# Linear Algebra for Quantum Mechanics

*Michael Fowler 9/12/06*

## Introduction

We've seen that in quantum mechanics, the state of an electron in some potential is given by a wave function $\psi(\vec{x},t)$, and physical variables are represented by operators on this wave function, such as the momentum in the *x*-direction $p_x = -i\hbar\partial/\partial x$. The Schrödinger wave equation is a *linear* equation, which means that if $\psi_1$ and $\psi_2$ are solutions, then so is $c_1\psi_1 + c_2\psi_2$, where $c_1$, $c_2$ are arbitrary complex numbers.

This linearity of the sets of possible solutions is true generally in quantum mechanics, as is the representation of physical variables by operators on the wave functions. The mathematical structure this describes, the linear set of possible states and sets of operators on those states, is in fact a *linear algebra* of operators acting on a *vector space*. From now on, this is the language we'll be using most of the time. To clarify, we'll give some definitions.

## What is a Vector Space?

The prototypical vector space is of course the set of real vectors in ordinary three-dimensional space, these vectors can be represented by trios of real numbers $(v_1, v_2, v_3)$ measuring the components in the *x*, *y* and *z* directions respectively.

*The basic properties of these vectors are*:

any vector multiplied by a number is another vector in the space, $a(v_1, v_2, v_3) = (av_1, av_2, av_3)$;

the sum of two vectors is another vector in the space, that given by just adding the corresponding components together: $(v_1 + w_1, v_2 + w_2, v_3 + w_3)$.

These two properties together are referred to as "*closure*": adding vectors and multiplying them by numbers cannot get you out of the space.

A further property is that there is a unique null vector $(0,0,0)$ and each vector has an additive inverse $(-v_1, -v_2, -v_3)$ which added to the original vector gives the null vector.

Mathematicians have generalized the definition of a vector space: a general vector space has the properties we've listed above for three-dimensional real vectors, but the operations of addition and multiplication by a number are generalized to more abstract operations between more general entities. The operators are, however, restricted to being commutative and associative.

Notice that the list of necessary properties for a general vector space does not include that the vectors have a magnitude—that would be an additional requirement, giving what is called a *normed* vector space. More about that later.

To go from the familiar three-dimensional vector space to the vector spaces relevant to quantum mechanics, first the real numbers (components of the vector and possible multiplying factors) are to be generalized to complex numbers, and second the three-component vector goes an $n$ component vector. The consequent $n$-dimensional complex space is sufficient to describe the quantum mechanics of angular momentum, an important subject. But to describe the wave function of a particle in a box requires an *infinite* dimensional space, one dimension for each Fourier component, and to describe the wave function for a particle on an infinite line requires the set of all normalizable continuous differentiable functions on that line. Fortunately, all these generalizations are to finite or infinite sets of complex numbers, so the mathematicians' vector space requirements of commutativity and associativity are always trivially satisfied.

We use Dirac's notation for vectors, $|1\rangle, |2\rangle$ and call them "kets", so, in his language, if $|1\rangle, |2\rangle$ belong to the space, so does $c_1|1\rangle + c_2|2\rangle$ for arbitrary complex constants $c_1$, $c_2$. Since our vectors are made up of complex numbers, multiplying any vector by zero gives the null vector, and the additive inverse is given by reversing the signs of all the numbers in the vector.

Clearly, the set of solutions of Schrödinger's equation for an electron in a potential satisfies the requirements for a vector space: $\psi(\vec{x},t)$ is just a complex number at each point in space, so only complex numbers are involved in forming $c_1\psi_1 + c_2\psi_2$, and commutativity, associativity, etc., follow at once.

## Vector Space Dimensionality

The vectors $|1\rangle, |2\rangle, |3\rangle$ are *linearly independent* if $c_1|1\rangle + c_2|2\rangle + c_3|3\rangle = 0$ implies $c_1 = c_2 = c_3 = 0$.

*A vector space is n-dimensional if the maximum number of linearly independent vectors in the space is n.*

Such a space is often called $\boldsymbol{V}^n(C)$, or $\boldsymbol{V}^{\boldsymbol{n}}(R)$ if only real numbers are used.

Now, vector spaces with finite dimension $n$ are clearly insufficient for describing functions of a continuous variable $x$. But they are well worth reviewing here: they are fine for describing quantized angular momentum, and they serve as a natural introduction to the infinite-dimensional spaces needed to describe spatial wavefunctions.

A set of $n$ linearly independent vectors in $n$-dimensional space is a *basis*—any vector can be written *in a unique way* as a sum over a basis:

$$|V\rangle = \sum v_i |i\rangle$$

You can check the uniqueness by taking the difference between two supposedly distinct sums: it will be a linear relation between independent vectors, a contradiction.

Since all vectors in the space can be written as linear sums over the elements of the basis, the sum of multiples of any two vectors has the form:

$$a|V\rangle + b|W\rangle = \sum (av_i + bw_i)|i\rangle.$$

## Inner Product Spaces

The vector spaces of relevance in quantum mechanics also have an operation associating a number with a pair of vectors, a generalization of the dot product of two ordinary three-dimensional vectors,

$$\vec{a}.\vec{b} = \sum a_i b_i.$$

Following Dirac, we write the inner product of two ket vectors $|V\rangle$, $|W\rangle$ as $\langle W|V\rangle$. Dirac refers to this $\langle\ |\ \rangle$ form as a "bracket" made up of a "bra" and a "ket".

An *important difference* from the simple three-dimensional real vector inner product is that we're usually dealing with *complex* vector spaces, and *we require*

$$\langle W|V\rangle = \langle V|W\rangle^*$$

where * denotes complex conjugate. This means that $\langle V|V\rangle$ is real. We require $\langle V|V\rangle$ to be *positive*, except for the null vector, for which it is zero. (These requirements do restrict the classes of vector spaces we are considering, but they are all satisfied by the spaces relevant to quantum mechanics.)

The *norm* of $|V\rangle$ is defined by

$$|V| = \sqrt{\langle V|V\rangle}.$$

If $|V\rangle$ is a member of $V^n(C)$, so is $a|V\rangle$, for any complex number $a$. We require the inner product operation to commute with multiplication by a number, so $\langle W|(a|V\rangle) = a\langle W|V\rangle$.

The complex conjugate of the right hand side is $a^*\langle V|W\rangle$. For consistency, the bra corresponding to the ket $a|V\rangle$ must be $\langle V|a^*$.

It follows that if

$$|V\rangle = \sum v_i |i\rangle, \quad |W\rangle = \sum w_i |i\rangle, \quad \text{then } \langle V|W\rangle = \sum v_i^* w_j \langle i|j\rangle.$$

## Constructing an Orthonormal Basis: the Gram-Schmidt Process

To have something better resembling the standard dot product of ordinary three vectors, we need $\langle i|j\rangle = \delta_{ij}$, that is, we need to construct an *orthonormal basis* in the space. There is a straightforward procedure for doing this called the Gram-Schmidt process. We begin with a linearly independent set of basis vectors, $|1\rangle, |2\rangle, |3\rangle, \ldots$ .

We first normalize $|1\rangle$ by dividing it by its norm. Call the normalized vector $|I\rangle$. Now $|2\rangle$ cannot be parallel to $|I\rangle$, because the original basis was of linearly independent vectors, but $|2\rangle$ in general has a nonzero component parallel to $|I\rangle$, equal to $|I\rangle\langle I|2\rangle$, since $|I\rangle$ is normalized. Therefore, the vector $|2\rangle - |I\rangle\langle I|2\rangle$ is perpendicular to $|I\rangle$, as is easily verified. It is also easy to compute the norm of this vector, and divide by it to get $|II\rangle$, the second member of the orthonormal basis. Next, we take $|3\rangle$ and subtract off its components in the directions $|I\rangle$ and $|II\rangle$, and normalize the remainder, and so on.

In an *n*-dimensional space, having constructed an orthonormal basis with members $|i\rangle$, any vector $|V\rangle$ can be written as a column vector,

$$|V\rangle = \sum v_i |i\rangle = \begin{pmatrix} v_1 \\ v_2 \\ . \\ . \\ v_n \end{pmatrix}, \text{ where } |1\rangle = \begin{pmatrix} 1 \\ 0 \\ . \\ . \\ 0 \end{pmatrix} \text{ and so on.}$$

The corresponding bra is $\langle V| = \sum v_i^* \langle i|$, which we write as a row vector with the elements complex conjugated, $\langle V| = (v_1^*, v_2^*, \ldots v_n^*)$. This operation, going from columns to rows and taking the complex conjugate, is called taking the *adjoint*, and can also be applied to matrices, as we shall see shortly.

The reason for representing the bra as a row is that the inner product of two vectors is then given by standard matrix multiplication:

$$\langle V|W\rangle = (v_1^*, v_2^*, \ldots, v_n^*)\begin{pmatrix} w_1 \\ \cdot \\ \cdot \\ w_n \end{pmatrix}.$$

(Of course, this only works with an orthonormal base.)

## The Schwartz Inequality

The Schwartz inequality is the generalization to any inner product space of the result $|\vec{a}\cdot\vec{b}|^2 \le |\vec{a}|^2|\vec{b}|^2$ (or $\cos^2\theta \le 1$) for ordinary three-dimensional vectors. The equality sign in that result only holds when the vectors are parallel. To generalize to higher dimensions, one might just note that two vectors are in a two-dimensional subspace, but an illuminating way of understanding the inequality is to write the vector $\vec{a}$ as a sum of two components, one parallel to $\vec{b}$ and one perpendicular to $\vec{b}$. The component parallel to $\vec{b}$ is just $\vec{b}\left(\vec{a}\cdot\vec{b}\right)/\left|\vec{b}\right|^2$, so the component perpendicular to $\vec{b}$ is the vector $\vec{a}_\perp = \vec{a} - \vec{b}\left(\vec{a}\cdot\vec{b}\right)/\left|\vec{b}\right|^2$. Substituting this expression into $\vec{a}_\perp \cdot \vec{a}_\perp \ge 0$, the inequality follows.

This same point can be made in a general inner product space: if $|V\rangle$, $|W\rangle$ are two vectors, then

$$|Z\rangle = |V\rangle - \frac{|W\rangle\langle W|V\rangle}{|W|^2}$$

is the component of $|V\rangle$ perpendicular to $|W\rangle$, as is easily checked by taking its inner product with $|W\rangle$.

$$\langle Z|Z\rangle \ge 0 \text{ gives immediately } \left|\langle V|W\rangle\right|^2 \le |V|^2|W|^2.$$

## Linear Operators

A *linear operator A* takes any vector in a linear vector space to a vector in that space, $A|V\rangle = |V'\rangle$, and satisfies

$$A\left(c_1|V_1\rangle + c_2|V_2\rangle\right) = c_1A|V_1\rangle + c_2A|V_2\rangle,$$

with $c_1$, $c_2$ arbitrary complex constants.

The *identity operator I* is (obviously!) defined by:

$$I|V\rangle = |V\rangle \text{ for all } |V\rangle.$$

For an *n*-dimensional vector space with an orthonormal basis $|1\rangle,\ldots,|n\rangle$, since any vector in the space can be expressed as a sum $|V\rangle = \sum v_i |i\rangle$, the linear operator is completely determined by its action on the basis vectors—this is all we need to know. It's easy to find an expression for the identity operator in terms of bras and kets.

Taking the inner product of both sides of the equation $|V\rangle = \sum v_i |i\rangle$ with the bra $\langle i|$ gives $\langle i|V\rangle = v_i$, so

$$|V\rangle = \sum v_i |i\rangle = \sum |i\rangle\langle i|V\rangle.$$

Since this is true for any vector in the space, it follows that that the identity operator is just

$$I = \sum_1^n |i\rangle\langle i|.$$

This is an important result that will reappear in many disguises.

To analyze the action of a general linear operator *A*, we just need to know how it acts on each basis vector. Beginning with $A|1\rangle$, this must be some sum over the basis vectors, and since they are orthonormal, the component in the $|i\rangle$ direction must be just $\langle i|A|1\rangle$.

That is,

$$A|1\rangle = \sum_1^n |i\rangle\langle i|A|1\rangle = \sum_1^n A_{i1}|i\rangle, \text{ writing } \langle i|A|1\rangle = A_{i1}.$$

So if the linear operator *A* acting on $|V\rangle = \sum v_i |i\rangle$ gives $|V'\rangle = \sum v_i'|i\rangle$, that is, $A|V\rangle = |V'\rangle$, the linearity tells us that

$$\sum v_i'|i\rangle = |V'\rangle = A|V\rangle = \sum v_j A|j\rangle = \sum_{i,j} v_j |i\rangle\langle i|A|j\rangle = \sum_{i,j} v_j A_{ij}|i\rangle$$

where in the fourth step we just inserted the identity operator.

Since the $|i\rangle$'s are all orthogonal, the coefficient of a particular $|i\rangle$ on the left-hand side of the equation must be identical with the coefficient of the same $|i\rangle$ on the right-hand side. That is,

$$v_i' = A_{ij} v_j.$$

Therefore the operator *A* is simply equivalent to *matrix multiplication*:

$$\begin{pmatrix} v_1' \\ v_2' \\ . \\ . \\ v_n' \end{pmatrix} = \begin{pmatrix} \langle 1|A|1\rangle & \langle 1|A|2\rangle & . & . & \langle 1|A|n\rangle \\ \langle 2|A|1\rangle & \langle 2|A|2\rangle & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ \langle n|A|1\rangle & \langle n|A|2\rangle & . & . & \langle n|A|n\rangle \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ . \\ . \\ v_n \end{pmatrix}.$$

It is important to note that a linear operator applied successively to the members of an orthonormal basis might give a new set of vectors which *no longer span the entire space*. To give an example, the linear operator $|1\rangle\langle 1|$ applied to any vector in the space picks out the vector's component in the $|1\rangle$ direction. It is called a *projection operator*. The operator $(|1\rangle\langle 1|+|2\rangle\langle 2|)$ projects a vector into its components in the subspace spanned by the vectors $|1\rangle$ and $|2\rangle$, and so on—if we extend the sum to be over the whole basis, we recover the identity operator.

*Exercise*: prove that the $n\times n$ matrix representation of the projection operator $(|1\rangle\langle 1|+|2\rangle\langle 2|)$ has all elements zero except the first two diagonal elements, which are equal to one.

There can be no inverse operator to a nontrivial projection operator, since the information about components of the vector perpendicular to the projected subspace is lost.

Since linear operators are represented by matrices, and their action on vectors is standard matrix multiplication, it follows that *applying two linear operators in succession is equivalent to matrix multiplication*—and, therefore, since matrices do not in general commute, nor do linear operators.

## The Adjoint Operator and Hermitian Matrices

As we've discussed, if a ket $|V\rangle$ in the *n*-dimensional space is written as a column vector with *n* (complex) components, the corresponding bra is a *row* vector having as elements the complex conjugates of the ket elements. $\langle W|V\rangle = \langle V|W\rangle^*$ then follows automatically from standard matrix multiplication rules, and on multiplying $|V\rangle$ by a complex number *a* to get $a|V\rangle$ (meaning that each element in the column of numbers is multiplied by *a*) the corresponding bra goes to $\langle V|a^* = a^*\langle V|$.

But suppose that instead of multiplying a ket by a number, we operate on it with a linear operator. What generates the parallel transformation among the bras? In other words, if $A|V\rangle = |V'\rangle$, can we find a corresponding operator that sends the bra $\langle V|$ to $\langle V'|$? This must be another linear operator, because *A* is linear, that is, if under *A* $|V_1\rangle \rightarrow |V_1'\rangle, |V_2\rangle \rightarrow |V_2'\rangle$ and

$|V_3\rangle = |V_1\rangle + |V_2\rangle$, then under $A$ $|V_3\rangle$ is required to go to $|V_3'\rangle = |V_1'\rangle + |V_2'\rangle$. Consequently, under the parallel *bra* transformation we must have $\langle V_1| \to \langle V_1'|$, $\langle V_2| \to \langle V_2'|$ and $\langle V_3| \to \langle V_3'|$, so the bra transformation is necessarily also linear. Recalling that the bra is an $n$-element *row* vector, the most general linear transformation sending it to another bra is an $n \times n$ matrix operating on the bra from the *right*.

This bra operator is called the *adjoint* of $A$, written $A^\dagger$. That is, the ket $A|V\rangle$ has corresponding bra $\langle V|A^\dagger$. In an orthonormal basis, using the notation $\langle Ai|$ to denote the bra $\langle i|A^\dagger$ corresponding to the ket $A|i\rangle = |Ai\rangle$, say,

$$\left(A^\dagger\right)_{ij} = \langle i|A^\dagger|j\rangle = \langle Ai|j\rangle = \langle j|Ai\rangle^* = A_{ji}{}^*.$$

So *the adjoint operator is the transpose complex conjugate*.

*Important*: for a *product* of two operators (prove this!),

$$(AB)^\dagger = B^\dagger A^\dagger.$$

**An operator equal to its adjoint $A = A^\dagger$ is called *Hermitian***, one equal to minus its adjoint $A = -A^\dagger$ is anti Hermitian (sometimes skew Hermitian). These operator types are in some ways generalizations of real and imaginary numbers. Any operator can be expressed as a sum of a Hermitian operator and an anti Hermitian operator,

$$A = \tfrac{1}{2}(A + A^\dagger) + \tfrac{1}{2}(A - A^\dagger).$$

This notion naturally extends to *vectors and numbers*: the adjoint of a ket is the corresponding bra, the adjoint of a number is its complex conjugate. This is useful to bear in mind when taking the adjoint of an operator which may be partially constructed of vectors and numbers, such as projection-type operators. The adjoint of a product of matrices, vectors and numbers is the product of the adjoints *in reverse order*. (Of course, for numbers the order doesn't matter.)

## Unitary Operators

An operator is *unitary* if $U^\dagger U = 1$. This implies first that $U$ operating on any vector gives a vector having the same norm, since the new norm $\langle V|U^\dagger U|V\rangle = \langle V|V\rangle$ and furthermore inner products are preserved, $\langle W|U^\dagger U|V\rangle = \langle W|V\rangle$. *The original orthonormal basis in the space must therefore under a unitary transformation go to another orthonormal basis*.

Conversely, any transformation that takes one orthonormal basis into another one is a unitary transformation. To see this, suppose that a linear transformation $A$ sends the members of the orthonormal basis $\left(|1\rangle_1, |2\rangle_1, \ldots, |n\rangle_1\right)$ to the different orthonormal set $\left(|1\rangle_2, |2\rangle_2, \ldots, |n\rangle_2\right)$, so

$A|1\rangle_1 = |1\rangle_2$, etc. Then the vector $|V\rangle = \sum v_i |i\rangle_1$ will go to $|V'\rangle = A|V\rangle = \sum v_i |i\rangle_2$, having the same norm, from $\langle V|V\rangle = \sum |v_i|^2$, and similarly $\langle W|V\rangle$ is invariant.

A unitary operation amounts to a rotation (possibly combined with a reflection) in the space. Evidently, since $U^\dagger U = 1$, the adjoint $U^\dagger$ rotates the basis back—it is the inverse operation, and so $UU^\dagger = 1$ also, that is, *U* and $U^\dagger$ commute.

## Determinants

We review in this section the *determinant* of a matrix, a function closely related to the operator properties of the matrix.

Let's start with $2\times 2$ matrices:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

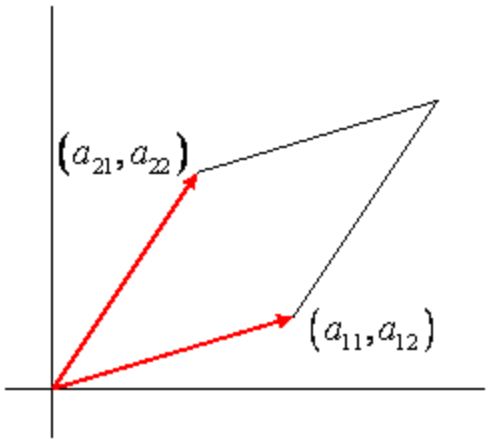
The determinant of this matrix is defined by:

$$\det A = |A| = a_{11}a_{22} - a_{12}a_{21}.$$

Writing the two rows of the matrix as vectors:

$$\vec{a}_1^R = (a_{11}, a_{12})$$
$$\vec{a}_2^R = (a_{21}, a_{22})$$

(*R* denotes row), $\det A = \vec{a}_1^R \times \vec{a}_2^R$ is just the *area* (with appropriate sign) of the parallelogram having the two row vectors as adjacent sides:




This is zero if the two vectors are parallel (linearly dependent) and is not changed by adding any multiple of $\vec{a}_2^R$ to $\vec{a}_1^R$ (because the new parallelogram has the same base and the same

height as the original—check this by drawing).

Let's go on to the more interesting case of $3\times3$ matrices:

$$A=\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

The determinant of $A$ is defined as

$$\det A = \varepsilon_{ijk} a_{1i} a_{2j} a_{3k}$$

where $\varepsilon_{ijk} = 0$ if any two are equal, +1 if *ijk* = 123, 231 or 312 (that is to say, an even permutation of 123) and –1 if *ijk* is an odd permutation of 123. Repeated suffixes, of course, imply summation here.

Writing this out explicitly,

$$\det A = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} - a_{31}a_{22}a_{13}.$$

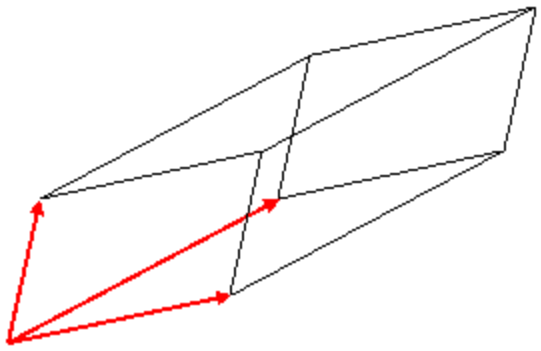
Just as in two dimensions, it's worth looking at this expression in terms of vectors representing the rows of the matrix

$$\vec{a}_1^{\,R} = (a_{11}, a_{12}, a_{13})$$
$$\vec{a}_2^{\,R} = (a_{21}, a_{22}, a_{23})$$
$$\vec{a}_3^{\,R} = (a_{31}, a_{32}, a_{33})$$

so

$$A=\begin{pmatrix} \vec{a}_1^{\,R} \\ \vec{a}_2^{\,R} \\ \vec{a}_3^{\,R} \end{pmatrix}, \quad \text{and we see that } \det A = (\vec{a}_1^{\,R} \times \vec{a}_2^{\,R}) \cdot \vec{a}_3^{\,R}.$$

This is the volume of the parallelopiped formed by the three vectors being adjacent sides (meeting at one corner, the origin).



This parallelepiped volume will of course be zero if the three vectors lie in a plane, and it is not changed if a multiple of one of the vectors is added to another of the vectors. That is to say, *the determinant of a matrix is not changed if a multiple of one row is added to another row*. This is because the determinant is linear in the elements of a single row,

$$\det\begin{pmatrix} \vec{a}_1^R + \lambda\vec{a}_2^R \\ \vec{a}_2^R \\ \vec{a}_3^R \end{pmatrix} = \det\begin{pmatrix} \vec{a}_1^R \\ \vec{a}_2^R \\ \vec{a}_3^R \end{pmatrix} + \lambda \det\begin{pmatrix} \vec{a}_2^R \\ \vec{a}_2^R \\ \vec{a}_3^R \end{pmatrix}$$

and the last term is zero because two rows are identical—so the triple vector product vanishes.

A more general way of stating this, applicable to larger determinants, is that for a determinant with two identical rows, the symmetry of the two rows, together with the *antisymmetry* of $\varepsilon_{ijk}$, ensures that the terms in the sum all cancel in pairs.

Since the determinant is not altered by adding some multiple of one row to another, if the rows are linearly dependent, one row could be made identically zero by adding the right multiples of the other rows. Since every term in the expression for the determinant has one element from each row, the determinant would then be identically zero. For the three-dimensional case, the linear dependence of the rows means the corresponding vectors lie in a plane, and the parallelepiped is flat.

The algebraic argument generalizes easily to $n \times n$ determinants: they are *identically zero if the rows are linearly dependent*.

The generalization from $3\times3$ to $n\times n$ determinants is that $\det A = \varepsilon_{ijk} a_{1i} a_{2j} a_{3k}$ becomes:

$$\det A = \varepsilon_{ijk\ldots p} a_{1i} a_{2j} a_{3k} \ldots a_{np}$$

where *ijk…p* is summed over all permutations of 132…*n*, and the $\varepsilon$ symbol is zero if any two of its suffixes are equal, +1 for an even permutation and −1 for an odd permutation. (*Note*: any permutation can be written as a product of swaps of neighbors. Such a representation is in general not unique, but for a given permutation, all such representations will have either an odd number of elements or an even number.)

An important theorem is that for a product of two matrices $A$, $B$ the determinant of the product is the product of the determinants, $\det AB = \det A \times \det B$. This can be verified by brute force for $2\times2$ matrices, and a proof in the general case can be found in any book on mathematical physics.

It can also be proved that *if* the rows are linearly independent, the determinant cannot be zero.

(*Here's a proof*: take an $n \times n$ matrix with the *n* row vectors linearly independent. Now consider the components of those vectors in the $n - 1$ dimensional subspace perpendicular to (1, 0, … ,0). These *n* vectors, each with only $n - 1$ components, must be linearly dependent, since there are more of them than the dimension of the space. So we can take some combination of the rows below the first row and subtract it from the first row to leave the first row (*a*, 0, 0, … ,0), and *a* cannot be zero since we have a matrix with *n* linearly independent rows. We can then subtract multiples of this first row from the other rows to get a determinant having zeros in the first column below the first row. Now look at the $n - 1$ by $n - 1$ determinant to be multiplied by *a*.
Its rows must be linearly independent since those of the original matrix were. Now proceed by induction.)

To return to three dimensions, it is clear from the form of

$$\det A = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} - a_{31}a_{22}a_{13}$$

that we could equally have taken the *columns* of *A* as three vectors, $A = (\vec{a}_1^{\,C}, \vec{a}_2^{\,C}, \vec{a}_3^{\,C})$ in an obvious notation, $\det A = (\vec{a}_1^{\,C} \times \vec{a}_2^{\,C}) \cdot \vec{a}_3^{\,C}$, and linear dependence among the *columns* will also ensure the vanishing of the determinant—so, in fact, linear dependence of the columns ensures linear dependence of the rows.

This, too, generalizes to $n \times n$: in the definition of determinant $\det A = \varepsilon_{ijk\ldots p} a_{1i} a_{2j} a_{3k} \ldots a_{np}$, the row suffix is fixed and the column suffix goes over all permissible permutations, with the appropriate sign—but the same terms would be generated by having the *column* suffixes kept in numerical order and allowing the row suffix to undergo the permutations.

## An Aside: Reciprocal Lattice Vectors

It is perhaps worth mentioning how the inverse of a $3 \times 3$ matrix operator can be understood in terms of vectors. For a set of linearly independent vectors $(\vec{a}_1, \vec{a}_2, \vec{a}_3)$, a reciprocal set $(\vec{b}_1, \vec{b}_2, \vec{b}_3)$ can be defined by

$$\vec{b}_1 = \frac{\vec{a}_2 \times \vec{a}_3}{\vec{a}_1 \times \vec{a}_2 \cdot \vec{a}_3}$$

and the obvious cyclic definitions for the other two reciprocal vectors. We see immediately that

$$\vec{a}_i \cdot \vec{b}_j = \delta_{ij}$$

from which it follows that the inverse matrix to

$$A = \begin{pmatrix} \vec{a}_1^{\,R} \\ \vec{a}_2^{\,R} \\ \vec{a}_3^{\,R} \end{pmatrix} \text{ is } B = \begin{pmatrix} \vec{b}_1^{\,C} & \vec{b}_2^{\,C} & \vec{b}_3^{\,C} \end{pmatrix}$$

.

(These reciprocal vectors are important in *x*-ray crystallography, for example. If a crystalline lattice has certain atoms at positions $n_1\vec{a}_1 + n_2\vec{a}_2 + n_3\vec{a}_3$, where $n_1$, $n_2$, $n_3$ are integers, the reciprocal vectors are the set of normals to possible planes of the atoms, and these planes of atoms are the important elements in the diffractive *x*-ray scattering.)

## Eigenkets and Eigenvalues

If an operator $A$ operating on a ket $|V\rangle$ gives a multiple of the same ket,

$$A|V\rangle = \lambda|V\rangle$$

then $|V\rangle$ is said to be an *eigenket* (or, just as often, *eigenvector*, or *eigenstate*!) of $A$ with *eigenvalue* $\lambda$.

Eigenkets and eigenvalues are of central importance in quantum mechanics: dynamical variables are operators, a physical measurement of a dynamical variable yields an eigenvalue of the operator, and forces the system into an eigenket.

In this section, we shall show how to find the eigenvalues and corresponding eigenkets for an operator $A$. We'll use the notation $A|a_i\rangle = a_i|a_i\rangle$ for the set of eigenkets $|a_i\rangle$ with corresponding eigenvalues $a_i$. (Obviously, in the eigenvalue equation here the suffix $i$ not summed over.)

The first step in solving $A|V\rangle = \lambda|V\rangle$ is to find the allowed eigenvalues $a_i$.

Writing the equation in matrix form:

$$\begin{pmatrix} A_{11}-\lambda & A_{12} & . & . & A_{1n} \\ A_{21} & A_{22}-\lambda & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ A_{n1} & . & . & . & A_{nn}-\lambda \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ . \\ . \\ v_n \end{pmatrix} = 0.$$

This equation is actually telling us that the *columns* of the matrix $A - \lambda I$ are linearly dependent! To see this, write the matrix as a row vector each element of which is one of its columns, and the equation becomes

$$(\vec{M}_1^C, \vec{M}_2^C, \ldots, \vec{M}_n^C) \begin{pmatrix} v_1 \\ . \\ . \\ . \\ v_n \end{pmatrix} = 0$$

which is to say

$$v_1 \vec{M}_1^C + v_2 \vec{M}_2^C = \ldots + v_n \vec{M}_n^C = 0,$$

the columns of the matrix are indeed a *linearly dependent* set.

We know that means the determinant of the matrix $A - \lambda I$ is zero,

$$\begin{vmatrix} A_{11} - \lambda & A_{12} & . & . & A_{1n} \\ A_{21} & A_{22} - \lambda & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ A_{n1} & . & . & . & A_{nn} - \lambda \end{vmatrix} = 0.$$

Evaluating the determinant using $\det A = \varepsilon_{ijk\ldots p} a_{1i} a_{2j} a_{3k} \ldots a_{np}$ gives an $n^{\text{th}}$ order polynomial in $\lambda$ sometimes called the *characteristic polynomial*. Any polynomial can be written in terms of its roots:

$$C(\lambda - a_1)(\lambda - a_2)\ldots(\lambda - a_n) = 0$$

where the $a_i$'s, the roots of the polynomial, and $C$ is an overall constant, which from inspection of the determinant we can see to be $(-1)^n$. (It's the coefficient of $\lambda^n$.) The polynomial roots (which we don't yet know) are in fact the eigenvalues. For example, putting $\lambda = a_1$ in the matrix, $\det(A - a_1 I) = 0,$ which means that $(A - a_1 I)|V\rangle = 0$ has a nontrivial solution $|V\rangle$, and this is our eigenvector $|a_1\rangle$.

Notice that the diagonal term in the determinant $(A_{11} - \lambda)(A_{22} - \lambda)\ldots(A_{nn} - \lambda)$ generates the leading two orders in the polynomial $(-1)^n \left(\lambda^n - (A_{11} + \ldots + A_{nn})\lambda^{n-1}\right)$, (and some lower order terms too). Equating the coefficient of $\lambda^{n-1}$ here with that in $(-1)^n (\lambda - a_1)(\lambda - a_2)\ldots(\lambda - a_n)$,

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} A_{ii} = \mathrm{Tr}\, A.$$

Putting $\lambda = 0$ in both the determinantal and the polynomial representations (in other words, equating the $\lambda$-independent terms),

$$\prod_{i=1}^{n} a_i = \det A.$$

So we can learn something about the eigenvalues directly from the determinant. But if the space is bigger than $2\times 2$, this is not enough to solve the problem.

We must solve the polynomial equation $\det(A-\lambda I)=0$ to find the set of eigenvalues, then use them to calculate the corresponding eigenvectors. This is done one at a time.

Labeling the first eigenvalue found as $a_1$, the corresponding equation for the components $v_i$ of the eigenvector $|a_1\rangle$ is

$$\begin{pmatrix} A_{11}-a_1 & A_{12} & \cdot & \cdot & A_{1n} \\ A_{21} & A_{22}-a_1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ A_{n1} & \cdot & \cdot & \cdot & A_{nn}-a_1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ v_n \end{pmatrix} = 0.$$

This looks like $n$ equations for the $n$ numbers $v_i$, but it isn't: remember the rows are linearly dependent, so there are only $n-1$ independent equations. However, that's enough to determine
the ratios of the vector components $v_1, \ldots, v_n$, then finally the eigenvector is normalized.

## Eigenvalues and Eigenstates of Hermitian Matrices

For a *Hermitian* matrix, it is easy to establish that the eigenvalues are always *real*. Taking (in this section) $A$ to be hermitian, $A=A^{\dagger}$, and labeling the eigenkets by the eigenvalue, that is,

$$A|a_1\rangle = a_1|a_1\rangle$$

the inner product with the bra $\langle a_1|$ gives $\langle a_1|A|a_1\rangle = a_1\langle a_1|a_1\rangle$. But the inner product of the *adjoint* equation (remembering $A=A^{\dagger}$)

$$\langle a_1|A = a_1{}^*\langle a_1|$$

with $|a_1\rangle$ gives $\langle a_1|A|a_1\rangle = a_1{}^*\langle a_1|a_1\rangle$, so $a_1 = a_1{}^*$, and all the eigenvalues must be real.

Furthermore, the eigenkets belonging to *different* eigenvalues are *orthogonal*, if

$$A|a_1\rangle = a_1|a_1\rangle$$
$$A|a_2\rangle = a_2|a_2\rangle$$

taking the adjoint of the first equation and taking its inner product with $|a_2\rangle$,

$$\langle a_1|A|a_2\rangle = a_1\langle a_1|a_2\rangle = a_2\langle a_1|a_2\rangle$$

so $\langle a_1|a_2\rangle = 0$ unless the eigenvalues are equal. If they *are* equal, then any linear combination of them is also an eigenvector with the same eigenvalue, and the Gram Schmidt procedure can be used to construct an orthonormal basis in this subspace as part of a full orthonormal basis.

A subspace spanned by eigenvectors having the same eigenvalue is termed *degenerate*.

## Diagonalizing a Hermitian Matrix

Diagonalizing a matrix means transforming to an orthonormal basis in which the only nonzero matrix elements are on the diagonal, $i = j$.

In fact, it is evident from the discussion at the end of the previous section that a hermitian matrix is diagonal in the orthonormal basis of its set of eigenvectors: $|a_1\rangle, |a_2\rangle, \ldots, |a_n\rangle$, since

$$\langle a_i|A|a_j\rangle = \langle a_i|a_j|a_j\rangle = a_j\langle a_i|a_j\rangle = a_j\delta_{ij}.$$

However, we are assuming that we know the matrix elements of $A$ in some other orthonormal basis, so to diagonalize $A$ we need to rotate from the initial orthonormal basis to one made up of the eigenkets of $A$.

Denoting the initial orthonormal basis in the standard fashion

$$|1\rangle = \begin{pmatrix}1\\0\\0\\.\\0\end{pmatrix}, \quad |2\rangle = \begin{pmatrix}0\\1\\0\\.\\0\end{pmatrix}, \quad |i\rangle = \begin{pmatrix}0\\\vdots\\1\\\vdots\\0\end{pmatrix} \ldots (1 \text{ in } i^{\text{th}} \text{ place down}), \quad |n\rangle = \begin{pmatrix}0\\0\\0\\.\\1\end{pmatrix}.$$

the elements of the matrix are $A_{ij} = \langle i|A|j\rangle$.

A transformation from one orthonormal basis to another is a *unitary* transformation, as discussed above, so we write it

$$|V\rangle \rightarrow |V'\rangle = U|V\rangle.$$

Under this transformation, the matrix element

$$\langle W|A|V\rangle \rightarrow \langle W'|A|V'\rangle = \langle W|U^{\dagger}AU|V\rangle.$$

So we can find the appropriate transformation matrix $U$ by requiring that $U^{\dagger}AU$ be diagonal with respect to the *original* set of basis vectors. (Transforming the operator in this way, leaving the vector space alone, is equivalent to rotating the vector space and leaving the operator alone. Of course, in a system with more than one operator, the same transformation would have to be applied to all the operators).

In fact, the unitary matrix $U$ we need is just composed of the normalized eigenkets of the operator $A$,

$$U = \left(|a_1\rangle, |a_2\rangle, \ldots, |a_n\rangle\right).$$

It is easy to see that this must be true:

$$AU = \left(a_1|a_1\rangle, a_2|a_2\rangle, \ldots, a_n|a_n\rangle\right)$$

so

$$\left(U^{\dagger}AU\right)_{ij} = \langle a_i|a_j|a_j\rangle = \delta_{ij}a_j, \text{ a diagonal matrix.}$$

(The repeated suffixes here are of course *not* summed over.)

To actually find the eigenkets, then, we need to go through the routine outlined in the previous section, that is, first find the eigen*values* by finding the zeros of the determinantal equation, $|A - \lambda I| = 0,$ then enter the roots $a_1, a_2, \ldots, a_n$ one at a time into the matrix equation

$$\begin{pmatrix} A_{11}-a_i & A_{12} & . & . & A_{1n} \\ A_{21} & A_{22}-a_i & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ A_{n1} & . & . & . & A_{nn}-a_i \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ . \\ . \\ v_n \end{pmatrix} = 0$$

to give $n$ - 1 linear equations for the $n$ unknown components of the eigenket $|a_i\rangle$.

Then

$$U^{\dagger}AU = \begin{pmatrix} a_1 & 0 & 0 & . & 0 \\ 0 & a_2 & 0 & . & 0 \\ 0 & 0 & a_3 & . & 0 \\ . & . & . & . & . \\ 0 & . & . & . & a_n \end{pmatrix}$$

If some of the eigenvalues are the same, the Gram Schmidt procedure may be needed to generate an orthogonal set, as mentioned earlier.

## Functions of Matrices

The same unitary operator $U$ that diagonalizes an Hermitian matrix $A$ will also diagonalize $A^2$, because

$$U^{-1}A^2U = U^{-1}AAU = U^{-1}AUU^{-1}AU$$

so

$$U^{\dagger}A^2U = \begin{pmatrix} a_1^2 & 0 & 0 & . & 0 \\ 0 & a_2^2 & 0 & . & 0 \\ 0 & 0 & a_3^2 & . & 0 \\ . & . & . & . & . \\ 0 & . & . & . & a_n^2 \end{pmatrix}.$$

Evidently, this same process works for any power of $A$, and formally for any function of $A$ expressible as a power series, but of course convergence properties need to be considered, and this becomes trickier on going from finite matrices to operators on infinite spaces.

## Commuting Hermitian Matrices

From the above, the set of powers of an Hermitian matrix all commute with each other, and have a common set of eigenvectors (but not the same eigenvalues, obviously). In fact it is not difficult to show that any two Hermitian matrices that commute with each other have the same set of eigenvectors (after possible Gram Schmidt rearrangements in degenerate subspaces).

If two $n \times n$ Hermitian matrices $A$, $B$ commute, that is, $AB = BA$, and $A$ has a *nondegenerate* set of eigenvectors $A|a_i\rangle = a_i|a_i\rangle$, then $AB|a_i\rangle = BA|a_i\rangle = Ba_i|a_i\rangle = a_iB|a_i\rangle$, that is, $B|a_i\rangle$ is an eigenvector of $A$ with eigenvalue $a_i$. Since $A$ is nondegenerate, $B|a_i\rangle$ must be some multiple of $|a_i\rangle$, and we conclude that $A$, $B$ have the same set of eigenvectors.

Now suppose $A$ *is* degenerate, and consider the $m \times m$ subspace $S_{a_i}$ spanned by the eigenvectors $|a_i,1\rangle, |a_i,2\rangle, \cdots$ of $A$ having eigenvalue $a_i$. Applying the argument in the paragraph above, $B|a_i,1\rangle, B|a_i,2\rangle, \cdots$ must also lie in this subspace. Therefore, if we transform $B$ with the same unitary transformation that diagonalized $A$, $B$ will not in general be diagonal in the subspace $S_{a_i}$, but it will be what is termed *block diagonal*, in that if $B$ operates on any vector in $S_{a_i}$ it gives a vector in $S_{a_i}$.

$B$ can be written as two diagonal blocks: one $m \times m$, one $(n-m)\times(n-m)$, with zeroes outside these diagonal blocks, for example, for $m = 2$, $n = 5$:

$$\begin{pmatrix} b_{11} & b_{12} & 0 & 0 & 0 \\ b_{21} & b_{22} & 0 & 0 & 0 \\ 0 & 0 & b_{33} & b_{34} & b_{35} \\ 0 & 0 & b_{43} & b_{44} & b_{45} \\ 0 & 0 & b_{53} & b_{54} & b_{55} \end{pmatrix}.$$

And, in fact, if there is only *one* degenerate eigenvalue that second block will only have nonzero terms on the diagonal:

$$\begin{pmatrix} b_{11} & b_{12} & 0 & 0 & 0 \\ b_{21} & b_{22} & 0 & 0 & 0 \\ 0 & 0 & b_3 & 0 & 0 \\ 0 & 0 & 0 & b_4 & 0 \\ 0 & 0 & 0 & 0 & b_5 \end{pmatrix}.$$

*B* therefore operates on two subspaces, one *m*-dimensional, one (*n*–*m*)-dimensional, *independently*—a vector entirely in one subspace stays there.

This means we can complete the diagonalization of *B* with a unitary operator that *only* operates on the $m \times m$ block $S_4$. Such an operator will also affect the eigenvectors of *A*, but that doesn't matter, because all vectors in this subspace are eigenvectors of *A* with the *same* eigenvalue, so as far as *A* is concerned, we can choose any orthonormal basis we like—the basis vectors will still be eigenvectors.

This establishes that any two *commuting* Hermitian matrices can be diagonalized at the same time. Obviously, this can never be true of *noncommuting* matrices, since all diagonal matrices commute.

## Example: Diagonalizing a Unitary Matrix

As previously stated, a unitary matrix is an operator that rotates an orthonormal basis into another orthonormal basis. It is clear that in three dimensions the axis through the origin about which the basis is rotated is an eigenvector of this transformation. It's less clear what the other two eigenvectors might be—or, equivalently, what are the eigenvectors corresponding to a two dimensional rotation of basis in a plane? The way to find out is to write down the matrix and diagonalize it.

The matrix

$$U(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

Note that the determinant is equal to unity. The eigenvalues are given by solving

$$\begin{vmatrix} \cos\theta - \lambda & \sin\theta \\ -\sin\theta & \cos\theta - \lambda \end{vmatrix} = 0 \text{ to give } \lambda = e^{\pm i\theta}.$$

The corresponding eigenvectors satisfy

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} u_1^{\pm} \\ u_2^{\pm} \end{pmatrix} = e^{\pm i\theta} \begin{pmatrix} u_1^{\pm} \\ u_2^{\pm} \end{pmatrix}.$$

The eigenvectors, normalized, are:

$$\begin{pmatrix} u_1^{\pm} \\ u_2^{\pm} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm i \end{pmatrix}.$$

Note that, in contrast to a Hermitian matrix, the eigenvalues of a unitary matrix do not have to be real. In fact, from $U^{\dagger}U = 1$, sandwiched between the bra and ket of an eigenvector, we see that any eigenvalue of a unitary matrix must have unit modulus—it's a complex number on the unit circle. With hindsight, we should have realized that one eigenvalue of a two-dimensional rotation had to be $e^{i\theta}$, the product of two two-dimensional rotations is given be adding the angles of rotation, and a rotation through $\pi$ changes all signs, so has eigenvalue -1. Note that the eigenvector itself is independent of the angle of rotation—the rotations all commute, so they must have common eigenvectors. Successive rotation operators applied to the plus eigenvector add their angles, when applied to the minus eigenvector, all angles are subtracted.