Part 4

Technical Appendix

Peter Fortune

Part 1 of this three part series reviewed the history, development, and interpretation of quantum mechanics. This was done in a nonmathematical fashion appropriate to a general background of the field.

Part 2 reviewed some of the details of quantum theoretical methods. The objective was to lay out the gist of the field with a minimal level of mathematics.

Part 3 reviews issues in Classical and Quantum Information Theory, focusing on Cryptology and Computing

Part 4 is the Technical Appendix

December 2012

Contents

Linear Algebra

This review of the basic mathematical concepts behind quantum mechanics begins with a review of the linear algebra of Cartesian (Euclidean) Vector Spaces. These are vector spaces underlying the graphs we have known and loved since high school. They consist of real numbers in a rectangular coordinate system. Such a space is referred to as an $\mathbf{R}^N$ space, where **R** indicates it is based on the set of all real numbers and N is the number of dimensions in the space.

Once we have the rules of linear algebra down, we turn to other mathematical concepts important to quantum mechanics. Hilbert Spaces—the spaces used in quantum mechanics—and the related concepts of observables and operators.

Scalars, Vectors and Matrixes

There are three fundamental concepts in linear algebra: a scalar, a vector, and a matrix. A scalar is a number, denoted herein by italics. A vector is a collection of scalars fashioned into a row or a column. Thus, a row vector with N scalars—called a 1 x N vector—is written as

(LA.1) $$\langle v| = [v_1\ v_2\ \ldots\ v_M]$$

where v_i (i= 1,…,M) are M scalars and $\langle v|$, called a *bra*, is the notation used in physics for a row vector; v can be any symbol you want to use for the vector—often ψ is used for a vector of quantum numbers, so the bra could be written $\langle \psi|$. The row vector has *dimension* 1 x M because it has 1 row and M columns ("dimension" is the number of rows is given first, so dimension = rows x columns).

A vector can also be written as a column vector, with (say) N rows and one column—an N x 1 vector. This would be a *ket*, as below

$$(\text{LA.2}) \qquad |w\rangle = \begin{bmatrix} w_1 \\ w_2 \\ .... \\ w_N \end{bmatrix}$$

where the notation $|w\rangle$ denotes a column vector referred to as "ket w". The column vector has dimension Nx1. Note that the vector is v or w, its components are scalars *v* or *w*, and the bra-ket notation simply tells us whether the vector is a row vector or a column vector.

A matrix is a collection of vectors, each of the same size or *dimension*; the dimension of a matrix is the number of rows times the number of columns: a matrix with M column vectors, each with N rows , is an N x M matrix containing NM entries (scalars, functions, whatever).

For example, suppose we have three separate 1xN column vectors denoted as $|v_1\rangle$, $|v_2\rangle$, and $|v_3\rangle$. The matrix V formed from those kets is

$$(\text{LA.3}) \qquad V = [|v_1\rangle \ |v_2\rangle |v_3\rangle] = \begin{bmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \\ \dots & \dots & \dots \\ v_{1N} & v_{2N} & v_{3N} \end{bmatrix}$$

Vector Operations

Vectors have many of the properties we are familiar with from simple arithmetic. Among them, vectors are additive if they have the same dimensions. The result of vector addition is the elementwise sum of the components, or

$$(\text{LA.4}) \qquad |v\rangle + |w\rangle = \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \dots + \dots \\ v_N + w_N \end{bmatrix}$$

The simplest vector operation is *scalar multiplication*—the multiplication of a vector by a scalar. To multiply a vector by a scalar you simply multiply each element of the vector by the scalar, as in

$$\text{(LA.5)} \qquad s|v\rangle = \begin{bmatrix} sv_1 \\ sv_2 \\ \dots \\ sv_N \end{bmatrix}$$

The effect is to change the magnitude of the vector (its length in vector space) but not its direction.

An often-used operation is calculating the *dot product* (also called the *inner product*). The dot product is the (post) multiplication of a bra by a ket, as in $\langle v||w\rangle$. Both vectors must *conform*, that is, have the same number of elements (M = N). Also note that in dot multiplication the row vector must come first—a 1xN vector can dot multiply an Nx1 vector, with the result being a scalar (which is simply a 1x1 vector); but an Nx1 vector cannot dot multiply a 1xN vector.

The dot product of (2) and (2'), $\langle v||w\rangle$, is a scalar obtained by multiplying each scalar in $\langle w|$ by the scalar in $|v\rangle$ that is in the same position (first *v* time first *w*, second *v* times second *w*, and so on), then adding the results.

$$\text{(LA.6)} \qquad \langle v||w\rangle = [v_1\ v_2\ \dots.\ v_N] \begin{bmatrix} w_1 \\ w_2 \\ \dots. \\ w_N \end{bmatrix} = v_1w_1 + v_2w_2 + \dots + v_Nw_N = \sum_1^N v_iw_i$$

The dot product tells us about the correlation between the components in the two vectors. If the dot product is large and positive, it says that v and w tend to move in the same direction; if it is large and negative, the vector values tend to move in opposite directions; if it is small, they tend to be only loosely correlated.

A particularly important result is when the dot product of v and w is zero. In that case v and w are said to be *orthogonal*, which is a fancy word for "perpendicular to each other." Orthogonal vectors allow the scalars in one vector to be measured without concern for the value on the other vector: a change in x carries no implications for the value of y. Orthogonal vectors form the *basis* of a vector space, a concept worth drawing out.

Also of particular interest is the dot product of a vector with itself, say $\langle v \| v \rangle$. This is formed by multiplying each component of v by itself, then adding the results to get the "sum of squares"

$$\text{(LA.7)} \qquad \langle v \| v \rangle = \begin{matrix} [v_1\ v_2\ \ldots.\ v_N] \\ \end{matrix} \begin{bmatrix} v_1 \\ v_2 \\ \ldots. \\ v_N \end{bmatrix} = v_1^2 + v_2^2 + \ldots + v_N^2 = \sum_1^N v_i^2$$

The last vector operation we will look at is the *tensor product*, often called the *outer product*. The tensor product of two vectors is $|v\rangle \otimes \langle w|$, that is, a column vector times a row vector. If the column vector is Nx1 and the row vector is 1xM, the tensor product is an NxM matrix. Thus

$$\text{(LA.8)} \qquad |v\rangle \otimes \langle w| = \begin{bmatrix} v_1 \\ v_2 \\ \ldots. \\ v_N \end{bmatrix} [w_1\ \ w_2\ \ldots.\ w_M] = \begin{bmatrix} v_1w_1 & v_1w_2\ \ldots.\ v_1w_M \\ v_2w_1 & v_2w_2 \ldots.\ v_2w_M \\ \ldots & \ldots\ \ldots. & \ldots \\ v_Nw_1 & v_Nw_2 & v_Nw_M \end{bmatrix}$$

This is done by multiplying each element in the column vector $|v\rangle$ by the row vector $\langle w|$; there are MN elements in this NxM matrix.

Matrices

A matrix is a collection of vectors, each of the same dimension. The size of a matrix is described as number of rows times number of columns. For example, a matrix with M column vectors, each with N rows , is an N x M matrix; it contains NM entries (scalars, functions, whatever).

For example, suppose we have three separate Nx1 column vectors denoted as $|v_1\rangle$, $|v_2\rangle$, and $|v_3\rangle$. The matrix V formed from those kets is

(LA.9) $$V = [|v_1\rangle, |v_2\rangle, |v_3\rangle] = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \\ \dots & \dots & \dots \\ v_{N1} & v_{N2} & v_{N3} \end{bmatrix}$$

This matrix is an NxN ("square") matrix formed by putting the second ket next to the first ket, the third ket next to the second ket, and so on. Suppose we form a matrix W in exactly the same way, as below

(LA.10) $$W = [|w_1\rangle, |w_2\rangle, |w_3\rangle] = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \\ \dots & \dots & \dots \\ w_{N1} & w_{N2} & w_{N3} \end{bmatrix}$$

Now we have two NxN matrices V and W. We can convert either of these to its *transpose*, by simply rotating the first column into the first row, the second column into the second row, and so on. For the transpose of W, denoted W*, the elements w_{ij} in W become w_{ji} in W*, i.e.,

(LA.11) $$W^* = \begin{bmatrix} w_{11} & w_{21} & w_{31} & \dots & w_{N1} \\ w_{12} & w_{22} & w_{32} & \dots & w_{N2} \\ w_{13} & w_{23} & w_{33} & \dots & w_{N3} \\ \dots & \dots & \dots & \dots & \dots \\ w_{1N} & w_{2N} & w_{3N} & \dots & w_{NN} \end{bmatrix}$$

Two matrices can be multiplied if they are conformable, that is, if the number of rows (or columns) of one is equal to the number of columns (or rows) of the other. V and W* are conformable because V is NxN and W* is NxN. If V postmultiplies W*, the result is the NxN matrix below. (Note that W* can not postmultiply V because the columns of W* and the number of rows of V are different.)

(LA.12)

$$W^*V = \begin{bmatrix} \langle w_1|v_1\rangle & \langle w_1|v_2\rangle & \dots & \langle w_1|v_N\rangle \\ \langle w_2|v_1\rangle & \langle w_2|v_2\rangle & \dots & \langle w_2|v_N\rangle \\ \langle w_3|v_1\rangle & \langle w_3|v_2\rangle & \dots & \langle w_3|v_N\rangle \\ \dots & \dots & \dots & \\ \langle w_N|v_1\rangle & \langle w_N|v_2\rangle & \dots & w_N|v_N\rangle \end{bmatrix}$$

$$= \begin{bmatrix} \sum_1^N w_{1i}v_{1i} & \sum_1^N w_{1i}v_{2i} & \dots & \sum_1^N w_{1i}v_{Ni} \\ \sum_1^N w_{2i}v_{1i} & \sum_1^N w_{2i}v_{2i} & \dots & \sum_1^N w_{2i}v_{Ni} \\ \sum_1^N w_{3i}v_{1i} & \sum_1^N w_{3i}v_{2i} & \dots & \sum_1^N w_{Ni}v_{Ni} \\ \dots & \dots & \dots & \dots \\ \sum_1^N w_{Ni}v_{1i} & \sum_1^N w_{Ni}v_{2i} & \dots & \sum_1^N w_{Ni}v_{Ni} \end{bmatrix}$$

Thus, matrix multiplication is simply repeated dot products of the row and column vectors.

We have seen that the tensor product of a scalar and a vector is simply a vector with each element in the original vector multiplied by the scalar. We now move on to the tensor product of two matrices (the tensor product of two vectors is formed the same way).

The tensor product of two matrices is formed by multiplying each *element* of the first matrix by the entire second matrix. The size of matrices (or vectors) in tensor products is irrelevant—there is no conformability requirement for tensor products. Suppose that A is an M x N matrix and B is an K x L matrix. The tensor product of A and B is

$$\text{(LA.13)} \qquad B\otimes A = \begin{bmatrix} b_{11}A & b_{12}A & b_{13}A & \dots & b_{1L}A \\ b_{21}A & b_{22}A & b_{23}A & \dots & b_{2L}A \\ b_{31}A & b_{32}A & b_{33}A & \dots & b_{3L}A \\ \dots & \dots & \dots & \dots & \dots \\ b_{K1}A & b_{K2}A & b_{K3}A & \dots & b_{KL}A \end{bmatrix}$$

This matrix is a KM x LN matrix. . To show it's expansion, assume that A and B are both 2x2 matrices. Then a 4x4 matrix is created:

$$\text{(LA.14)} \quad B\otimes A = \begin{bmatrix} b_{11}A & b_{12}A \\ b_{21}A & b_{22}A \end{bmatrix} = \begin{bmatrix} b_{11}a_{11} & b_{11}a_{12} & b_{12}a_{11} & b_{12}a_{12} \\ b_{11}a_{12} & b_{11}a_{22} & b_{21}a_{12} & b_{21}a_{22} \\ b_{21}a_{11} & b_{21}a_{12} & b_{21}a_{11} & b_{22}a_{12} \\ b_{21}a_{12} & b_{21}a_{22} & b_{22}a_{12} & b_{22}a_{22} \end{bmatrix}$$

Just as matrices can be multiplied by matrices, so matrices can be multiplied by vectors (which are, in fact, just matrices with one column or one row). Suppose we want to multiply the 1xN row vector $\langle w| = [w_1\ w_2\ \dots\ w_N]$ by the NxN vmatrix W, as below

$$\text{(LA.15)} \quad \langle w|Z = [w_1\ w_2\ \dots\ w_N] \begin{bmatrix} z_{11} & z_{21} & z_{31} & \dots & z_{N1} \\ z_{12} & z_{22} & z_{32} & \dots & z_{N2} \\ z_{13} & z_{23} & z_{33} & \dots & z_{3N} \\ \dots\dots & \dots\dots & \dots\dots & \dots & \dots\dots \\ z_{1N} & z_{2N} & z_{3N} & \dots & z_{NN} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_1^N v_{1i}vz_{1i} & \sum_1^N v_{1i}vz_{2i} & \dots & \sum_1^N v_{1i}vz_{Ni} \end{bmatrix}$$

This is the dot product of *w* with each column vector in Z, that is,

Determinants, Adjoints, and Inverses

We denote the *determinant* of matrix **A** by **|A|**, the *transpose* of the matrix by $\mathbf{A^T}$, the *adjoint* of the matrix by $\mathbf{A^+}$, and the inverse of the matrix by $\mathbf{A^{-1}}$. These are very related properties in linear algebra.

Determinant of a Matrix

Consider the following 3x3 matrix:

$$\text{(LA.16)} \qquad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

The determinant of **A** is a scalar formed by a composite of multiplication and addition of the elements in **A**. The formal approach is called *expansion by cofactors*. The *cofactor* of a matrix element $\boldsymbol{a}_{ij}$, denoted $\pm C_{ij}$, is the determinant of the matrix remaining after both the row and the column of $\boldsymbol{a}_{ij}$ have been eliminated from the matrix; the cofactor is positive if i + j is even and negative if it is odd.

For example, the cofactor of $\boldsymbol{a_{22}}$ is:

$$\text{(LA.17)} \qquad A = \begin{matrix} a_{11} & \not{a}_{12} & a_{13} \\ \not{a}_{21} & \not{a}_{22} & \not{a}_{23} \\ a_{31} & \not{a}_{32} & a_{33} \end{matrix}$$

The row and column with $\boldsymbol{a_{22}}$ are struck out, leaving $C_{22} = \begin{vmatrix} \boldsymbol{a_{11}} & a_{13} \\ a_{31} & \boldsymbol{a_{33}} \end{vmatrix}$ as the cofactor of $\boldsymbol{a_{22}}$. The cofactor (like any determinant) is evaluated as the product of numbers on the NW-SE direction less the product of numbers on the SW to NE direction, i.e., $C_{22} = \boldsymbol{a}_{11}\boldsymbol{a}_{33} - \boldsymbol{a}_{31}\boldsymbol{a}_{13}$.

Expansion by cofactors requires selecting *any* row (or column) and computing the cofactors for each of those elements (taking careful note of the cofactor's sign). Then each element in the selected row or column is multiplied by its cofactor and the results are summed. Thus, if we choose the first row of A the determinant is $|\mathbf{A}| = \boldsymbol{a}_{11}C_{11} + \boldsymbol{a}_{12}(-C_{12}) + \boldsymbol{a}_{13}C_{13}$. The cofactors for the row are

(LA.18) $$+C_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = a_{22}a_{33} - a_{32}a_{33}$$

$$-C_{12} = -\begin{vmatrix} a_{21} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} = -(a_{21}a_{23} - a_{31}a_{33})$$

$$C_{13} = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} = (a_{21}a_{32} - a_{31}a_{22})$$

And the determinant of A is

(LA.19) $$|\mathbf{A}| = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13}$$

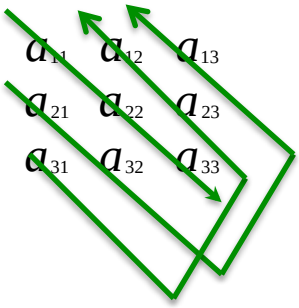
$$= a_{11}(a_{22}a_{33} - a_{32}a_{33}) - a_{12}(a_{21}a_{23} - a_{31}a_{33}) + a_{13}(a_{21}a_{32} - a_{31}a_{22})$$

Note that each element in a matrix has a cofactor, and that expansion by cofactors can be done using single any row or column.

A less formal but quicker approach is shown below. The operation goes as follows

- *Multiply each triplet of elements on the diagonals counterclockwise starting with the main diagonal, then moving to the next lower diagonal, and so on, then add each triplet multiple together.*

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = +a_{11}a_{12}a_{13}+a_{21}a_{32}a_{13}+a_{31}a_{23}a_{12}$$



• *Multiply each triplet of elements clockwise starting with the minor diagonal (SW-NE)l, then moving to the next diagonal and so on. Subtract each triplet.*

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = -a_{31}a_{22}a_{13}-a_{32}a_{23}a_{11}-a_{33}a_{12}a_{21}$$

• *Add the two results together to calculate the determinant*

$$|\mathbf{A}| = (+a_{11}a_{12}a_{13}+a_{21}a_{32}a_{13}+a_{31}a_{23}a_{12}) - (a_{31}a_{22}a_{13} + a_{32}a_{23}a_{11} + a_{33}a_{12}a_{21})$$

A numerical example is the 3x3 matrix

$$|\mathbf{A}| = \begin{bmatrix} 3 & 1 & 5 \\ 2 & 4 & 6 \\ 1 & 7 & 10 \end{bmatrix} = 120+70+6-20-39-20 = 117$$

Matrix determinants are extremely important with many applications in physics. We will soon see some uses.

Matrix Transpose

The transpose of a matrix, $\mathbf{A}^{\mathbf{T}}$, is easily derived by pivoting columns to become rows: the first column becomes the first row, the second column becomes the second row, and so on. Thus element a_{ij} becomes element a_{ji}.

(LA.20) *If* $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{23} & a_{33} \end{bmatrix}$ *then* $\mathbf{A}^{\mathbf{T}} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$

Matrix Adjoint

The adjoint of a matrix, denoted $\mathbf{A}^{+}$, is the transposed matrix of cofactors. First, form the matrix of cofactors **C**, then transpose it to get

$$\text{(LA.21)} \quad \mathbf{A}^{+} = \mathbf{C}^{\mathrm{T}} = \begin{bmatrix} C_{11} & -C_{12} & C_{13} \\ -C_{21} & C_{22} & -C_{23} \\ -C_{31} & -C_{32} & C_{33} \end{bmatrix}^{\mathrm{T}}$$

$$= \begin{bmatrix} C_{11} & -C_{21} & -C_{31} \\ -C_{12} & C_{22} & -C_{32} \\ C_{13} & -C_{23} & C_{33} \end{bmatrix}$$

A matrix is said to be *self-adjoint* if $\mathbf{A} = \mathbf{A}^{+}$. We will see later that *Hermitian matrices* are an important example of self-adjoint matrices in a complex vector (Hilbert) space.

Matrix Inverse

We've just completed all the pieces to the grand finale: inverting a matrix. A matrix inverse, denoted $\mathbf{A}^{-1}$, is a matrix for which $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$, where **I** is the *identity matrix* or *unit matrix* (a diagonal matrix with 1's on the principal diagonal and zeros elsewhere. Thus

$$\text{(LA.22)} \qquad \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Is a 3x3 identity matrix. Note that all identity matrices are square.

Computing an inverse matrix is straightforward now that we have all the pieces. The inverse matrix is the adjoint of the matrix divided by the determinant of the matrix, or

(LA.23) $$\mathbf{A}^{-1} = \mathbf{A}^{+}/|\mathbf{A}|$$

Note that the matrix can *not* be inverted if $|\mathbf{A}| = 0$. In this case some columns or rows of **A** are linearly dependent, meaning that there are redundant columns or rows. This brings us to the topic of the rank of a matrix.

Matrix Rank

We have seen that each matrix has a dimension—the number of rows times the number of columns, say NxM. The dimension is easily discovered simply by inspecting the matrix.

A more subtle characteristic is the matrix *rank*. The rank measures how much independent information the matrix contains. Consider our square NxN matrix. If all the rows and vectors are *linearly independent* the matrix has full rank, that is, rank N. In a matrix with full rank, there are N^2 pieces of information.[1]

But now suppose that one row is linearly related to another row. For example, each element in row 1, $\boldsymbol{a}_{1j}$, is related to the corresponding element in the 3rd row by the equation $\boldsymbol{a}_{1j} = a + \boldsymbol{a}_{3j}$. This means that one of those two rows carries exactly the same information (up to a linear transformation) as the other row—there is a redundant (informationless) row. In this case the matrix is not of full rank: its rank is N-1. There can be more than one linearly dependent row or column in a matrix: if k rows or columns are linearly dependent, the matrix rank is N – k.

When a matrix is not of full rank it means that some of its information is useless because it replicates other information. This means that we overstate the amount of datain matrix with less that full rank. To correct this, one needs to investigate the sources of linear dependency and eliminate the excess rows or columns, reducing the matrix dimensions to a size that has full information.

[1] If the matrix is NxM (not square), its maximum rank is the lesser of M or N.

Fortunately there are signs to warn the data user. If the determinant of an NxN matrix is zero, it has rank < N; that is a slam dunk test. But how does one know how much redundancy there is in the data? One way is to compute the *eigenvalues* of the matrix: the rank of a matrix is equal to the number of *non*zero eigenvalues. So if a square NxN mtrix has N-k nonzero eigenvalues, you know that there are k redundant rows or columns. Note that *the number of redundant rows is always equal to the number of redundant columns*, so we can use either columns or rows to examine linear dependence within a matrix.

Which brings us to eigenvalues…

Eigenvalues and Eigenvectors

Our last linear algebra topic is eigenvectors and eigenvalues. These have many applications in the physical sciences.

Eigenvalues emerge when we want to take a general matrix **A** and diagonalize it so that all information is in the principle diagonal, and zeros are in the off-diagonal positions. The numbers on the principle diagonal are the eigenvalues of the matrix and the columns of the matrix are the eigenvectors

Suppose we have our 3x3 matrix **A** and we want to diagonalize it. This means solving the linear equations

(LA.24) $$\mathbf{A}|x\rangle = \lambda\mathbf{I}|x\rangle$$

where **A** is an NxN matrix, $|x\rangle$ is an Nx1 column vector, **I** is the NxN identity matrix and λ is a scalar. The vector $|x\rangle$ is called an *eigenvector* of the matrix and the scalar λ is an *eigenvalue* of the matrix. The task is to take a known matrix A and to find its eigenvalue(s) and eigenvector(s).

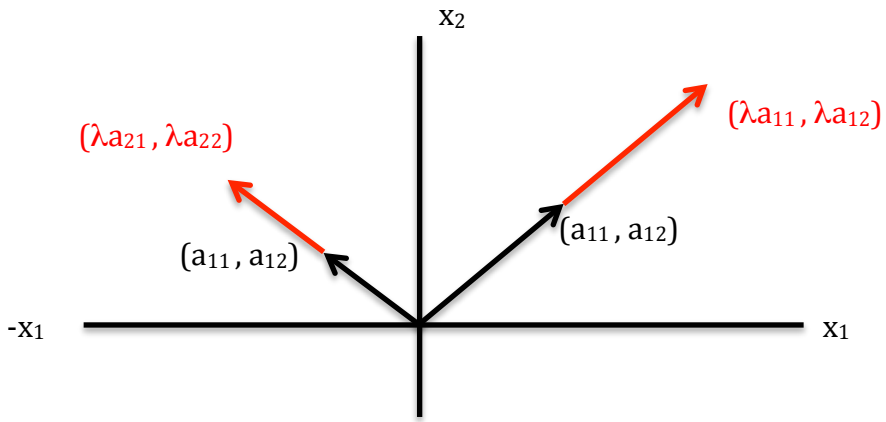
What is this equation system doing? If **A** is a 2x2 matrix, LA.24 It takes its two vectors and extends the length of each by the same proportion without

changing the directions. To see this, suppose that A is a 2x2 matrix and $|x\rangle$ is a 2x vector, so the equation system is

(LA. 25) $$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} |x\rangle = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} |x\rangle$$

that is, $$a_{11}x_1 + a_{12}x_2 = \lambda x_1$$
$$a_{21}x_1 + a_{22}x_2 = \lambda x_2$$

The result is shown below. The original vectors (black) are each extended in length by a factor of 2, marked in red (they could also be reversed in direction, pointing SW and SE).



Let's continue with our two equation system. The equations can also be written (using $|0\rangle$ as a column vector of zeros)

(LA.26) $$(\mathbf{A} - \lambda\mathbf{I})|x\rangle = |0\rangle$$

that is, $$(a_{11} - \lambda)x_1 + a_{12}x_2 = 0$$
$$a_{21}x_1 + (a_{22}-\lambda)x_2 = 0$$

To solve for the eigenvector, invert $(\mathbf{A} - \lambda\mathbf{I})$ and solve $|x\rangle = (\mathbf{A} - \lambda\mathbf{I})^{-1}|0\rangle$. Ooops! That gives the trivial solution $|x\rangle = |0\rangle$—the eigenvector is all zeros—but we know that the eigenvectors are generally not all zeros! This makes no sense until you realize that it is telling you something important—that the system makes

sense only when matrix (**A** - λ**I**) can *not* be inverted, that is, it does *not* have full rank!

Recall that when a matrix does not have full rank it's determinant must be zero. Thus

(LA.27) $$|(\mathbf{A} - \lambda\mathbf{I})| = 0 \Rightarrow (a_{11} - \lambda)(a_{22}-\lambda) - a_{12}\, a_{21} = 0$$
$$\Rightarrow \lambda^2 - (a_{11}+a_{22})\lambda + (a_{11}a_{22} - a_{12}\, a_{21}) = 0$$

This last equation is called the *characteristic polynomial of **A*** and it can be solved for the eigenvalue. It turns out that there will be two roots to this quadratic equation; each root is an eigenvalue of **A**. In general, if **A** is an NxN matrix not of full rank there will be an N^{th} order characteristic polynomial equation in λ with N roots (eigenvalues): some of these eigenvalues will be zero (when rows are linearly dependent), others may be duplicates, and still others will be unique roots; but there will be N roots.

Consider the very simplest case: $a_{12}= a_{21} = 0$. The matrix (**A** - λ**I**) is then diagonal and simple inspection reveals that *the two eigenvalues must be $\lambda_1= a_{11}$ and $\lambda_2 = a_{22}$.* In the more general case the characteristic polynomial is a quadratic equation be solved using . The two eigenvalues are

(LA.28) $$\lambda_1, \lambda_2 = \tfrac{1}{2}\{-(a_{11}+a_{22}) \pm \sqrt{[(a_{11}+a_{22})^2 - 4(a_{11}a_{22} - a_{12}\, a_{21})}\}$$

Manual solution is extraordinarily tedious for the very general case of an NxN matrix, but fortunately we have computers to do the scut work. But however large the matrix, the method is essentially the same as shown for a 2x2 matrix.

There are many uses of eigenvalues. One is in determining the rank of the matrix **A**. It turns out that *the number of redundant rows or columns is equal to the number of zero eigenvalues*. So if you want to find the rank of an NxN matrix, compute the N eigenvalues and count the number that are zero. If that number is k the rank of **A** is N – k.

An important use in quantum mechanics is in the interpretation of the amplitudes in a superposition of quantum states: The eigenvalues of the

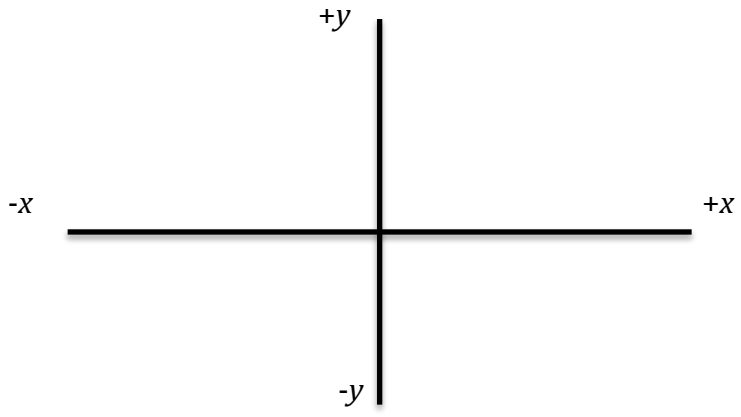
Schroedinger probability wave's Hamiltonian matrix are the amplitudes of the quantum system's superposition.

What about the eigen*vectors*? These can also be found, but we forego that pleasure.

Basis Vectors and Vector Spaces: $\mathbb{R}$-Space

Suppose we have two 2x1 vectors [1 0] and [0 1] (written as rows to conserve space). These two vectors have a zero dot product and are, therefore, orthogonal. These are the *basis vectors* for a two-dimensional Cartesian[2] two-dimensional vector space, denoted an $\mathbb{R}^2$ vector space. The axes of this space are at 90° angles with each other—movement along one axis does not necessarily change position on the other axis. If a third dimension exists, there are three orthogonal basis vectors [1 0 0], [0 1 0], and [0 0 1] required to form the three-dimensional $\mathbb{R}^3$ vector space.

An $\mathbb{R}^2$ vector space is shown below. This is the familiar graph that you encountered even before high school. It is formed by the two basis vectors [1 0] and [0 1] where we call the axes *x* and *y*, denoting each point in the space as (*x*, *y*).

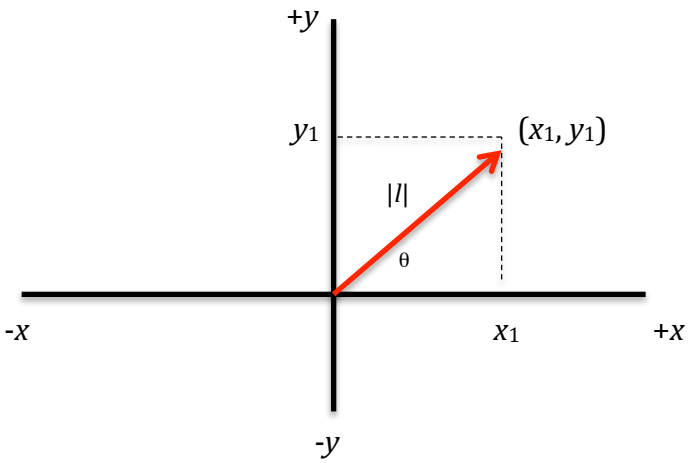


Two-Dimensional Vector Space

[2] Cartesian vector spaces are named after Rene Descartes, the 17th century French philosopher and mathematician who invented analytic geometry (the marriage of algebra and Euclidean geometry). Descartes was the first to use vector spaces (his were derived to apply to spaces consistent with the axioms of Euclidean geometry).

Suppose we stick to two dimensions and focus on the point (x_1, y_1) on the graph below. The red arrow is the vector describing that point (yes, vectors have a visual representation). Let's look at its parts.

The red arrow is the equation $y = ax$—a straight line with slope a passing through the center (origin)—that terminates at (x_1, y_1). It can be represented either as an arrow, as shown above, or as a point like (x_1, y_1). That vector has two characteristics: *magnitude* and *direction*. Its magnitude, denoted $|l|$, also called its *norm*, is the vector's length. By the Pythagorean Theorem, $|l| = \sqrt{(x_1^2 + y_1^2)}$. Its direction is the angle that it makes with the horizontal axis, $\theta = \arctan(y_1/x_1)$.



A Vector in $\mathbb{R}^2$

This vector could be derived for any number of dimensions, though we can't draw a figure in more than three dimensions. In three dimensions it would start at the origin and go straight to the point (x_1, y_1, z_1). It's magnitude would be $|l| = \sqrt{(x_1^2 + y_1^2 + z_1^2)}$ and its direction would be described by two angles, one between the x and z axes, the other between the y and z axes.

Combining Vector Spaces

Suppose we have two separate three-dimensional vector spaces, $\mathbb{R}^3_1$ and $\mathbb{R}^3_2$. Each has three basis vectors, arranged below as 3x3 matrices. Each basis

matrix has a 1 as the element on the principle diagonal and zeros elsewhere. Such a matrix is called a *unit matrix*.

(LA.29)

$$\mathbb{R}^3_1 \qquad\qquad \mathbb{R}^3_2$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad\qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Because these are separate vector spaces, points in one do not translate to points in the other: a space traveller in $\mathbb{R}^3_1$ can report his galactic position to anyone in his space, but an observer in $\mathbb{R}^3_2$ can make no sense of it. Suppose we want to combine them into a single space, $\mathbb{R}^3_3$ so that an observer in one space can make his location known to observers in the other space. Is there a way of combining the two vector spaces?

Well, yes there is, and in this case it is straightforward. It is done by forming the *direct sum* of $\mathbb{R}^3_1$ and $\mathbb{R}^3_2$. The direct sum of two matrices, denoted by the symbol $\oplus$ and written $\mathbb{R}^3_1 + \mathbb{R}^3_2$, creates a *block diagonal matrix* with $\mathbb{R}^3_1$ in the northwest quadrant, $\mathbb{R}^3_2$ in the southeast quadrant, and zeros in the northeast and southwest quadrants, as below (the dashed lines delineate the four blocks in the matrix).

(LA.30)

$$\mathbb{R}^3_3$$

$$\mathbb{R}^3_1 \oplus \mathbb{R}^3_2 = \left[\begin{array}{ccc:ccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

Like its component matrices, $\mathbb{R}^3_3$ is a unit matrix. Its columns are the basis vectors for, respectively, the variables *x, y, z, u, v,* and *w.*

Using this method vector spaces of any size can be combined into a single vector space. This operation is rarely done in Cartesian vector spaces in the

physical sciences the combination of *Hilbert vector spaces* is a standard procedure. We have gone through this process to set a foundation for our discussion of Hilbert Spaces.

Hilbert Spaces, Observables, and Operators

When we reviewed linear algebra we used Cartesian vector space in which each point is measured as a real number, basis vectors are orthogonal (at right angles), and there are a finite number of dimensions. This is sufficient to understand the basics of quantum mechanics, but not sufficient to *do* quantum mechanics. Quantum particles live in a very different world, a world called *Hilbert Space*, denoted as $\mathbb{H}^N$ and named after David Hilbert, an early 20th century mathematician.

Hilbert Spaces

It turns out that a Cartesian vector space satisfies the requirements of a Hilbert Space: All Cartesian Spaces are Hilbert Spaces, but not all Hilbert Spaces are Cartesian Spaces. Some major points of similarity or dissimilarity between Cartesian and Hilbert spaces are

- ***Basis Vectors:*** in both Cartesian and Hilbert spaces the basis vectors are orthogonal (at 90° to each other).
- ***Dimensionality***: Cartesian spaces have a finite number of dimensions while Hilbert Spaces typically have a (countably) infinite number of dimensions, though finite dimensions are sometimes used for expository purposes. Cartesian spaces are usually analyzed with discrete differences or differential/integral calculus. But because Hilbert spaces have an infinite number of dimensions they are analyzed using infinitesimal methods like integral and differential calculus and partial differential equations.
- ***Content***: The content of Cartesian space is the set of real numbers; the content of Hilbert space can be real numbers, complex numbers, or even functions. This makes Hilbert spaces suitable for analyzing a broad range physical phenomena,

particularly those containing cycles. For example, the harmonics introduced by standing waves of violin string vibrations, and the evolution of Schroedinger probability waves, are examined using spaces of complex numbers, i.e. Hilbert spaces.

- ***Inner Products***: Both Cartesian and Hilbert spaces are "inner product spaces," meaning that the inner product (dot product) between two points measures the distance between two points. Thus, the *distance metric* for (x_1, y_1) and (x_2, y_2) is $\sqrt{[(x_2 - x_1)^2 + (y_2 - x_1)^2]}$ in both Cartesian and Hilbert spaces.
- ***Completeness***: Both Cartesian and Hilbert spaces are "complete" in the mathematical sense that sequences like sums converge "in the limit." This means that sums of infinite series and integration over infinite values are both meaningful operations.
- **Combining Spaces:** Cartesian spaces are *additive* spaces, combined using the direct sum of the basis matrices. Thus, two 3-dimensional Cartesian spaces, each with a 3x3 identity matrix as the basis matrix, are combined to create one 6-dimensional Cartesian space with a 6x6 identity matrix as the basis matrix (see LA.30). Hilbert spaces are *product* spaces combines as tensor products, so two 3x3 Hilbert spaces, $\mathbb{H}^3_1$ and $\mathbb{H}^3_2$, each with a 3x3 basis matrix, becomes a 9x9 basis matrix, as below:

$$\mathbb{H}^3_1 \otimes \mathbb{H}^3_2 = I_3 \otimes I_3 = I_9 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Note that both combinations of Cartesian spaces and combinations of Hilbert spaces result in identity matrices, but the resulting vector space dimensions are different: a combination of an MxM Cartesian space with an NxN Cartesian Space is an (M+N)x(M+N) Cartesian space, but a combination of an MxM Hilbert space with an NxN Hilbert Space is an MNxMN Hilbert Space.

"Observables"

An *observable* in quantum mechanics, also called an *observable operator*, is a function that acts on a quantum matrix to give a measurement of physical (observable) characteristics of the system, like energy, momentum, position, angular momentum, and so on.

Observables are *Hermitian Matrices*, with several properties:

- They are self-adjoint, that is $\mathbf{A} = \mathbf{A}^+$
- They have real eigenvalues, interpreted as the amplitudes of each state
- For each eigenvalue there is at least one eigenvector
- The eigenvectors form an orthonormal basis for an eigensystem—a Hilbert space incorporating the eigenvalues.

Perhaps the most prominent observables are those giving rise to the system's *Hamiltonian*, which measures the energy of the quantum system. The Hamiltonian for a single particle in one dimension (x) has the following three observables:

The Hamiltonian of a Quantum Particle

$$H = \frac{p^2}{m} + V(m)$$

$$p = -i\hbar \frac{\partial}{\partial x}$$

$$x = -i\hbar \frac{\partial}{\partial p}$$

H is the Hamiltonian defining total energy, V is potential energy, *m* is mass, *x* is position, and *p* is momentum. $\hbar$ is the reduced Planck Constant ($\hbar = h/2\pi$).

The momentum observable $p = -i\hbar \frac{\partial}{\partial x}$ and the position operator $x = -i\hbar \frac{\partial}{\partial p}$ contain the symbols $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial p}$. These are the partial derivatives with respect to position and momentum of the elements in matrix H. This is why observables are also called observable operators—the do not stand alone; rather, they operate on a vector or matrix. These three inter-related observables describe *Schroedinger's Wave Equation* for the special case of a single particle, one spatial dimension and time-independence.

The Hamiltonian and the momentum and position observables describe the energy of the particle: H is the total energy, V is the potential energy, and $\frac{p^2}{m}$ is the particles kinectic energy. Note that in the case of a particle traveling at the speed of light, $p^2 = (mc)^2$ so the Hamiltonian is $H = mc^2 + V$ where H is total energy, mc^2 is Einstein's kinectic energy ($E = mc^2$) and V is the particle's potential energy.

"Operators"

As noted repeatedly above, quantum superpositions can be manipulated by linear matrices called *operators*. Operators are not associated with measurements;

rather, they transform the quantum state superposition without measurement that would create decoherence (wave collapse).

In quantum mechanics operators are *unitary matrices:* a *unitary matrix* (**U**) is a matrix for which the product of the matrix and its adjoint is an identity matrix, that is, $\mathbf{U}\mathbf{U}^{+} = \mathbf{I}$. The effect of a unitary operator is to rotate the quantum system's vectors in Hilbert space. Important examples are the three *Pauli Matrices* that rotate spin along the z, x, and y axes. The Pauli matrices are both Hermitian and Unitary, hence the square of each is an identity matrix ($\mathbf{U}^2 = \mathbf{I}$).

Pauli "Spin" Matrices

$$\sigma_z = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \sigma_x = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \qquad \sigma_y = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Other important unitary matrices are those that create quantum logic gates. For example, the "Controlled-Not" matrix (C-Not), shown below for a three-qubit system, reverses the second qubit if the first qubit is 1. The result is the *mod 2* addition $q_1 \oplus q_2 \Rightarrow q_3$, or

C-NOT

$\mathbf{q_1}$	$\mathbf{q_2}$		$\mathbf{q_3}$
0	0	$\Rightarrow$	0
0	1	$\Rightarrow$	1
1	0	$\Rightarrow$	1
1	1	$\Rightarrow$	0

The C-NOT writes the *modulo 2* sum of the q_1 and q_2 qubits to the q_3 qubit. Thus, q_3 contains $q_1 \oplus q_2$ *mod 2*.

The C-NOT gate is implemented by the specific Hermitian unitary matrix

$$\mathbf{U}_{\text{C-NOT}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

and the entire operation is

$$\mathbf{U}_{\text{C-NOT}}\,(\,|0>\otimes|1>) \;=\; \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \end{matrix} \begin{matrix} \mathbf{q_2} \\ \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \end{matrix} = \begin{matrix} \mathbf{q_3} \\ \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \end{matrix}$$

Another important unitary matrix is the *Hadamard transform*, used to create a Hadamard logic gate. This operates on a single qubit using the following Hadamard matrix:

$$\mathbf{U}_{\text{H}} = \sqrt{½}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

The Hadamard transform on the superposition (|0 > + |1 >) is

$$\mathbf{U}_{\text{H}}(|0> + |1>) \;= \sqrt{½}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}(|0> + |1>)$$

$$= \sqrt{½}(|0> + |1>) + \sqrt{½}(|0> - |1>)$$

Yet another operator is the phase shift matrix $\mathbf{U}_{\text{PH}}$:

$$\mathbf{U}_{\text{PH}} = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{bmatrix}$$

used to shift the phase of complex numbers. A phase-shift transformation is

$$\mathbf{U}_{\text{PH}}(|0> + |1>) \;= |0> + e^{i\phi}|1>$$

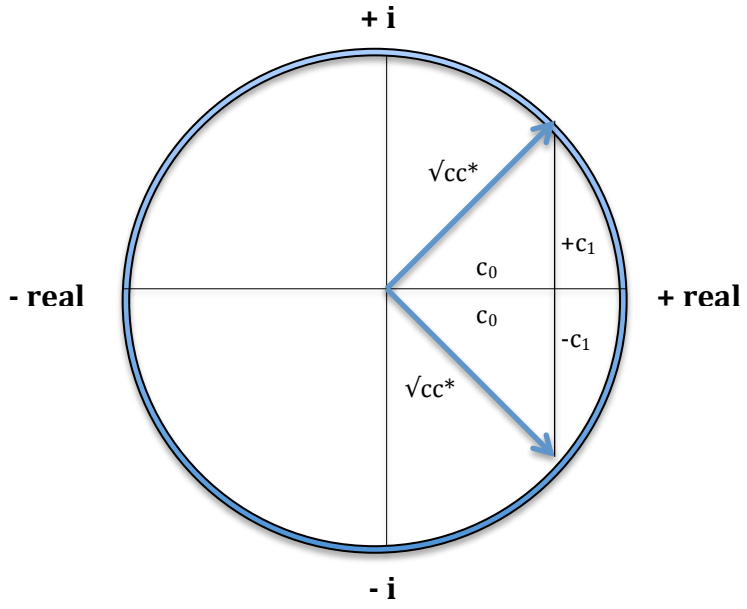
resulting in a complex number with the |1 > state vector phase-shifted relative to the |0 > state vector. If, for example, $|0> = \begin{bmatrix} \cos(\theta) \\ 0 \end{bmatrix}$ and $|1> = \begin{bmatrix} 0 \\ \sin(\theta) \end{bmatrix}$, we have the familiar complex number $\cos(\theta) + e^{i\phi}\sin(\theta)$.

Complex Numbers

At several points we've noted that quantum mechanics makes liberal use of complex numbers: the amplitudes in quantum superpositions are often complex numbers, and the squares of those complex numbers are the probabilities attached to the associated quantum state; the Feynman clocks used to describe QED are complex numbers; the probability waves that suffuse quantum amplitudes are derived from complex numbers. We have been able to convey the spirit of quantum mechanics without direct resort to complex numbers, but in this appendix we outline their nature and their implications.

A complex number is a number that consists of a real part and an imaginary part, say $c = c_0 + c_1 i$, with c_0 the real part and $c_1 i$ the imaginary part. The imaginary part is the product of a real number, c_1, and the imaginary number i ($i = \sqrt{-1}$). For each complex number there is a *conjugate complex* $c^* = c_0 - c_1 i$.

(CN.1)



Complex Number and its Complex Conjugate

The figure above shows a complex number and its conjugate as hands on a clock. The complex number is the arrow (vector) sloping upward at angle θ; the complex conjugate slopes downward at angle -θ. By the Pythagorean Theorem we know that the length of the clock hand is $|c| = \sqrt{(c_0^2 + c_1^2)}$; |c| is called the *modulus* of the complex number. The modulus is the square root of the product of a complex number and its conjugate, i.e., $\sqrt{cc^*} = \sqrt{(c_0^2 + c_1^2)}$ is the length of the arrow. Note that all angles are in measured in radians, r, with $r = 2\pi(\theta°/360°)$ being the conversion between degrees and radians.

Thus, the clock above shows all there is to know about a single complex number: the length of the hand, |c|, is the amplitude of the associated quantum event in a superposition, and $|c|^2$ is the probability of that quantum state. The angle of the hand, called the *argument* of the complex number, is the *phase* of the cycle: an angle of 0° (r = 0 radians) is the start of a cycle, an angle of 90° ($r = \pi/2$ radians) is ¼ of the way through a cycle, an angle of 180° ($r = \pi$ radians) is ½ of the way through a cycle, and an angle of 270° ($r = 3\pi/2$ radians) is ¾ of the way through a cycle, and an angle of 360° (r = 0 radians) returns the cycle to its original starting point. ℍ

As a cycle proceeds, the complex number hand sweeps counter-clockwise on the clock, starting at the 3 o'clock position (θ = 0), while the complex conjugate's hand sweeps clockwise. Because all of the information in the one is shown in the other, we can dispense with consideration of the complex conjugate and focus on the complex number.

The complex number can be written in other forms, which are often easier to manipulate. This is shown in the box below. The first alternative form—the *trigonometric form*—converts the complex number to a sine function using simple trigonometry of the clock figure above. The second form—the *polar form*—converts the complex number to the polar coordinates |c| and θ. All three forms are equivalent, so the one used is a matter of convenience.

(CN.2)

Alternative Forms of Complex Numbers

A complex number and its conjugate can be written as

Trigonometric Form

- $c_0 + c_1 i = |c|\{\cos(\theta) + i\sin(\theta)\}$

  $c_0 - c_1 i = |c|\{\cos(\theta) - i\sin(\theta)\}$

Polar Form

- $c_0 + c_1 i = |c| \cdot \exp\{+i\theta\}$

  $c_0 - c_1 i = |c| \cdot \exp\{-i\theta\}$

———

Note: $|c|$ is the *modulus* ($|c| = \sqrt{cc^*}$), and $\exp\{i\theta\}$ is the Euler constant ($\exp = 2.7128\ldots$) raised to the power $i\theta$.

The phase of the complex number, θ, can be calculated as the arctangent of c_1/c_0, that is, the angle for which c_1/c_0 is the tangent. This requires an adjustment for the quadrant that the clock hand is in. The adjustments are shown below.

Phase Angle of a Complex Number

$\theta = \arctan(c_1/c_0)$	if $c_0 > 0$ and $c_1 > 0$	(Quadrant I)
$\theta = \pi/2$	if $c_0 = 0$ and $c_1 > 0$	
$\theta = \arctan(c_1/c_0) + \pi/2$	if $c_0 < 0$ and $c_1 > 0$	(Quadrant II)
$\theta = \arctan(c_1/c_0) - \pi/2$	if $c_0 < 0$ and $c_1 < 0$	(Quadrant III)
$\theta = \arctan(c_1/c_0) - \pi/2$	if $c_0 > 0$ and $c_1 < 0$	(Quadrant IV)

The arithmetic of complex numbers is fairly straightforward. Suppose we have two complex numbers, $c_0 + c_1 i$ and $d_0 + d_1 i$, with phases θ_c and θ_d respectively. The modulus of the numbers is |c| and |d|. The following rules apply

(CN.3)

Complex Number Arithmetic

Addition and Subtraction

- $(c_0 + c_1 i) + (d_0 + d_1 i) = (c_0 + d_0) + (c_1 + d_1)i$
  $= |c+d|\{\cos(\theta_{c+d}) + i\sin(\theta_{c+d})\}$
  $= |c+d|\exp(i\theta_{c+d})$
- $(c_0 + c_1 i) - (d_0 + d_1 i) = (c_0 - d_0) + (c_1 - d_1)i$
  $= |c-d|\{\cos(\theta_{c-d}) + i\sin(\theta_{c-d})\}$
  $= |c-d|\exp(i\theta_{c-d})$

Multiplication and Division

- $(c_0 + c_1 i)(d_0 + d_1 i) = (c_0 d_0 - c_1 d_1) + (c_0 d_1 + c_1 d_0)i$
  $= |c||d|\{\cos(\theta_c + \theta_d) + i\sin(\theta_c + \theta_d)\}$
  $= |c||d|\exp\{i((\theta_c + \theta_d)\}$
- $(c_0 + c_1 i)/(d_0 + d_1 i) = [(c_0 d_0 + c_1 d_1) + (c_1 d_0 - c_0 d_1)i]/|d|^2$
  $= (|c|/|d|)\{\cos(\theta_c - \theta_d) + i\sin(\cos(\theta_c - \theta_d)\}$
  $= (|c|/|d|)\exp\{i((\theta_c - \theta_d)\}$

Exponentiation

- $(c_0 + c_1 i)^n = |c|^n\{\cos(n\theta_c) + i\sin(n\cos(\theta_c)\}$
  $(|c|^n)\exp\{in(\theta_c)\}$

Note: $|c+d| = \sqrt{(c_0 + d_0)^2 + (c_1 + d_1)^2}$ $\theta_{c+d} = \arctan\{(c_1 + d_1)/(c_0 + d_0)\}$

Phase Shifts and Imaginary Numbers

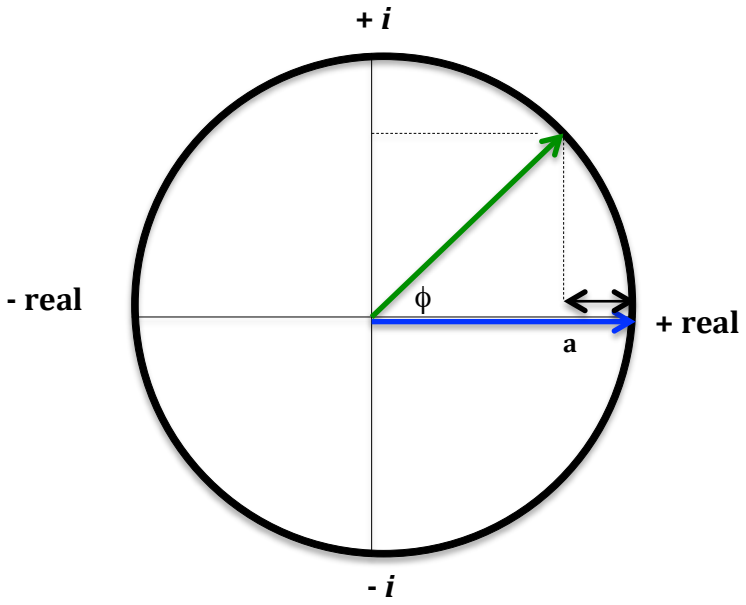
We have seen that a complex number can be written as |c|[cos (θ) + *i*sin(θ)] with real part cos (θ) and imaginary part *i*sin(θ). The evaluation of the imaginary part is plagued by mystery, so here we will try to demystify it a bit.

What role does *i* play in this function? Suppose we rewrite the function describing the cycle as |c|[cos (θ) + $e^{i\phi}$sin(θ)]sin(θ)] with angle ϕ indicating a phase shift and angle θ (as before) indicating the position in the general cycle: θ = 0 is a new cycle beginning, θ = 90° is halfway through a cycle, and so on. It turns out that when ϕ = 90 then $e^{i\phi} = i$. Thus, the complex number |c|[cos (θ) + *i*sin(θ)] describes a cycle with a 90° phase shift. The phase shift alters the value attached to the sine part of the number relative to the cosine part.

Let's pursue this a bit further. Consider the complex number $e^{i\phi}$sin(θ) where ϕ is a phase angle. A unit circle (radius = 1) is shown below. The blue vector represents the contribution of the sine portion when θ = 0 and ϕ =0, that is, the cycle is just beginning and there is no phase shift. Because e^{i0}= 1, sin(0) = 0, and cos(0) = 1, that point is (0, 1) its blue vector's end point is at (01, 0).

Now introduce a phase shift of ϕ = π/4 radians (45°) while θ remains at zero. The phase shift rotates the vector counterclockwise by 45° to the position of the green vector. The *real* part of the complex number is now **a**, which is less than 1.

(CN.4)



The phase shift has reduced the real value of the complex number even though the position in the cycle (measured by θ) has not changed. We can proceed in this fashion to determine the effect on the real value at other phase shifts: when $\phi = \pi/2$ radians (90°) the real part is zero; when $\phi = \pi$ radians (180°) the real part is -1; when $\phi = 3\pi/2$ radians (270°), the real part disappears again; when $\phi = 2\pi$ radians (360°) the real part is 1.

Thus, a phase shift will influence the amplitude of the real part of the complex number at each stage in the sine function's cycle. This is important because *only the real part of the complex number can be measured*. So by affecting the real part of the complex number at each θ phase shifts affect the size of the sine function's influence on observed cyclical phenomena.

As noted above, the amount of phase shift has a specific relationship with the imaginary number *i*. That relationship is shown in the table below. Thus, the effect of phase shifts is to cause the real value to "skip" to another place in the cycle, just as a scratch on an old phonograph record causes the needle to skip to another position.

Effect of Phase Shift on Cos (θ) + $e^{i\phi}$Sin(θ)

Phase Shift Angle (ϕ) *degrees*	*radians*	**Phase Shift** $e^{i\phi}$	**Complex Number** **
0°	0	1	**cos (θ) + sin(θ)**
45°	π/4	$e^{i\pi/4)}$	**[cos (θ) + sin(θ)]** + *isin(θ)*
90°	π/2	$e^{i\pi/2)}$	**cos (θ)** + *i*sin(θ)
135°	3π/4	$e^{i3\pi/4)}$	**[cos (θ) - sin(θ)]** + *isin(θ)*
180°	π	$e^{i\pi}$	**cos (θ) - sin(θ)**
225°	5π/4	$e^{i5\pi/4)}$	**[cos (θ) - sin(θ)]** + *isin(θ)*
270°	3π/2	$e^{i3\pi/2)}$	**cos (θ)** - *isin(θ)*
315°	7π/4	$e^{i7\pi/4)}$	**[cos (θ) + sin(θ)]** - *isin(θ)*
360°	2π	$e^{i2\pi}$	**cos (θ) + sin(θ)**

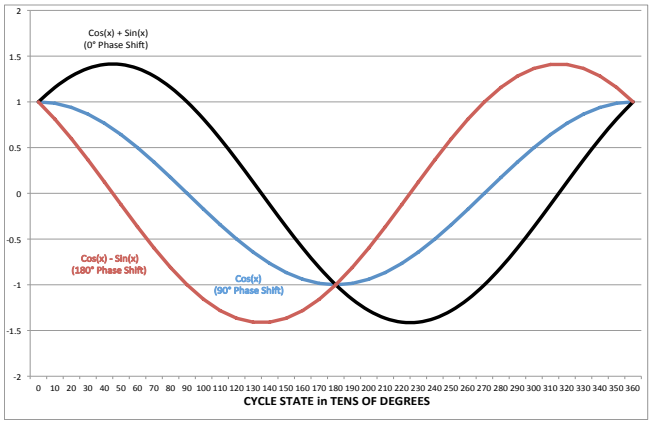
* *e* is Euler's constant, *e* = 2.7183, and $e^{i\pi} = -1$.

** The real part of the number is shown in bold font

The figure below shows the value of the real part of the complex number over a full cycle.

The Real Part of Cos(θ) + $e^{i\phi}$Sin(θ)

At Different Phase Shifts



The horizontal axis is the stage of the cycle as measured by θ, and the vertical axis is the amplitude of the function at each θ. Each of the three wavy lines is for different phase shifts: the black line is the value of cos(θ) + sin(θ), which is the real value of the function when the phase shift is ϕ = 0°, 45°, or 315°; the blue line is the value of cos(θ) which describes a phase shift of ϕ = 90° or 270°; the red line is cos(θ) + sin(θ), describing a phase shift of ϕ = 135°, 180°, or 225°.

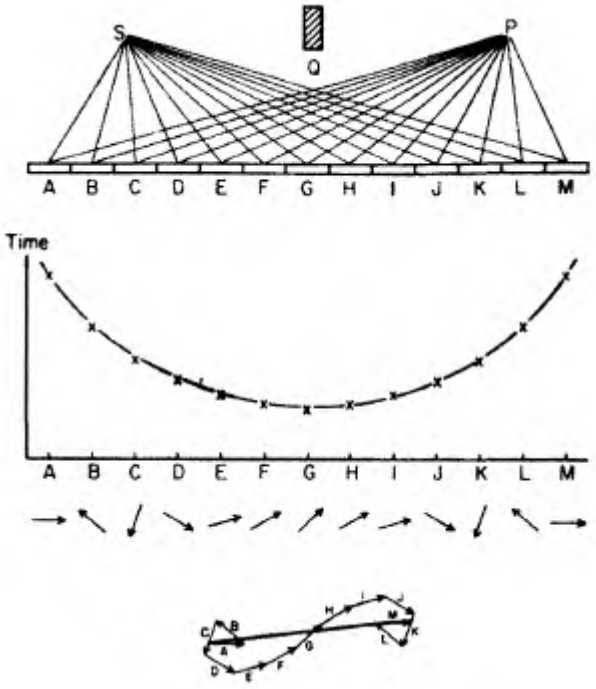
Feynman Clocks as Complex Numbers

Richard Feynman invented the Feynman Clock as a metaphor for the complex mathematics underlying quantum mechanics. The clock hand has two parts: its length, or modulus, measures the amplitude of a particle's wave function; its argument the position in the wave cycle. The motion of quantum particles can be seen as a sequence of Feynman clocks.

In Part 2 we discussed Feynman's example of His Sum Over Histories approach to quantum electrodynamics. We summarize that section below, then discuss its relationship to complex numbers.

The Sum Over Histories approach says that all possible outcomes in a quantum superposition occur simultaneously until a measurement is made. Feynman asks what this implies for the path taken by a photon of light. The textbook answer is high school physics is that (a) light always takes a straight-line path between point A and point B because that is the shortest and fastest route, or (b) light always reflects at the angle of incidence, which is another way of saying the same thing.

Feynman's theory of quantum electrodynamics rejects this view: a photon takes *every possible path* between two points, even the strangest of paths is taken, as when a flashlight's rays go around the moon to shine on the object in front of you. But each path has a different amplitude and, therefore, a different probability—the around-the moon path *can* occur, but it is highly improbable; the shortest path *can* occur and it is much more probable.



To demonstrate this Feynman used the figure reproduced above to show that *light does not travel a path that gives the shortest time—a straight line; it takes an infinite number of paths (All Histories), some of them very strange. But the strange paths happen infrequently, and the most probable paths are those closest to least-time paths*. That figure shows thirteen possible paths (A – M) of photons of a specific frequency (color) that travel from point S (a photon gun) to point P (a photon detector). Many photons will not be detected at all—their paths simply pass by the detector—but that's OK because Feynman's goal is to measure the probability that a photon that *is* detected takes a specific path. So he supposes that any detected photon must take one of the thirteen paths.

Paths A and M are the longest (and strangest) paths, with photons reflecting backwards! Path G is the shortest path. A clock hand is used to show the "time" taken for each path. The clock does not directly measure time (though time can be calculated using the number of wavelengths and the speed of light); it is calibrated to show the number of wavelengths (or cycles) the photon goes through as it travels from photon gun to photon detector—one wavelength is one full rotation of the hand and a partial wavelength is a partial rotation of the hand.

The clock is really a complex number, written $|l|e^{i\theta}$, where θ is the angle in radians and $|l|$ is the length of the clock hand. Each photon wave begins with the clock hand at 3 o'clock (θ = 0°). It rotates counter clockwise as the cycle proceeds, returning to 3 o'clock after a full rotation of 360° or 60 "minutes." The hand's final direction is determined at the instant the photon on that path is detected. That direction (θ) is measured *modulo 60*: the direction at the end of the last cycle is the remainder from the total number of "minutes" traveled divided by 60. For example, if the hand is at 9 o'clock then the photon has 30 "minutes" or ½ wavelength in the last cycle.

To put a fine point on it, suppose that the clock hand rotates for n + x cycles, where x is a partial cycle. Then 60(n + x) "minutes" have passed while the photon

goes from the emitter to the detector. The number of full cycles is 60n and the remainder—the partial last cycle—is 60x minutes. If 1,830 "minutes," have passed 30.5 cycles have completed (1830 mod 60 is 30.5) so ½ (or 30 "minutes") of the last cycle have been completed. The clock started at 3 o'clock so it must now point to 9 o'clock.[3] Using the same method we can see that one "minute" is equivalent to $\pi/30$ radians. Fifteen minutes is $\frac{1}{2}\pi$ so a counterclockwise rotation of the hand by 15 minutes sets the clock at $e^{i\pi/2}$. This is ¼ of a rotation, or 12 o'clock.

There is a clock for each possible path and the amplitude of each path, when squared, gives the probability that the photon will take that particular path. The probability that a photon will hit the detector is obtained by adding the clock hands (vectors) together, hence the name "sum of all histories." This is done by laying the hands end to end, each pointing in its final direction: starting with A, attaching the tail of B to the head (arrow of A, attaching the tail of C to the head of B, and so on.

The result is the worm-like pattern of arrows at the bottom of Feynman's diagram. Note that at the extreme paths (A-C and K-M) the vectors tend to point in different directions. This means that they are offsetting—there is "destructive interference"—and it is clear that the net amplitude of paths A-C (and K-M) is very small. But the vectors for the middle paths E-I point in the same general direction, showing "constructive interference." The overall amplitude for a photon reaching the detector is the line obtained by connecting the start of path A's clock hand to the end of path M's hand. The square of that overall amplitude is the probability that a photon will be detected. Virtually all of the probability of detection arises from paths E-I. Thus, light does not travel in a straight line that gives the shortest path—it can take an infinite number of paths—but the path with the highest probability is the shortest path.

Just for fun, let's calculate the hand position for a path on which a photon is detected after traveling ½ meter, or 500 million nanometers (nm). Light photons

[3] Precisely the same result will occur if we use radians rather than minutes. There are 2π radians in a cycle so 121π radians have passed. The remainder *mod* (2π) *is* π--the hand is pointing at 180° from its start, or at 9 o'clock.

have a specific wavelength (λ) and velocity (v), and the frequency of the wave is $f = v/\lambda$. A photon path can be measured in wavelengths. Suppose the photons are red light with wavelength 650 nm (nanometers), velocity at light speed (300 million nm/nsec), and a frequency of 462×10^{14} terahertz (462 trillion cycles per second or 462,000 cycles per nsec). Then the photon travels for 769,230.77 wavelengths (cycles). The remainder—77 percent of a cycle—amounts to 46 "minutes" on the clock, so the hand points toward 6 o'clock when the photon is detected.

In his demonstration Feynman uses the same hand length $|l|$ for all paths. This assumes that the amplitude (hence the probability) of each path is the same though directions will differ. This is a useful assumption for his purposes but only a crude approximation. It turns out that the hand length (the size of the clock} also changes as the photon moves along its path. The change in $|l|$ is proportional to the time taken to complete the path. The exact formula is that $|l|$ increases in direct proportion to the *action* of the particle's motion, defined as: $\Delta|l|/|l| = mx^2/2ht$.

Modular Arithmetic

Modular arithmetic focuses on the remainder left over after a division operation. Consider the division 8/3. Our standard arithmetic would compute this as 2.6667 or $2\frac{2}{3}$: a whole number (integer 2) plus the fraction 2/3. But modular arithmetic would compute this as 2—the remainder after the whole number is discarded.

In modular arithmetic the operation would be described as "$8 \equiv 2 \bmod(3)$," meaning that when 8 is divided by 3 the remainder is 2 (or that both 8 and 2 have the same remainder when divided by 3). The relationship between 8 and 2 is said to be "congruent (the symbol ≡ means "is congruent to," not "is equal to").

Modular arithmetic is not an alien concept; it appears in everyday experience. The most frequently encountered way is when you look at a clock. A 12-hour clock reports the hours as 1, 2, 3, …., 11, 0 (we call the end of the cycle "12" but in fact it is zero because it starts the cycle all over again. Suppose we start our daily cycle at zero on the stroke of midnight. The first 11 hours are 1, 2, …., 11 and then at "12" we start again at zero and go 1, 2, ….11 through the afternoon hours. Thus, each hour reading is **h** = **t** mod(12), with h being the number of elapsed hours in a 24-hour day and **t** being clock time (the elapsed time since the last reset). A 24-hour clock (military time) is simply **t** measured from as **h** = **t**mod(24).

There are two equivalent ways to express the congruence on modular arithmetic:

If **a** ≡ **b** mod(**n**), then

- **a** - **b** is exactly divisible by **n**, leaving remainder zero.
- both **a** and **b**, when divided by **n**, have the same remainder

Of particular interest in binary arithmetic—hence in computer science—is *modulo 2 arithmetic*. Binary numbers have a cycle of two—each bit has two values 0 and 1. Any higher number restarts the cycle for that bit at 0. For example, in the text we saw that simple addition 1 + 1 is zero with a 1 carried over. But in *modulo 2*

addition, denoted by the symbol ⊕, 1 ⊕ 1 = 0 because when added the decimal result is 2 and that is exactly divisible by 2, leaving a remainder of zero; that is, 1 + 1 = 0 mod(2). There is no carryover because each bit position stands on its own. In mod(2) addition the rule is: *if there are an even number of 1's in a column, the sum is 0; if there are an odd number of 1's the sum is 1*. So modulo-2 arithmetic is a natural for binary numbers.

Thus, the simple addition in the text

$$\begin{array}{rl} & 10011110 \quad (= 158) \\ + & \underline{00010001 \quad (= \ 17)} \\ & 10101111 \quad (= 175) \end{array}$$

has the answer 1000111 in ⊕ (mod 2 addition).

Note that just as ⊕ and + are different operations with different results, so mod 2 multiplication (⊗) and standard multiplication (×) also differ. Thus

$$\begin{array}{rl} & 111 \ (= 7) \\ \times & \underline{011} \ (= 3) \\ & 111 \\ & \underline{111} \\ & 10101 \end{array} \qquad\qquad \begin{array}{rl} & 111 \ (= 7) \\ \otimes & \underline{011} \ (= 3) \\ & 111 \\ & \underline{111} \\ = & 1001 \end{array}$$