta_0$ to get 5-63 and then 5-64 and hence the above.

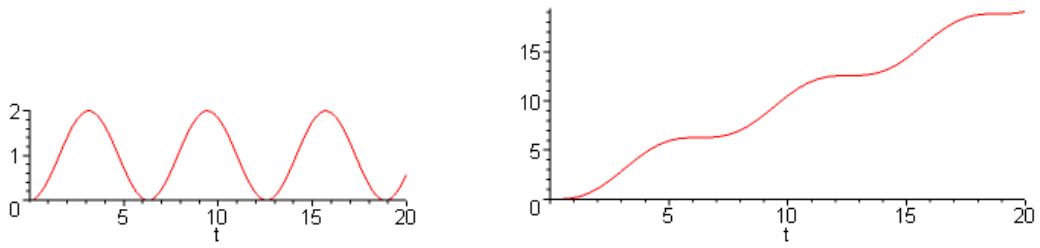
The precession situation is described now by 5-66. This says that $\omega_{precession} = \dot{\phi}$ always has the same sign, but varies between 0 and a max value of β/α. You can see this in the rightmost picture on page

168. So unlike the nutation situation in which the "angle", at least in u(t), is a trig of time, here it is the precession *frequency* which is a trig of time. We get

$\dot{\phi} = (\beta/2a)[1 - \cos(at)] = (\beta/a)\sin^2[(a/2)t]$

$\Rightarrow \varphi(t) = (\beta/2a)[\, t - (1/a)\sin(at)\,]$

and here are some plots with a=1 and β/2a = 1 so β=2: [$\dot{\phi}(t)$ on the left, φ(t) on the right]



Clearly the average $\omega_{precession} = \langle\dot{\phi}\rangle = (\beta/2a) = Mg\ell/I_3\omega_z$ and we then see that $\omega_z \rightarrow \infty$ takes the precession frequency down to 0. So, a very very fast top or gyroscope will precess very slowly. As it slows down, expect faster precession. We already saw how fast reduces the amplitude of nutation, but according to 5-65 it increases the frequency of nutation .

At this point on page 171, G gives a summary of what we have learned about the drop-the-top case. He notes that the fast-top small nutation usually goes away in practice from frictional effects. Some have called this **pseudoregular precession**, meaning there would be some nutation if not for the friction.

How might you start the top off to really have no nutation so you have true "regular precession" ? This requires that $u_1 = u_2$ so f(u) has a double zero. He shows that this is the same as having $\ddot{\theta}(0) = 0$. He shows that, for this to happen, the values of $\dot{\phi}(0)$ and $\dot{\psi}(0)$ and θ(0) must satisfy 5-70 or 5-69. For example, we have to rule out $\dot{\phi}(0) = 0$. If you pick some of these items, then there will in general be two solutions for $\dot{\phi}(0)$ that will work. These give two different precession rates with no nutation. Details are not pursued. But in general, to start a top off at some $\theta_0 > 0$ with no nutation, you need to push it in just the right way (two choices), or it will nutate on you. Yes, it was an interesting question.

Heavy top started vertically (p 173). This means certainly that $u_1 = \cos\theta_1 = 1$ in the p 168 pictures. The question is: what happens when you just let it go? Since $L_\psi = L_\varphi$ in this case, you get a = b. G also shows that α = β. And he shows that with such relations between the various constants, the f(u) curve in fact has a double zero at u=1, and the third root may or may not be in the physical range depending on the values of ω_x, $Mg\ell$, I_1 and I_3. He shows that if $\omega_z > \omega_{critical}$, then that third zero is out of range and the top just sits there. For ω_z less than this critical speed, the third zero u_3 moves in range and the top then does regular nutation between u = 1 and u_3. That is, it bobs always up to θ=0 and then down to some $\theta = \theta_3$. This explains what we have all seen. You start a top, it stays vertical but loses speed from friction. When

ω falls below critical, it starts to bob (nutate) and precess. It moves from the sleeping state to the wobbly state. As that zero moves more to the left, the nutation amplitude increases, etc etc.

Some Applications (p174).

(1) Precession of the Equinoxes. Because the earth is tilted relative to the sun and is oblate, the gravitational pull of the sun is more on the near side than the far side (page 175 picture pencil) and this puts a torque on the earth. The torque direction varies during the year I think (he does not say this, and in my model there would be no torque right at the equinoxes). The moon also does this, and the whole thing is a big mess. The effect is weak, hence the 26,000 year period, and the effect is obviously irregular. Called astronomical nutation. This is just one paragraph, so not much detail.

(2) Gyroscopes. This is a torque-free mounting of a top in a 3D gimbals system. As you move the external holder, the top keeps pointing in the same direction. The toys we have all seen do not have this full mounting, they are really just tops with a holder which lets you easily spin them up. I should have one but I don't. With the toy, the whole thing acts as a top.

(3) Gyrocompass. Not enough detail is given in this small paragraph to merit notes. I think modern ones are just gyroscopes with digital logic, but older ones were mounted in the x-y earth plane and worked that way somehow. I would need to go read about this in some specialized source.

5.8 Precession of charged bodies in a magnetic field (176). Here we get a nice derivation of Larmor precession. For a rotating classical charge distribution of, say, all electrons, you can compute **M** and you find that **M** = (e/2mc)**L** as in 5-76. Precession is uniform, there is no nutation. In the last paragraph, G says that |**M**| might depend on orientation in some situation and cause us to need a small correction. This is very obscure, but G wrote a little paper about it and threw it in. Here is the abstract of this paper, which says there is a small nutation involved in the classical solution.

The Classical Motion of a Rigid Charged Body in a Magnetic Field

Herbert Goldstein
Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received June 1, 1950)

A frequently used classical analog to the quantum precession of a magnetic moment in a magnetic field is the motion in a uniform static magnetic field of a symmetrical **rigid body charged** so that the ratio of charge to mass density is uniform. In general, the figure axis of such a magnetic top **both precesses and nutates**, and the mechanism producing the nutation is qualitatively different from that responsible for the nutation of the familiar gravitational top. Unless the body is spherical the angular momentum vector also nutates in the course of the precession, a phenomenon caused by the deviation of the torque on the body from the customary, but here incorrect, expression **M×B**. The uniform precession frequency of the magnetic top is *not* the Larmor frequency but involves in addition terms depending on higher powers of the magnetic field, a correction which corresponds to the quadratic Zeeman effect. ©1951 *American Association of Physics Teachers*

I read all the book reference notes, lots of stuff. Amused to see that Klein and Sommerfeld during 1897-1910 wrote a huge 4 volume Theorie des Kreisels, Theory of Tops. This was the kind of stuff people worked on in the late Victorian era. There was no QM to divert efforts. The pure history of all this stuff is

in itself interesting, you could do this forever. Comment on boomerangs. Arnold Sommerfeld was certainly a heavy lifter in the physics history, and a serious teacher and documenter.

Question: Did I ever read this chapter in my class? I certainly don't remember it. I have squiggled blue ink and also orange underlining on page 143-158, but not on the remaining pages. Probably Sydney Coleman made us read part of this but then we moved on? He probably wanted to get us into the Hamiltonian formalism rather than do all this ancient top stuff.

****

Chapter 6. Special Relativity in Classical Mechanics (p 185)

****

6.1 The basic program(185). Postulates: #1: speed of light c is the same in all inertial frames; #2: Laws of physics are the same in all inertial frames, you cannot tell whether you are moving constant v or not in any absolute sense. Therefore, Galilean velocity addition must be wrong. The program then has two parts: (1) find the correct transformation law and resulting velocity addition law; (2) adjust equations so they are covariant, ie, have the same form in all inertial frames (including boosts).

6.2 The Lorentz Transformation (187). Consider spherical light wave in two frames to get 6-7. This gives one condition of the transformation. To get the required minus sign, define x_4= ict and write as 6-8. Ie, don't use a metric tensor, just do it this way (the unpopular way in my era). Then the transform matrix is going to be orthogonal $AA^T = 1$ and not unitary (since it will contain complex elements), as in 6-10. Pick a boost direction z and reduce the 4x4 desired matrix to p 189 bottom, block diagonal. Then consider the origin of one system as seen by the other, and you end up with 6-5 where Goldstein does not use $\gamma = 1/\sqrt{1-\beta^2}$. The main conclusions are in 6-17 where we see how things change along the boost direction, both z and t are affected, not x and y. These results are of course independent of the metric or notation chosen. G then quickly shows **length contraction** and **time dilation**, though he uses the older term time dilatation which is 40x less popular on the web. If you measure a rod which is at rest in another frame, it appears shorter by factor $1/\gamma$ along the boost direction. A clock in another frame appears to run slower by factor γ, time dilation. The **rule for addition** is easily found by applying two boost matrices in sequence, and the result is $\beta' = (\beta_1 + \beta_2)/(1 + \beta_1\beta_2)$. If both starting betas are 1, so is the sum.

6.3 The covariance of equations (194). If a_μ is a 4-vector, then $a_\mu a_\mu$ is a scalar. Lorentz scalars are called **world scalars** by G. A nice example is to define $d\tau^2 \equiv dx_\mu dx_\mu$ and call τ the **proper time**. In a rest frame of a clock, $dx_i = 0$ and then $d\tau = dt$. In another frame, a pair of clock ticks are spaced by $\mathbf{dt = \gamma\, d\tau}$ so this is consistent with our time dilation notion.

If X_μ is some 4-vector, then the sign of X•X can be plus, minus, or zero and the vector is then called **space-like**, **time-like** or **light-like** (G does not use that last term).

So now we start to "fix things up". First idea is to define $\boxed{u_\nu \equiv dx_\nu/d\tau}$ which will be a 4-vector. We will then have $u_i = (dx_i/(dt/\gamma)) = \gamma\, dx_i/dt = \gamma\, v_i$. For low speed, they are the same. It turns out that $u_0 = i\gamma c$ to transform correctly, and then $u_\nu u_\nu = -c^2$ = a world scalar of course, and time-like.

Next, the **wave equation** 6-26 we are all used to can be written $\partial_\mu\partial_\mu\, \psi = 0$ and he shows that ∂_μ is a 4-vector operator and ψ is then a scalar, hurray.

6.4 What happens to F = ma ? (199) This is the big question of course. Newton says $d(mv_i)/dt = F_i$ so we are tempted to fix this up to say $d(mu_\nu)/d\tau = K_\nu$ where the LHS = a 4-vector, and we hunt for a RHS 4-vector which will reduce to F_i at low speed. At this point G takes two alternate paths.

First path: consider the E&M force which you suspect is already relativistically correct. We know the force F_i on a particle from E&M force and we showed earlier in the book, page 21, that you could write it as shown in 6-31 which seems a bit of a mess. But then you imagine a 4-vector $A_\mu = (A_i, c\varphi)$, though justification is not given, then you can write F_i as in 6-33. Notice 6-32. Well, we are led to postulate a K_μ as shown below 6-34, and then we know that $F_i = K_i/\gamma$. So at γ=1, it gives the correct RHS. So here it is:

$$d(mu_\nu)/d\tau = K_\nu = (q/c)\,(\partial_\mu[\, u_\nu A_\nu] - \partial_\tau A_\mu\,)$$

Both sides are clearly 4-vectors, and both sides have the right $d(mv_i)/dt = F_i$limit for low speed. So we have a good candidate at least for the E&M force. But what about for other forces, like gravity say? For the EM force, we have found that $K_i = \gamma F_i$, so maybe this is true for any force. Fiddlings:

$$d(mu_\nu)/d\tau = K_\nu \qquad dt = \gamma\, d\tau \qquad \gamma d(mu_i)/dt = K_i = \gamma F_i \quad \text{so } d(mu_i)/dt = F_i \text{ at low speed.}$$

The K_ν force he calls the **Minkowski force**.

Second path: just assume that $p_i = \gamma m v_i = m u_i$. Then get $dp_i/dt = K_i/\gamma = F_i$ and then Newton's law looks the same. In general of course $\boxed{p_\mu = \gamma m v_\mu = m u_\mu}$ = the momentum 4-vector of a particle and its mass is a scalar.

Now at mid page 202. We have found that $K_i = \gamma F_i$ and he then shows what $K_4 = (i\gamma/c)$ **F•v** which is the rate of doing work by the force. If we define kinetic energy T as $\boxed{T = \gamma mc^2}$, then $dT/dt =$ **F•v** which rescues the idea that you do work to add to T. The non-rel limit is then $T = mc^2 + \frac{1}{2} mv^2$ as we expect.

Next, using $p_\nu = mu_\nu$ we find that $p_4 = iT/c$ so now **p** and T are in the same 4-vector, and now conservation of p_μ implies both conservations of momentum and KE.

He then makes some comments about inelastic collisions and changes in mc^2. In a non-rel inelastic collision, momentum is but energy is not conserved unless you include heat. Comments here are no very clear since heat is really photons.

6.5 Relativistic Lagrangian situation (205). G shows that a possible relativistic L is $L = -mc^2/\gamma - V$ if we want to use the *traditional* Euler-Lagrange equations. This gives $p_i = \partial L/\partial v_i = \gamma m v_i$ which we know is the correct conserved relativistic 3-momentum, and we get $dp_i/dt = F_i = -\partial_i V$ as the correct Newton's law. Also, if we use the same definition of H shown bottom page 206, we find that H = T+V = E, where $T = \gamma mc^2$ as we found on page 203. Finally, if we want to add an EM field's V, we use

$$V = q(\varphi - \mathbf{A}\bullet\mathbf{v}/c)$$

If we write this out, we get

$$L = -mc^2/\gamma + (q/c\gamma)\, A_\mu u_\mu$$

but these terms are not really Lorentz scalars because of the γ factors which are only rotational scalars. Still, this L gives the right canonical momentum as shown 6-52.

6.6 Covariant Lagrangian situation (207). The previous section is really only half-baked because we continued to use the non-covariant E-L equations, though we found a candidate Lagrangian that made some things come out right. The correct covariant formulation requires 6-53 for the Lagrangian integrated over proper time τ with 4-vector arguments for position and velocity. This Lagrangian is called L' and is different from the L one of the last section, and the covariant E-L equations of 6-54 are also different. What then is the correct L' ? The answer is this:

$$L' = \tfrac{1}{2}\, m u_\nu u_\nu + (q/c) u_\nu A_\nu$$

which is clearly a Lorentz scalar, and $p_\nu = \partial L'/\partial u_\nu = m u_\nu + (q/c)A_\nu$ which we know must be the correct canonical momentum.

If we then apply our new covariant E-L equation, we get the results shown top of page 210, which is basically this:

$$d(m u_\nu)d\tau = K_\nu$$

which agrees with our earlier result. K_ν is the correct EM force on a particle. So here is the covariant form of Newton's Law, at least for a particle in the presence of an EM force. We don't really know what to do in the presence of gravitational forces.

The final shot here is to show that the correct form for KE is T^2 shown in 6-59. For a free particle this was $T = \gamma mc^2$, but here you see how the presence of **A** modifies things. Notice that φ does not affect T. Easy to show that if **A**=0, you get $T = \gamma mc^2$ again.

G has some final comments: special relativity **does not work for action-at-a-distance forces** like gravity, is messy in **accelerating systems** (probably messier fictitious forces), and everything becomes a **mess for rigid body motion** which is completely non-covariant. There are a few books on this subject, but not many. It must be pretty messy if G did not attempt it. I think there is a problem with the whole idea of a rigid body in relativity, since distance between two points is boost dependent, etc. I cannot find a quick web source on this subject. I did find an interesting discussion of what happens when you continue to accelerate a "rigid" wheel up to relativistic speeds.

SR is of course a huge subject and this chapter is merely touching the surface, commenting on how earlier mechanics stuff gets generalized or adjusted. One needs to solve many problems to get an understanding of things here.

****

Chapter 7. The Hamilton Equations of Motion (p 215)

****

7.1 Legendre Transformations and the Hamilton Equations of Motion (215) The Legendre trick is amazingly simple, I don't think I ever understood it before. Start with 7-3 and say you want to replace dx with some other du differential and some new fellow on the LHS. So just define g = f-ux and that does it! When you say g = f - ux, you mean to throw out x and install u. So in this case, you start with df = udx + vdy, and you end up with dg = - xdu +vdy. Yes, there is a minus sign. Each "coefficient" can be

interpreted as a partial derivative in the very obvious manner. So in the first equation, we know that u = $\partial f/\partial x$ but in the second equation u is one of the variables. So this would be a change in independent variables from (x,y) to (u,y).

So this is exactly what we are going to do. In the analogy, f = L(q,$\dot{q}$,t) and g = H(q,p,t) where we are throwing out $\dot{q}$ in favor of p as independent variables. There is an overall minus sign so g' = -g is really used here. So 7-8 makes complete sense in terms of it being nothing more than a change of independent variables. Here are the starting and ending positions:

$dL = A\,dq + B\,d\dot{q} + C\,dt \qquad A = \partial L/\partial q \qquad B = \partial L/\partial \dot{q} = p \qquad C = \partial L/\partial t \qquad (1)$

$dH = A'\,dq + B'\,dp + C'\,dt \qquad A' = \partial H/\partial q \qquad B' = \partial H/\partial p \qquad C' = \partial H/\partial t \qquad (2)$

But we have H =$\dot{q}$p - L so we have more information:

$$dH = \partial H/\partial p\,dp + \partial H/\partial \dot{q}\,d\dot{q} \{ - \partial L/\partial \dot{q}\,d\dot{q} - \partial L/\partial q\,dq - \partial L/\partial t\,dt \} \qquad (3)$$

$$= \dot{q}\,dp + p\,d\dot{q} \{ - \partial L/\partial \dot{q}\,d\dot{q} - \partial L/\partial q\,dq - \partial L/\partial t\,dt \} \qquad (4)$$

The 2nd and 3rd terms cancel due to the definition of p, and we are left with

$$dH = \dot{q}\,dp - \partial L/\partial q\,dq - \partial L/\partial t\,dt \qquad (5)$$

But now we use the E-L equation to replace $\partial L/\partial q$ with $\dot{p}$ and we have

$$dH = \dot{q}\,dp - \dot{p}\,dq - \partial L/\partial t\,dt \qquad 7\text{-}11 \qquad (6)$$

Now let's do some comparisons. Comparing (2) and (6) we can say

$$A' = \partial H/\partial q = -\dot{p}$$

$$B' = \partial H/\partial p = \dot{q}$$

$$C' = \partial H/\partial t = -\partial L/\partial t$$

and these last three equations are now called Hamilton's Equations of Motion. Not clear at this point that we have gained anything by doing all this.

7.2 Cyclic coordinates and Routh's procedure (218) In the Lagrangian formulation, if you have a cyclic coordinate q_n, L can still be a non-trivial function of $\dot{q}_n$ so you have not removed this nth coordinate from the problem completely. In the Hamiltonian formulation, if q_n is cyclic, it does not appear in H (or L), and also $p_n = \alpha$ = a constant, so solving for $\dot{q}_n = \partial H/\partial p_n = \partial H/\partial \alpha$ is a completely separated problem, so you can regard p_n and q_n as both being removed from the problem in the Hamiltonian situation.

In passing, G mentions the Routh procedure which is simply this: write $H' = \Sigma^{s} \dot{q}p - L$ but only include the s cyclic coordinates in the sum. Then this H' will be H for the s cyclic coordinates, but will be -L for the other n-s coordinates. You then have a Lagrangian problem in only n-s coordinates and you can get the s other things from the $\dot{q}_k = \partial H/\partial p_k = \partial H/\partial \alpha$ method. This H' is called R, the **Routhian**. It is an interesting idea, but I don't think G is going to really do anything with it. If you really like doing Lagrangian dynamics, this method reduces the problem you have to solve to n-s variables + s simple things to solve. You have one foot in each camp!

7.3 Conservation Theorems and Physical Significance of H (220) Equation 7-19, which G easily derives, shows that $dH/dt = \partial H/\partial t = -\partial L/\partial t$. The whole formalism so far (started in E-L equations) requires that V not be a function of velocity. In addition, if the $\mathbf{r}_n$ equations don't have t, then I guess $\partial L/dt = 0$ and then $dH/dt = 0$ which means H = constant of the motion we call "energy".

Example 1: particle in central potential V(r). We first write H in terms of r and θ and p_r and p_θ. We examine the Hamilton equations of motion (4 items), and we duplicate results we got in the previous chapter.

Example 2: Compute H for a relativistic particle in this central potential, get 7-20 as T + V.

Now let's throw in our velocity-dependent EM potential for V, and we get H as in 7-22 which is the first "recipe" formula for H. This is the non-rel version. Do it again in 7-23 for the rel version. In each case, we arrive at the H result by using the recipe of adding -q**A**/c to **p**.

Covariant Form: Now, as we did in the relativity chapter, we conjecture that $H' = p_\mu u_\mu - L'$ where L' was the covariant Lagrangian we used in that chapter. The resulting H' is as in 7-26 for particle in an EM potential. We then goof around with the math to show explicitly that $dH/dt = \partial H/\partial t$, so this is true even in the presence of this special case velocity-dependent potential.

We are reminded that in 1950, no one had a covariant formulation for non-EM forces!

7.4 Hamilton's Equations from a variational principle (225). First, we start with our usual L variational thing, put in the replacement $H = p\dot{q} - L$ and write $\dot{q}dt = dq$ and we get 7-28' which he names the **modified Hamilton's principle**, the unmodified one being the usual L thing. then on page 226 we go through a long series of steps, including the usual parts where $\partial q/\partial \alpha = 0$ at the two endpoints. When the dust settles on page 227, instead of an Euler-Lagrange equation, we get the two Hamilton equations! The point is made that in the L world, $\dot{q}$ really depended on q, just a time derivative, whereas in the H world, our variables p and q really are independent. In the L world then you sort of have n variables q to worry about, but you get a second order ODE to solve (EL) for n variables q. In the H world, you have 2n variables to worry about (the p and q) but the Ham equations are first order and there are 2n of them. This is perhaps an easier problem to deal with.

7.5 Principle of Least Action (228) The $p\dot{q}$ sum in $H = p\dot{q} - L$, when integrated over time, is called the **action**. So the claim of least action is that, when H is conserved, the solution to your problem occurs when the integral $\Delta(\text{action}) = \Delta\left(\int \Sigma_i p_i \dot{q}_i dt \right) = 0$, but this Δ is a new kind of "variation" that differs from the δ we used in all our earlier variational work in this book. This caused me much pain on first reading, I will now try to detangle it.

Suddenly G is using new language. He refers now to the δ variations as "virtual displacements", two new words really. It is true that in the δ case, those η functions could be arbitrary apart from being tacked down at the ends, so if you call these η functions "displacements", or perhaps you refer to this

combination as a "displacement" : $q_i(t,\alpha) = q_i(t) + \alpha\eta_i(t)$, then we are OK on terms. " the varied path in the δ-process need not correspond to a possible path of motion for the system", and H may not be conserved during such a variation.

Now in the Δ variation, you are not doing a "virtual thing" (virtual means a δx with dt=0 recall). He suddenly starts talking about a "succession of displacements" each including dt. This is where G just plain becomes unclear, and I will need now an outside source.

As an aside, I was right about the word "action", as seen here:

Let the system occupy, at the instants t_1 and t_2, positions defined by two sets of values of the co-ordinates, $q^{(1)}$ and $q^{(2)}$. Then the condition is that the system moves between these positions in such a way that the integral

$$S = \int_{t_1}^{t_2} L(q, \dot{q}, t)\, dt \qquad (2.1)$$

takes the least possible value.† The function L is called the *Lagrangian* of the system concerned, and the integral (2.1) is called the *action*.

I have found the Chand source which is very similar to Goldstein, he probably cribbed from G.

Let's start by ignoring the various interpretation comments of G on page 228. and instead just look at the math. The idea is that we think of $q = q(t,\alpha)$ as in the δ world, but we also now thing of $\alpha = \alpha(t)$, though I don't know what that means. We have $q_i(t,\alpha) = q_i(t) + \alpha(t)\ \eta_i(t)$ I guess. This then leads to the extra term such that we have

$$\Delta q = d\alpha(dq/d\alpha) \qquad \text{where} \qquad dq/d\alpha = \partial q(\alpha,t)/\partial\alpha + \partial q(\alpha,t)/\partial t\ dt(\alpha)/d\alpha$$

Now I suppose we have to say $\partial q(\alpha,t)/\partial t = \dot{q}$ so we have

$$dq/d\alpha = \partial q(\alpha,t)/\partial\alpha + \dot{q}(\alpha,t)\ dt(\alpha)/d\alpha$$

Then

$$\Delta q = d\alpha(dq/d\alpha) = d\alpha\ [\partial q(\alpha,t)/\partial\alpha + \dot{q}(\alpha,t)\ dt(\alpha)/d\alpha\ ] = d\alpha\ \partial q(\alpha,t)/\partial\alpha + \dot{q}(\alpha,t)\ d\alpha\ dt(\alpha)/d\alpha$$

then he notes that the first term is our older δq, and in the second term $\Delta t \equiv d\alpha\ dt(\alpha)/d\alpha$. So we have in the end

$$\Delta q = \delta q + \dot{q}\ \Delta t$$

This Δ thing is a certain operator applied onto q(t,α).

Now suppose we go up another level and we have some f(q(t,α), t). Then the claim is this:

$$\Delta f = \partial f/\partial q\ \Delta q + \partial f/\partial t\ \Delta t = \partial f/\partial q\ [\delta q + \dot{q}\ \Delta t] + \partial f/\partial t\ \Delta t$$

$$= \partial f/\partial q\ \delta q + \{\ \partial f/\partial q\ \dot{q} + \partial f/\partial t\} \Delta t$$

So far, then, we are just playing with this Δ operator, and we just did a Δ-variation of f(q,t).

Derivation of the fact that the correct dynamics result in $\Delta A = 0$

Now instead of doing this with a general function, let's do it to "the action"

$$A = \int p\dot{q}dt = \int (L+H)dt = \int Ldt + H(t_2-t_1)$$

if we assume H = constant independent of time, like a conserved energy. Now apply Δ. The first term gives: (here, I is the integral of L, whatever it might be, so $L = \dot{I}$ for use soon below)

$$\Delta \int Ldt = \Delta [\, I(t_2) - I(t_1) \,] \qquad = \Delta I(t_2) - \Delta I(t_1) \qquad (230\ a)$$

Then we have

$$\Delta A = \Delta \left(\int Ldt + H(t_2-t_1) \right) = \Delta I(q,t_2) - \Delta I(q,t_1) + H(\Delta t_2 - \Delta t_1) \qquad \text{// no written}$$

It must be that Δ is a linear operator so it can do these things. So now mid page 230. Next, consider:

$$\Delta I(q,t_2) = \delta I(q,t_2) + \dot{I}\, \Delta t_2 \qquad \text{where I(q) is being treated like our f(q,t).}$$

Thus,

$$\Delta I(q,t_2) = \delta I(q,t_2) + L(q,t_2)\, \Delta t_2$$

and therefore, from above,

Now go back and consider: $f = \int Ldt$ so that $\dot{f} = L$

$$\Delta \int L(q,t)\, dt = \delta \int L(q,t)\, dt + L\, L(q,t)\, \Delta t$$

$$\Delta \int Ldt = \Delta I(t_2) - \Delta I(t_1) = [\, \delta I(q,t_2) - \delta I(q,t_1)] + [\, L(q,t_2)\, \Delta t_2 - L(q,t_1)\, \Delta t_1] \qquad (230\ c)$$

Now the first two terms are $\delta \int L(q,t)dt$ for this reason:

$$\delta \{ \int L(q,t)dt \} = \delta \{ I(q,t_2) - I(q,t_1) \} = \delta I(q,t_2) - \delta I(q,t_1)$$

Then we finally arrive at 7-36 which says

$$\Delta \left(\int L(q,t)\, dt \right) = \delta \left(\int L(q,t)\, dt \right) + \left[L(q,t_2)\, \Delta t_2 - L(q,t_1)\, \Delta t_1 \right] \qquad (7\text{-}36)$$

So here you see the new contribution on the right. Not only might the endpoints like t_1 vary with α, but L can be different at the two times, since L is probably not "conserved" in time. So two contributions to this extra term.

Our next task is to work on the $\delta \left(\int L(q,t)\, dt \right)$ term above as on bottom of page 230. We first work entirely in the δ world, using the E-L results, but then in the last step we use this reversed deal,

$$\delta q = \Delta q - \dot{q}\, \Delta t$$

and this takes us to the bottom result d on page 230. Now, why does Δq vanish at the endpoints? I guess this is part of the setup of things, the way we had $\delta q = 0$ at the endpoints of the δ variation. We impose this condition on our "variations of the path" between the endpoints. Now since we are allowed to have variations at the two ends which are different, we might have $\Delta t_1 = d\alpha\, dt_1/d\alpha$ be different from what it is at the other end. So only the second terms survives in equation p230 d. Now use the definition of p_i in this equation d and you have arrived at 7-37

$$\delta \left(\int L(q,t)\, dt \right) = -\left[p(t_2)\dot{q}(t_2)\, \Delta t_2 - p(t_1)\dot{q}(t_1)\, \Delta t_1 \right] \qquad (7\text{-}37)$$

Now of course we want to install this result into 7-36 to get:

$$\Delta \left(\int L(q,t)\, dt \right) = -\left[p(t_2)\dot{q}(t_2)\, \Delta t_2 - p(t_1)\dot{q}(t_1)\, \Delta t_1 \right] + \left[L(q,t_2)\, \Delta t_2 - L(q,t_1)\, \Delta t_1 \right]$$

$$= \left[L(q,t_2) - p(t_2)\dot{q}(t_2) \right] \Delta t_2 - \left[L(q,t_1) - p(t_1)\dot{q}(t_1) \right] \Delta t_1$$

$$= \left[-H_2 \right] \Delta t_2 - \left[-H_1 \right] \Delta t_1 = -H \left(\Delta t_2 - \Delta t_1 \right) \qquad \text{// since H conserved}$$

But above we had,

$$\Delta A = \Delta \left(\int L dt + H(t_2 - t_1) \right) = \Delta \left(\int L dt \right) + H \left(\Delta t_2 - \Delta t_1 \right)$$

and we thus end up with

$$\Delta A = \Delta \left(\int p\dot{q} dt \right) = 0$$

Since we have used the E-L equations along the way (the dynamics of the system), our conclusion is that the dynamic solution of our problem causes action = 0 ! QED,

Now some examples. Consider a non-rel particle $T = \frac{1}{2} mv^2 = \frac{1}{2} \dot{q}p = \frac{1}{2} A$, so action $A = 2T$. Then our new rule says $\Delta(\int T dt) = 0$. Suppose there are no forces on our particle so T = constant. Then we get that $\Delta A = T \Delta(\int dt) = T[\Delta t_2 - \Delta t_1] = 0$. Can write this as $\Delta(t_2 - t_1) = 0$. Now $t_2 - t_1$ = "time of transit" for a particular "path". I guess this means a path going from point 1 to point 2 in space. So for the non-rel free particle, the claim is that the solution path is the one which minimizes the time of transit! This will appear in optics somewhere.

Now back to $\Delta(\int T dt) = 0$ if there are forces say from a potential V. On page 232 we first express this idea:

$(ds)^2 = (dx_1)^2 + (dx_2)^2 + (dx_3)^2$ so that ds is the arclength along the path.

We can replace dt with ds as shown top page 232 and we end up easily with 7-40. So our solution path is going to result in this

$$\Delta \left(\int \sqrt{H-V}\, ds \right) = 0$$

The solution path minimizes the arclength weighted by this function!

Now G wants to generalize this result to more than 1 particle. What do we use for ds in that case? We first assume no t in the r-equations, some word for that I forget. then we know that T is quadratic in the velocities, and the weighting matrix turns out to be the metric tensor he is calling m_{ik} here. Then it turns out that the correct generalization to many particles is this:

$(d\rho)^2 = \mathbf{dq} \bullet \mathbf{m} \bullet \mathbf{dq} = dq^T m\, dq$ where $dq = (dq_1, dq_2, ...dq_n)$.

where for run I use both matrix notation and dyadic notation. The claim is that for normal coordinate systems we use a lot (orthogonal curvilinear), m is diagonal s we then get

$(d\rho)^2 = \Sigma_i\, m_{ii}\, (dq_i)^2$ orthogonal curvilinear

and in cartesians we would have

$(d\rho)^2 = \Sigma_i\, 1\, (dq_i)^2 = \Sigma_i\, ((dx_i)^2 + (dx_i)^2 + (dx_i)^2$

Now divide by $(dt)^2$ to get

$(d\rho/dt)^2 = (d\mathbf{q}/dt) \bullet \mathbf{m} \bullet (d\mathbf{q}/dt) = \dot{q} \bullet m \bullet \dot{q} = 2T$ // from 232-a

This lets you again replace dt not with ds but with the more general dρ and we get

$$\Delta \left(\int \sqrt{H-V}\, d\rho \right) = 0 \qquad \text{// for multiple particles} \qquad (7\text{-}44)$$

This is called Jacobi's Form of the Least Action Principle.

Now, you can observe that $\sqrt{T} \sim v$, speed at any point on the curve. So in general, when we are staying that $\Delta \left(\int \sqrt{H-V}\, d\rho \right) = \Delta \left(\int \sqrt{T}\, d\rho \right) = 0$, we are saying $\Delta \left(\int v\, d\rho \right) = 0$. This says that when you integrate the particle's speed along the solution path, you get a minimum. This is basically an average speed along the path and it is going to be as large as possible. Of course one way to do this is to take the shortest path which G refers to as a geodesic of the space.

Notice that Jacobi's Form is really about paths in space, time does not appear, so you might get things like orbits from this form. You usually introduce some parametrization variable like θ where

$$(d\rho/d\theta)^2 = (d\mathbf{q}/d\theta) \bullet \mathbf{m} \bullet (d\mathbf{q}/d\theta) \quad \text{so } d\rho = d\theta \sqrt{(d\mathbf{q}/d\theta) \bullet \mathbf{m} \bullet (d\mathbf{q}/d\theta)}$$

as shown in 7-45. This θ is a spatial thing only, not a time thing. So if we replace dρ with this dθ, our Jacobit result 7-44 becomes 7-45 and if we take the entire integrand to be F, then the claim is the Δ variation is the same as a δ one (I guess since time not involved), and then

$$\Delta \left(\int F d\theta \right) = 0 = \delta \left(\int F d\theta \right)$$

and this as usual gives you E-L equations, but now derivatives are relative to θ, not t.

Now finally, back to G's comments on page 228. Why is any of the varied paths an allowed physical path? G just fails to make this clear to me and I suspect to other readers.

The web points out this alternative way of writing the action

$$\Delta A = \Delta \left(\int p\dot{q} dt \right) = \Delta \left(\int pdq \right) = 0 = \Delta \left(\int \mathbf{p} \bullet d\mathbf{q} \right)$$

where now it is little clearer that you are integrating on paths through the multi-particle phase space. The dt form makes this less clear somehow.

There is an excellent website http://www.eftaylor.com/leastaction.html#variational which talks about why presentations like Goldstein's are so painful to students. No simple examples are given, and the subject is never really used for anything later in the book. It is just wedged in there for completeness. I looked at some of the papers on the above site and they are most excellent indeed. So once again, I can go and "learn more" about a Goldstein subject if I want, but I would prefer to continue and finish his book. \I do realize that the action is VERY important for many reasons, and it would be a better thing to study later than, say, rigid body motion.

Comment on the word "action". As discussed here http://en.wikipedia.org/wiki/Action_(physics), several definitions are in use, and now they have standard names. The action Goldstein uses is now called the "abbreviated action" and is written S_0 and we have

$$S_0 = A = \int p\dot{q} dt = \int pdq \qquad \text{" abbreviated action"}$$

The action most people now refer to by this word is called S , the "action" or "classical action",

$S = \int L dt$ " action" or "classical action"

// later we learn that S is the same as Hamilton's Principle Function "S"

In a situation of constant H, we know from above that

$$A = \int p\dot{q}dt = \int (L+H)dt = \int L dt + H(t_2-t_1)$$

which says

$S_0 = S + E(t_2-t_1)$ "abbreviated action"

// later we learn that S_0 is Hamilton's Characteristic Function "W"

so they are certainly closely related to each other. Goldstein defines Hamilton's Principle as

$\delta S = 0$ $S = \int L dt$

****

Chapter 8. Canonical Transformations (p 237)

****

8.1 The Equations of Canonical Transformation (237). If you could somehow get ALL coordinates q_i to be cyclic, then a problem has the trivial linear in time solution shown in 8.2. The purpose of canonical transformation from p,q,H to some P,Q,K is to increase the number of cyclic coordinates! If you just transform from q_i to some $Q_i(q_i)$, that is called a point transformation, but in general $Q_i(q_i,p_i)$ is available. Of course not only to you need some P and Q, you need your new Hamiltonian here called K such that in the new coordinates and with the new Ham, the Ham equations will still be true, in which case we have done a "**canonical" transformation**. This can be guaranteed by making sure the δ variation as shown in the modified Ham principle is 0 in the new system of Q,P,K. That means that the ($p\dot{q}$ - H) and ($P\dot{Q}$ - K) must differ by dF/dt because then the variation integrands will differ by the constant $F(t_2) - F(t_1)$ which has zero variation under the δ operator. F is called the **generating function** of the transformation.

We then consider four different ways to argument-up the function F, called F_1 through F_3 as shown page 240. G then considers $F_1(q,Q,t)$ first. We get a set of three equations. The first two involve partials of F_1, and the third (the same in all four cases) says that $K = H + \partial F/\partial t$.

He goes on then to consider the F_2 form where we replace Q by P as an independent variable, and the obvious method is to use a Legendre transformation on F_1 to get F_2 as shown in 8-10, and in this new

form we also get three equations. He then continues to the forms F_3 and F_4 (where he does a double L T on F_1), so in each case we get our little set of 3 equations, which is really 2n+1 equations.

In the final section, G asks how we might "add" the idea of a transformation on the time, since we know this is going to happen in special relativity. The idea here is to write our standard modified Ham variation thing, but replace time t with some kind of spatial path marker called θ (arbitrarily). Then can interpret the second term as saying there is an n+1st coordinate called $q_{n+1} = t$ with conjugate momentum $p_{n+1} = -H$. Then you can merge the last term into the first as in equation α on page 243. Then when you do a canonical transform to P, Q, you still have your n+1 terms, and $Q_{n+1} = T$, $P_{n+1} = -K$. In this case he suggests a generating function G of the F_2 type which gives the F_2 type equations shown in β p 243. In this manner, then, we are able to treat time t on an equal footing with the other coordinates. Obviously you could consider all four forms for G, but probably he has picked G_2 because that one will be used later in the book I get.

8.2. Examples of canonical transformations (244)

Example 1: $F_2 = q_i P_i$ results in P=p and Q=q, an "identity canonical transformation"

Example 2: $F_2 = f_i(q_j)\ P_i$ makes $Q_i = f_i(q_i)$, and $p_i = P_i\ f_i'(q_j)$ and this is the "point transformation" idea where no p are mixed into Q. A linear example is $f_i = a_{ik}q_k$ so $f_i'(q_j) = a_{ij}$ so that in fact $p_i = a_{ij}P_i$ or $\mathbf{p} = A^T\mathbf{P}$. So here an orthogonal transformation is an example of a point transformation.

Example 3: $F_1 = q_k Q_k$ results in p = Q and P = -q so basically you get a swap. This emphasizes that there is no real distinction between coordinates and momenta, those now are just "words".

Example 4: Start with the harmonic oscillator usual H(p,q) as in 8-28, then use the strange generatort which says $F_1 = (m\omega/2)\ q^2 \cot Q$. Result is that $H = \omega P$ which is cyclic in Q. This leads to the usual solution of the HO as 8-30, very strange.

In all four of these examples, there was no t dependence in the F_i generator, so K = H.

8.3 The integral invariants of Poincare (247). There are a sequence of such invariants, and the simplest has the form

$$J_1 = \int\int d\mathbf{q} \bullet d\mathbf{p} \quad \text{where} \qquad d\mathbf{q} \bullet d\mathbf{p} = [dq_1dp_1 + dq_2dp_2 + ... dq_ndp_n]$$

What we want to prove is that

$$\int\int d\mathbf{q} \bullet d\mathbf{p} = \int\int d\mathbf{Q} \bullet d\mathbf{P}$$

so J_1 is *invariant under a canonical transformation*. The integrand is rewritten, without proof, as

$$d\mathbf{q} \bullet d\mathbf{p} = \Sigma_i\ dq_i dp_i = \Sigma_i\ Jac(q_i,p_i: u,v)\ dudv$$

Goldstein then uses as an example an $F_2(q,P,t)$ generator function to show that, in fact

$$\Sigma_i\ Jac(q_i,p_i: u,v) = \Sigma_i\ Jac(Q_i,P_i: u,v)$$

which then proves that J_1 is in fact an invariant. This proof uses obvious facts about determinants and the simple facts about the F_2 type generator function. It is quite simple, actually, though lots of notation.

When this is all done, he says there are in fact a whole series of such invariants. In 8-36 he writes what he calls J_2 , but the summation intended unclear. If it is meant to be a sum over i and k, then he is claiming that

$$J_2 = \int\int\int\int (d\mathbf{q} \bullet d\mathbf{p})^2$$

but I don't think that is what his Σ means! If that is what it meant, then you would arrive here

$$J_n = \int\int\int \ldots . \int (d\mathbf{q} \bullet d\mathbf{p})^n$$

but it seems to me this is not the same as the J_n shown which is this

$$J_n = \int\int\int \ldots . \int \prod_i dq_i dp_i = \int\int\int \ldots . \int dq_1 dq_2 \ldots . dp_1 dp_2 \ldots ..$$

I scoured the web, but could only find someone who cloned Goldstein's ambiguous results. This is a very obscure topic, so naturally it's not out there, or there may be some other name it goes by. The last item says that your *phase space volume* is invariant under a canonical transformation. We could probably prove this one using the same methods given earlier, the Jacobian would be a larger one. Hopefully G will only use J_1 or J_n in anything further he does.

8.4 Lagrange and Poisson Brackets as canonical invariants (250). This is the Big Section of Goldstein's book which supplies the embarkation point from classical to quantum mechanics. There is lots of trivial derivation stuff here, but let's quote the results and keep our eye on the ball:

• define the **Lagrange (curly) brackets** like this:

$$\{u,v\}_{q,p} \equiv \Sigma_i \left(\partial q_i/\partial u \, \partial p_i/\partial v - \partial p_i/\partial u \, \partial q_i/\partial v \right)$$

• Notice that this is in fact the Jacobian $\partial(q_i, p_i)/\partial(u,v)$ which we have already shown to be an **invariant** under canonical transformations. Therefore, we know that

$$\{u,v\}_{q,p} = \{u,v\}_{Q,P} = \{u,v\} \qquad \text{// just write as } \{u,v\} \text{ with no labels}$$

• Now look at special cases for u and v and you find that (the fundamental Lagrange relations)

$$\begin{array}{lll} \{q_i,q_j\}_{q,p} = 0 & \Rightarrow & \{q_i,q_j\} = 0 \\ \{p_i,p_j\}_{q,p} = 0 & \Rightarrow & \{p_i,p_j\} = 0 \\ \{q_i,p_j\}_{q,p} = \delta_{ij} & \Rightarrow & \{q_i,p_j\} = \delta_{ij} \end{array}$$

• now define the **Poisson (square) brackets** like this:

$$[u,v]_{q,p} \equiv \Sigma_i \left(\partial u/\partial q_i \; \partial v/\partial p_i - \partial u/\partial p_i \; \partial v/\partial q_i\right)$$

which has the tops and bottoms of the Lagrange brackets all switched.

• Now, suppose somebody hands you *any* set of 2n independent functions of the form $u_j(\mathbf{q}, \mathbf{p})$. Then Goldstein proves this fact:

$$\Sigma_\ell \; \{ u_\ell , u_i \}_{q,p} [u_\ell , u_j]_{q,p} = \delta_{ij} \qquad \text{or} \qquad \Sigma_\ell \; L_{\ell i} P_{\ell j} = \delta_{\ell j} \qquad \text{or} \qquad LP = 1$$

where we define Lagrange and Poisson matrices L and P as shown. Thus, $P = L^{-1}$ which supports one's feeling that the two brackets are sort of inverses of each other just from how they were defined. We suspect at this point that $[u_\ell , u_j]_{q,p} = [u_\ell , u_j]_{Q,P}$ but G has not quite shown this yet. Here is how I might show this: let's just add q,p labels to our matrices, then since we know L is invariant, we get

$$P_{q,p} = (L_{q,p})^{-1} = (L_{Q,P})^{-1} = P_{Q,P}$$

So from now on, no need to put any labels on either Lagrange or Poisson brackets.

• Now let's pick the following special case situation; $u_i = q_i$ and $u_j = p_j$ and the overall set of $\{u_\ell\}$ is just the full set of all the p's and all the q's. Then the LP=1 theorem just quoted becomes

$$[p_i, p_j] = 0$$

By doing similar special cases, we get all three results we anticipate:

$$[q_i,q_j] = 0$$
$$[p_i,p_j] = 0$$
$$[q_i,p_j] = \delta_{ij}$$

and these are exactly the same as the three results we got for the Lagrange curly brackets above.

• When you put the actual canonical variables p and q inside either brackets, you get the **fundamental Lagrange or Poisson bracket relationships**, so the above triplet of results is "fundamental".

• Now imagine F(**p**,**q**) and G(**p**,**q**). Goldstein first shows the following

$$[F,Q_k]_{q,p} = -\partial F/\partial P_k \qquad [F,P_k]_{q,p} = +\partial F/\partial Q_k$$

where P and Q are some coordinates which are canonical to q and p (were they not, you could not do some of the intermediates steps). Then he ends up showing this

$$[F,G]_{q,p} = [F,G]_{Q,P}$$

where F and G are any functions you want. But I think I already knew this invariance.

• I note in passing that the above results are true for any Q, P set, and in particular

$[F,q_k]_{q,p} = -\partial F/\partial p_k = [F,q_k]$ $\qquad$ $[F,p_k]_{q,p} = +\partial F/\partial q_k = [F,p_k]$

• One more interesting fact: if c is a "constant" not depending on p or q, then [u,c] = 0 for any u.

• In a footnote, he says that in quantum mechanics, one will have this "correspondence"

$i\hbar$ [F,G] Poisson bracket $\leftrightarrow$ [F,G] commutator of two operators.

for example, $[x,p_x]_{QM} = i\hbar\,[x,p_x]_{PB} = i\hbar$

• Notice that we are probably going to use Poisson brackets more than Lagrange ones because the Poisson ones have our "functions of interest" in the numerator and have the "variables" in the denominator and this is how things normally appear. Probably we will never use Lagrange brackets themselves for anything.

8.5 Equations of motion in Poisson bracket notation (255). Now we continue on this pre-QM avenue, and we combine our "canonical Hamilton's equations" with our knowledge of Poisson brackets. Right off the bat now we have:

$-\partial F/\partial p_k = [F,q_k] \quad => \qquad [q_k,F] = \partial F/\partial p_k$

so what happens if F = H, the Hamiltonian. Then

$[q_k,H] = \partial H/\partial p_k = \dot{q}_k$

What happens with the reverse:

$+\partial F/\partial q_k = [F,p_k] \quad => \qquad [p_k,F] = -\partial F/\partial q_k$

so we get

$[p_k,H] = -\partial H/\partial q_k = \dot{p}_k$

Thus, I would say the "Poissonator" of either p or q with H "generators" the time derivative of the item in question! G then shows this is true for any function u(**p**,**q**; t) and we have this very famous result:

$du/dt = \dot{u} = [u,H] + \partial u/\partial t$

In this book, the "quantities of interest" u will not have explicit time dependence so we will have

du/dt = [u,H] // as you install various quantities "u", you get "equations of motion"

Thus, if you can find a u such that [u,H]= 0, then you have found a constant of the motion! This provides a test.

The last section here derives the Jacobi Identity which I know well from commutator theory, and I am not at all surprised to find that it is true using Poissonators, but Goldstein dutifully provides a page long proof which is not bad. The application of interest of this Jacobi Identity is very clear:

• If u and v are both constants of the motion, then [u,v] is also a constant of the motion. This comes from Jacobi just setting w = H. The result seems awfully reasonable. It is called Poisson's Theorem, but one imagines Poisson has lots of theorems in the larger picture.

8.6 Infinitesimal contact transformations, constants of the motion, symmetry (258) Here we just consider a very small transformation from p,q to some P,Q, and we ask what the "generator" G might look like for such a transformation. He uses the word "contact transformation" for some reason, but he just means any canonical transformation. The smallness parameter is called ε. We use the F_2 form and have the generator as εG(q,P) [see p 240] , but to first order in ε we can use εG(q,p). Our usual F_2 equations then tell us that $\delta p_i = -\varepsilon \partial G/\partial q_i$ and $\delta q_i = \varepsilon \partial G/\partial p_i$ which is 8-64, used much later.

We could consider ε = dt and G = H, the Ham, and he makes a big deal of saying that we can interpret the transformation two ways. The way of interest now is that in dt, we are going to actually move a little big in phase space, so this is not a virtual transformation, nor are we just relabeling the same point from p,q to P,Q. The Hamiltonian is thus the "generator" of time movement of the system in phase space. I am used to generator being a quantum operator, but here it is the classical generator of a canonical transformation.

So here is the big idea: the Hamiltonian can act to transform your system in incremental steps from t back to t_o where your system had some initial values q_{i0} and p_{i0} and in some sense, then, you have moved from q(t), p(t) back to Q = q(0), P=p(0) and now in these new coordinates, all Q and P are "constants". I am very used to this in QM, but here we are doing it in CM with Poisson brackets.

Now look at 8-67 and this time we have taken u = H instead of G = H, and G is still just some G. If G commutes (" Poissons", classically speaking) with H, then δH = 0. This is just another way of saying that if a generator of some sort Poissons with H, then H does not change as G is varied. This is what is meant by a symmetry!

Two examples are then given. In the first, suppose we have the only transformation change being $\delta q_5 = \varepsilon$. Then from 8-64b we must have $G = p_5$. First, we see that p_5 is therefore the "generator" of a translation in the q_5 direction. Second, if p_5 Poissons with H, then δH = 0 from 8-67 and p_5 is then a symmetry of H. So this example is linear momentum.

The second example is going to be angular momentum. If we take G as in 8-71 ($G = L_z$), we find that we have generated a rotation about the z axis. And if L_z Poissons with H, then it is a symmetry. Again, we would say that $G = L_z$ is the "generator" of rotations about the z axis.

So far, we have nothing exponentiated to build up a finite transformation.

8.7 The angular momentum Poisson bracket relations (263). Consider 8-66 δu = ε[u,G] and let's imagine u as some vector function **F**, and G as our rotation generator **n** • **L** for rotations about the **n** axis, so we then have $\delta \mathbf{F} = d\theta [\, \mathbf{F}, \mathbf{n} \bullet \mathbf{L} \,]$. In the case of $\mathbf{n}=\hat{\mathbf{z}}$ this can be written out as in 8-75 which shows that, for some abstract **F**, you get both a change in functional form [hence A'] *and* you get a change in value [vector is rotated] . But certain quantities **F** will have the property that A' = A, etc, and **L** is an

example of such a quantity. In general, any vector composed from the **p** and **q** of the system will have this special property of same functional form. In this case (only) on page 264 G shows that we can write $\delta\mathbf{F} = d\theta\ \mathbf{n} \times \mathbf{F}$ and then the two $d\theta$ cancel and we get $[\ \mathbf{F},\ \mathbf{n} \bullet \mathbf{L}\ ] = \mathbf{n} \times \mathbf{F}$ which is 8-77. This is then the transformation rule for a certain class of "vectors **F**" under the generator shown. Remember, these are Poissonators, not commutators. In particular, we have $[\ \mathbf{L},\ \mathbf{n} \bullet \mathbf{L}\ ] = \mathbf{n} \times \mathbf{L}$ and this gives the usual commutatation relations for angular momentum as in 8-80 but again, Poissonators and no $i\hbar$! It is then no surprise that $[\ \mathbf{L}^2,\ \mathbf{n} \bullet \mathbf{L}\ ] = 0$. Our previous Poisson's theorem says that if L_x and L_z are both constants of the motion, then L_z must also be so. So somehow we have the notions, within classical mechanics, of vector and a scalar objects. Somehow I think this is more easily done within the framework of group theory and representations, but Goldstein has done things pretty much by brute force here in 1950. I wonder what the newer editions do? G mentions that, when he was at Harvard, Schwinger showed him that all these commutator-like results existed within classical mechanics via the Poissson brackets.

8.8 Liouville's Theorem (206). This theorem relates to statistical mechanics and states that, in a statistical ensemble of systems, the density of systems (ie, number of systems in the ensemble per unit volume of phase space) is constant if you look at any specific point in phase space as time mores forward. Because the systems within an ensemble are not perfectly known, they "swarm" around some target point in phase

****

Chapter 9. Hamilton-Jacobi Theory (p 273)

****

9.1 The Hamilton-Jacobi Identity for the Hamilton's Principle Function (273). In the last chapter we discussed CT's and part of that was $K = H + \partial F/\partial t$. Suppose you arrange F such that K = 0. Then in the transformed world P,Q,K you would have $\dot{P} = \dot{Q} = 0$ from Hamilton's equations there. This would seem to imply that in this new world, all the P and all the Q are constants of the motion! If we go with an F_2 form, then $p = \partial F/\partial q$ and our requirement K = 0 becomes $H(q,\partial F/\partial t) + \partial F/\partial t = 0$ as shown in 9-3. Can we find such an F? This equation is called **the Hamilton-Jacobi equation (HJE)**, and solutions are called S, **Hamilton's Principle Function.**

G notes that since there are n+1 first-order partial derivatives in this HJ equation, there will be n+1 integration constants. One will be an overall adder to S which does not affect the HJ equations, so we then have n constants he calls α_i. A solution S will then be $S(\mathbf{q}, \boldsymbol{\alpha}, t)$. Remember, no matter what values you gives these constants, you get a solution S. Suppose then we choose $\alpha_i = P_i$ which we know are constants in the Q,P,K world. Our F_2 equations 8-11 tell us:

$$p_i = \partial S/\partial q_i \qquad \text{and} \qquad Q_i = \partial S/\partial P_i = \partial S/\partial \alpha_i \equiv \beta_i$$

So we emphasize things are constants by saying $P_i = \alpha_i$ and $Q_i = \beta_i$. Look at this last equation:

$$\partial S(\mathbf{q}, \boldsymbol{\alpha}, t)/\partial \alpha_i \equiv \beta_i$$

You ought to be able to solves this for $\mathbf{q}(\boldsymbol{\alpha},\boldsymbol{\beta},t)$, a process G calls turning the above equation *inside out*. So he has just outlined a possible path to solve the problem for the q_i which is of course what you want.

Although it is not useful for the calculation method just outlined, G shows that

$S = \int L\, dt + const$

So basically S is the true "action" as used in Hamilton's Principle. So this letter S is consistent with the website noted above, which refers to $\int pdq$ as S_0, the *abbreviated* action, not S. This final fact is not "useful" here for finding S because you cannot integrate L unless you know the q(t), even if you know the form of L.

9.2 A HJE example: the harmonic oscillator (277). We know H(p,q) for the HO, one dimension only. Replace $p = \partial S/\partial q$ and write out the HJ 9-3 which is then 9-12. As a solution, try the form shown in 9-13 and you end up with 9-14 which is simpler than 9-12. Integrate to find W and then you know S, at least as an integral. We could do the integral, but we only care about partial of S so why bother. Note: we intend that $\alpha = P$, as in our general writeup. Ie, we select α to be P. Then β is as shown top page 278, and this time we do the integral to get 9-16. This is of the form $\partial S(\mathbf{q}, \boldsymbol{\alpha}, t)/\partial \alpha_i \equiv \beta_i$ which we want to turn inside out to get the result $\mathbf{q}(\boldsymbol{\alpha},\boldsymbol{\beta},t)$ which now appears in 9-17 which we recognize as the HO solution. We then consider the normal starting conditions for a HO at t=0 and find that α = as in 9-18 = total energy E. G now shows trivially why $\alpha = E$: namely, the HJ equation says $H = -\partial S/\partial t = +\alpha$. Of course in our starting situation, we know that $\beta = P = p_0 = 0$. In the final section, we write out again our integral S having replaced our constants α and β, and we find indeed that S is the integral of L, as expected. So this is a very good example and well worth three pages of the book.

Aside for next section: In the L world, if L was cyclic in some q_i, then the corresponding p_i was a constant. In the H world, as shown page 217 from the Hamilton's equations, this rule is symmetric. Cyclic in p_i means q_i is constant, and cyclic in q_i means p_i is constant

9.3 Hamilton's Characteristic Function W (279) We now specialize to the common situation in which H is not an explicit function of time. We define W by $S = W(\neq t) - \alpha_1 t$ so that $\partial F/\partial t = -\alpha_1$ and then the HJE says $H = \alpha_1$, so we know that α_1 is the conserved total energy E. We then work directly in terms of W and we think of W as generating its own F_2 type canonical transformation to a world of P,Q and K. We select $\alpha_i = P_i$ and here are the F_2 equations for $F_2 = W$:

$p_i = \partial W/\partial q_i$ $\quad$ $Q_i = \partial W/\partial P_i = \partial W/\partial \alpha_i$ $\quad$ $P_i = \alpha_i$

$K = \alpha_1 = P_1$ = cyclic in all the P_i and Q_i (except P_1), so the only non-constant is Q_1.

On page 281 G writes all the Q_i in terms of the β_i and all the $P_i = \alpha_i$.

Aside: Equations 9-22 A here are going to be heavily used in upcoming sections so it is good to understand them. They say this:

$\dot{Q}_i = \partial K/\partial \alpha_i$ [8.11b] $= \delta_{i,1}$

Our CT has taken us to a place where everything is cyclic except P_1 which says Q_1 is the only non-constant coordinate and is in fact linear in t. The solution to the above is

$Q_i(t) = \delta_{i,1}\, t + \beta_i$ where $\beta_i = \partial W/\partial \alpha_i$ = the constants shown page 275 9-7

While we are here, we should also explain the next step. Suppose we can form an independent set of constants $\gamma_i(\boldsymbol{\alpha})$. Then we can think of $K(Q_i, \alpha_i(\gamma_j))$. I think we are supposed to visualize this as some new function I will for the moment call $K'(Q_i, \gamma_i)$ where now we have $P'_i = \gamma_i$ as our new momenta conjugate to the same Q_i. Our Hamilton's equations are now:

$\dot{Q}_i = \partial K'/\partial \gamma_i$ $\partial K'/\partial Q_i = -\dot{\gamma}_i = 0.$

The first equation here can be written

$$\dot{Q}_i = \partial K'/\partial \gamma_i = \partial K/\partial \alpha_j \;\; \partial \alpha_j/\partial \gamma_i = \delta_{j,1} \;\; \partial \alpha_j/\partial \gamma_i = \partial \alpha_1/\partial \gamma_i = \text{a constant} \equiv \nu_i$$

So this $\partial K'/\partial \gamma_i = \nu_i$ is the meaning of the page 282 A, G does not show a prime on K (nor later does he put it on P_i). This of course leads to the solution:

$$Q_i = \nu_i t + \beta_i$$

where we are defining the β_i afresh right here.

With this all in mind, we can look at the right column on page 283. In my notation, I would have put primes on K and P_i in the case $\gamma \neq \alpha$, but I guess this is just a semantic thing.

In the action-angle variable section, we will have an example where $\gamma_i = J_i$.

Note added: In this γ_i world, we have $K'(Q_i, \gamma_i) = \alpha_1(\gamma_i)$. so ALL the new momenta $P'_i = \gamma_i$ appear in our expression $\alpha_1(\gamma_i)$ for the Hamiltonian. That means in general that NONE of the conjugate coordinates Q_i will be constants in time. As shown above, we know that $\dot{Q}_i$ *is* a constant ν_i, so the Q_i are then all linear functions of t (in the most general case).

Going back, above we said HJE says $H = \alpha_1$ but this is $H(q_i, \partial W/\partial q_i) = \alpha_1$, so this is what the HJE looks like in terms of W.

So, how now do you solve for the original $q_i(t)$ and $p_i(t)$? You have done your CT to the world of almost all P and Q being constants (or perhaps Q linear in time). We can compute the various constants α and β in terms of $q_i(0)$ and $p_i(0)$ once we find solutions for $q_i(t)$ and $p_i(t)$. But this takes the usual work, as shown in the HO example. You need to get some expression for W, then you can solve "inside out" 9-22b for the q_i and so on. Since t is not involved, we end up with "orbit equations", and we will be seeing

examples soon. Finally, G shows that W is none other than the indefinite integral of the abbreviated action S_0.

At this point, we get a very long table which summarizes how you solve a problem using S or W. The solution methods are not the same, because S and W don't generate the exact same CT. For constant H, you can use either method, but probably the W method is simpler. For general H, you have to use the S method. In the first S method, all the P_i and Q_i end up constants γ and β where you could take $\gamma = \alpha$ as a special case.

So, we now have a great formalism, we need now to apply it to something! The HO example was really an example of the W method!

Review of the Table:

1. Let's first look at the left side which can be used for H(p,q,t). We find a generating function $F_2 = S$ [S = Hamilton's Principle Function] which makes K = 0. Then Hamilton's equations tell us that $\dot{Q}_i$ and $\dot{P}_i = 0$ so that *all the coordinates and all the momenta* are constants of the motion, so we write $Q_i = \beta_i$ and $P_i = \alpha_i$. [p 283 top left] We then have to compute S by solving the HJ equation $H(q, \partial S/\partial q,t) + \partial S/\partial t = 0$. We can always shift from the α_i to some γ_i as noted above. The CT equations $p_i = \partial S/\partial q_i$ tell us nothing new because they are "built in" to the formalism already. But the other CT equations $Q_i = \partial S/\partial \gamma_i = \beta_i$ [p 284] can be "inverted" to find solutions for the $q_i(t)$ and we solve our problem. Again this final equation has the following form: $\partial S(q_i,\gamma_i,t)/\partial \gamma_i = \beta_i$, which we solve to get $q_i(\gamma_i,\beta_i,t)$. And S = the classical action. The text however provides no time-dependent H examples where this formalism is used.

2. Now we look at the right side of the table for case H(p,q) and no time.

{ Certainly with our three assumptions outlined in Chapter 1 above [conservative potential, scleronomous coordinates q_i in relation to the r_j , and non-holonomic constraints] we can say that H = E = conserved total energy. Explicit time would enter H say if you had a time-dependent potential, and one might expect this to do work on your system and make the energy be not conserved.}

So we write $S = W-\alpha_1 t$ and work with the CT generated by W [W = Hamilton's Characteristic Function] . We then have $K = H = \alpha_1 = P_1$, so the new Ham is cyclic in all coordinates and all momenta except P_1. We have $P_i = \alpha_i$. and $Q_i = \partial W/\partial \alpha_i = \delta_{i,1} t + \beta_i$. If we consider a different but independent set of constants $\gamma_i(\alpha_j)$, then we can restate these results as $P'_i = \gamma_i$ and $Q'_i = \partial W/\partial \gamma_i = \nu_i t + \beta_i$ where now $H = K = \alpha_1(\gamma_i)$ is non-cyclic in all the momenta P'_i and as a result, all the Q'_i have this linear time dependence. We have $\nu_i = \partial K/\partial \gamma_i = \partial H/\partial \gamma_i = \partial \alpha_1/\partial \gamma_i$. In this case, we normally omit the primes on the P_i and Q_i. We then have to compute W by solving the HJ equation $H(q, \partial W/\partial q,t) = \alpha_1$. The CT equations $p_i = \partial W/\partial q_i$ tell us nothing new because they are "built in" to the formalism already. But the other CT equations $Q_i = \partial W/\partial \gamma_i = \nu_i t + \beta_i$ can be "inverted" to find solutions for the $q_i(t)$ and we solve our problem. Again this final equation has the following form: $\partial W(q_i,\gamma_i,t)/\partial \gamma_i = \nu_i t + \beta_i$, which we solve to get $q_i(\gamma_i, \beta_i, \nu_i, t)$. And W = the abbreviated action. The text does an example of this method using the harmonic oscillator.

9.4 Separation of Variables in the HJE (284) By "separable" G means that you can write W down as a sum of functions W_i each of which depends only on its own coordinate q_i, all the constants α_j, and no

explicit time. So I accept the definition of separability as equation A on page 284, but I do not accept 9-23 because I am unable to derive it nor can I see what the H_i would be. Are we supposed to have $\Sigma_i H_i = H$? I don't think we shall need this equation, and perhaps later I can interpret it.

At this point, on page 285, we review again the case of conserved H. We can think of t as a coordinate, and so we "separate" W as shown mid page into W + S_2 where S_2 only contains t, separating it from the rest. This is just a way to interpret what we have already done.

Now suppose we have only one non-cyclic coordinate q_1. We "consider" a separable W of the special very simple form shown top page 286 which I interpret to say $W = \Sigma_i W_i(q_i, P_1 \ldots P_n)$. Equation 9-25a is our standard result of the previous work. Since H is cyclic in all but q_1, the *other* q_i don't appear in H and the p_i appear as constants α_i, so our equation p180 B is stated as 9-25b. We imagine that we solve this for W_1. The other W_i are obtained by trivially integrating 9-25a to get $W_i = \alpha_i q_i$. Thus, our full W has the form shown in 9-26. So in fact we end up with

$$W = W_1(q_1, P_1 = \partial W_1/\partial q_1, \alpha_2 \ldots \alpha_n) + \Sigma_{i\neq 1} W_i(q_i, \alpha_i) = W_1 + \Sigma_{i\neq 1} q_i \alpha_i$$

Notice that W_1 is a function of q_1 and all the P_i (mostly α_i), whereas the other W_i are functions only of q_i and $P_i = \alpha_i$. The sum above in W gets all the cyclic coordinates! If you think of $q_1 = t$ and $\alpha_1 = -H$ then you can think of the term $-\alpha_1 t$ (as the above sum) we used when *only* t was cyclic as a special case of the sum in 9.26, as in 9-19.

Example: Let's do central force V(r) in a plane with coordinates r and φ. Write H as equation C, done, and verified at page bottom. Cyclic in φ, but not r, so W_1 will be for the one non-cyclic r. Then 9-27 is application of 9-26 where φ is the only cyclic. The HJE says $H(q, \partial W/\partial q, t) = \alpha_1$, but if we install the fact that $W = W_1(r) + \alpha_\varphi \varphi$, the HJE becomes $H(r, \partial W_1/\partial r, t) = \alpha_1$. It is cyclic in φ. We can solve for W_1 as claimed, so Eq D then states the entire W for this problem! Ie, we have computed Hamilton's Characteristic function $W(r,\varphi; \alpha_1, \alpha_\varphi)$. Page 281 already gives us the general solution in the W formalism and so our example we can now compute $t+\beta_1$ as $\partial W/\partial \alpha_1$ which we directly do to get 9-29a. This gives us t = t(r) which we got with lots of work back in Chapter 3, where we used E and ℓ for α_1 and α_φ. Our second equation β_2 as $\partial W/\partial \alpha_\varphi$ of 9-22b then gives the orbit equation we got earlier! It is an orbit equation because time does not appear anywhere in it, and φ is sitting by itself so we get $\varphi = \varphi(r)$.

This example, G claims, shows how fast you can get to a problem's solution in separable cases. Very good!

9.5 Action-angle variables (288) Phase space trajectories (paths) are not something I have dealt with very much. Let's do a simple HO example:

$$q(t) = A \sin(\omega t + \varphi)$$
$$\dot{q}(t) = -A\omega \cos(\omega t + \varphi)$$
$$p(t) = -mA\omega \cos(\omega t + \varphi)$$

Then we find that $q^2 + p^2/(m\omega)^2 = A^2$. The phase space path is thus an ellipse centered at the origin. This would be an example of a closed orbit. We are periodic in both p and q, and with the same frequency, and this is the **libration** situation. A more general libration path is shown top left of page 289, again closed.

Aside: For the above HO I computed $J = \oint p dq = \oint p\dot{q}\, dt = \pi m A^2 \omega$. The meaning of the line integral is the area of the enclosed region, see my separate doc on that subject, and this is true even for arbitrary convex loops.

The next example is the simple pendulum as analyzed on page 289. For low energy, it swings back and forth and we get **libration**. Probably for very small energy, the phase space path is an ellipse since same as HO. As you increase energy, you eventually get just enough to let it be perfectly "up" and swing back and forth between this as a limit on each side. The pendulum in this case is not doing small motion and so is not a HO but something messier, and the path is shown to have cusps at π and $-\pi$. It would be a good problem to derive the shape of this curve and all those on the inside. If you increase E still more, it will go around forever and θ increases without limit, and you then get a "**rotation**" type phase space path. Another example is shown page 289 top right. The word "rotation" is apropo since θ keeps increasing. Obviously the rotation orbit is not closed.

In a multi-variable problem, you can think about the orbit or phase-space path in each separate $p_i q_i$ plane. The overall motion is "periodic" if it is periodic (libration or rotation) in all of these separate planes. A good example is the 3D HO with non-rational ratios. In each pq plane we get a simple elliptical closed motion with a closed orbit, while in the actual 3-space the orbit never closes. So be careful in what space you define an "orbit". This last situation is an example of **conditional periodicity**.

Now, we suddenly define, for each coordinate/momentum pair, $J_i = \oint p_i dq_i$ which is integral once around the closed path for closed orbit, or over one "cycle" for the rotation case. For cyclic where p = constant, curve is a straight line as per top 289 and we arbitrarily define a cycle as $\Delta q = 2\pi$, probably since this is often an angle. Notice that J_i has units of angular momentum which motivates the J symbol.

Clarification: see separate doc on line integrals of this type and why they give area inside loop.

Next, we tie this notion of periodic phase space orbits and the definition of the J_i with our last section on separable HJE work when H not a function of t (the right column of the big table). We assume that we have found some coordinates where the W function is separable, so we can write 9-32, which is really an equation for the phase space orbit $p_i = f(q_i)$. So throw this into J_i and get 9-35. You see that J_i is a function of the constants α_i so is itself a constant, and we are going to use the J_i's as the γ_i's of our previous Aside. That is, imagine that $P'_i = J_i$.

Now, instead of calling the coordinate conjugate to the J_i by the name Q'_i , we call them w_i . Since the J_i are angular momentum in dimension, the w_i must have dimensions of angle.

$J_i = \oint p_i dq_i$ dim = ang mom, looks like abbreviated action = "action variables"

$w_i = \partial W/\partial J_i$ like $Q_i = \partial W/\partial \gamma_i$ in right side table page 284 top

= conjugate to J_i with dim - angle = "angle variables"

$= \nu_i(J_i)\, t + \beta_i$! [since $\dot{\omega}_i =$ top right table page 284]

I would say at this point that you could compute the change in w_i over one libration period τ to get

$$\Delta w_i \equiv w_i(t+\tau) - w(t) = \nu_i(J_i)\,\tau$$

so if we knew Δw_i and ν_i, we would know the frequency of libration or rotation, namely

$$\text{frequency} = 1/\tau = \nu_i(J_i)/\,\Delta w_i.$$

We can compute Δw_i as shown bottom of page 292 and we find it is 1 if we integrate around one cycle in the i-plane (and not surprisingly we get 0 if we integrate around a cycle in some other plane). I was perhaps expecting $\Delta w_i = 2\pi$ since it is an angle, but it must be an angle measured in "cycles". We therefore have learned that

$$\text{frequency} = 1/\tau = \nu_i(J_i) \qquad \text{where } \nu_i(J_i) = \partial K/\partial J_i \quad \text{top right side 283}$$

So we seem to have a procedure here: (1) compute $W(\alpha)$ from the HJE; (2) use that to compute $J(\alpha)$ from 9-35: (3) invert that to get $H = \alpha(J)$; (4) then $\nu = \partial H/\partial J$.

G now does the HO as an example, starting bottom page 293. The first step is to compute W, but we already did that back on page 277. Notice that α appearing in W is the α of $H = \alpha$ and of $S = W - \alpha t$. The second step is then to compute J from 9-35 which is a line integral as shown page 293 bottom. When written with the square root, it is implied that you take the + branch on the top of the ellipse and – on the bottom, but this is swept under the rug. The θ parameterization removes this confusion allowing a simple computation of the integral which gives $J = 2\pi\alpha\sqrt{m/k}$ which we then "invert" to get $\alpha = (J/2\pi)\sqrt{k/m} = H$, and in the last step $\nu = \partial H/\partial J = \sqrt{k/m}\,/2\pi = \omega_{HO}/2\pi$ as expected.

The main point is this: Even in a messy problem we cannot solve or don't want to solve, we might be able to easily find W, compute J, invert it, and learn the frequency of periodic motion of our solution! We might not know the r(t) or even the orbit equation $\varphi(r)$, but we will know the period. That is the point of action-angle variables! We take our original problem and do a canonical transformation to a new system which has $K = H$, and $P_i = J_i$ and $Q_i = w_i = \nu_i\, t + \beta_i$ and ν_i is the resulting frequency.

Comments: this is not a simple subject because we are building higher levels of a huge towering construction involving the Hamiltonian, Hamilton's equations, canonical transformations to coordinates where almost everything is cyclic, then we separate out time and define the W function, then we require that it be separable, then we go to these action-angle variables as the last step.

9.6 Further properties of action-angle variables (294) If we stay in a particular q_k-p_k plane, and if we know that the motion is periodic with frequency ν_k, then we know that the motion can only be composed of this fundamental frequency and all harmonics. Thus, we can expand as in 9-45a:

$$q_k(t) = \Sigma_j\, a_j\, e^{2\pi i j w_k} \qquad \text{where } w_k = \nu_k t + \beta_k$$

The inverse is given by 9-47

$$a_j = \int_0^1 dw_k\, q_k\, e^{-2\pi i j w_k}$$

Here is my proof of this fact,

$$a_j = \int_0^1 dw_k\, q_k\, e^{-2\pi i j w_k} \quad = \int_0^1 dw_k\, [\Sigma_{j'}\, a_{j'}\, e^{2\pi i j' w_k}]\, e^{-2\pi i j w_k}$$

$$= \Sigma_{j'}\, a_{j'} \left\{ \int_0^1 dw_k\, e^{2\pi i (j'-j) w_k} \right\} = \Sigma_{j'}\, a_{j'}\, \delta_{(j'-j),0} = a_j$$

For a rotation, assume that on each cycle q_k increases by some fixed amount q_{0k}. Then

$$q'_k = q_k - w_k q_{0k}$$

increases by 0 on each cycle, since w_k increases by 1 per cycle. So q'_k is periodic and we can do the expansion in 9-45b just duplicating what we did in 9-45a, so fine.

If we combine the motion on all the subplanes, and we then move back from the n generalized coordinates q_i to our actual Cartesian coordinates x_i, we will get a mix of all the planar frequency effects and we then have the Fourier series shown in 9-48 which, in general, gives a non-periodic motion in Cartesian space, although we were periodic in all of the sub-planes. This situation was called **conditional periodicity** earlier, and now G also calls it **multiply periodic**.

Before continuing, on page 296 we have a little digression. Suppose you define W' as shown. If any $q_i = w_i$ does a complete cycle, we know its characteristic function W increases by J_i because this is what 9-35 on page 291 tells us. This is one form of the definition of W_i. A complete cycle for a w_i means it increases by 1. Thus, the function W' = W(q,p) - **w•J** is periodic in each w_i and you could if you wanted Fourier expand this thing the way we did the x_i on page 295. G comments that W' as so written is a Legendre transform thing, and you get W'(q,w) which is the F_1form. This W' is another transformation that maps you from q,p to w,J, but I don't see why this is useful yet.

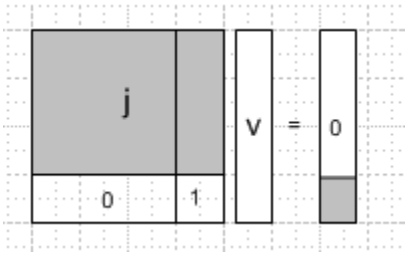
Now on p 297 we start into the whole issue of how you handle things when at least some of the frequencies in some of the sub-planes q_k-p_k are harmonically related. You might have $3\nu_1 + 5\nu_2 = 0$, say. If some frequencies are related like this, the problem is "**degenerate**". If there are $m \leq n-1$ relations of this form, you have **m-fold degeneracy**, and if m = n-1 you have **complete degeneracy**. G uses the strange word **commensurability** to refer to the conditions of degeneracy.

The problem is treated on page 298 and I change to a better notation here. The idea is that we are going to do a transformation from w,J to some new w',J' where all the m degenerate frequencies ν' will be 0, and this means that none of the corresponding J' will appear in the H' Hamiltonian, which I guess is going to make it a simpler Hamiltonian to deal with. This is all proven right here:

First, write

$\Sigma_{i=1}^{n}\ j_{ki}\ \nu_i = 0$ for k = 1...m // these are the harmonic relations, assume m of them

Here we are using some elements of a square nxn matrix which I called ***j.*** The above equation appears as this way, and I have arbitrarily added a chunk of 0 and some identity matrix to this thing:



There are only m zeros in the right column, only m equations

Now consider the proposed F_2 transformation which we can write as:

$F_2 = \mathbf{J'} \bullet \boldsymbol{j} \bullet \mathbf{w} = (\mathbf{w} \bullet \boldsymbol{j}) \bullet \mathbf{J'}$ = 9-54

This transform does this, where I am showing the traditional and current names of things

q_i	p_i	Q_i	P_i
w_i	J_i	w'_i	J'_i

Since the above F_2 has the form $F_2 =$ **f** (w) • **J'** = **f**(q) • **P** , we see that this matches the form of 8-19 on page 244 and thus this is indeed a "point" transformation.

Next, our 8-11b equation tells us:

$Q_j = \partial F_2/\partial P_j$ => $w'_i = \partial F_2/\partial J'_i = \partial\ (\mathbf{J'} \bullet \boldsymbol{j} \bullet \mathbf{w})\ /\partial\ J'_i = (\boldsymbol{j} \bullet \mathbf{w})_i$ = 9-55

=> **w'** = ***j*** • **w** // identity for the last non-harmonic ones

This shows how the q's are shuffled into Q's by this point transformation. Notice how the last terms in this vector ***j*** • **w** use the 0 and 1 portion of the ***j*** matrix and the bottom part of the vector w' is just a replication of the original w vector.

Next, our 8-11a equation page 241 tells us:

$p_j = \partial F_2/\partial q_j$ => $J_i = \partial F_2/\partial\ w_i = \partial\ (\mathbf{J'} \bullet \boldsymbol{j} \bullet \mathbf{w})\ /\partial\ w_i = (\ \mathbf{J'} \bullet \boldsymbol{j}\ )_i = (\ \boldsymbol{j}^T \bullet \mathbf{J'})_i$

=> **J** = $\boldsymbol{j}^T$ • **J'** 9-57 // identity for the last non-harmonic ones

Notice that we are not claiming we can invert matrix ***j.*** Probably in general we cannot.

Now we can look at the new set of frequencies ν_i'

$$\mathbf{v}' = \dot{\mathbf{w}}' = \boldsymbol{j} \bullet \dot{\mathbf{w}} = \boldsymbol{j} \bullet \mathbf{v}$$

This is the full matrix version of what we showed graphically above. We can see at once that the first m components of the **v**' vector are 0 due to our degeneracy relations above, while the last n-m are not. So, the point is that this "point canonical transformation" has mapped the degenerate frequencies all to the value 0, and has left the non-degenerate ones "as they were".

Now consider this Hamilton's equation (and K = H as usual since no t dep)

$$\dot{Q}_i = \partial H/\partial P_i \qquad \Rightarrow \qquad \dot{w}'_i = \partial H/\partial J'_i = v'_i = 0 \text{ for the first m values of i}$$

This means that K is cyclic in the first m values of the generalized momenta J'_i.

9.7 The Kepler problem in action-angle variables (299) You have to imagine the little 3D picture of a particle at location r,θ,φ. There is a lot of derivation work to do here, and a fair dose of confusion in the names of symbols. Imagine that a Kepler ellipse lies in the grey plane shown at the bottom of page 107, and the position of a planet is described by θ, φ (to establish the plane) and then ψ and r (planet position in the plane).

Part I: The basics

First we know we can make our little orthogonal system $\hat{\theta}$, $\hat{\varphi}$ and $\hat{r}$ and I have page of useful stuff in my math binder. **Ambiguity** arises in the definition of physical versus canonical things like velocity and momenta. I will use upper case today for physical, and lower case for canonical. Then we can say

$$V_r = \dot{r} \qquad V_\theta = r\dot{\theta} \qquad V_\varphi = r\sin\theta\,\dot{\varphi}$$

The last two are very obvious. The radius for the θ circle really is r, but for the φ circle is rsinθ. We then have these physical momenta

$$P_r = m\,V_r = m\dot{r} \qquad P_\theta = mV_\theta = m\,r\dot{\theta} \qquad P_\varphi = mV_\varphi = m\,r\sin\theta\,\dot{\varphi}$$

and of course

$$P^2 = P_r^2 + P_\theta^2 + P_\varphi^2 = m^2\,[\,\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\,\dot{\varphi}^2\,]$$

And so $H = P^2/2m - k/r$ and we get the desired result

$$H = (m/2)\,[\,\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\,\dot{\varphi}^2\,] - k/r \qquad \text{// see } \mathbf{9\text{-}58}$$

While we are using these physical things, we can do one more thing:

$$\mathbf{L}/r = \hat{\mathbf{r}} \times \mathbf{P} = \hat{\mathbf{r}} \times [\,P_r\hat{r} + P_\theta\hat{\theta} + P_\varphi\hat{\varphi}\,] = P_\theta\,\hat{\varphi} - P_\varphi\hat{\theta}$$

where we use our reference sheet for the cross product of the unit vectors. Therefore

$$L^2/r^2 = P_\theta^2 + P_\varphi^2 = (m r\dot{\theta})^2 + (m r\sin\theta\, \dot{\varphi})^2 = m^2r^2 [\dot{\theta}^2 + \sin^2\theta\, \dot{\varphi}^2]$$

Now we can rewrite H as follows, using the above,

$$H = (m/2) [\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\, \dot{\varphi}^2] - k/r$$
$$= (1/2m) [m^2 \dot{r}^2 + m^2r^2\dot{\theta}^2 + m^2 r^2\sin^2\theta\, \dot{\varphi}^2] - k/r$$
$$= (1/2m) [m^2 \dot{r}^2 + L^2/r^2] - k/r$$
$$= (1/2m) [p_r^2 + L^2/r^2] - k/r$$

where I have used ahead of time the canonical momenta p_r found below. Thus, we have derived equation C on page 300 except L^2 is presented as p^2 which is immensely confusing, but OK, we shall see soon why this is done.

Now enough physical stuff, we turn now to canonical quantities. We have

$$v_r = \dot{r} \qquad v_\theta = \dot{\theta} \qquad v_\varphi = \dot{\varphi}$$

where I am just defining these as time derivatives, but they are not used anywhere and are dangerous. Now here is the Lagrangian

$$L = T\text{-}V = (m/2) [\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\, \dot{\varphi}^2] + k/r$$

We can then directly compute the canonical momenta from $p = \partial L/\partial \dot{q}$ to get

$$p_r = m\dot{r} \qquad p_\theta = mr^2\dot{\theta} \qquad p_\varphi = m r^2\sin^2\theta\, \dot{\varphi} \qquad \text{// } \mathbf{9\text{-}59}$$

$$P_r = m V_r = m\dot{r} \qquad P_\theta = mV_\theta = mr\dot{\theta} \qquad P_\varphi = mV_\varphi = m r\sin\theta\, \dot{\varphi}$$

so

$$P_r = m\dot{r} = p_r \qquad P_\theta = mr\dot{\theta} = p_\theta/r \qquad P_\varphi = mr\sin\theta\, \dot{\varphi} = p_\varphi/(r\sin\theta)$$

Notice that we do not have p = mv using canonical quantities. No one talks about canonical v. I replicated the physical P line from above for comparison, and you see how the physical and canonical momenta are related. While we are here, let's re-express L^2 as follows, continuing from above,

$$L^2/r^2 = P_\theta^2 + P_\varphi^2 = (m r\dot{\theta})^2 + (m r\sin\theta\, \dot{\varphi})^2 = m^2r^2 [\dot{\theta}^2 + \sin^2\theta\, \dot{\varphi}^2]$$

$= (p_\theta/r)^2 + [p_\varphi/(r\sin\theta)]^2 = (1/r^2)\,[\,p_\theta^2 + p_\varphi^2/\sin^2\theta]$

so we can then identify

$L^2 = [\,p_\theta^2 + p_\varphi^2/\sin^2\theta]$

Looking back at 9-60 we have the following, which will be used a few lines below:

$H = (1/2m)\,[\,p_r^2 + L^2/r^2\,] - k/r$ which is page 300 C, but he uses p^2 in place of L^2

Now we can solve these for the dotted quantities and put back into H of 9-58 and we get **9-60** which is H in terms of the canonical coordinates and momenta

$H(q_i, p_i) = (1/2m)\,[\,p_r^2 + p_\theta^2/r^2 + p_\varphi^2/(r^2\sin^2\theta)\,] - k/r = (1/2m)\,[\,p_r^2 + L^2/r^2\,] - k/r$

where this last form appears on page 300 as equation C.

At once we see H is cyclic in φ so p_φ = constant = α_φ in a minute here.

To get the Hamilton-Jacobi equation, you use $p_i = \partial W/\partial q_i$ and set $H = \alpha_1 = E$. This trivially gives **9-61**. We then try to separate W into its three parts to see if things will work out OK as in **9-62**. This means we replace each W with W_i in its derivative in 9-61.

First, the φ part requires that $\partial W_\varphi/\partial\varphi$ = constant = α_φ (which is p_φ). Reading top of page 300 makes us agree with 9-63b where another constant is defined, α_θ, which contains α_φ as shown. Our purely radial equation is then 9-63c in terms of our three constants, one hidden.

Now G wants to associate each of our "constants" with something "conserved". We already know that p_φ = conserved = α_φ . Canonical momenta to an angle is always an ang mom, so we find that angular momentum about the φ axis is conserved. Of course we know that **L** overall is conserved, but ignore that fact for the moment.

Looking back we see we have

$\alpha_\theta^2 = [\,p_\theta^2 + p_\varphi^2/\sin^2\theta] = L^2 \equiv p^2$ so $\alpha_\theta = L = p =$ the total momentum magnitude

And of course $\alpha_1 = E$ is also conserved. So E, L_z and L^2 are conserved quantities.

Comment on the alphas: Let's back up a moment and remember a few things:

$p_i = \partial W_i/\partial q_i$ says

$p_r = \partial W_r/\partial q_r$ $p_\theta = \partial W_\theta/\partial\theta$ $p_\varphi = \partial W_\varphi/\partial\varphi$

It happens that in our H, φ does not appear, so p_φ is a constant called α_φ. The other two p_i are NOT constants. You have to solve the page 300 equations for them. The solutions are shown top of page 301 as the integrand arguments. The alpha called α_θ is not particular p_i or P_i, it is L which is called p. When we

do our HJ transformation below, we end up with $P_i = J_i$ and $Q_i = w_i$. It turns out that α_θ is a linear combination of the J's as shown bottom page 301.

Part II: Computing the $J_{\underline{i}}$ action variables

These are stated bottom of page 300 and are correctly written out top of page 301. As usual, you have to about branches of the square root at least in the last equation.

Computation of J_φ is trivial and we get $J_\varphi = 2\pi\alpha_\varphi$.

Computation of J_θ is going to require a trick of Van Vleck. If we look back to our "planar" central force work on page 60, we find that we could write

$$T = (m/2)\,[\dot{r}^2 + r^2\dot{\psi}^2\,] \qquad \text{// as in 3-6}$$

where to avoid confusion, we change that name of that page 60 angle from θ to ψ. But in line 3-8 we shows that $m\,r^2\dot{\psi} = L$, the total angular momentum magnitude in this problem, which we are agreeing for the moment to call "p" or better , p_ψ. The simple "p" notation reminds us that this canonical momentum p_ψ is the same as the total angular momentum L, so a subscript is confusing perhaps.

Thus from the planar analysis we know that

$$T = (m/2)\,[\dot{r}^2 + r^2\dot{\psi}^2\,] = (m/2)\,[\dot{r}^2 + (r^2\dot{\psi})^2/r^2\,] = (m/2)\,[\dot{r}^2 + p^2/r^2\,]$$

But I am off the track. Recall from page 54 that, with conservative forces, we know

$$2T = \Sigma_i p_i \dot{q}_i$$

So let's write this in our two systems:

$$2T = p_r\dot{r} + p_\theta\dot{\theta} + p_\varphi\dot{\varphi}$$

$$2T = p_r\dot{r} + p_\psi\dot{\psi} = p_r\dot{r} + p\dot{\psi} = p_r\dot{r} + p\dot{\psi}$$

so comparing we have

$$p_r\dot{r} + p_\theta\dot{\theta} + p_\varphi\dot{\varphi} = p_r\dot{r} + p\dot{\psi}$$

which says

$$p_\theta\dot{\theta} + p_\varphi\dot{\varphi} = p\dot{\psi} \qquad \text{or} \qquad p_\theta d\theta + p_\varphi d\varphi = p d\psi$$

This is the fact we need to have, a pretty fast derivation, might take a while geometrically. We then have

$$p_\theta d\theta = p d\psi - p_\varphi d\varphi$$

Therefore, $J_\theta = \oint p_\theta d\theta = \oint p d\psi - \oint p_\varphi d\varphi = 2\pi p - 2\pi p_\varphi = 2\pi(\alpha_\theta - \alpha_\varphi)$

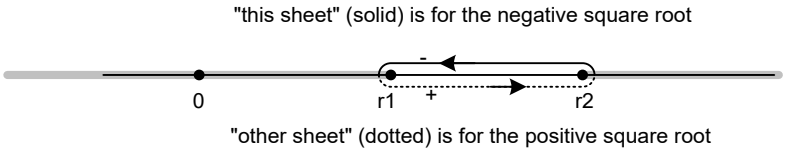
So all this last stuff was a little trick to do the integral 9-65b without using an integral table.

Computation of J_r is going to be more serious. It is written out in 9-68 and we have now to do this integral. Another trick shall be invoked, this one from Sommerfeld, our friend. Since E < 0, let's write the square root this way:

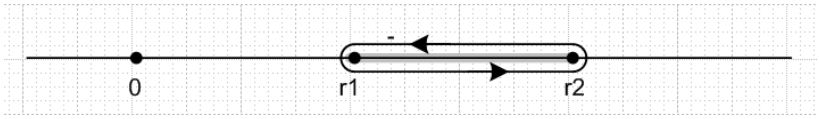
$$\sqrt{-a + 2B/r - C/r^2} = (1/r)\sqrt{-ar^2 + 2Br - C} = (1/r)\sqrt{(-a)(r^2 - 2Br/a + C/a)}$$

$$= (1/r)\sqrt{(-a)(r-r_1)(r-r_2)} = (1/r)\sqrt{(a)(r-r_1)(r_2-r)} = (\sqrt{a}/r)\sqrt{(r-r_1)(r_2-r)}$$

Notice that $r_1 r_2 = C/a$. When r is in the range $r_1 < r < r_2$ the above square root is real. This is good since this is basically p_r during an orbit. Here is what the branch situation is in the complex r plane: We arrange to do the outbound path on the + sheet, and the return inbound path on the – sheet.



The function is real analytic on the interval. This is not the p_r-r space orbit picture, this is the complex r plane by itself. Because the contour is as shown, we can deform the branch cuts so that they meet and then the entire contour will be on the – sheet, like so:



Now we can deform the contour and pick up the residue at r=0 and another at r=∞. This is shown somewhat poorly at the top of page 302 where the regions labeled positive and negative square roots don't make sense to me. Those are sheets, not line segments. But the contour senses are correct for the deformation.

So, now replace -a with A and define

$$f(r) = -\sqrt{A + 2B/r - C/r^2} \qquad \text{// the – sheet function}$$

At the origin, a Laurent series expansion would show the behavior $f(r) = -\sqrt{-C}/r$ so CCW residue there would be just $-\sqrt{-C}$, just as G claims page 303 #1. Now the great circle part can be handled as is, or we

can convert it to a contour around the origin in z = 1/r space. Doing this changes the rotation sense of the contour, but then dz = -dr/r^2 makes a cancelling minus sign, so you get dr = – dz/z^2 and we end up with formula 3 including the correct sign shown (same sign as 1). This has a double pole at the origin, and the rule for residue there is f'(z=0) where f = – $\sqrt{A + 2Bz - Cz^2}$ so f' = – (1/f) (B-Cz) → – B/$\sqrt{A}$. Now both contours required are backwards, need CW as shown, not CCW, so get extra minus on both residues, and we end up with

$$\oint f(r)\, dr = 2\pi i \left[+ \sqrt{-C} + B/\sqrt{A} \right] \qquad \text{// as claimed in 5.}$$

We then identify A = 2mE B = mk $C = (J_\theta + J_\varphi)^2 / (4\pi^2)$

so integral becomes

$$J_r = \oint f(r)\, dr = 2\pi i \left[+ \sqrt{-C} + B/\sqrt{A} \right]$$

$$= 2\pi i \left[i\sqrt{C} - i\, mk/\sqrt{-2mE} \right] = -2\pi\sqrt{C} + 2\pi mk/\sqrt{-2mE}$$

$$= -(J_\theta + J_\varphi) + \pi k\sqrt{2m}/\sqrt{-E} \qquad \text{// as in 9-69}$$

where we have always used this idea: $\sqrt{-x} = +i\sqrt{x}$. Obviously more could be said about that, but since we got the right answer, fine. Now solve for E to get

$$(J_r + J_\theta + J_\varphi)^2 = \pi^2 k^2\, 2m/(-E)$$

$$H = E = -2\pi^2 mk^2/(J_r + J_\theta + J_\varphi)^2 \qquad \text{as in 9.70}$$

This result seems rather amazingly symmetric to me. Well, remember how you get the frequencies:

$$\nu_i = \partial H/\partial J_i$$

so the fact that all appear as a simple sum means that the three derivatives will all be the same, and so all three frequencies will be the same (as we expect for Kepler orbits).

Next, we are going to do the extra transformation to Q,P = w', J' to zero out two of the frequencies ν_i. We use exactly the little matrix formalism presented earlier and I have verified every equation on page 304, mostly there in pencil. We know that we get $\nu_1' = \nu_2' = 0$ and $\nu_3' = \nu_3$. As shown on page 300 bottom, the order of things is 1,2,3 = φ, θ, r (probably he did this to get the tough J_r to be treated last). The new J' are now displayed. We know that the first two cannot appear in H because $\nu'_i = \partial H/\partial J_i$ (p298) = 0 for 1,2. This is seen to be correct, and only J_3' appears as shown 9-75.

At this point, G enters a 3-page commentary section with very important remarks which I will try to put in my own words.

• First, in the action-angle world, you end up with $w_i = \nu_i t + \beta_i$. I failed to realize how important this statement is. No matter how messy your original coordinate functions $r_i(t)$ might be for an exact solution to the problem, these "angle" coordinates at most do simple linear motion in t ! It is rather amazing that you can transform a tough problem to a set of "coordinates" where the motion is nothing more than linear in time.

• When we get to the final world here with $w_i' = \nu_i' t + \beta_i'$ using F_2 of 9-54, we have $\nu_1' = \nu_2' = 0$, which says that $w_1' = \beta_1'$ and $w_2' = \beta_2'$ and both these angle variables are constants for Kepler! G points out that since $J_1' = J_1 = J_\varphi$, we can think of w_1' as some fixed angle like the φ shown on page 107, the position of the line of nodes. Similarly, $J_2' = J_\varphi + J_\theta = 2\pi L = 2\pi p$ is related to total angular momentum L, so its angle variable w_2' must be a ψ angle in page 107. It is a "fixed angle" β_2' which one often takes to mark the position of the perihelion (closest approach) of an orbit. Finally, since $J_1'/J_2' = p_\varphi/p = \cos\theta$ so we get a connection to the third polar angle.

So his point is that the two w_i' and this ratio are constants which are equivalent to the two Euler angles which define the motion plane, and the third ψ which tells you a point in that plane.

• Action-angles are currently in use for working with planetary orbits. Called Delaunay elements, I think they involve only r and ψ and you work in the orbital plane, I have seen F and G mentioned. When you involve complex orbital perturbations, you can characterize them as causing the normally static angle variables to start moving slowly according to $w_i' = \nu_i' t + \beta_i'$ where the 1 and 2 ν_i' are no longer exactly zero due to the perturbation. So this is even currently a significant role for action-angle tools. (Maybe since 1950 computers have replaced all this analysis?)

• In the Bohr atom, it turns out that the path to quantization is simply to say that the actions J_i are quantized in multiples of Plank h. Sommerfeld's "royal road to quantization". Thus, if we take $J_3' = nh$, we get the Bohr formula for energy levels showing n as the principle quantum number. Very cool. Then since J_2' is related to total angular momentum $\boldsymbol{\ell}$, we can write $J_2' = \boldsymbol{\ell}h$ and then quantized $\boldsymbol{\ell}$ gets into our hydrogen atom model. In planets, this is called k and affects precession of the perihelion. Then add a fixed B field for Zeeman effect and you get $J_1' = mh$ and that is your magnetic quantum number. At each step, you break a degeneracy and you see it all happening in these action-angle variables.

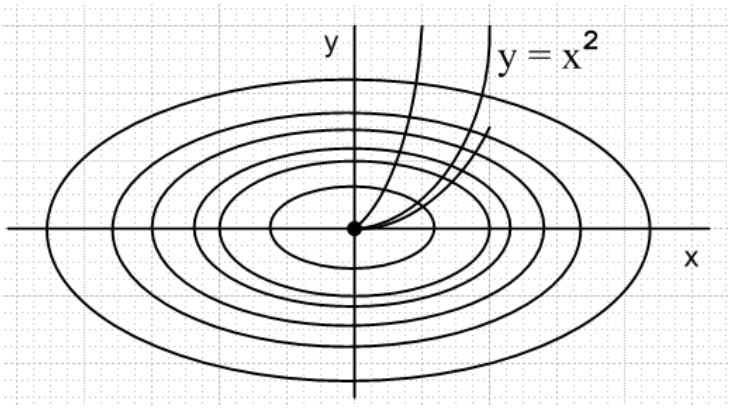
• In the early days of the old Bohr model, everyone was using action-angle as a "daily tool". But then when the full quantum theory was discovered, and always required perturbation theory, it was always easier to do it in quantum and not classical mechanics, so action-angle went by the wayside. It remains now in astronomy and in text books like this one!

9.8 Hamilton-Jacobi theory, geometric optics, and wave mechanics (307)

Preamble: Before starting here, consider a set of ellipses made simply by scaling z, so we have

$$z = f(x,y) = ax^2 + by^2$$

For example, if we set z = 1 and a = $1/A^2$ and b = $1/B^2$, we get the "standard ellipse" with semi-major axis A and so on. In 3D, our function z = $ax^2 + by^2$ is an "oval cup" made of every larger ellipses stacked on top of each other (called a paraboloid). If we look from above and project all these ellipses onto the x-y plane, we get our family of concentric ellipses of interest. We can then examine the curves which are normal to all the ellipses. Here is a picture:



It turns out that you find the normal curves by solving the equation dy/dx = f_y/f_x . In our case, the solution to the equation is y = K $x^{b/a}$ where constant K determines the location of the curve. Some of these are drawn above for K = 2, 1 and ½ in the case b/a = 2. You can roughly see in this crude drawing that all intersections are right angles.

It took me several days (!) to figure out the above little detail, see "surface geometry questions.doc".

I am now going to skip to the optics discussion starting page 310 equation 9-87 which is the wave equation we all know and love. The plane wave solutions are written for constant index n. At this point, due to my unfamiliarity with optics, we define **"geometric optics"** to be the system of equations you get when you assume that n varies, but varies slowly compared to light wavelength λ. Probably this subset of "optics" is not going to know about diffraction fringes, but will know about Snell's law and so on. I have a whole unread optics book by Born and Wolf, by the way. Another gaping hole in my physics knowledge.

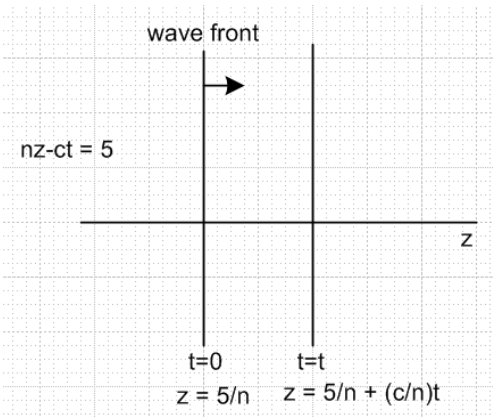
So, suppose we imaging that if n varies slowly, we will still have the "plane wave form", but things will vary slowly in phase and amplitude. That is, assume A and L are real and the form is

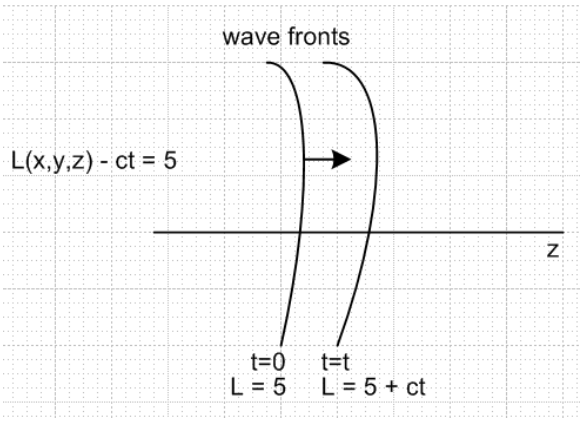
$\varphi(\mathbf{r},t) = \exp [A(\mathbf{r}) + ik_o (L(\mathbf{r}) - ct)]$ where $k_o = \omega/c$ as if we had n= 1 (vaccuum)

Meanwhile, our wave equation solution for constant n has this form

$\varphi(\mathbf{r},t) = \exp [0 + ik_o (n \hat{\mathbf{k}}\bullet\mathbf{r} - ct)]$ normally seen as $\exp [i (\mathbf{k}\bullet\mathbf{r} - \omega t)]$ $k = nk_o$

We would then say that for constant n, the attenuation factor $A(\mathbf{r}) = 0$ and $L(\mathbf{r}) = n \hat{\mathbf{k}}\bullet\mathbf{r} = nz$ for example, if k is in the z direction. You can see that L is some kind of phase thing. In the plane wave, we would write phase = (kz-ωt) and plane wave propagation of a particular "wave front" has (kz-ωt) = constant. We can also write this as (nz-ct) = constant since $\hat{\mathbf{k}}\bullet\mathbf{r}$= z for our special case. We then have some kind of "phase wave front" which is moving along at velocity u = c/n. To see this, consider the left figure:





Again, we are looking at a "surface" nz – ct = constant and it moves to the right for our plane wave.

What is going to happen when n varies slowly? The amplitude of φ will vary due to A(r), but more importantly for our purposes, the "wave fronts" will be something like that shown on the right. These are now surfaces on which L-ct = constant phase. We know that L(x,y,z) = ct => a changing shape as t changes, so that is why we have drawn the two surfaces with different shapes. From our several days spent on surface geometry, we know the following about the situation on the right, where imagine that t = dt = very small:

• We could define phase P such that $P(\mathbf{r},t) = L(\mathbf{r}) - ct$. Our left wavefront then has the equation $P(\mathbf{r},0) = 5$. The right wavefront is then $P(\mathbf{r},t) = 5$. Of course this is a differently shaped curve, but all the wavefronts have the property that P = 5, surfaces of "constant phase". As we let time go by, the surface $P(\mathbf{r},t) = 5$ takes on the various shapes shown. Thus, at time t, P(r,t) aligns with the surface L = 5+ct.

• the normal at some point on the t=0 (L=5) surface is given by $\mathbf{n} = \nabla_{3D}L(\mathbf{r})/ |\nabla_{3D}L(\mathbf{r})|$. This of course points do the next wavefront at time t=dt.

• if the distance between the wave fronts is ds, we know that the wave front velocity at some point is given by u = ds/dt

• When dealing with the case g(x,y,z) = 0, we said that $g(\mathbf{r}+d\mathbf{r}) = g(\mathbf{r}) + \nabla_{3D}g \bullet d\mathbf{r} = \nabla_{3D}g \bullet d\mathbf{r} = dg$. So if we go off in the normal direction away from g, we have $dg = |\nabla_{3D}g(\mathbf{r})|\, ds$ since $d\mathbf{r} = \hat{\mathbf{n}}\, ds$.

• If we think of $g(\mathbf{r}) = L(\mathbf{r}) - 5$ in our left wavefront above, then we have $dL = |\nabla_{3D}L(\mathbf{r})|\, ds$.

• But we know that dL = cdt because our wavefronts are surfaces where (L-ct) = 5.

• Combining the last three items, namely that u = ds/dt and $dL = |\nabla_{3D}L(\mathbf{r})|\, ds$ and dL = cdt, we find

$$u = u(\mathbf{r}) = ds/dt = [dL/|\nabla_{3D}L(\mathbf{r})|] / dt = c/|\nabla_{3D}L(\mathbf{r})|$$

$$\mathbf{u(r)} = \{ c/|\nabla_{3D}L(\mathbf{r})|\}\, \hat{\mathbf{n}} = \{ c/|\nabla_{3D}L(\mathbf{r})|\} [\, \nabla_{3D}L(\mathbf{r})/ |\nabla_{3D}L(\mathbf{r})| \,] = c\nabla_{3D}L(\mathbf{r})/ |\nabla_{3D}L(\mathbf{r})|^2$$

So this is pretty good, we know the wavefront velocity at any point as a scalar and as a vector!

The question now arises in geometric optics: How are we going to solve for the function L(**r**) ? By starting with the wave equation 9-87 and inserting our form for φ involving A and L, G shows us that we get the equations shown bottom of page 311. But in the geometric optics limit where δn >> λ, G shows that we can neglect the first two terms in (9-93a), and we end up with this being the equation we need to solve for L

$$(\nabla_{3d}L(\mathbf{r}))^2 = n^2(\mathbf{r}) \qquad \text{or} \qquad |\nabla_{3D}L(\mathbf{r})| = n(\mathbf{r})$$

Notice this is not Laplace's equation, it is this: (there are no second derivatives)

$$(\partial L/\partial x)^2 + (\partial L/\partial y)^2 + (\partial L/\partial z)^2 = n^2(\mathbf{r})$$

This is called the **eikonal equation** and I have never worked with it, but I can see that it "controls" geometric optics. The function L here (called S in Born and Wolf) is called "the eikonal".

---

[1]The Eikonal equation is derived about 150 years ago by Sir William Rowan Hamilton. The word *Eikonal* was introduced in 1895 by H. Burns. It comes from the Greek word "εικων", from which the modern word *icon* derives. The equation's title is descriptive because it controls the formation of images in optical systems.

I cannot locate H. Burns on the web, Hamilton is *the* Hamilton.

OK, so I get the picture. In geometric optics, you need to solve this eikonal equation for L(r), and then you can draw your nice curved wavefronts and look at your normal paths I called h(x) in my surface notes.

Note Added: Let's go back and review the above a bit, because things are still hazy. I want to write the little expo φ lots of ways now and "gather up" all the little micro facts on this subject:

$$\varphi = \exp[\, i\,(\mathbf{k}\bullet\mathbf{r} - \omega t)] \qquad \text{// familiar to me}$$

But we have $k = nk_0$ where n is some index ≠ 1, and k_0 is what k would be if n= 1, ie, in vacuo. So now

$$\varphi = \exp[\, i\,(n\,\mathbf{k}_0\bullet\mathbf{r} - \omega t)] = \varphi = \exp[\, i\,(n\,k_0\,\hat{\mathbf{k}}_0\bullet\mathbf{r} - \omega t)] = \exp[\, i\,(k_0\ \{n\,\hat{\mathbf{k}}_0\bullet\mathbf{r}\} - \omega t)]$$

Now it is the curly bracket thing that is defined to be L, so we do one more step

$$\varphi = \exp[\, i\,(k_0\ \{n\,\hat{\mathbf{k}}_0\bullet\mathbf{r}\} - \omega t)] = \exp[\, i\,(k_0 L(\mathbf{r}) - \omega t)] \qquad L(\mathbf{r}) = n\,\hat{\mathbf{k}}_0\bullet\,\mathbf{r}$$

Now in optics problems, ω is a constant. Think of Snell refraction at a boundary between medium and vacuum. On both sides, ω is the same, so we can say $\omega = \omega_0$ and $\nu = \nu_0$ if we want. On each side we would have [u = phase velocity symbol, wavefront speed]

$u = \lambda/T = \lambda\nu$ $\quad u_0 = \lambda_0/T_0 = \lambda_0\nu_0$ $\quad$ but $\nu = \nu_0$ so

$u/\lambda = u_0/\lambda_0$ and of course we have

$k = 2\pi/\lambda$ and $k_0 = 2\pi/\lambda_0$. $\quad$ and $u_0 = c$ $\quad \omega/k = \lambda/T = u$ $\quad \omega/k_0 = \lambda_0/T_0 = c$
$\omega = 2\pi/T$ and $\omega_0 = 2\pi/T_0$

Now let's check our eikonal equation for constant index: (note that $\nabla(\mathbf{A}\bullet\mathbf{r}) = \mathbf{A}$ for constant $\mathbf{A}$)

$\nabla L(\mathbf{r}) = \nabla \{ n \hat{\mathbf{k}}_0 \bullet \mathbf{r} \} = n \hat{\mathbf{k}}_0 \Rightarrow$ $\quad (\nabla L(\mathbf{r}))^2 = n^2$ $\quad$ = eikonal equation.

Here is one more way to write φ where we use $\omega = ck_0$ then we get

$\varphi = \exp[i (k_0 L(\mathbf{r}) - \omega t)] = \exp[i (k_0 L(\mathbf{r}) - k_0 ct)] = \exp[i k_0 (L(\mathbf{r}) - ct)]$

And let's do yet one more variant of the above ωt form

$\varphi = \exp[i (k_0 L(\mathbf{r}) - \omega t)] = \exp\{ i 2\pi([k_0/2\pi] L(\mathbf{r}) - [\omega/2\pi]t)\} = \exp\{ i 2\pi(L(\mathbf{r})/\lambda_0 - \nu t) \}$

So, what is the connection between this optics stuff and our functions S and W of the Hamilton-Jacobi classical mechanics world? Well, our general HJE says

$H(q_i, p_i = \partial S/\partial q_i) + \partial S/\partial t = 0.$ $\quad$ 9-3 on page 274

If H = E, constant, then we know that $\partial S/\partial t = E$. This leads us to define

$S = W(q_i) - Et$

Notice that $\partial S/\partial q_i = \partial W/\partial q_i$ for the normal coordinates. Thus, the HJE in terms of W becomes

$H(q_i, p_i = \partial W/\partial q_i) = E$ $\quad$ // E is usually called α_1 by G

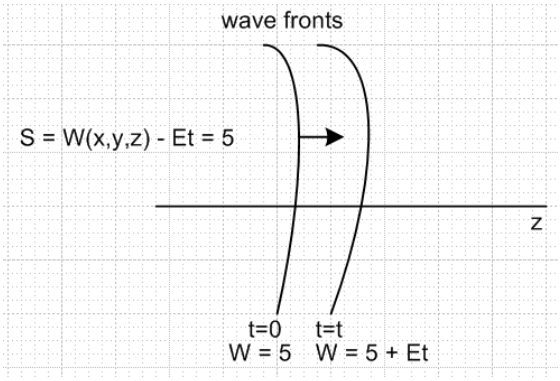
Now, suppose you have a single particle in a potential. Then we have $H = p^2/2m + V$ and the HJE then says this:

$(\nabla W)^2/2m + V = E$ $\quad$ or $\quad (\nabla W)^2 = 2m(E-V(r))$

So compare this to our eikonal equation of geometric optics:

$(\nabla L(\mathbf{r}))^2 = n(\mathbf{r})$
$(\nabla W(\mathbf{r}))^2 = 2m(E-V(\mathbf{r}))$

Since the forms are exactly the same, we can carry over our "wavefront" discussion from the optics world to the W world. I can relabel my picture above as follows



We are thus making the following replacement:

W-world	optics world
$S(\mathbf{r},t) = W(\mathbf{r}) - Et$	$P(\mathbf{r},t)$ = the phase = $L(\mathbf{r}) - ct$
$W(\mathbf{r})$	$L(\mathbf{r})$
E	c
$f = 2m(E-V(\mathbf{r})) = 2mT$	$f = n^2(\mathbf{r})$
$(\nabla W(\mathbf{r}))^2 = 2m(E-V(\mathbf{r})) = 2mT$	$(\nabla L(\mathbf{r}))^2 = n^2(\mathbf{r})$

Finally we understand Goldstein's opening picture on page 308. The surface $S(\mathbf{r},t) = 5$ (he calls it a) is going to move to the right as t increases, taking the position of the various $W = 5 + Et$. So somehow *associated with* our little mass particle in a potential is a function S(r,t) which has wavefronts like S(r,t)=5 which move in time and take on the shapes of the surfaces $W = a + Et$.

What on earth does this mean? On page 309 more facts are uncovered.

• the velocity of the wavefront is $u = c/|\nabla_{3D} L(\mathbf{r})|$ which becomes $u = E/|\nabla W| = E/\sqrt{2m(E-V(\mathbf{r}))}$ But we can replace $E-V = T = \frac{1}{2} mv^2$ for our particle, and then $u = E/(mv) = E/p$. This shows the strange result that the slower your particle goes (speed v), the faster the S=constant wavefronts move (speed u). From page 280 we know that $p_i = \partial W/\partial q_i$ so that $\mathbf{p} = \nabla W$. But we know that ∇W = normal to our wavefronts, so the particle momentum must follow a path like my 2D h(x) which is normal to all the wavefronts. Such a possible path is shown by the arrow in the page 308 figure.

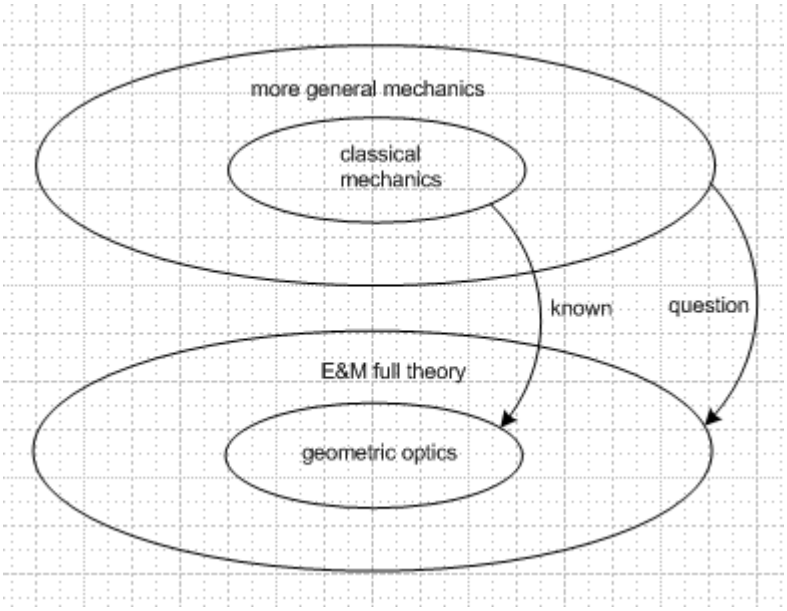
Say this again: what I call "the normal curves" to the wavefronts are each a potential spatial path for our particle's momentum **p**. So imagine an animation of the page 308 figure where the particle moves slowly along the trajectory, but as it does, the S = a specific surface moves very quickly to the right through its positions, passing the particle by.

Somehow, we suddenly have some kind of "wave theory" of the motion of a classical particle. At the same time, we have some kind of "particle trajectory theory" for light. And this has nothing to do with quantum mechanics.

What about least action? In mechanics on page 231 we showed that the "abbreviated action" = S_o = A was given by the integral of $p\dot{q}\,dt$, but for a particle this is $mv^2 = 2T$ so we have integral of 2Tdt. But we know that $T = \frac{1}{2} m v^2$ and v = ds/dt along path, so find $dt = \sqrt{m/2T}\,ds$. We can then write our action as an integral of $2Tdt = 2T\sqrt{m/2T}\,ds = \sqrt{2mT}\,ds$, as shown page 232 equation 7-40.

So, in classical mechanics our principle of least action means to extremize $\int \sqrt{2mT}\,ds$, probably we want to minimize this integral to solve our problem. From our table above, we see that in optics it is n that corresponds to $\sqrt{2mT}$, so geometrics optics solutions will minimize $\int n ds$. This is **Fermat's Principle** of geometric optics.

Big Question. This question could have occurred to people of the late 1800's who knew all the stuff described above, but who never heard of quantum mechanics. We have seen that "classical mechanics", as embodied in the Hamilton-Jacobi formalism, can be fully described by a function W which seem to define some kind of wavefronts which move in time, and we have seen that these waves satisfy a differential equation which corresponds to the "geometric limit" of E&M, it its optics incarnation. We have seen this particle-wave duality in classical mechanics, but the particles are the "senior partner" since nobody knows what to "do with" those waves in classical mechanics. The question is this: is there some more *general* theory of mechanics which corresponds to the full E&M theory?



At this point, we need to go back to page 3-9 and derive some simple facts:

- For our single particle, $\sqrt{2mT} = p = mv$.

• We already showed that the wavefront velocity is $u = E/|\nabla W| = E/\sqrt{2mT} = E/p$.

• $\mathbf{p} = \nabla W$ which we get either from the HJ formalism, or from above.

Question Added: For these mysterious mechanics-associated waves, what is the frequency and the wavelength of the wavefronts? The frequency is just something we assume, $\nu = \nu_0$ as noted earlier in the Snell comment, and then $\lambda = u/\nu = E/(p\nu)$ where $p = mv$ still.

NOW, *suppose* the correct "wave function" associated with a classical particle, the thing we have been talking about all this time, involved a constant $\hbar = h/(2\pi)$ and we had this form (just suppose),

$$\psi(\mathbf{r},t) = \exp[\, i\, S(\mathbf{r},t)/\hbar]$$

where $\psi(\mathbf{r},t)$ is the thing that has the wavefronts. Then we could write this as

$$\psi(\mathbf{r},t) = \exp[\, i\, 2\pi S(\mathbf{r},t)/h] = \exp\{\, i\, 2\pi\, [\, W(r) - Et]\, /h\} = \exp\{\, i\, 2\pi\, [\, W(r)/h - (E/h)t]\}$$

Let's now compare this to our last φ form above, namely

$$\varphi(\mathbf{r},t) = \exp\{\, i\, 2\pi(L(\mathbf{r})/\lambda_0 - \nu t)\, \}$$

These two forms align if we were to make this assumption

$$W(r)/h \leftrightarrow L(\mathbf{r})/\lambda_0$$
$$E/h \leftrightarrow \nu$$

Just above we said that the frequency of the classical mechanics waves "could be anything" and then λ depended on what we selected. Suppose we suddenly assume that the frequency of the classical mechanics waves is not arbitrary, but is in fact given by $\nu = E/h$. The thing above with $\leftrightarrow$ is just a correspondence between two analogous theories, but now we are making a very much stronger claim, that the frequency of the classical waves really is $\nu = E/h$. What would follow from this assumption? Well, what is our λ ?

$$\lambda = u/\nu = (E/p)/(E/h) = h/p$$

which looks a lot like the "deBroglie wavelength" of a physical particle.

Now go back to our supposed form for $\psi(\mathbf{r},t)$ given above which we rewrite here as

$$\psi(\mathbf{r},t) = \exp[\, i\, S(\mathbf{r},t)/\hbar] = \exp\{(\, i/\hbar)\, [\, W(r) - Et]\}$$

The wave equation for φ contains the index n, but we know

$$n = c/u = c/[E/|\nabla W|] = |\nabla W|\, c/\, E$$

Therefore (ignore this section, it went the wrong way)

$$(n/c)^2 = |\nabla W|^2 / E = (\nabla W)^2 / E^2$$

and then the wave equation for φ, when applied to our ψ, becomes

$$\nabla^2\psi - \{ (\nabla W)^2 / E^2 \}\ddot{\psi} = 0$$

Now we can go off and compute $\nabla^2\psi$ and we get

$$\nabla^2\psi = [(i/\hbar)^2 (\nabla W)^2 + (i/\hbar) \nabla^2 W] \psi$$

and we also know that

$$\ddot{\psi} = (-iE/\hbar)^2\psi$$

so our wave equation becomes

$$[(i/\hbar)^2 (\nabla W)^2 + (i/\hbar) \nabla^2 W] \psi - \{ (\nabla W)^2 / E^2 \} (-iE/\hbar)^2\psi = 0$$

or

$$[(i/\hbar)^2 (\nabla W)^2 + (i/\hbar) \nabla^2 W - (\nabla W)^2 * (i/\hbar)^2] \psi = 0$$

and the $(\nabla W)^2$ cancel and we are left with

$$\nabla^2 W \psi = 0$$

I am nonplussed by this result, it is not what I was expecting. One way to satisfy the equation would be to have $\nabla^2 W = 0$ which is a Laplace equation for W.

So let's back up a few steps and do things differently. We have

$$n = c/u \Rightarrow (n/c) = 1/u = (p/E)$$

and

$$\psi(\mathbf{r},t) = \exp\{ i 2\pi [W(\mathbf{r})/h - (E/h)t]\} = \exp\{ i [2\pi W(\mathbf{r})/h - \omega t]\}$$

so our time derivative term in the above is

$$- (n/c)^2 \ddot{\psi} = - 1/u^2 \partial^2\psi/\partial t^2 = - (p/E)^2 (- \omega^2\psi) = (p/E)^2 4\pi^2 \nu^2 \psi = (p\nu/E)^2 4\pi^2 \psi = (p\nu/h\nu)^2 4\pi^2 \psi$$

$$= (p/h)^2\, 4\pi^2\, \psi \;\; [= (1/\lambda^2)\, 4\pi^2\, \psi] \;\; = p^2\, 4\pi^2/h^2\, \psi \;\; = 2mT\, /\hbar^2\, \psi \;\; = 2m(E\text{-}V)/\hbar^2\psi$$

where you just verify each tiny algebraic step as shown. Our wave equation was

$$\nabla^2\psi - \; (n/c)^2\, \ddot{\psi} \;\; = 0$$

which now becomes

$$\nabla^2\psi + [2m(E\text{-}V)/\hbar^2]\psi \;\; = 0$$

which we can rewrite as

$$\hbar^2\nabla^2\psi/2m \;\; + (E\text{-}V)\psi = 0$$

and once more:

$$([\, -\, i\hbar]^2\nabla^2/2m \;\; + V)\psi = \;\; E\, \psi$$

In quantum mechanics, we know that $\hat{\mathbf{p}} \;\; =-\;\; i\hbar\nabla$ so we have arrived at

$$[\hat{\mathbf{p}}^2/2m + V]\, \psi = E\psi$$

which is of course **the Schrodinger equation**, time independent.

What assumptions and actions did we take in the previous section?

(1) We assumed that our "wave function" had the form $\psi(\mathbf{r},t) = \exp[\, i\, S(\mathbf{r},t)/\hbar]$ and satisfied the normal wave equation. This wave function form is completely within classical mechanics, but the constant $\hbar$ could be any number. This wave function is just what makes those surfaces of constant S, so nothing fancy here. And waves always satisfy the wave equation. So apart from $\hbar$, this is all classical.

(2) We assumed that our classical waves for a particle have a *specific frequency* that is $\nu = E/h$ and that h = a certain number. This does NOT come out of classical mechanics, it is a completely new idea. In classical mechanics, any frequency would be OK.

(3) Our motivation $\nu = E/h$ was based on a *comparison* between classical mechanics and the geometric optics limit of EM wave theory , the comparison being this:

$$\psi(\mathbf{r},t) = \exp\{\, i\, 2\pi\, [\, W(\mathbf{r})/h - (E/h)t]\, \}$$
$$\varphi(\mathbf{r},t) = \exp\{\, i\, 2\pi\, [\, L(\mathbf{r})/\lambda_0 - \quad \nu\, t]\, \}$$

Our comparison was encouraged by the fact that both $L(\mathbf{r})$ and $W(\mathbf{r})$ satisfy the same eikonal equation form in the limit of small wavelength which is the "geometric optics" limit for EM waves, and which is the non-quantum limit for mechanics (ie, classical mechanics).

(4) The wave function *is* the "exponentiated action" $\psi(\mathbf{r},t) = \exp[\, i\, S(\mathbf{r},t)/\hbar]$, an idea that will have lots of implications I think.

So, the proposal here is that maybe there is some general mechanics theory which has this wave equation and has some waves ψ and if you make constant h be "very small" then for normal physical objects for which we do classical mechanics, we will be in the geometric optics limit. Since $\lambda = h/mv$, we will be in the "geo optics limit" (ie, "classical mechanics"), if h is small and/or m is large.

I think if Hamilton were to read what I have just written, he would be stunned to learn that the constant h is actually one of the constants of the universe and really does exist. Hamilton knew about the analogy between geo optics and classical mechanics, but he did not make the above conjectural steps leading to the conjecture of some kind of "wave function" ψ analogous to the EM potential φ, and an analogous equation shown above : the Schrodinger equation. It says Hamilton in 1834 knew about eikonal optics if not the full Maxwell equations, and I can only presume the wave equation was well known by this time. G points out that Hamilton had no motivation to look for a "more general theory" of mechanics because in his time, it was thought to be complete.

We now think we know about that more general theory, and we call it "wave mechanics". In this theory there is a constant h which give the above equation. There are "waves" associated with particles which are surfaces of constant S (constant action), as HJ theory showed. The waves $\psi(\mathbf{r})$ are complex numbers, not just real, and are associated with the probability of a particle's location. A particle does have an associated wavelength which is $\lambda = h/p$ exactly as HJ theory says, and this is the "deBroglie wavelength". Small particles do "interfere" with each other, just as in the wave theory of light. There is a connection $E = h\nu$ between a particle's energy E and the frequency of the associated waves. This applies also to photons of light, but this fact was as yet unknown. Atomic physics could not be solved until this more general theory was discovered. Atomic physics served as the motivation to find the QM theory. That is because atomic systems have very small-mass particles and are thus not in the geometric optics limit.

G refers to "quantum mechanics" sometimes as "wave mechanics". Maybe the WM name is more associated with the Schrodinger equation approach to QM. Quantized spin for example is part of QM but has nothing to do with wave mechanics.

History Added:

Hamilton (1805-1865) This suggests that the Hamilton's equations of motion and the Hamilton-Jacobi equation are dated around 1834, here is some history from a Dover $9 book I might buy: (Struik, A Concise History of Mathematics).

Jacobi's lectures on dynamics lead us to another mathematician whose name is often linked with that of Jacobi, William Rowan Hamilton (not to be confused with his contemporary, William Hamilton, the Edinburgh philosopher). He lived his whole life in Dublin, where he was born of Irish parents. He entered Trinity College, became Royal Astronomer of Ireland in 1827 at the age of twenty-one, and held this position until his death in 1865. As a boy he learned continental mathematics—still a novelty in the United Kingdom—by studying Clairaut and Laplace, and showed his mastery of the novel methods in his extremely original work on optics and dynamics. His theory of optical rays (1824) was more than merely a differential geometry of line congruences; it was also a theory of optical instruments and allowed Hamilton to predict the conical refractions in biaxial crystals. In this paper appeared his "characteristic function," which became the guiding idea of the *General Method in Dynamics*, published in 1834–35. Hamilton's idea was to derive both optics and dynamics from one general principle. Euler, in his defense of Maupertuis, had already shown how the stationary value of the action integral could serve this purpose. Hamilton, in accordance with this suggestion, made optics and dynamics two aspects of the calculus of variations. He asked for the stationary value of a certain integral and considered it as a function of its limits. This was the "characteristic" or "principal" function, which satisfied two partial differential equations. One of these partial differential equations, which is usually written

$$\frac{\partial S}{\partial t} + H\left(\frac{\partial S}{\partial q}, q\right) = 0,$$

was specially selected by Jacobi for his lectures on dynamics and is now known as the Hamilton-Jacobi equation. This has obscured the importance of Hamilton's characteristic function, which had the central place in his theory as a means of unifying mechanics and mathematical physics. It was rediscovered by the Leipzig astronomer Heinrich Bruns in 1895 in the case of geometrical optics and as *eikonal* has shown its use in the theory of optical instruments.

The part of Hamilton's work on dynamics which has passed into the general body of mathematics is, in the first place, the "canonical" form, $\dot{q} = \partial H/\partial p$, $\dot{p} = -\partial H/\partial q$, in which he wrote the equations of dynamics. Canonical form and the Hamilton-Jacobi differential equation enabled Lie to establish the relation between dynamics and contact transformations. Another idea of Hamilton, since equally accepted, was the derivation of the laws of physics and mechanics from the variation of an integral. Modern relativity, as well as quantum mechanics, has based itself on "Hamiltonian functions" as its underlying principle.

The year 1843 was a turning point in Hamilton's life. In that year he found the quaternions, to the study of which he devoted the latter part of his life. We shall discuss this discovery later.

Hamilton, William Rowan. **1833** [1931]. "On a General Method of Expressing the Paths of Light, and of Planets, by the Coefficients of a Characteristic Function." In *The Mathematical Papers of Sir William Rowan Hamilton. Vol. 1.* Cambridge: Cambridge University Press.

****

Chapter 10. Small Oscillations (p 318)

****

This chapter is more complicated that it looks at first blush due to the tricky matrix aspects. There are lots of fine details all of which G mentions, but I do prefer to use my full matrix notation where possible to avoid the subscript details.

10.1 Formulation of the Problem (318). We want to examine small oscillations around an equilibrium point which is a point where $\partial V/\partial q_k = 0$ for all coordinates, Normally we think of this as a stable minimum situation, but it could be a min for some and a max for other coordinates. Fine. At the EQ point, set V = 0 by fiat. Then a Taylor expansion of V has the form 10-3 where the η_i are deviations of coordinates q_i from EQ. That is to say, we have $V \approx \frac{1}{2}\ \boldsymbol{\eta}^T V \boldsymbol{\eta} = \frac{1}{2}\ \boldsymbol{\eta} \bullet \mathbf{V} \bullet \boldsymbol{\eta}$ in dyadic which I won't use.

The kinetic energy is less obvious. Back on page 23 we showed that if the Cartesian-to-generalized equations have no time in them (scleronomous), then we know $T = \frac{1}{2}\ \dot{\mathbf{q}}^T M \dot{\mathbf{q}}$ where

$$M_{ij} = \tfrac{1}{2} \Sigma_k m_k \partial \mathbf{r}_k/\partial q_i \bullet \partial \mathbf{r}_k/\partial q_j = M_{ij}(\mathbf{q})$$

again, see page 23. One should really call this a kinetic energy tensor. Notice this fact:

$$M_{ii} = \tfrac{1}{2} \Sigma_k m_k \partial \mathbf{r}_k/\partial q_i \bullet \partial \mathbf{r}_k/\partial q_i = \tfrac{1}{2} \Sigma_k m_k \left| \partial \mathbf{r}_k/\partial q_i \right|^2 > 0 \qquad \text{(c1)}$$

You cannot have $M_{ii} = 0$ because that would require that $\mathbf{r}_k$ was not a function of q_i for all your particles, which means that q_i is not even a coordinate of the problem -- it is irrelevant.

In our small oscillations analysis, we approximate by saying $M_{ij}(\mathbf{q}) = M_{ij}(\mathbf{q}_0) \equiv T_{ij}$. Then we have $T \approx \frac{1}{2}\ \dot{\boldsymbol{\eta}}^T T \dot{\boldsymbol{\eta}}$ near EQ.

By their definitions, V and T are symmetric matrices, and they are real, hence Hermitian.

So, assume equalities now and write $V = \frac{1}{2}\ \boldsymbol{\eta}^T V \boldsymbol{\eta}$ and $T = \frac{1}{2}\ \dot{\boldsymbol{\eta}}^T T \dot{\boldsymbol{\eta}}$. We then have

$$L = T\text{-}V = \tfrac{1}{2}\ \dot{\boldsymbol{\eta}}^T T \dot{\boldsymbol{\eta}} - \tfrac{1}{2}\ \boldsymbol{\eta}^T V \boldsymbol{\eta}$$

and the Lagrange equations are then

$$d/dt(\nabla_{\dot{\eta}} L) - \nabla_{\eta} L = 0 \qquad \Rightarrow \qquad T\ddot{\boldsymbol{\eta}} + V\boldsymbol{\eta} = \mathbf{0}$$

What exactly is meant by the notation $\boldsymbol{\eta}$? It means $(\eta_1, \eta_2 ..... \eta_n)$ which are offsets on the generalized coordinates $(q_1, q_2 ..... q_n)$ from their EQ values $(q0_1, q0_2 ..... q0_n)$. So $\boldsymbol{\eta}$ is a vector of coordinates. Notice that all the coordinates are tangled together in the above ODE, so messy to solve.

10.2 The eigenvalue equation and the principle axis transformation(321) We then try solutions of the form $\boldsymbol{\eta}(t) = C\ \mathbf{a}\ e^{-i\omega t}$ which leads to

$$(V - \omega^2 T)\ \mathbf{a} = \mathbf{0} \qquad \text{or} \qquad (V - \omega^2 T)\ \boldsymbol{\eta} = \mathbf{0}$$

and we know this can only have non-zero solutions if $\det(V - \omega^2 T) = 0$ [so inverse does not exist], and this in turn determines the eigenvalues ω_k^2.

Now, for a *particular* eigenvalue ω_k^2 [k is just a label!] we can write the above as

$\boldsymbol{\eta}(t) = \mathbf{f_k}(\omega_k, t) = C_k\, \mathbf{a_k} \exp(-i\omega_k t)$ $\qquad$ $(V - \omega_k^2 T)\, \mathbf{a_k} = \mathbf{0}$ $\qquad$ or $\qquad$ $(V - \omega_k^2 T)\, \boldsymbol{\eta} = \mathbf{0}$

At this point, I want to head off a notational confusion. We put a solution label on our $\mathbf{f_k}$ here to indicate it is for a particular value of k. But we do NOT put a label on $\boldsymbol{\eta}$ because this is a vector of coordinates like x,y,z. If you have a particular solution to something, you say $x = \sin\omega_k t$ you don't say $x_k = \sin\omega_k t$. In other words, the notation $\boldsymbol{\eta}_k$ is non-sensical, so don't be caught using it!

Digression on T: Let's first think about matrix T all on its own. Since real symmetric, we know we can diagonalize it with a real unitary matrix which we will call B. Imagine we have done this so that

$B^T T B = D$ $\qquad$ where $BB^T = 1$ so $B^{-1} = B^T$. $\qquad$ D = a diagonal matrix $\qquad$ $TB = BD$

The columns of the matrix B we can call $\mathbf{b_k}$ and we know they are linearly independent or can be chosen so if there are degenerate eigenvalues.

Aside: We have $(\mathbf{b_k})_i = B_{ik}$. The fact that B is unitary tells us that

for any k: $\qquad 1 = (BB^T)_{kk} = B_{ki}B_{ki} = (\mathbf{b_k})_i\,(\mathbf{b_k})_i = \mathbf{b_k} \bullet \mathbf{b_k} = |\mathbf{b_k}|^2$

so it happens that our columns are normalized vectors as well as being orthogonal.
In other words, B is really a rotation.

These vectors $\mathbf{b_k}$ are the normalized eigenvalues of this eigenvalue problem:

$T\mathbf{b_k} = D_{kk}\mathbf{b_k}$ $\qquad$ $D_{kk} > 0$ from Second Digression below

Now define a new set of (non-normalized) eigenvectors by rescaling each separately as follows:

$\mathbf{f_k} \equiv \mathbf{b_k}/\sqrt{D_{kk}} \;\Rightarrow\; (\mathbf{f_k})_i = (\mathbf{b_k})_i/\sqrt{D_{kk}} \;\Rightarrow\; F_{ik} = B_{ik}/\sqrt{D_{kk}}$

and let F be the matrix with these eigenvectors as columns. Then consider

$(F^T T F)_{km} = F_{ik}\, T_{ij} F_{jm} = (1/\sqrt{D_{kk}})(1/\sqrt{D_{mm}})\, B_{ik}\, T_{ij} B_{jm} = (1/\sqrt{D_{kk}})(1/\sqrt{D_{mm}})(B^T T B)_{km}$

$= (1/\sqrt{D_{kk}})(1/\sqrt{D_{mm}})\, D$ $\qquad$ = a diagonal matrix $= \delta_{km}$

The diagonal elements are all 1, so we have shown that $F^T T F = 1$. Since the eigenvectors are not normalized, we cannot say that F is a rotation which we could say about B above. But we have found (that is, we know exists) a matrix F whose columns are linearly independent which brings T to this special diagonal form of unity!

We can write F = BS where S is real diagonal scaling matrix $S_{ij} = \delta_{ij}\sqrt{D_{ii}}$. Then we have

$F^T T F = (BS)^T T\, BS = S^T B^T T B\, S = S^T D\, S = 1$

Thus it is fair to say that F is matrix operation where we first scaled the axes, and then rotate.

Second Digression on T : We know that KE = ½ $\dot{\mathbf{q}}^T T \dot{\mathbf{q}}$ and that T is symmetric. We know we can diagonalize T with a unitary matrix B as above to get $B^T T B = D$. We can think of B of taking us to some new generalized coordinates and velocities $\dot{\mathbf{q}} = B\dot{\mathbf{q}}'$ and we find that KE = ½ $\dot{\mathbf{q}}'^T D \dot{\mathbf{q}}' = ½ \Sigma_k D_{kk} \dot{q}'^2_k$. But the kinetic energy associated with a generalized velocity $\dot{q}'_k$ must be positive, it cannot be 0 or negative. If it is zero, then q_k' is not really a generalized coordinate, it is some extra baggage the means nothing, so get rid of it. We conclude therefore that we must have $D_{kk} > 0$. This fact was used above when we rescaled the eigenvectors. Had we had $D_{kk} = 0$, we would have a problem there.

Therefore, we know that the diagonal elements of our original physical matrix T are positive, as noted in comment (c1) above, and we know that in coordinates where T is diagonal, the diagonal values are also positive.

Third Digression on T: In general, we write $T(\dot{\mathbf{q}}) = \dot{\mathbf{q}}^T T \dot{\mathbf{q}}$. Think of $\dot{\mathbf{q}}$ as a vector in a vector space of n dimensions. We know that for any possible selection of this vector in this space the kinetic energy T must be > 0. In M&M this thing was called a "quadratic form" $x^T T x$ and the matrix T in this case was called "positive definite". If you select the vector $\dot{\mathbf{q}}$ to be along the k-axis and to be a unit vector, then we have $T(\dot{\mathbf{q}}) = T_{kk}$, so we know all the diagonal elements of T are > 0. If we do a congruence on T, the result is still positive definite, which is easy to show if we let $T' = C^T T C$

$\mathbf{x}^T T' \mathbf{x} = \mathbf{x}^T C^T T C\, \mathbf{x} = (C\mathbf{x})^T T\, (C\mathbf{x}) = \mathbf{y}^T T \mathbf{y} > 0$ // for all $\mathbf{y}$

Now back to the main flow. We have this equation to solve

$(V - \omega^2 T)\, \mathbf{a} = \mathbf{0}$ or $V\mathbf{a} = \omega^2 T \mathbf{a}$

In order to have non-zero solutions **a**, we must have $\det(V - \omega^2 T) = 0$ and this will determine the eigenvalues we will call ω_k^2 which we shall show later are real and positive for a potential minimum situation.

Write the eigenvalue equation in matrix form using A and Λ. Assume we have found n distinct eigenvalues and corresponding solutions a_k . We then have

$V\mathbf{a}_k = T\mathbf{a}_k\, \omega_k^2$

But write the mth component of this equation:

$V_{mi}(\mathbf{a}_k)_i = T_{mi}(\mathbf{a}_k)_i\, \omega_k^2$

and then use $(\mathbf{a}_k)_i = A_{ik}$ to get

$$V_{mi} A_{ik} = T_{mi} A_{ik} \omega_k^2 = T_{mi} A_{ik'} \delta_{k'k} \omega_k^2 = T_{mi} A_{ik'} \Lambda_{k'k}$$

where we have defined a diagonal matrix Λ whose diagonal elements are the ω_k^2. Then we have shown that

$$VA = TA\Lambda \qquad \text{// an all-matrix equation.}$$

Notice that the diagonal matrix ends up *on the right* here, that is why we moved ω_k^2 over there early.

Show that $A^\dagger TA$ is real and diagonal. This is highly non-obvious! Here T is in general not diagonal, but it is symmetric and positive definite, and A is constructed from the solutions to our eigenvalue problem $(V - \omega^2 T)\ \mathbf{a} = \mathbf{0}$, *without regard to normalization*, $(\mathbf{a}_k)_i = A_{ik}$. That is, our claim is true no matter how the eigenvectors $\mathbf{a}_k$ are normalized.

Apply $A^\dagger$ to both sides of $VA = TA\Lambda$ to get

$$A^\dagger VA = A^\dagger TA\Lambda = (A^\dagger TA)\Lambda$$

Take the Hermitian adjoint of the above equation and use the fact that $V^\dagger = V$ same for T,

$$A^\dagger VA = \Lambda^* (A^\dagger TA)$$

Subtract these two matrix equations to get

$$0 = (A^\dagger TA)\Lambda - \Lambda^* (A^\dagger TA)$$

Lemma 1: Claim $A^\dagger TA$ is real.
Proof: Let $B = A^\dagger TA$ and let $A = X + iY$ where X and Y are real. Then we have

$$B = A^\dagger TA = (X + iY)^\dagger T (X + iY) = X^T TX - Y^T TY + i \{ - Y^T TX + X^T TY \}$$

Define $R = - Y^T TX + X^T TY$, the contents of the {..}. Then

$$R = - Y^T TX + X^T TY$$
$$R_{ij} = T_{mn} (-Y^T_{im} X_{nj} + X^T_{im} Y_{nj}) = T_{mn} (-Y_{mi} X_{nj} + X_{mi} Y_{nj}) = 0$$

The 0 follows because T_{mn} is contracted with an antisymmetric tensor. Thus we have

$$B = A^\dagger TA = X^T TX - Y^T TY = \text{real}$$

So far then we have show than $(A^\dagger TA)\Lambda = \Lambda^* (A^\dagger TA)$ where $(A^\dagger TA)$ is real.

Lemma 2: Claim $B\Lambda = \Lambda^* B$ with B real implies that Λ must be real.

Proof:

$$B_{im}\Lambda_{mj} = \Lambda^*_{im}B_{mj} \quad => \quad B_{ij}\,\Lambda_{jj} = \Lambda^*_{ii}B_{ij} \quad => \quad B_{ij}\,(\Lambda_{jj}-\Lambda^*_{ii}) = 0$$

Setting i = j we get $B_{ii}\,(\Lambda_{ii}-\Lambda^*_{ii}) = 0 \Rightarrow \Lambda_{ii}=\Lambda^*_{ii}$ => Λ is real if $B_{ii} > 0$. In our case, we know that $B_{ii} > 0$ for each i because $B=A^\dagger TA$ is a kinetic energy form, see previous comments.

Now we know that $B\Lambda = \Lambda B$ which says $[B, \Lambda] = 0$, where $B = A^\dagger TA$.

Lemma 3: if a matrix commutes with an ev's diff diagonal matrix, it must be diagonal as well.

Proof: $B_{im}\Lambda_{mj} = \Lambda_{im}B_{mj} \Rightarrow B_{ij}\Lambda_{jj} = \Lambda_{ii}B_{ij} \Rightarrow B_{ij}(\Lambda_{jj} - \Lambda_{ii}) = 0$ If the diagonal elements of Λ are all different, then for $i \neq j$ this tells us that $B_{ij} = 0$ so B must be diagonal. I ignore the case where there are identical diagonal elements in Λ because things are messy enough here. but G does talk about how this is handled. It is a GSO issue where we can choose our A to make things work, and so choosing A will result in a diagonal B.

So we now conclude that $B = A^\dagger TA$ is real and diagonal, where A is constructed from the eigenvectors of the problem $(V - \omega^2 T)\,\mathbf{a} = \mathbf{0}$ and this is true no matter how the $\mathbf{a_k}$ are normalized.

Show you can select normalization of the a_k such that $B = A^\dagger TA = 1$. Since B is already diagonal and positive definite, all we have to do is apply a matrix S which stretches the axes appropriately to make this work. Call this diagonal matrix S and $S_{ij} = \delta_{ij}/\sqrt{B_{ii}}$. Then use A' = AS and we get

$$B' = A'^\dagger TA' = (AS)^\dagger T\,(AS) = S^\dagger\, A^\dagger TA\, S = SBS = 1.$$

Again, we knew that B was diagonal with positive diagonal elements, so the last step works.

Show that matrix A' normalized in this way diagonalizes V. We had

$$VA' = TA'\Lambda \Rightarrow$$
$$A'^\dagger VA' = A'^\dagger TA'\,\Lambda = \Lambda \qquad \text{QED}$$

The main point here is that we have shown that we can find a matrix A' which diagonalizes both T and V in this way. T goes to 1, and V goes to Λ. The transformation A' that does this is called the **principle axis transformation** for our problem. We saw this idea before with the inertia tensor.

A simple way to find the eigenvectors **a**. We had this eigenvalue equation:

$$(V - \omega^2 T)\,\mathbf{a} = \mathbf{0}$$

We know that we can diagonalize T, it has positive diagonal elements and of course $\det(T) \neq 0$. This must then be true of the undiagonalized T. Therefore we know that T^{-1} exists. So apply it to the above equation to get

$(T^{-1}V - \omega^2)\,\mathbf{a} = \mathbf{0}$ $(Q - \omega^2)\,\mathbf{a} = \mathbf{0}$ $Q\,\mathbf{a} = \omega^2\,a$ $Q = T^{-1}V$

Now we have a "normal" eigenvalue problem with matrix Q and we just solve as usual for eigenvalues and vectors. Maple can do this just fine, see section 10.4 below.

10.3 Frequencies of free vibration and normal coordinates (329). Now go back to our original equation which was this:

$$(V - \omega_k^2 T)\,\boldsymbol{\eta} = \mathbf{0} \qquad \text{or} \qquad (V - \omega_k^2 T)_{mi}\,\eta_i = \mathbf{0}$$

Suppose we define some *new coordinates* this way:

$$\boldsymbol{\eta} = A\,\zeta \qquad \zeta \equiv A^{-1}\boldsymbol{\eta}$$

Then we have

$$(V - \omega_k^2 T)\,A\zeta = \mathbf{0}$$

Apply A^T on the left to get this vector equation:

$$(\Lambda - \omega_k^2)\,\zeta = \mathbf{0} \qquad (\Lambda - \omega_k^2)_{ij}\,\zeta_j = \mathbf{0} \qquad (\Lambda_{jj} - \omega_k^2)\,\zeta^{(k)}{}_j = 0$$

This is an trivial equation to solve, and we find that, for a specific ω_k, the eigenvectors are:

$$\zeta^{(k)}{}_i = \delta_{ki}\,\zeta_k \qquad \text{or} \qquad \zeta^{(k)} = (0,0,....\zeta_k,0,0..)$$

This is a "mode" in which only the kth coordinate is non-zero! This is called a "**normal mode**" and we have one for each eigenvalue ω_k^2. Because $\boldsymbol{\eta} = A\,\zeta$, it is likely that all the coordinates of the $\boldsymbol{\eta}$ vector are going to be non-zero. So in a normal mode, you might have lots of offset coordinates moving, even though you have only one normal coordinate moving.

But what is our one ζ_k coordinate *doing* in this normal mode? Well, we know these things so far,

$$\boldsymbol{\eta}(t) = \mathbf{f_k}(\omega_k, t) = C_k\,\mathbf{a_k}\exp(-i\omega_k t)$$
$$\boldsymbol{\eta} = A\,\zeta$$

Thus we can say

$$\zeta = (A^{-1})\,\boldsymbol{\eta} = (A^{-1})\,C_k\,\mathbf{a_k}\exp(-i\omega_k t) = C_k\exp(-i\omega_k t)\,[(A^{-1})\,\mathbf{a_k}]$$

If we look at a particular coordinate, we find

$$\zeta_i = C_k\exp(-i\omega_k t)\,[(A^{-1})_{ij}\,(\mathbf{a_k})_j] = C_k\exp(-i\omega_k t)\,[(A^{-1})_{ij}\,A_{jk}]$$

$$= C_k\exp(-i\omega_k t)\,\delta_{ik}$$

so doing it this way we again find out that only ζ_k is moving, and here is what it is doing

$\zeta_k = C_k \exp(-i\omega_k t)$
$\zeta_i = 0 \qquad i \neq k$

so our normal mode coordinate ζ_k is doing **simple harmonic motion**.

What are T and V and L in these new coordinates?

$$T = \tfrac{1}{2}\,\dot{\boldsymbol{\eta}}^T T \dot{\boldsymbol{\eta}} = \tfrac{1}{2}\,\dot{\boldsymbol{\zeta}}^T A^T T\, A\,\dot{\boldsymbol{\zeta}} = \tfrac{1}{2}\,\dot{\boldsymbol{\zeta}}^T \dot{\boldsymbol{\zeta}} = \tfrac{1}{2}\,\Sigma_i\ \dot{\zeta}_i^2$$

$$V = \tfrac{1}{2}\,\boldsymbol{\eta}^T V \boldsymbol{\eta} = \tfrac{1}{2}\,\zeta^T A^T V\, A\,\zeta = \tfrac{1}{2}\,\zeta^T \Lambda \zeta = \tfrac{1}{2}\,\Sigma_i\ \omega_i^2 \zeta_i^2$$

$$L = \tfrac{1}{2}\,\Sigma_i\ (\dot{\zeta}_i^2 - \omega_i^2 \zeta_i^2)$$

and the Lagrange equations become the first equation set below for i = 1..N.

$$\ddot{\zeta}_i + \omega_i^2 \zeta_i = 0$$
$$T_{ij}\ddot{\boldsymbol{\eta}}_j + V_{ij}\,\boldsymbol{\eta}_j = 0 \qquad \text{ie} \qquad T\ddot{\boldsymbol{\eta}} + V\,\boldsymbol{\eta} = \mathbf{0}$$

which is a simple harmonic oscillator for this coordinate at frequency ω_i, confirming what we just found out by different means. Notice how *simple* the ODE appears in normal coordinates. We have copied from above what things looked like in the original physical coordinates and we see again thatthe coordinates are tangled together in each equation.

So, a general solution to the problem then will be a superposition of SHM in each of these normal coordinates. A general solution then has

$\zeta_i = C_i \exp(-i\omega_i t)$ going on in each of the normal coordinates.

$\boldsymbol{\zeta}$ = a column vector of the above!

If you want a general form then for η, you have $\boldsymbol{\eta} = A\,\boldsymbol{\zeta}$ or

$$\eta_i = \Sigma_j A_{ij}\,\zeta_j = \Sigma_j\ A_{ij}\,C_j \exp(-i\omega_j t)$$

which can then have a mixture of all the normal mode frequencies! If only one normal mode k is activated, then you would have

$$\eta_i = A_{ik}\,\zeta_k = A_{ik}\,C_k \exp(-i\omega_k t)$$

and then all the η components are going at the same frequency. Notice that there are no harmonics of any of the normal modes, just the fundamental of each one. This is in the small oscillation limit.

How did we know the ω_k^2 were positive? Go back to $(V - \omega_k^2 T)\,\boldsymbol{\eta} = \mathbf{0}$ and write as

$$V\,\boldsymbol{\eta} = \omega_k^2\, T\,\boldsymbol{\eta}$$

Therefore, for a solution $\boldsymbol{\eta}(t) = \mathbf{f_k}(\omega_k, t) = C_k\,\mathbf{a_k}\exp(-i\omega_k t)$, we have

$$\boldsymbol{\eta}^T V\,\boldsymbol{\eta} = \omega_k^2\,\boldsymbol{\eta}^T T\,\boldsymbol{\eta}$$

But from page 319 the LHS is really δV away from EQ and, if we are at a minimum of V, then LHS is positive. We already know RHS is positive real because T is a positive definite matrix (digression #3). So as long as we are at a stable minimum of the potential defined as $\delta V = \boldsymbol{\eta}^T V\,\boldsymbol{\eta} > \mathbf{0}$, then we know that the ω_k^2 are real and positive, hence justifying the name. If we are NOT at a minimum, then ω_k = imaginary and we have *exponential instable behavior* in all normal coordinates.

Comparison with the inertia tensor situation. In that case, we considered $I = \mathbf{n}\bullet I\bullet\mathbf{n}$ where I was the moment of inertia about rotation axis $\mathbf{n}$. We diagonalized I by a rotation R so $R^T I R = I_{diag}$ because I was a Hermitian matrix, so I_{diag} has real positive diagonal elements . We could interpret this rotation R as taking us to a new rotated set of coordinates $\mathbf{r}'$ in which $\mathbf{n}$ appears as $\mathbf{n}'$ according to $\mathbf{n} = R\mathbf{n}'$. Then we find that $I = \mathbf{n}\bullet I\bullet\mathbf{n} = R\mathbf{n}'\bullet I\bullet R\mathbf{n}' = \mathbf{n}'\bullet I_{diag}\bullet\,\mathbf{n}' = \Sigma_i n'_i{}^2\,(I_{diag})_{ii} = I_{11}n'_x{}^2 + I_{22}n'_y{}^2 + I_{33}n'_z{}^2$. In these new coordinates, the ellipsoid of constant I is aligned with the axes. We referred to R as the principle axis transformation.

In the *current* context, we have $2T = \dot{\boldsymbol{\eta}}\bullet T\bullet\dot{\boldsymbol{\eta}}$. We find a combined rotation/scaling transformation we called $A' = AS$ above and define $\boldsymbol{\eta} = A'\,\zeta$ to get $2T = \dot{\zeta}\bullet 1\bullet\dot{\zeta} = \dot{\zeta}^T\dot{\zeta}$. We had to add the scaling so that A' would also diagonalize V as well as T. So this combination rotation/scaling matrix A' is still called a principle axis transformation. In the $\dot{\boldsymbol{\eta}}$ space, we have some kind of kinetic energy ellipsoid at some weird orientation. The rotation A (assume orthonormal a_k) rotates us to some axes in which this ellipsoid is aligned with the axes. Then the scaling reduces this aligned ellipsoid to a sphere:

$$2T = \dot{\boldsymbol{\eta}}\bullet T\bullet\dot{\boldsymbol{\eta}} = (A'\dot{\zeta})\bullet T\bullet(A'\dot{\zeta}) = \dot{\zeta}\bullet(\mathbf{A'}^T T A')\bullet\dot{\zeta} = \dot{\zeta}\bullet(\mathbf{S}^T(\mathbf{A}^T T A)S)\bullet\dot{\zeta} = \text{rotate, then scale} = \dot{\zeta}^T\dot{\zeta}$$

10.4 Free vibrations of a linear triatomic molecule (333). For this problem, we may quickly set up our basic matrices as follows

```
V := matrix([[k,-k,0],[-k,2*k,-k],[0,-k,k]]);
```

$$V := \begin{bmatrix} k & -k & 0 \\ -k & 2k & -k \\ 0 & -k & k \end{bmatrix}$$

```
T := matrix([[m,0,0],[0,M,0],[0,0,m]]);
```

$$T := \begin{bmatrix} m & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & m \end{bmatrix}$$

The mass tensor T is pretty obvious since center mass is of size M, others m. The V tensor follows quickly from the page 333 equations.

We want now to find the eigenvectors for our equation

$(V - \omega^2 T)\,\mathbf{a} = \mathbf{0} \quad \Leftrightarrow \quad (T^{-1}V - \omega^2)\,\mathbf{a} = \mathbf{0} \quad (Q - \omega^2 1)\,\mathbf{a} = \mathbf{0}$

We do the right equation in Maple as follows:

$Q = T^{-1}V \qquad (Q - \omega^2) = G \qquad G\mathbf{a} = 0$

```
TI := inverse(T);
```

$$TI := \begin{bmatrix} \frac{1}{m} & 0 & 0 \\ 0 & \frac{1}{M} & 0 \\ 0 & 0 & \frac{1}{m} \end{bmatrix}$$

```
Q := evalm(TI&*V);
```

$$Q := \begin{bmatrix} \frac{k}{m} & -\frac{k}{m} & 0 \\ -\frac{k}{M} & 2\frac{k}{M} & -\frac{k}{M} \\ 0 & -\frac{k}{m} & \frac{k}{m} \end{bmatrix}$$

```
G := evalm(Q - omega^2);
```

$$G := \begin{bmatrix} \frac{k}{m} - \omega^2 & -\frac{k}{m} & 0 \\ -\frac{k}{M} & 2\frac{k}{M} - \omega^2 & -\frac{k}{M} \\ 0 & -\frac{k}{m} & \frac{k}{m} - \omega^2 \end{bmatrix}$$

```
det(G);
```

$$\frac{(-k + \omega^2 m)\,\omega^2\,(2km + Mk - M\omega^2 m)}{m^2 M}$$

The determinant shows the expected eigenvalues including $\omega = 0$ as shown by G in 10-52 We proceed to get the eigenvectors as follows:

```
eigenvalues(Q);
```

$$\frac{k}{m}, 0, \frac{k(2m+M)}{Mm}$$

```
e :=eigenvectors(Q);
```

$$e := \left[\frac{k}{m}, 1, \{[-1, 0, 1]\}\right], [0, 1, \{[1, 1, 1]\}], \left[\frac{k(2m+M)}{Mm}, 1, \left\{\left[1, -2\frac{m}{M}, 1\right]\right\}\right]$$

The eigenvalues (here fore ω^2) agree with G 10.52. The eigenvectors are *not* normalized in such a way that $A^TTA=1$, but they are normalized here in a way that makes things very clear. Then center eigenvector is [1,1,1] which is the overall translation. The left eigenvector is the contrary motion of page 336 (b) of G. And the right eigenvector is page 336 (c).

We now want to construct our matrix A from these eigenvectors such that $A^TTA = 1$. This requires some Maple fiddling, but we manage to get our eigenvalues to be the columns of a matrix A, where we have added a scaling factor λ_i for each column (left side below)

```
R := evalm(AT&*T&*A);
```

$$A := \begin{bmatrix} \lambda 1 & \lambda 2 & -\lambda 3 \\ \lambda 1 & -2\frac{\lambda 2\, m}{M} & 0 \\ \lambda 1 & \lambda 2 & \lambda 3 \end{bmatrix} \qquad R := \begin{bmatrix} 2\lambda 1^2 m & 0 & 0 \\ 0 & 2\lambda 2^2 m + 4\frac{\lambda 2^2 m^2}{M} & 0 \\ 0 & 0 & 2\lambda 3^2 m + \lambda 3^2 M \end{bmatrix}$$

We hope to select scaling values such that $A^TTA = 1$. We first compute this product directly as on the right above. Using the solve function, we make Maple tell us the values for the λi and we get

```
lambda1 := L1;
```

$$\lambda 1 := \frac{\sqrt{2}\sqrt{m(2m+M)M}}{2mM+4m^2}$$

```
lambda2 := L2;
```

$$\lambda 2 := \frac{1}{2}\frac{\sqrt{2}}{\sqrt{m}}$$

```
lambda3 := L3;
```

$$\lambda 3 := \frac{1}{\sqrt{2m+M}}$$

```
R := simplify(evalm(R&*&*()));
```

$$R := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

And this does indeed make R = 1, as shown on the right above, and we can also display the updated A matrix

```
A := simplify(evalm(A&*&*()));
```

$$A := \begin{bmatrix} -\frac{1}{2}\frac{\sqrt{2}}{\sqrt{m}} & \frac{1}{2}\frac{\sqrt{2}\sqrt{m(2m+M)M}}{m(2m+M)} & \frac{1}{\sqrt{2m+M}} \\ 0 & -\frac{\sqrt{2}\sqrt{m(2m+M)M}}{(2m+M)M} & \frac{1}{\sqrt{2m+M}} \\ \frac{1}{2}\frac{\sqrt{2}}{\sqrt{m}} & \frac{1}{2}\frac{\sqrt{2}\sqrt{m(2m+M)M}}{m(2m+M)} & \frac{1}{\sqrt{2m+M}} \end{bmatrix}$$

Note that $A_{31} = 1/\sqrt{2m}$, while $A_{32} = 1/\sqrt{2m(1+2m/M)}$. These agree with the 9 matrix values that G gives on page 336 in equation (10-56a,b,c), though his columns are shuffled relative to mine.

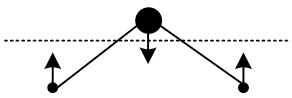
Finally, we can show that our matrix A does in fact diagonalize the V matrix as the formalism claims:

```
simplify(evalm(AT&*V&*A));
```

$$\begin{bmatrix} \frac{k}{m} & 0 & 0 \\ 0 & \frac{k(2m+M)}{Mm} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and the diagonal elements are the three ω_k^2 and this is our matrix Λ as discussed above.

G goes on to discuss the fact that this three point-mass system really has 9 coordinates in 3D space, so removing translations there are 6 modes. The thing can rotate non-trivially with point masses in two dimensions, so that leaves 4 modes, and we found 2 already, so there are 2 more. They are the same mode in two dimensions, so really only one new item to think about, it is this one:



So overall there are 2+2 = 4 vibrational modes. We have not solved this last problem for the eigenfrequency, but it does not seem too hard.

10.5 Forced vibrations and dissipative forces (338). Suppose we just start all over again and throw in some frictional forces to get

$$T\ddot{\boldsymbol{\eta}} + F\dot{\boldsymbol{\eta}} + V\boldsymbol{\eta} = \mathbf{0}$$

Put in $\boldsymbol{\eta}(t) = C\,\mathbf{a}\,e^{-i\omega t}$ to get

$$(V - i\omega F - \omega^2 T)\,\boldsymbol{\eta} = 0$$

Treat this as an eigenvalue equation, det=0 will determine the ω_k. We then have

$(V - i\omega_k F - \omega_k^2 T)\,\mathbf{a}_k = 0$ or $(V - \gamma_k F + \gamma_k^2 T)\,\mathbf{a}_k = 0$

Goldstein does not to on to actually solve this problem in a general sense.

Digression: Let's try to follow the former path and see where it leads.

Write the eigenvalue equation in matrix form using A and Λ. Assume we have found n distinct eigenvalues ω_k and corresponding solutions $\mathbf{a}_k$. We then have

$$V\mathbf{a}_k - iF\mathbf{a}_k\omega_k - T\mathbf{a}_k\,\omega_k^2 = \mathbf{0}$$

But write the mth component of this equation:

$$V_{mi}(\mathbf{a}_k)_i - iF_{mi}(\mathbf{a}_k)_i\,\omega_k - T_{mi}(\mathbf{a}_k)_i\,\omega_k^2 = 0$$

and then use $(\mathbf{a}_k)_i = A_{ik}$ to get

$$V_{mi}\,A_{ik} - iF_{mi}\,A_{ik}\,\omega_k - T_{mi}\,A_{ik}\,\omega_k^2 = 0$$

$$V_{mi}\,A_{ik} - iF_{mi}\,A_{ik'}\delta_{k'k}\,\omega_k - T_{mi}\,A_{ik'}\,\delta_{k'k}\,\omega_k^2 = 0$$

$$VA - iFA\Omega - TA\Omega^2 = 0 \qquad \Omega^2 = \Lambda$$

where we have defined a diagonal matrix Ω whose diagonal elements are the ω_k. So we have shown that

$VA = TA\,\Omega^2 + iFA\Omega$ // an all-matrix equation.

Notice that the diagonal matrix ends up *on the right* here, that is why we moved ω_k and ω_k^2 over there early.

Show that $A^\dagger TA$ is real and diagonal ???

Apply $A^\dagger$ to both sides of $VA = TA\,\Omega^2 + iFA\Omega$ to get

$$A^\dagger VA = A^\dagger TA\,\Omega^2 + i\,A^\dagger FA\Omega$$

Take the Hermitian adjoint of the above equation and use the fact that $V^\dagger = V$ same for T and F,

$$A^\dagger VA = \Omega^{2*}A^\dagger TA + i\,\Omega^* A^\dagger FA$$

Subtract these two matrix equations to get

$$0 = [(A^\dagger TA)\,\Omega^2 - \Omega^{2*}\,(A^\dagger TA)\,] + i\,[\,(A^\dagger FA)\,\Omega - \Omega^*(A^\dagger FA)]$$

As we did before, write A = X + iY where X and Y are real. Exactly as in Lemma1 above, we can show that

$B = A^\dagger TA$ = real [and it is also $B^\dagger = B$ so $B^T = B$ so B is real and symmetric]

so the above equation then says

$$0 = [B\,\Omega^2 - \Omega^{2*}\,B\,] + i\,[B\,\Omega - \Omega^*\,B] \qquad B = A^\dagger TA = \text{real}$$

Unlike before, because there are two complex terms above, we cannot say Ω = real. Instead, write

$\omega_k = p_k + i\,q_k$ $\qquad\qquad$ $\Omega = P + iQ$ $\qquad\qquad$ P and Q are diagonal matrices

$\omega_k^2 = p_k^2 - q_k^2 + 2ip_kq_k$ $\qquad\qquad$ $\Omega^2 = P^2 - Q^2 + 2iPQ$

Stuff this into the above to get

$$0 = [B\,\Omega^2 - \Omega^{2*}\,B\,] + i\,[B\,\Omega - \Omega^*\,B]$$
$$= [B\,(P^2 - Q^2 + 2iPQ) - (P^2 - Q^2 - 2iPQ)\,B\,] + i\,[B\,(P + iQ) - (P - iQ)\,B]$$

$$= \{\,B\,(P^2 - Q^2) - (P^2 - Q^2)B - BQ + QB\,\} + i\,\{\,2BPQ + 2PQB + BP - PB\,\} = 0$$

But now B, P and Q are all real, so we can write this as two separate equations:

$$B\,(P^2 - Q^2) - (P^2 - Q^2)B - BQ - QB = 0$$
$$2BPQ + 2PQB + BP - PB = 0$$

Suppose we find P and Q from our eigenvalue determinant. Rewrite the above as

$[\,B,(P^2 - Q^2] - [B,Q]_+ = 0$ $\qquad$ $\Omega = P + iQ$ $\qquad$ (1)

$2[B,PQ]_+ + [B,P] = 0$ $\qquad$ $B = A^\dagger TA$ = real $\qquad$ (2)

All matrices are symmetric here. For such matrices we know that

$[Q,R]^T = -[Q,R]$ $\qquad$ and $[Q,R]_+ = +[Q,R]_+$

So let's transpose both equations above to get

$-[\,B,(P^2 - Q^2] - [B,Q]_+ = 0$ $\qquad$ (3)

$2[B,PQ]_+ - [B,P] = 0$ $\qquad$ (4)

Then write down (1) - (3) and (2) - (4) to get

$$[\,B,(P^2 - Q^2] = 0$$

$$[B,P] = 0$$

In either case, we see that B is commuting with a diagonal matrix, so B must be diagonal by Lemma 3, assuming as before that all eigenvalues are different. For CC pairs and other degeneracy we will need to do GSO as before, so let's ignore this for now.

So, we have indeed shown that $B = A^\dagger TA$ is real and diagonal, just as in the former problem. We then want to do the same trick as before and write $A' = AS$ to get $B' = A'^\dagger TA' = 1$. Let's assume we have done this, and let's drop the prime on A from now on. We had as our starting point:

$$VA = TA\,\Omega^2 + iFA\Omega$$

Apply $A^\dagger$ from the left to get

$$A^\dagger VA = A^\dagger TA\,\Omega^2 + i\,A^\dagger FA\Omega = \Omega^2 + i\,A^\dagger FA\Omega$$

$$\Omega^2 + i\,A^\dagger FA\,\Omega - A^\dagger VA = 0$$

Well, all I can say is this

$$\Omega^2 - A^\dagger\{\,V - iFA\Omega A^{-1}\}A = 0$$

so we can argue that transformation A brings the matrix $V - iFA\Omega A^{-1}$ to diagonal form. But that is pretty meaningless I'm afraid. Go back to the previous,

$$\Omega^2 + i\,A^\dagger FA\,\Omega - A^\dagger VA = 0$$

This is $2N^2$ real nonlinear equations for the $2N^2$ real unknowns in A_{ij} .

We might just as well go back to our starting point

$$(V - i\omega_k F - \omega_k^2 T)\,\mathbf{a}_k = 0 \qquad \text{or} \qquad (V - i\omega_k F - \omega_k^2 T)\,A = 0$$

and regard it as $2N^2$ real linear equations for the $2N^2$ real parameters of A. Of course only possible solutions when the ω_k eigenvalues are used.

STOP. No doubt the theory for doing all this is presented somewhere, maybe using the quadratic forms stuff. I don't want to reinvent that whole bag here, and since Goldstein omitted it, it is likely a bit messy. I might guess that in general there are no "normal modes" in the general case.

If it happens that F is diagonalized by the same A that we had in our original problem with no friction, then you get "normal modes" as before with equations as in 10-66. This results in the shifted eigenvalues as shown in 10-68 so we get some damping. So this situation is easy.

Now G talks about a general system under the effect of a **"driving force"** and we then have 10-74. This gives 10-75 which is the equation I have been dealing with above, but the RHS is not 0 so you can use

Cramer's Rule to solve it! The whole thing runs at the driving frequency ω and we will get resonant behavior from D(ω) shown in the Cramer's solution 10-76. Factoring this thing gives the complex frequencies which result in resonant peaks which are finite due to the damping. This is more the kind of problem you want to solve when light is shined on atoms.

G closes with some comments about how this theory applies to EE work. Instead of talking about that, he is going to shift attention to continuous systems in the last chapter.

****

Chapter 11. Continuous Systems and Fields (p 347)∂

****

11.1. Transition from discrete to continuous system (347). Our model here is to imagine an **elastic rod** as made of a cross section of little masses on springs, like a muscle fiber sheaf, and we can think about just one of these fibers, though two are shown in the figure. We start with constants **m,a,k** and we think of those longitudinal modes we just studied in the last chapter. Since the rod is of finite length, we start with a finite number of masses, perhaps L/a, We are going to take the **limit a→0** and see what happens.

The finite Lagrangian is as shown 11-3,4 in terms of those differential-from-rest longitudinal position variables η_i. You can see that m/a → μ, linear mass density of one fiber. And the paren is going to be the derivative dη/dx squared, and so ka → Y, the **Young modulus** so called. As a shrinks, k must grow. The positions were η_i which we could express as $\eta(x_i)$ where $x_i = 0 + i*a$. This becomes η(x) and we have a longitudinal displacement then at any value of x.

Now, 11-5 shows the EL equations, and you get two Δη terms from separate terms in the series, those adjacent to point x_i. The a^2 denominator is correct, and the two terms on the right will become -Y times η_{xx} as seen in 11-7 and we end up with a **1D wave equation**, no surprise, **velocity** of Y and μ. The L has the first derivative η_x squared as shown, coming directly from 11-4. Notice $\Sigma_i a$ becomes $\int dx$. So this integral in 11-6 can be thought of as just summing over all the particles in the object. For each value of x, η(x) is a different "generalized coordinate". The thing inside the integral is a Lagrangian density $\mathcal{L}$. So this model would correctly predict longitudinal waves in the rod, but G does not mention that.

11.2 The Lagrangian formulation for continuous systems (350). In our example we saw $\mathcal{L}(\eta_t,\eta_x)$. Had we used $V = kx^3$ instead of kx^2, we would have gotten η_x^3 in 11-6. This form is coming somehow from the nearest-neighbor power law interaction. If V = ksin(x), you could still figure it out from trig expansion formulas and I think in all cases result will be a function of η_x. So all these comments are justification for the general form of a $\mathcal{L}$ density as presented in 11-10. Notice in 11-8 that L is a spatial integral only of $\mathcal{L}$. If you were to include dt as well, then you are talking "action", and that is coming next.

Now we do a variational business on the action in 11-11, but G does not use the word "action" for object I. The summation characteristics are all constant in terms of variation (here integrals) so $\delta\mathcal{L}$ in 11-12 comes from variations in the coordinates η etc, not variations in the summation indices. He might have used the notation $\partial\mathcal{L}/\partial\eta_x$ in 11-12 but he would have to write out all three terms including $\partial\mathcal{L}/\partial\eta_y$. The trick is going to be **parts integration** on the second two terms to make them linear in δη which is an arbitrary variation of the coordinates η(x) at all points in the spacetime volume (but no variation at the

endpoints so parts will go away). He shows the $\dot{\eta}$ parts in equation α. The parts object would be the product $\partial\mathcal{L}/\partial\dot{\eta}_i\ \delta\eta$ and $\delta\eta = 0$ at the surface boundary so parts yield 0. Then he does the η_x parts in 11-14 and 11-15 swinging around the ∂_x operator. Of course the usual minus signs, so the last two terms in 11.16 are negative. As desired, δη is isolated and arbitrary, so the effective EL equations or Hamilton's Equations are as in 11-17. I might write this as

$$\partial_t\,(\partial_{\dot{\eta}}\mathcal{L}) - [\ \partial_\eta\mathcal{L}\ - (\nabla\bullet\nabla_\eta)\ \mathcal{L}\ ] = 0$$

We can't call these EL equations because L does not appear in them, and we have no derivation path. So right now these are just "**Hamilton's Principle equations on motion**". Since everything is a function of **r**, you can think of there being a separate equation here for each η(**r**) coordinate.

If we were to think of elastic rod in 3D, then we would have had η^x and η^y and η^z coordinates at each location **r**, and this is then the j index in 11-19. So yes, his notation is well justified. I use superscripts.

G then introduces the "**functional derivative**" which picks up the change in $\mathcal{L}$ from change in $\eta^i(x)$ AND from $\nabla\ \eta^i(x)$. That is to say, if we somehow change $\eta^i(x)$, then it is likely we will also change its spatial derivatives $\partial_k\eta^i(x)$, and we know from 11-10 that $\mathcal{L}$ is a function explicitly of both these things. so this fancy derivative $\delta L/\delta\eta^i$ picks up all these terms. So 11-19, are both very clear to me. Notice that by definition, L appears inside the $\delta L/\delta\eta$, but the RHS contains $\mathcal{L}$.

As G points out, this $\delta L/\delta\eta$ thing picks up the ugly gradient terms so you don't have to keep writing them, and then everything becomes simple again. We get 11-22 for example, and equations of motion once again look just like EL equations in 11-23, where we are just inserting our definitions 11-19,20. We now see why we put L in these things, because the results come out very simple. We have recovered our EL equation form!

In passing, he notes that we can write the 11-17 Hamilton's equations in **covariant form** as in 11-24 which certainly is impressive, and if $\mathcal{L}$ is a scalar, you are cooking with relativistic gas.

G wraps up the section computing these new objects for our simple longitudinal rod, and now he points out this is the wave equation and quotes the velocity in 11-25. So far so good!

11.3 Sound vibrations in a gas (355). This will be our second real-world example of using the Lagrangian density formalism to solve a problem. Unfortunately, this involves dynamics of gases, and my knowledge of "fluid dynamics" is very poor, though I do have a Schaum book on the subject, an early chapter of which deals with the basic equations of fluid dynamics. Luckily, Goldstein helps us out and derives most things.

One must understand up front the meaning of **η**(**r**,t), even this is not trivial to me. You must picture somehow the gas at complete rest and at position **r** you imagine a tiny volume, perhaps a cube or a sphere, that is perhaps differentially small. In a high speed dynamic situation such as a high-frequency sound wave passing through the gas, at some time t the quantity **η**(**r**,t) tells you how much and in what direction that tiny volume is "displaced" from its rest position. The assumption here is that this displacement is temporary and there is some restoring force from the nearby gas that wants to make the little volume move back to its rest position. G does not bring this subject up, but it must be a stable equilibrium situation. In a longitudinal wave travelling in z, I guess the displacements are in the z direction, and the little volumes are pushed around by pressure variations.

So fine. We think of each little microvolume as a generalized coordinate -- a sort of particle which a position and a velocity which would be $\dot{\boldsymbol{\eta}}$(**r**,t). Now G likes to use μ as gm/cm^3 density instead of ρ. We

know that the mass in our little volume is going to be $\mu\Delta V$ and the energy of this mass $\frac{1}{2}(\mu\Delta V)\,\dot{\eta}^2$. If we leave out the ΔV, we are talking kinetic energy "per unit volume" and he uses $\mathcal{T}$ (script T) as the "kinetic energy density". So we like equation α on page 355. The μ_0 means the at-rest density and we are assuming that we are not hugely varying this with our disturbance.

Obtaining the corresponding potential density $\mathcal{V}$ takes G a full 3 pages! I will outline the steps here

(1) First, I think V_o is meant to be some small but finite volume of gas which near constant $\mathcal{V}$. Then the claim is that $V_o\,\mathcal{V}$ is the potential energy in this volume which arises from work done PdV integrated as the volume say increases to $V_0 + \Delta V$, where now ΔV is perhaps distributed around the boundary of V_o.

(2) We need a formula for P(V), and we assume a linear fit as shown 11-28:

$$P(V) = P_o + (dP/dV)(V-V_0) \qquad \Rightarrow \qquad \text{integral } PdV = P_o\,\Delta V + \tfrac{1}{2}\,(dP/dV)(\Delta V)^2$$

(3) Now we need a formula for dP/dV so we need some kind of gas law equation. Because we are doing relatively high frequency, heat does not flow, we are adiabatic, and we have PV^γ = constant, which I vaguely remember from somewhere. So the derivative is as shown 11-31 and we can plug that in.

(4) Now he introduces a "dimensionless density change variable" $\sigma \equiv \Delta\mu/\mu_0$ which is just fine with me. We know that $\mu = V/M$ and $\mu_0 = V_0/M$, and here we have another implicit assumption. It is that our volume V contains a certain fixed number of gas molecules with total mass M, so the volume sort of moves with the particles in some sense. That is, the mass M of our volume is fixed, so that $\mu = V/M \Rightarrow \Delta\mu = \Delta V/M$. Then

$$\Delta V = M\,\Delta\mu = M\,(\mu_0\,\sigma) = (M\,\mu_0)\,\sigma = V_o\sigma\,.$$

So we can replace ΔV in both places in 11-28 with $V_o\sigma$ to get 11-34.

(5) The next problem is to relate σ to $\boldsymbol{\eta}$. The quantity $\mu_0\boldsymbol{\eta}$ must be the amount of mass that is "living on the wrong side of the border" at some time t and at locality $\mathbf{r}$, per unit of border area, and dimensions are correctly gm/cm^2. I don't think it would be right to call this a mass "flow" because there is no nothing of speed here, just how much is over the border at our instant of interest. If we now consider a different FIXED and FINITE volume and use the same symbol V for this volume, then I agree that the total mass displaced out of this volume is going to be $\int \mu_0\boldsymbol{\eta} \bullet \mathbf{dA}$ = mass displaced out of volume if d$\mathbf{A}$ points out from the volume. This means the total mass in the volume $\int \mu\, dV$ is low by this amount, so

$$\Delta\left(\int \mu\, dV\right) = -\int \mu_0\boldsymbol{\eta} \bullet \mathbf{dA} = -\mu_0 \int \boldsymbol{\eta} \bullet \mathbf{dA} = -\mu_0 \int \text{div}\,\boldsymbol{\eta}\, dV \qquad \text{by def of divergence}$$

But the LHS is the same as $\left(\int \Delta\mu\, dV\right)$ since $\Delta\mu$ is the local mass density change from EQ. Thus we conclude that

$\Delta\mu = -\mu_0 \text{ div } \boldsymbol{\eta}$ => $\Delta\mu/\mu_0 = \sigma = -\text{div } \boldsymbol{\eta}$

and so we have related σ to **η** as requested. Somehow this says that the dimensionless outflow of stuff from a tiny volume reduces the dimensionless density -- continuity. But again, it is not a flow, it is the amount caught outside at time t must equal the amount not present on the inside. So this is 11.36.

(6) We then replace σ by $-\nabla\bullet\boldsymbol{\eta}$ both places it appears in 11-34 to get 11-37, and combining this with our KE result, we get $\mathcal{L}$ as shown in 11-38, and sure enough, we see spatial derivatives of η appearing.

(7) We now want to use 11-18 to get our equations of motion from $\mathcal{L}$. The necessary derivatives are computed in 11-39 and it is at this point that we see that the linear $\nabla\bullet\boldsymbol{\eta}$ won't contribute to the equations of motion. The reason is that the equations say we need to do $\partial_{\mathbf{x}}$ on $\delta_{\mathbf{ik}}$ which is a constant. So the simplified $\mathcal{L}$ is in 11-40 and we arrive at 11-41 which is a strange looking equation indeed in **η**. But apply div to both sides and it becomes 11-42 which is the wave equation in $\sigma \equiv \Delta\mu/\mu_0$. We thus have a wave in the mass density, a sine wave in same. The pressure is not so simple due to adiabatic gas law 11-30, and out pops our sound velocity as shown, a function of γ, P_o and μ_o. Higher pressure means faster sound, also lower density means faster sound (but in EQ, these two quantities are related by $PV = \mu RT/M_m$ for ideal gas, say, where M_m is the molar mass).

So in conclusion, this serves as another Lagrangian Density problem example.

11.4 The Hamiltonian Formulation (355) No surprise that this comes next. We go back to the discrete spring and mass chain and compute H in the discrete case in 11-46 and then go to the limit in 11-47. We define the canonical momentum *density* π, and get 11-48 which relates the densities $\mathcal{H}$, $\mathcal{L}$ and π exactly we you would expect, this for a single coordinate function η. In the 3D case we have a dot product of **π** with $\dot{\boldsymbol{\eta}}$ sitting in there. It is all as expected.

We then do the variation dH and shuffle all the parts around as usual and we end up with 11-52. Then we do dH a different way using the functional derivatives we know understand, and we end up with 11-53. But if we use the Lagrange equations, two terms cancel here giving 11-54. If we then compare this to 11-52, we end up with Hamilton's Equations of motion shown 11-55 page 362. As expected, they contain the functional derivatives just the way we had things with L. We can write these out as in 11-56 which we will use for calculations.

So much for the H formalism! We now go back to our sound in gas example and do it the H way. We find that $\mathcal{H} = \mathcal{T}+\mathcal{V}$ as expected. One Hamilton's equation repeats our π definition, the other recovers our equation of motion for the gas, same as we got by the L method.

Next, we show that if the $\mathcal{H}$ has not explicit time dependence, H is conserved, again as expected. He is just turning his big crank.

Next, we consider an unknown object G which is the integral of a density which I guess we have to assume is a function of η and π in the same convective sense that H is. We find easily that

$dG/dt = [G,H] + \partial G/\partial t$

where these are Poisson brackets but we have a huge sum over the indices which here are continuous space. So this shows that if G "commutes" with H, it is a constant of the motion.

We are reminded that this is a sort of global symmetry (conservation rule) and there are local ones like continuity conditions as well, the gauge condition $\partial_\mu A^\mu = 0$ comes to mind.

11.5 Descriptions of Fields (364). In our sound example in the elastic rod, $\eta(x)$ was a scalar field, but it worked in a medium of matter. In E&M, we have fields like A_i, φ which don't need any "matter" to "**undulate in".** This must have been a big shock in the physics world when it was finally realized (but see Einstein's comments on the ether as seen in general relativity, separate doc I saved).

Next topic: **any $\mathcal{L}$ is acceptable if it gives the right equations of motion**, even if does not have the traditional form of T - V. For example, 11-61 provides an alternative to the $\mathcal{L}$ of 11-40, but it gives the same equations of motion for the sound in gas problem. That is, we get the same wave equation.

Next: G presents an $\mathcal{L}$ in 11-65 which, when we apply the Lagrange equations, **generates Maxell's equations**. This is shown in complete detail, I did it all, not too hard. You have to get used to $\partial(\partial\eta/\partial x)$ kind of things. Just assuming the **A** and φ fields (potentials) satisfies the first two Maxwell's, so we only have to do the meaty Maxwell's which show interaction with charge and current.

Next: What happens when we install $\mathbf{j} = \rho\mathbf{v}$ into the picture, which relates to physical charged particles? If we start with our vetted E&M $\mathcal{L}$ shown in 11-65, and install this j/ρ relation caused by actual charged particles, we get 11-72. But this $\mathcal{L}$ includes only the "charge effects" of these particles, not "inertial effects". That is to say, 11-72 has particles of charge q_i, and their motion (such as velocity) affects the equations of motion, as in Maxwell's equations. But if you want to include inertial effects of these particles, you have to "throw in" the usual KE term and this gives 11-73 = 11-73'. We can then interpret "the middle term" as the interaction between the particle world and the light world, and this of course is the famous $q_i v_\mu A^\mu = j_\mu A^\mu$ interaction term which we use in QED. He says nothing about "perturbation theory" here in regard to this interaction term. But we see that $E^2 - B^2$ is the $\mathcal{L}$ for E&M and this must be $F_{\mu\nu}F^{\mu\nu}$ which is a world scalar, and of course so is $j_\mu A^\mu$.

So G launches us in this way off into QED and mentions the construction of field Lagrangian densities as the job of **the particle theorist of the 1950 era**. They were getting into **π mesons** for example, and were trying the Yukawa potential perhaps, or scalar field theory, I cannot remember right now.

The closing comment is that you need (he feels) to have a mechanical description of something before you can "quantize it". The L and H methods of this book are the paths you follow.