

Mathematics Review A

I. Matrices

A. Vectors

A vector is an object having n -components

$$\mathbf{x} = (x_1, x_2, \dots, x_n).$$

These components may represent, for example, the cartesian coordinates of a particle (in which case, $n=3$) or the cartesian coordinates of N particles (in which case, $n=3N$). Alternatively, the vector components may have nothing what so ever to do with cartesian or other coordinate-system positions.

The numbers x_i are called the components of the vector \mathbf{x} in the directions of some n elementary unit vectors:

$$\begin{aligned} \mathbf{x} &= x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + x_3 \mathbf{e}_3 + \dots + x_n \mathbf{e}_n \\ &= x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ \vdots \\ \vdots \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \vdots \\ \vdots \\ \vdots \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \vdots \\ \vdots \\ \vdots \end{pmatrix} + \dots + x_n \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ \vdots \\ 1 \end{pmatrix} \end{aligned}$$

The unit vectors \mathbf{e}_i , whose exact definition, meaning and interpretation depend on the particular application at hand, are called basis vectors and form the elements of a basis. They are particularly simple to work with because they are orthogonal. This means that their dot products vanish $\mathbf{e}_i \cdot \mathbf{e}_j = 0$, unless $i = j$. If $i = j$, then the scalar or dot product is unity (it is usually convenient, but not necessary, to use bases that are normalized so $\mathbf{e}_i \cdot \mathbf{e}_i = 1$). The shorthand way of representing this information is to write

$$\mathbf{e}_i \cdot \mathbf{e}_j = \langle \mathbf{e}_i \mathbf{e}_j \rangle = \delta_{ij},$$

where δ_{ij} is called the Kronecker delta function defined by:

$$\delta_{ij} = 0, \text{ if } i \neq j, \text{ and}$$

$$\delta_{ij} = 1 \text{ if } i = j.$$

The above equation for \mathbf{x} provides an example of expressing a vector as a linear combination of other vectors (in this case, the basis vectors). The vector \mathbf{x} is expressed as

a linear combination of the unit vectors \mathbf{e}_i , and the numbers x_i are the coefficients in the linear combination. Another way of writing this within the summation notation is:

$$\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i .$$

The idea of a linear combination is an important idea that will be encountered when we discuss how a matrix operator affects a linear combination of vectors.

B. Products of Matrices and Vectors

If $\underline{\mathbf{M}}$ is an $n \times n$ matrix with elements M_{ij} , (the first subscript specifies the row number and the second subscript specifies the column number), then the product $\underline{\mathbf{M}} \mathbf{x} = \mathbf{y}$ is a vector whose components (when subscripts i, j, k , etc. appear, they can take any value $1, 2, \dots, n$ unless otherwise specified) are defined as follows:

$$y_k = \sum_{j=1}^n M_{kj} x_j$$

The vector components y_k can be understood as either the components of a new vector \mathbf{y} in the directions of the original basis \mathbf{e}_i ($i=1, 2, \dots, n$) or as the components of the old vector \mathbf{x} in the directions of new bases.

There are always these two ways to view a matrix acting on a vector:

1. The operation can be thought of as transforming the vector into a different vector. This view is called the active view (vector in a different place), and is the interpretation we will use most often.

2. The operation can be thought of as expressing the same vector in terms of a different coordinate system or basis. This view is called the passive view.

Some examples may help to clarify these perspectives:

For the matrix-vector product

$$\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax \\ y \end{pmatrix}$$

the active interpretation states that the vector is scaled in the x direction by an amount a . In the passive interpretation, the original vector is written in terms of new bases $(a^{-1}, 0)$ and $(0, 1)$:

$$\begin{pmatrix} x \\ y \end{pmatrix} = ax \begin{pmatrix} a^{-1} \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \end{pmatrix} .$$

As another example, consider the following matrix multiplication:

$$\underline{\mathbf{M}} \mathbf{x} = \begin{pmatrix} \text{Cos} & -\text{Sin} \\ \text{Sin} & \text{Cos} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \text{Cos} - y \text{Sin} \\ x \text{Sin} + y \text{Cos} \end{pmatrix}$$

In the active interpretation, the vector whose cartesian and polar representations are:

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r\cos \\ r\sin \end{pmatrix},$$

is rotated by an angle to obtain:

$$\begin{aligned} \underline{\mathbf{M}}\mathbf{x} &= \begin{pmatrix} \cos & -\sin \\ \sin & \cos \end{pmatrix} \begin{pmatrix} r\cos \\ r\sin \end{pmatrix} \\ &= \begin{pmatrix} (r\cos \cos - r\sin \sin) \\ (r\cos \sin + r\sin \cos) \end{pmatrix} \\ &= \begin{pmatrix} r\cos(\cos^2 + \sin^2) \\ r\sin(\sin \cos + \cos \sin) \end{pmatrix}. \end{aligned}$$

In the passive interpretation, the original vector \mathbf{x} is expressed in terms of a new coordinate

system with axes rotated by $-\theta$ with new bases $\begin{pmatrix} \cos \\ -\sin \end{pmatrix}$ and $\begin{pmatrix} \sin \\ \cos \end{pmatrix}$.

$$\begin{aligned} \begin{pmatrix} x \\ y \end{pmatrix} &= (x\cos - y\sin) \begin{pmatrix} \cos \\ -\sin \end{pmatrix} + (x\sin + y\cos) \begin{pmatrix} \sin \\ \cos \end{pmatrix} \\ &= \begin{pmatrix} x(\cos^2 + \sin^2) + y(\sin \cos - \sin \cos) \\ y(\cos^2 + \sin^2) + x(\sin \cos - \sin \cos) \end{pmatrix} \\ &= \begin{pmatrix} x \\ y \end{pmatrix} \end{aligned}$$

As a general rule, active transformations and passive transformations are inverses of each other; you can do something to a vector or else do the reverse to the coordinate system. The two pictures can be summarized by the following two equations:

(i.) $\underline{\mathbf{M}}\mathbf{x} = \mathbf{y}$ states the active picture, and

(ii.) $\mathbf{x} = \underline{\mathbf{M}}^{-1}\mathbf{y}$ states the passive picture.

C. Matrices as Linear Operators

Matrices are examples of linear operators for which

$$\underline{M}(\underline{ax} + \underline{by}) = a\underline{Mx} + b\underline{My},$$

which can easily be demonstrated by examining the components:

$$\begin{aligned} [\underline{M}(\underline{ax} + \underline{by})]_i &= \sum_k M_{ik}(ax_k + by_k) \\ &= a \sum_k M_{ik}x_k + b \sum_k M_{ik}y_k \\ &= a(\underline{Mx})_i + b(\underline{My})_i. \end{aligned}$$

One can also see that this property holds for a linear combination of many vectors rather than for the two considered above.

We can visualize how the action of a matrix on arbitrary vectors can be expressed if one knows its action on the elementary basis vectors. Given the expansion of \underline{x} in the \underline{e}_i ,

$$\underline{x} = \sum_i x_i \underline{e}_i,$$

one can write

$$\underline{Mx} = \sum_i x_i \underline{Me}_i.$$

Using the fact that all of the components of \underline{e}_i are zero except one, $(\underline{e}_i)_i = 1$, we see that

$$(\underline{Me}_i)_k = \sum_j M_{kj}(\underline{e}_i)_j = M_{ki}$$

This equation tells us that the i -th column of a matrix, \underline{M} , contains the result of operating on the i -th unit vector \underline{e}_i with the matrix. More specifically, the element M_{ki} in the k -th row and i -th column is the component of \underline{Me}_i in the direction of the \underline{e}_k unit vector. As a generalization, we can construct any matrix by first deciding how the matrix affects the elementary unit vectors and then placing the resulting vectors as the columns of the matrix.

II. Properties of General $n \times n$ (Square) Matrices

The following operations involving square matrices each of the same dimension are useful to express in terms of the individual matrix elements:

1. Sum of matrices; $\underline{A} + \underline{B} = \underline{C}$ if $A_{ij} + B_{ij} = C_{ij}$
2. Scalar multiplication; $c\underline{M} = \underline{N}$ if $c M_{ij} = N_{ij}$

3. Matrix multiplication; $\underline{A}\underline{B} = \underline{C}$ if
$$C_{ij} = \sum_k A_{ik} B_{kj}$$

4. Determinant of a matrix;
The determinant is defined inductively

for $n = 1$
$$A = \det(A) = A_{11}$$

for $n > 1$
$$A = \det(A) = \sum_j A_{ij} \det(a_{ij}) (-1)^{(i+j)}$$
 ; where
 i
 $j=1,2,\dots,n$ and a_{ij} is the minor matrix obtained by deleting the i th row and j th column.

5. There are certain special matrices that are important to know:

A. The zero matrix;
$$O_{ij} = 0 \text{ for } i = 1,2,\dots,n \text{ and } j = 1,2,\dots,n$$

B. The identity matrix;
$$I_{ij} = \delta_{ij}$$

(Note $(\underline{I} \underline{M})_{ij} = \sum_k \delta_{ik} M_{kj} = M_{ij}$, so $\underline{I} \underline{M} = \underline{M} \underline{I} = \underline{M}$)

6. Transpose of a matrix; $(M^T)_{ij} = M_{ji}$

7. Complex Conjugate of a Matrix; $(M^*)_{ij} = M^*_{ij}$

8. Adjoint of a Matrix; $(M^{-1})_{ij} = M^*_{ji} = (M^T)^*_{ij}$

9. Inverse of a Matrix; if $\underline{N} \underline{M} = \underline{M} \underline{N} = \underline{I}$ then $\underline{N} = \underline{M}^{-1}$

10. Trace (or Character) of Matrix;
$$\text{Tr}(\underline{M}) = \sum_i M_{ii}$$

(sum of diagonal elements)

III. Special Kinds of Square Matrices

If a matrix obeys certain conditions, we give it one of several special names. These names include the following:

A. Diagonal Matrix: $D_{ij} = d_i \delta_{ij} = d_j \delta_{ij}$

B. Real Matrix: $\underline{M} = \underline{M}^*$ or $M_{ij} = M_{ij}^*$ (real elements)

C. Symmetric Matrix: $\underline{M} = \underline{M}^T$ or $M_{ij} = M_{ji}$ (symmetric about main diagonal)

D. Hermitian Matrix: $\underline{M} = \underline{M}^*$ or $M_{ij} = M_{ji}^*$

E. Unitary Matrix: $\underline{M} = \underline{M}^{-1}$

F. Real Orthogonal $\underline{M} = \underline{M}^T = \underline{M}^{-1}$

IV. Eigenvalues and Eigenvectors of a Square Matrix

An eigenvector of a matrix, \underline{M} , is a vector such that

$$\underline{M}\mathbf{v} = \lambda \mathbf{v}$$

where λ is called the eigenvalue. An eigenvector thus has the property that when it is multiplied by the matrix, the direction of the resultant vector is unchanged. The length, however, is altered by a factor λ . Note that any multiple of an eigenvector is also an eigenvector, which we demonstrate as follows:

$$\underline{M}(a\mathbf{v}) = a\underline{M}\mathbf{v} = a \lambda \mathbf{v} = \lambda(a\mathbf{v}).$$

Hence, an eigenvector can be thought of as defining a direction in n-dimensional space. The length (normalization) of the eigenvector is arbitrary; we are free to choose the length to be anything we please. Usually we choose the length to be unity (because, in quantum mechanics, our vectors usually represent some wavefunction that we wish to obey a normalization condition).

The basic eigenvalue equation can be rewritten as

$$(\underline{M} - \lambda \underline{I})\mathbf{v} = 0$$

or, in an element-by-element manner, as

$$\begin{aligned} (M_{11} - \lambda)v_1 + M_{12}v_2 + M_{13}v_3 + \dots + M_{1n}v_n &= 0 \\ M_{21}v_1 + (M_{22} - \lambda)v_2 + M_{23}v_3 + \dots + M_{2n}v_n &= 0 \end{aligned}$$

$$M_{n1}v_1 + M_{n2}v_2 + M_{n3}v_3 + \dots + (M_{nn} - \lambda)v_n = 0.$$

If you try to solve these n equations for all of the elements of the \mathbf{v} vector ($v_1 \dots v_n$), you can eliminate one variable using one equation, a second variable using a second equation, etc., because the equations are linear. For example you could solve for v_1 using the first equation and then substitute for v_1 in the second equation as you solve for v_2 , etc. Then when you come to the nth equation, you would have n-1 of the variables expressed in terms of the one remaining variable, v_n .

However, you find that you cannot use the remaining equation to solve for the value of v_n ; the last equation is found to appear in the form

$$(C - \lambda) v_n = 0$$

once the v_1, v_2, v_3 , and v_{n-1} are expressed in terms of v_n . We should not really have expected to solve for v_n since, as we saw above, the length of the vector \mathbf{v} is not determined from the set of eigenvalue equations. You find that the only solution is $v_n = 0$, which then implies that all of the other $v_k = 0$ because you expressed them in terms of v_n , unless the eigenvalue λ is chosen to obey $\lambda = C$.

Upon analyzing what has gone into the C element, one finds that the v_k ($k=1,2,3,\dots,n-1$) were eliminated in favor of v_n by successively combining rows of the $(\underline{M} - \lambda \underline{I})$ matrix. Thus, $(C - \lambda)$ can vanish if and only if the last row of $(\underline{M} - \lambda \underline{I})$ is a linear combination of the other $n-1$ rows. A theorem dealing with determinants states that the rows of a matrix are linearly dependent (i.e., one row is a linear combination of the rest) if and only if the determinant of the matrix is zero. We can therefore make the eigenvalue equation have a solution \mathbf{v} by adjusting λ so the determinant of $(\underline{M} - \lambda \underline{I})$ vanishes.

A. Finding the Eigenvalues

In summary, to solve an eigenvalue equation, we first solve the determinantal equation:

$$\det(\underline{M} - \lambda \underline{I}) = 0.$$

Using the definition of a determinant, one can see that expanding the determinant results in an n^{th} order polynomial in

$$a_n(\lambda) \lambda^n + a_{n-1}(\lambda) \lambda^{n-1} + \dots + a_1(\lambda) \lambda + a_0(\lambda) = 0,$$

where the coefficients $a_i(\lambda)$ depend on the matrix elements of \underline{M} . A theorem in algebra shows that such an equation always has n roots some or all of which may be complex (e.g., a quadratic equation has 2 roots). Thus there are n different ways of adjusting the parameter λ so the determinant vanishes. Each of these solutions is a candidate for use as λ in subsequently solving for the \mathbf{v} vector's coefficients. That is, each λ value has its own vector \mathbf{v} .

B. Finding the Eigenvectors

One can substitute each of these n values ($\lambda_k, k = 1,2,\dots,n$) back into the eigenvalue equation, one at a time, and solve for the n eigenvectors $\mathbf{v}(k)$. By using one of the λ_k values, the n^{th} equation is guaranteed to be equal to zero, so you use $n-1$ of the equations to solve for $n-1$ of the components in terms of the n^{th} one. The eigenvectors will then be determined up to a multiplicative factor which we then fix by requiring normalization:

$$v_i^*(k)v_i(k) = \langle \mathbf{v}(k) | \mathbf{v}(k) \rangle = 1.$$

This expression now defines the dot or inner product (Hermitian inner product) for vectors which can have complex valued components. We use this definition so the dot product of a complex valued vector with itself is real.

In summary, we have shown how to solve the n equations:

$$M_{ij}v_j(k) = \lambda_k v_i(k); k = 1, 2, \dots, n.$$

for the eigenvalues of the matrix M_{ij} and the corresponding normalized eigenvectors $\mathbf{v}(k)$, $k = 1, 2, \dots, n$. Now let us work an example that is chosen to illustrate the concepts we have learned as well as an additional complication, that of eigenvalue degeneracy.

C. Examples

Consider the following real symmetric matrix:

$$\underline{M} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & \frac{5}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{5}{2} \end{pmatrix}$$

The set of eigenvalue-eigenvector equations has non-trivial ($\mathbf{v}(k) \neq 0$ is "trivial") solutions if

$$\underline{M} - \lambda_k \underline{I} = 0.$$

In our example, this amounts to

$$\begin{pmatrix} 3 - \lambda_k & 0 & 0 \\ 0 & \frac{5}{2} - \lambda_k & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{5}{2} - \lambda_k \end{pmatrix} = 0,$$

or

$$(3 - \lambda_k) \begin{vmatrix} \frac{5}{2} - \lambda_k & \frac{1}{2} \\ \frac{1}{2} & \frac{5}{2} - \lambda_k \end{vmatrix} = 0,$$

or,

$$(3 - \lambda_k) \left(\frac{5}{2} - \lambda_k \right)^2 - \frac{1}{4} = 0,$$

or,

$$(3 - k) [k^2 - 5k + 6] = 0 = (3 - k)(k - 3)(k - 2).$$

There are three real solutions to this cubic equation (why all the solutions are real in this case for which the M matrix is real and symmetric will be made clear later):

$$1) \quad \lambda_1 = 3, 2) \quad \lambda_2 = 3, \text{ and } 3) \quad \lambda_3 = 2.$$

Notice that the eigenvalue $\lambda_k = 3$ appears twice; we say that the eigenvalue $\lambda_k = 3$ is doubly degenerate. $\lambda_k = 2$ is a non-degenerate eigenvalue.

The eigenvectors $\mathbf{v}(k)$ are found by plugging the above values for λ_k into the basic eigenvalue equation

$$\underline{\mathbf{M}} \mathbf{v}(k) = \lambda_k \mathbf{v}(k)$$

For the non-degenerate eigenvalue ($\lambda_3 = 2$), this gives:

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & \frac{5}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{5}{2} \end{pmatrix} \begin{pmatrix} v_1(3) \\ v_2(3) \\ v_3(3) \end{pmatrix} = 2 \begin{pmatrix} v_1(3) \\ v_2(3) \\ v_3(3) \end{pmatrix} .$$

The following algebraic steps can then be followed:

- i. $3v_1(3) = 2v_1(3)$ implies that $v_1(3) = 0$,
- ii. $\frac{5}{2} v_2(3) + \frac{1}{2} v_3(3) = 2v_2(3)$, and
- iii. $\frac{1}{2} v_2(3) + \frac{5}{2} v_3(3) = 2v_3(3)$.

The last two equations can not be solved for both $v_2(3)$ and $v_3(3)$. To see the trouble, multiply equation iii. by 5 and subtract it from equation ii. to obtain

$$-12 v_3(3) = 2v_2(3) - 10v_3(3),$$

which implies that $v_3(3) = -v_2(3)$. Now, substitute this result into the equation ii. to obtain

$$\frac{5}{2} v_2(3) + \frac{1}{2} (-v_2(3)) = 2v_2(3),$$

or,

$$2v_2(3) = 2v_2(3).$$

This is a trivial identity; it does not allow you to solve for $v_2(3)$.

Hence, for this non-degenerate root, one is able to solve for all of the v_j elements in terms of one element that is yet undetermined.

As in all matrix eigenvalue problems, we are able to express (n-1) elements of the eigenvector $\mathbf{v}(k)$ in terms of one remaining element. However, we can never solve for this one last element. So, for convenience, we impose one more constraint (equation to be

obeyed) which allows us to solve for the remaining element of $\mathbf{v}(\mathbf{k})$. We require that the eigenvectors be normalized:

$$\langle \mathbf{v}(\mathbf{k}) | \mathbf{v}(\mathbf{k}) \rangle = \sum_a \mathbf{v}_a^*(\mathbf{k}) \mathbf{v}_a(\mathbf{k}) = 1.$$

In our example, this means that

$$v_1^2(3) + v_2^2(3) + v_3^2(3) = 1,$$

or,

$$0^2 + v_2^2(3) + (-v_2(3))^2 = 1,$$

which implies that $v_2(3) = \pm \frac{1}{\sqrt{2}}$.

So,

$$v_3(3) = \pm \frac{1}{\sqrt{2}},$$

and finally the vector is given by:

$$\mathbf{v}(3) = \pm \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

Note that even after requiring normalization, there is still an indeterminacy in the sign of $\mathbf{v}(3)$. The eigenvalue equation, as we recall, only specifies a direction in space. The sense or sign is not determined. We can choose either sign we prefer.

Finding the first eigenvector was not too difficult. The degenerate eigenvectors are more difficult to find. For $\lambda_1 = \lambda_2 = 3$,

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & \frac{5}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{5}{2} \end{pmatrix} \begin{pmatrix} v_1(1) \\ v_2(1) \\ v_3(1) \end{pmatrix} = 3 \begin{pmatrix} v_1(1) \\ v_2(1) \\ v_3(1) \end{pmatrix}.$$

Again, the algebraic equations can be written as follows:

- i. $3v_1(1) = 3v_1(1)$; this tells us nothing!
- ii. $\frac{5}{2} v_2(1) + \frac{1}{2} v_3(1) = 3v_2(1)$, and

$$\text{iii. } \frac{1}{2} v_2(1) + \frac{5}{2} v_3(1) = 3v_3(1).$$

If we multiply equation iii. by 5 and subtract it from equation ii., we obtain:

$$-12v_3(1) = -15v_3(1) + 3v_2(1),$$

or

$$3v_3(1) = 3v_2(1),$$

which implies that $v_3(1) = v_2(1)$. So far, all we have is $v_3(1) = v_2(1)$; we don't know $v_1(1)$ nor do we know either $v_3(1)$ or $v_2(1)$, and we have used all three equations.

Normalization provides one more equation $v_1^2(1) + v_2^2(1) + (v_2(1))^2 = 1$, but we are still in a situation with more unknowns (2) than equations (1).

One might think that by restricting our eigenvectors to be orthogonal as well as normalized, we might overcome this difficulty (however, such is not the case, as we now show).

For our vectors, the constraint that the nondegenerate vector $\mathbf{v}(3)$ be orthogonal to the one we are trying to find $\mathbf{v}(1)$, amounts to

$$\langle \mathbf{v}(3) \mathbf{v}(1) \rangle = 0$$

$$v_1(3)^* v_1(1) + v_2(3)^* v_2(1) + v_3(3)^* v_3(1) = 0$$

$$0 v_1(1) \pm \frac{1}{\sqrt{2}} v_2(1) - \frac{1}{\sqrt{2}} (v_2(1)) = 0.$$

We see that $\mathbf{v}(3)$ and $\mathbf{v}(1)$ are already orthogonal regardless of how $v_2(1)$ and $v_3(1)$ turn out. This is shown below to be guaranteed because $\mathbf{v}(1)$ and $\mathbf{v}(3)$ have different eigenvalues (i.e., two eigenvectors belonging to different eigenvalues of any symmetric or hermitian matrix must be orthonormal). Hence, this first attempt at finding additional equations to use has failed.

What about the two degenerate eigenvectors $\mathbf{v}(1)$ and $\mathbf{v}(2)$? Are they also orthonormal? So far, we know that these two eigenvectors have the structure

$$\mathbf{v}(1) = \begin{pmatrix} v_1(1) \\ v_2(1) \\ v_3(1) \end{pmatrix}, \text{ with } 1 = v_1^2(1) + 2v_2^2(1).$$

If we go through all of the above steps for $\mathbf{v}(2)$ with $\ell = 2$, we will find that this vector obeys exactly the same set of equations

$$\mathbf{v}(2) = \begin{pmatrix} v_1(2) \\ v_2(2) \\ v_3(2) \end{pmatrix}, \text{ with } 1 = v_1^2(2) + 2v_2^2(2).$$

We showed above that $\langle \mathbf{v}(1) \mathbf{v}(3) \rangle = 0$, and it is easy to show that $\langle \mathbf{v}(2) \mathbf{v}(3) \rangle = 0$ because the elements of $\mathbf{v}(2)$, thus far, obey the same equations as $\mathbf{v}(1)$.

If we also wish to make the two degenerate eigenvectors orthogonal to one another

$$\langle \mathbf{v}(1) \mathbf{v}(2) \rangle = 0,$$

then we obtain additional relationships among our yet-undetermined vector amplitudes. In particular, we obtain

$$v_1(1)v_1(2) + v_2(1)v_2(2) + v_3(1)v_3(2) = 0,$$

or,

$$v_1(1)v_1(2) + 2v_2(1)v_2(2) = 0.$$

Although progress has been made, we still have four unknowns $v_1(1), v_2(1); v_1(2), v_2(2)$ and only three equations:

$$0 = v_1(1) v_1(2) + 2v_2(1) v_2(2) = 0,$$

$$1 = v_1(1) v_1(1) + 2v_2(1) v_2(1) = 1, \text{ and}$$

$$1 = v_1(2) v_1(2) + 2v_2(2) v_2(2) = 1.$$

It appears as though we are stuck again. We are; but for good reasons. We are trying to find two vectors $\mathbf{v}(1)$ and $\mathbf{v}(2)$ that are orthonormal and are eigenvectors of $\underline{\mathbf{M}}$ having eigenvalue equal to 3. Suppose that we do find two such vectors. Because $\underline{\mathbf{M}}$ is a linear operator, any two vectors generated by taking linear combinations of these two vectors would also be eigenvectors of $\underline{\mathbf{M}}$. There is a degree of freedom, that of recombining $\mathbf{v}(1)$ and $\mathbf{v}(2)$, which can not be determined by insisting that the two vectors be eigenfunctions. Thus, in this degenerate-eigenvalue case our requirements do not give a unique pair of eigenvectors. They just tell us the two-dimensional space in which the acceptable eigenvectors lie (this difficulty does not arise in nondegenerate cases because one-dimensional spaces have no flexibility other than a sign.).

So to find an acceptable pair of vectors, we are free to make an additional choice. For example, we can choose one of the four unknown components of $\mathbf{v}(1)$ and $\mathbf{v}(2)$ equal to zero. Let us make the choice $v_1(1) = 0$. Then the above equations can be solved for the other elements of $\mathbf{v}(1)$ to give $v_2(1) = \pm \frac{1}{\sqrt{2}} = v_3(1)$. The orthogonality between $\mathbf{v}(1)$ and $\mathbf{v}(2)$ then gives $0 = 2 \pm \frac{1}{\sqrt{2}} v_2(2)$, which implies that $v_2(2) = v_3(2) = 0$; the remaining equation involving $\mathbf{v}(2)$ then gives $v_1(2) = \pm 1$.

In summary, we have now found a specific solution once the choice $v_1(1) = 0$ is made:

$$\mathbf{v}(1) = \pm \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \quad \mathbf{v}(2) = \pm \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \text{and } \mathbf{v}(3) = \pm \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$$

Other choices for $\mathbf{v}(1)$ will yield different specific solutions.

V. Properties of Eigenvalues and Eigenvectors of Hermitian Matrices

The above example illustrates many of the properties of the matrices that we will most commonly encounter in quantum mechanics. It is important to examine these properties in more detail and to learn about other characteristics that Hermitian matrices have.

A. Outer product

Given any vector \mathbf{v} , we can form a square matrix denoted $|\mathbf{v}(i)\rangle\langle\mathbf{v}(i)|$, whose elements are defined as follows:

$$|\mathbf{v}(i)\rangle\langle\mathbf{v}(i)| = \begin{pmatrix} \mathbf{v}_1^*(i)\mathbf{v}_1(i) & \mathbf{v}_1^*(i)\mathbf{v}_2(i) & \dots & \mathbf{v}_1^*(i)\mathbf{v}_n(i) \\ \mathbf{v}_2^*(i)\mathbf{v}_1(i) & \mathbf{v}_2^*(i)\mathbf{v}_2(i) & \dots & \mathbf{v}_2^*(i)\mathbf{v}_n(i) \\ \dots & \dots & \dots & \dots \\ \mathbf{v}_n^*(i)\mathbf{v}_1(i) & \mathbf{v}_n^*(i)\mathbf{v}_2(i) & \dots & \mathbf{v}_n^*(i)\mathbf{v}_n(i) \end{pmatrix}$$

We can use this matrix to project onto the component of a vector in the $\mathbf{v}(i)$ direction. For the example we have been considering, if we form the projector onto the $\mathbf{v}(1)$ vector, we obtain

$$\begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix},$$

for $\mathbf{v}(2)$, we get

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and for $\mathbf{v}(3)$ we find

$$\begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} .$$

These three projection matrices play important roles in what follows.

B. Completeness Relation or Resolution of the Identity

The set of eigenvectors of any Hermitian matrix form a complete set over the space they span in the sense that the sum of the projection matrices constructed from these eigenvectors gives an exact representation of the identity matrix.

$$\sum_i |\mathbf{v}(i)\rangle\langle\mathbf{v}(i)| = \mathbf{I}.$$

For the specific matrix we have been using as an example, this relation reads as follows:

$$\begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} [1 \ 0 \ 0] + \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} =$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} .$$

Physically, this means that when you project onto the components of a vector in these three directions, you don't lose any of the vector. This happens because our vectors are orthogonal and complete. The completeness relation means that any vector in this three-dimensional space can be written in terms of $\mathbf{v}(1)$, $\mathbf{v}(2)$, and $\mathbf{v}(3)$ (i.e., we can use $\mathbf{v}(1), \mathbf{v}(2), \mathbf{v}(3)$ as a new set of bases instead of $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$).

Let us consider an example in which the following vector is expanded or written in terms of our three eigenvectors:

$$\mathbf{f} = \begin{pmatrix} 7 \\ -9 \\ 12 \end{pmatrix} = a_1 \mathbf{v}(1) + a_2 \mathbf{v}(2) + a_3 \mathbf{v}(3)$$

The task at hand is to determine the expansion coefficients, the a_j . These coefficients are the projections of the given \mathbf{f} vector onto each of the $\mathbf{v}(i)$ directions:

$$\langle \mathbf{v}(1) \mathbf{f} \rangle = a_1 ,$$

since,

$$\langle \mathbf{v}(1) \mathbf{v}(2) \rangle = \langle \mathbf{v}(1) \mathbf{v}(3) \rangle = 0.$$

Using our three \mathbf{v}_i vectors and the above \mathbf{f} vector, the following three expansion coefficients are obtained:

$$a_1 = \begin{matrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 7 & -9 & 12 \end{matrix} = \frac{3}{\sqrt{2}}$$

$$\langle \mathbf{v}(2) \mathbf{f} \rangle = 7$$

$$\langle \mathbf{v}(3) \mathbf{f} \rangle = -\frac{21}{\sqrt{2}}$$

Therefore, \mathbf{f} can be written as:

$$\mathbf{f} = \begin{matrix} 7 \\ -9 \\ 12 \end{matrix} = \frac{3}{\sqrt{2}} \begin{matrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{matrix} + 7 \begin{matrix} 1 \\ 0 \\ 0 \end{matrix} - \frac{21}{\sqrt{2}} \begin{matrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{matrix}$$

This works for any vector \mathbf{f} , and we could write the process in general in terms of the resolution of the identity as

$$\mathbf{f} = \underline{\mathbf{I}}\mathbf{f} = \sum_k |\mathbf{v}(k)\rangle \langle \mathbf{v}(k)| \mathbf{f}\rangle = \sum_k |\mathbf{v}(k)\rangle a_k ,$$

This is how we will most commonly make use of the completeness relation as it pertains to the eigenvectors of Hermitian matrices.

C. Spectral Resolution of M

It turns out that not only the identity matrix $\underline{\mathbf{I}}$ but also the matrix $\underline{\mathbf{M}}$ itself can be expressed in terms of the eigenvalues and eigenvectors. In the so-called spectral representation of $\underline{\mathbf{M}}$, we have

$$\underline{M} = \sum_k |\mathbf{v}(k)\rangle \langle \mathbf{v}(k)|$$

In the example we have been using, the three terms in this sum read

$$\begin{array}{ccccccc}
 & 0 & 0 & 0 & & 0 & 0 & 0 & & 3 & 0 & 0 \\
 & & & & 1 & 0 & 0 & & & & & \\
 3 & 0 & \frac{1}{2} & \frac{1}{2} & + 3 & 0 & 0 & 0 & + 2 & 0 & \frac{1}{2} & -\frac{1}{2} & = & 0 & \frac{5}{2} & \frac{1}{2} & = \underline{M} \\
 & 0 & \frac{1}{2} & \frac{1}{2} & & 0 & 0 & 0 & & 0 & -\frac{1}{2} & \frac{1}{2} & & 0 & \frac{1}{2} & \frac{5}{2}
 \end{array}$$

This means that a matrix is totally determined if we know its eigenvalues and eigenvectors.

D. Eigenvalues of Hermitian Matrices are Real Numbers

A matrix can be expressed in terms of any complete set of vectors. For $n \times n$ matrices, a complete set is any n linearly independent vectors. For a set of vectors $|\mathbf{k}\rangle$, $k = 1, 2, \dots, n$, the matrix elements of \underline{M} are denoted M_{jk} or $\langle \mathbf{j} | \underline{M} | \mathbf{k} \rangle$. If the matrix is Hermitian then

$$\langle \mathbf{j} | \underline{M} | \mathbf{k} \rangle = \langle \mathbf{k} | \underline{M} | \mathbf{j} \rangle^*$$

If the vectors $|\mathbf{k}\rangle$ are eigenvectors of \underline{M} , that is, if $\underline{M} |\mathbf{k}\rangle = \lambda_k |\mathbf{k}\rangle$, then the eigenvalues are real. This can be shown as follows:

$$\langle \mathbf{k} | \underline{M} | \mathbf{k} \rangle = \lambda_k \langle \mathbf{k} | \mathbf{k} \rangle = \langle \mathbf{k} | \underline{M} | \mathbf{k} \rangle^* = \lambda_k^* \langle \mathbf{k} | \mathbf{k} \rangle,$$

so $\lambda_k = \lambda_k^*$. This is a very important result because it forms the basis of the use of Hermitian operators in quantum mechanics; such operators are used because experiments yield real results, so to connect with experimental reality, only hermitian operators must appear.

E. Nondegenerate Eigenvectors of Hermitian Matrices are Orthogonal

If two eigenvalues are different, $\lambda_k \neq \lambda_j$, then

$$\langle \mathbf{k} | \underline{M} | \mathbf{j} \rangle = \lambda_j \langle \mathbf{k} | \mathbf{j} \rangle = \langle \mathbf{j} | \underline{M} | \mathbf{k} \rangle^* = \lambda_k^* \langle \mathbf{j} | \mathbf{k} \rangle^* = \lambda_k \langle \mathbf{k} | \mathbf{j} \rangle,$$

which implies that $(\lambda_k - \lambda_j) \langle \mathbf{k} | \mathbf{j} \rangle = 0$. Since, by assumption, $\lambda_k \neq \lambda_j$, it must be that $\langle \mathbf{k} | \mathbf{j} \rangle = 0$. In other words, the eigenvectors are orthogonal. We saw this earlier in our example when we "discovered" that $\mathbf{v}(3)$ was automatically orthogonal to $\mathbf{v}(1)$ and to $\mathbf{v}(2)$.

If one has degenerate eigenvalues, $\lambda_k = \lambda_j$ for some k and j , then the corresponding eigenvectors are not automatically orthogonal to one another (they are orthogonal to other eigenvectors), but the degenerate eigenvectors can always be chosen to be orthogonal. We also encountered this in our earlier example.

In all cases then, one can find n orthonormal eigenvectors (remember we required $\langle \mathbf{k} | \mathbf{k} \rangle = 1$ as an additional condition so that our amplitudes could be interpreted in terms of probabilities). Since any vector in an n -dimensional space can be expressed as a linear combination of n orthonormal vectors, the eigenvectors form a complete basis set. This is why the so-called resolution of the identity, in which the unit matrix can be expressed in terms of the n eigenvectors of \underline{M} , holds for all hermitian matrices.

F. Diagonalizing a Matrix using its Eigenvectors

The eigenvectors of \underline{M} can be used to form a matrix that diagonalizes \underline{M} . This matrix \underline{S} is defined such that the k^{th} column of \underline{S} contains the elements of $\mathbf{v}(k)$

$$\underline{S} = \begin{pmatrix} v_1(1) & v_1(2) & \dots & v_1(n) \\ v_2(1) & v_2(2) & \dots & v_2(n) \\ v_3(1) & v_3(2) & \dots & v_3(n) \\ \vdots & \vdots & \ddots & \vdots \\ v_n(1) & v_n(2) & \dots & v_n(n) \end{pmatrix} .$$

Then using the above equation one can write the action of \underline{M} on \underline{S} as

$$M_{ij} S_{jk} = \sum_j M_{ij} v_j(k) = \sum_j v_i(k) = S_{ik} \epsilon_k .$$

Now consider another matrix $\underline{\epsilon}$ which is a diagonal matrix with diagonal elements ϵ_k , i.e.

$$\epsilon_{ik} = \epsilon_k \delta_{ik} .$$

One can easily show using the δ_{jk} matrix that

$$\sum_j S_{ij} \delta_{jk} \epsilon_k = S_{ik} \epsilon_k ,$$

since the only non-zero term in the sum is the one in which $j = k$. Thus, comparing with the previous result for the action of \underline{M} on \underline{S} ,

$$M_{ij} S_{jk} = \sum_j S_{ij} \delta_{jk} \epsilon_k .$$

These are just the i,k th matrix elements of the matrix equation

$$\underline{M} \underline{S} = \underline{S} \underline{\epsilon} .$$

Let us assume that an inverse \underline{S}^{-1} of \underline{S} exists. Then multiply the above equation on the left by \underline{S}^{-1} to obtain

$$\underline{S}^{-1} \underline{M} \underline{S} = \underline{S}^{-1} \underline{S} \underline{\Lambda} = \underline{I} \underline{\Lambda} = \underline{\Lambda}.$$

This identity illustrates a so-called similarity transform of \underline{M} using \underline{S} . Since $\underline{\Lambda}$ is diagonal, we say that the similarity transform \underline{S} diagonalizes \underline{M} . Notice this would still work if we had used the eigenvectors $\mathbf{v}(k)$ in a different order. We would just get $\underline{\Lambda}$ with the diagonal elements in a different order. Note also that the eigenvalues of \underline{M} are the same as those of $\underline{\Lambda}$ since the eigenvalues of $\underline{\Lambda}$ are just the diagonal elements λ_k with eigenvectors \mathbf{e}_k (the elementary unit vectors)

$$\underline{\Lambda} \mathbf{e}_k = \lambda_k \mathbf{e}_k, \text{ or, } (\underline{\Lambda} \mathbf{e}_k)_i = \sum_j \lambda_{kj} \delta_{ji} = \lambda_{ki} = \lambda_k (\mathbf{e}_k)_i.$$

G. The Trace of a Matrix is the Sum of its Eigenvalues

Based on the above similarity transform, we can now show that the trace of a matrix (i.e., the sum of its diagonal elements) is independent of the representation in which the matrix is formed, and, in particular, the trace is equal to the sum of the eigenvalues of the matrix. The proof of this theorem proceeds as follows:

$$\begin{aligned} \sum_k \lambda_k &= \text{Tr}(\underline{\Lambda}) = \sum_k (\underline{S}^{-1} \underline{M} \underline{S})_{kk} = \sum_{kij} (\underline{S}_{ki}^{-1} M_{ij} \underline{S}_{jk}) \\ &= \sum_{ij} M_{ij} \sum_k \underline{S}_{jk} \underline{S}_{ki}^{-1} = \sum_{ij} M_{ij} \delta_{ji} = \sum_i M_{ii} = \text{Tr}(\underline{M}) \end{aligned}$$

H. The Determinant of a Matrix is the Product of its Eigenvalues

This theorem, $\det(\underline{M}) = \det(\underline{\Lambda})$, can be proven, by using the theorem that $\det(\underline{AB}) = \det(\underline{A}) \det(\underline{B})$, inductively within the expansion by minors of the determinant

$$\begin{aligned} \lambda_1 \lambda_2 \lambda_3 \dots \lambda_n &= \det(\underline{\Lambda}) = \det(\underline{S}^{-1} \underline{M} \underline{S}) = \det(\underline{S}^{-1}) \det(\underline{M}) \det(\underline{S}) \\ &= \det(\underline{M}) \det(\underline{S}^{-1}) \det(\underline{S}) \\ &= \det(\underline{M}) \det(\underline{S}^{-1} \underline{S}) \\ &= \det(\underline{M}) \det(\underline{I}) = \det(\underline{MI}) = \det(\underline{M}). \end{aligned}$$

I. Invariance of Matrix Identities to Similarity Transforms

We will see later that performing a similarity transform expresses a matrix in a different basis (coordinate system). If, for any matrix \underline{A} , we call $\underline{S}^{-1} \underline{A} \underline{S} = \underline{A}'$, we can show that performing a similarity transform on a matrix equation leaves the form of the equation unchanged. For example, if

$$\underline{A} + \underline{B} \underline{C} = \underline{P} \underline{Q} \underline{R},$$

performing a similarity transform on both sides, and remembering that $\underline{S}^{-1} \underline{S} = \underline{I}$, one finds

$$\begin{aligned} \underline{S}^{-1}(\underline{A} + \underline{B} \underline{C})\underline{S} &= \underline{S}^{-1} \underline{P} \underline{Q} \underline{R} \underline{S} = \\ \underline{S}^{-1} \underline{A} \underline{S} + \underline{S}^{-1} \underline{B} (\underline{S} \underline{S}^{-1}) \underline{C} \underline{S} &= \underline{S}^{-1} \underline{P} (\underline{S} \underline{S}^{-1}) \underline{Q} \underline{S} \underline{S}^{-1} \underline{R} \underline{S}, \end{aligned}$$

or,

$$\underline{A}' + \underline{B}' \underline{C}' = \underline{P}' \underline{Q}' \underline{R}'.$$

J. How to Construct \underline{S}^{-1}

To form \underline{S}^{-1} , recall the eigenvalue equation we began with

$$\underline{M} \mathbf{v}(k) = \lambda_k \mathbf{v}(k).$$

If we multiply on the left by \underline{S}^{-1} and insert $\underline{S}^{-1} \underline{S} = \underline{I}$, we obtain

$$\underline{S}^{-1} \underline{M} (\underline{S} \underline{S}^{-1}) \mathbf{v}(k) = \lambda_k \underline{S}^{-1} \mathbf{v}(k),$$

or,

$$\mathbf{v}(k)' = \lambda_k \mathbf{v}(k)'.$$

According to the passive picture of matrix operations, $\mathbf{v}(k)' = \underline{S}^{-1} \mathbf{v}(k)$ is the old vector \mathbf{v} expressed in terms of new basis functions which are given by the columns of \underline{S} . But the rows of \underline{S}^{-1} , or equivalently, the columns of \underline{S} have to be orthonormal vectors (using the definition of an inverse):

$$\sum_j \underline{S}^{-1}_{ij} \underline{S}_{jk} = \delta_{ik}.$$

However, the columns of \underline{S} are just the eigenvectors $\mathbf{v}(k)$, so the rows of \underline{S}^{-1} must also have something to do with $\mathbf{v}(k)$.

Now consider the matrix \underline{S}^{-1} . Its elements are $(\underline{S}^{-1})_{kj} = \underline{S}_{jk}^*$, so $(\underline{S}^{-1})_{kj} = \mathbf{v}_j^*(k)$, as a result of which we can write

$$(\underline{S}^{-1})_{ik} = \sum_j S_{ij} S_{jk} = \sum_j \mathbf{v}_j^*(i) \mathbf{v}_j(k) = \langle \mathbf{v}(i) | \mathbf{v}(k) \rangle = \delta_{ik}.$$

We have therefore found \underline{S}^{-1} , and it is \underline{S}^{-1} . We have also proven the important theorem stating that any Hermitian matrix \underline{M} , can be diagonalized by a matrix \underline{S} which is unitary

($\underline{S}^{-1} = \underline{S}^{-1}$) and whose columns are the eigenvectors of \underline{M} .

VI. Finding Inverses, Square Roots, and Other Functions of a Matrix Using its Eigenvectors and Eigenvalues

Since a matrix is defined by how it affects basis vectors upon which it acts, and since a matrix only changes the lengths, not the directions of its eigenvectors, we can form other matrices that have the same eigenvectors as the original matrix but have different eigenvalues in a straightforward manner. For example, if we want to reverse the effect of \underline{M} , we create another matrix that reduces the lengths of the eigenvectors by the same amount that \underline{M} lengthened them. This produces the inverse matrix. We illustrate this by using the same example matrix that we used earlier.

$$\underline{M}^{-1} = \sum_k k^{-1} |\mathbf{v}(i)\rangle \langle \mathbf{v}(i)| \quad (\text{this is defined only if all } k \neq 0)$$

$$\begin{pmatrix} 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} + \frac{1}{3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & -\frac{1}{2} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{10}{24} & -\frac{2}{24} \\ 0 & -\frac{2}{24} & \frac{10}{24} \end{pmatrix} = \underline{M}^{-1}$$

To show that this matrix obeys $\underline{M}^{-1} \underline{M} = \underline{1}$, we simply multiply the two matrices together:

$$\begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{10}{24} & -\frac{2}{24} \\ 0 & -\frac{2}{24} & \frac{10}{24} \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & \frac{5}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{5}{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

An extension of this result is that a whole family of matrices, each member related to the original \underline{M} matrix, can be formed by combining the eigenvalues and eigenvectors as follows:

$$f(\underline{M}) = \sum_i |\mathbf{v}(i)\rangle \langle \mathbf{v}(i)| f(k_i),$$

where f is any function for which $f(x)$ is defined. Examples include $\exp(\underline{M})$, $\sin(\underline{M})$, $\underline{M}^{1/2}$, and $(\underline{I}-\underline{M})^{-1}$. The matrices so constructed (e.g., $\exp(\underline{M}) = \sum_i \exp(\lambda_i) |\mathbf{v}(i)\rangle \langle \mathbf{v}(i)|$) are proper representations of the functions of \underline{M} in the sense that they give results equal to those of the function of \underline{M} when acting on any eigenvector of \underline{M} ; because the eigenvectors form complete sets, this implies that they give identical results when acting on any vector. This equivalence can be shown most easily by carrying out a power series expansion of the function of \underline{M} (e.g., of $\exp(\underline{M})$) and allowing each term in the series to act on an eigenvector.

VII. Projectors Revisited

In hindsight, the relationships developed above for expressing functions of the original matrix in terms of its eigenvectors and eigenvalues are not unexpected because each of the matrices is given in terms of the so-called projector matrices $|\mathbf{v}(i)\rangle \langle \mathbf{v}(i)|$. As we saw, the matrix $\sum_i |\mathbf{v}(i)\rangle \langle \mathbf{v}(i)|$ behaves just like \underline{M} , as demonstrated below:

$$\sum_i |\mathbf{v}(i)\rangle \langle \mathbf{v}(i)| \sum_j |\mathbf{v}(j)\rangle = \sum_i |\mathbf{v}(i)\rangle \langle \mathbf{v}(i)| = \sum_j |\mathbf{v}(j)\rangle$$

The matrix $|\mathbf{v}(i)\rangle \langle \mathbf{v}(i)|$ is called a projector onto the space of eigenvector $|\mathbf{v}(i)\rangle$ because its action on any vector $|\mathbf{f}\rangle$ within the class of admissible vectors (2- or 4-dimensional vectors are in a different class than 3- or 1- or 7- or 96-dimensional vectors)

$$(|\mathbf{v}(i)\rangle \langle \mathbf{v}(i)|) |\mathbf{f}\rangle = |\mathbf{v}(i)\rangle (\langle \mathbf{v}(i)|\mathbf{f}\rangle)$$

gives $|\mathbf{v}(i)\rangle$ multiplied by the coefficient of $|\mathbf{f}\rangle$ along $|\mathbf{v}(i)\rangle$.

This construction in which a vector is used to form a matrix $|\mathbf{v}(i)\rangle \langle \mathbf{v}(i)|$ is called an "outer product". The projection matrix thus formed can be shown to be idempotent, which means that the result of applying it twice (or more times) is identical to the result of applying it once $P P = P$. This property is straightforward to demonstrate. Let us consider the projector $\underline{P}_i = |\mathbf{v}(i)\rangle \langle \mathbf{v}(i)|$, which, when applied twice yields

$$|\mathbf{v}(i)\rangle \langle \mathbf{v}(i)| |\mathbf{v}(i)\rangle \langle \mathbf{v}(i)| = |\mathbf{v}(i)\rangle \langle \mathbf{v}(i)| = |\mathbf{v}(i)\rangle \langle \mathbf{v}(i)|$$

Sets of projector matrices each formed from a member of an orthonormal vector set are mutually orthogonal, (i.e., $\underline{P}_i \underline{P}_j = 0$ if $i \neq j$), which can be shown as follows:

$$\begin{aligned} |\mathbf{v}(i)\rangle \langle \mathbf{v}(i)| |\mathbf{v}(j)\rangle \langle \mathbf{v}(j)| &= |\mathbf{v}(i)\rangle \langle \mathbf{v}(i)|_{ij} \langle \mathbf{v}(j)| \\ &= |\mathbf{v}(i)\rangle \langle \mathbf{v}(j)| = 0 \quad (i \neq j) \end{aligned}$$

VIII. Hermitian Matrices and The Turnover Rule

The eigenvalue equation:

$$\underline{M}|\mathbf{v}(i)\rangle = \lambda_i|\mathbf{v}(i)\rangle,$$

which can be expressed in terms of its indices as:

$$M_{kj}v_j(i) = \lambda_i v_k(i),$$

is equivalent to (just take the complex conjugate):

$$v_j^*(i)M_{kj}^* = \lambda_i^* v_k^*(i),$$

which, for a Hermitian matrix \underline{M} , can be rewritten as:

$$v_j^*(i)M_{jk} = \lambda_i^* v_k^*(i),$$

or, equivalently, as:

$$\langle \mathbf{v}(i) | \underline{M} = \lambda_i \langle \mathbf{v}(i) |.$$

This means that the $\mathbf{v}(i)$, when viewed as column vectors, obey the eigenvalue identity $\underline{M}|\mathbf{v}(i)\rangle = \lambda_i|\mathbf{v}(i)\rangle$. These same vectors, when viewed as row vectors (and thus complex conjugated), also obey the eigenvalue relation, but in the "turn over" form $\langle \mathbf{v}(i) | \underline{M} = \lambda_i \langle \mathbf{v}(i) |$. For example, in the case we have been studying, the first vector obeys

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{2} & \frac{5}{2} \end{pmatrix} = \begin{pmatrix} 0 & \frac{3}{\sqrt{2}} & \frac{3}{\sqrt{2}} \\ 0 & \frac{5}{2} & \frac{1}{2} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = 3 \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

As a general rule, a hermitian matrix operating on a column vector to the right is equivalent to the matrix operating on the complex conjugate of the same vector written as a row vector to the left. For a non-Hermitian matrix, it is the adjoint matrix operating on the complex conjugate row vector that is equivalent to the original matrix acting on the column vector on the right. These two statements are consistent since, for a Hermitian matrix the adjoint is identical to the original matrix (often, Hermitian matrices are called "self-adjoint" matrices for this reason) $\underline{M} = \underline{M}^\dagger$.

IX. Connection Between Orthonormal Vectors and Orthonormal Functions.

For vectors as we have been dealing with, the scalar or dot product is defined as we have seen as follows:

$$\langle \mathbf{v}(i) \mathbf{v}(j) \rangle = \sum_k v_k^*(i) v_k(j) .$$

For functions of one or more variable (we denote the variables collectively as x), the generalization of the vector scalar product is

$$\langle \mathbf{f}(i) \mathbf{f}(j) \rangle = \int f_i^*(x) f_j(x) dx .$$

The range of integration and the meaning of the variables x will be defined by the specific problem of interest; for example, in polar coordinates for one particle, $x = r, \theta$, and for N particles in a general coordinate system $x = \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$.

If the functions $f_i(x)$ are orthonormal;

$$\langle \mathbf{f}(i) \mathbf{f}(j) \rangle = \delta_{ij} ,$$

and complete, then a resolution of the identity holds that is similar to that for vectors

$$\int_i |\mathbf{f}(i)\rangle \langle \mathbf{f}(i)| = \delta(x-x') = \int_i f_i(x) f_i^*(x') dx$$

where $\delta(x-x')$ is called the Dirac delta function. The function is zero everywhere except at x' , but has unit area under it. In other words,

$$\begin{aligned} \int (x-x') g(x') dx' = g(x) &= \int_i |\mathbf{f}(i)\rangle \langle \mathbf{f}(i)| \mathbf{g}\rangle = \int_i f_i(x) \int f_i^*(x') g(x') dx' \\ &= \int_i f_i(x) a_i , \end{aligned}$$

where a_i is the projection of $g(x)$ along $f_i(x)$. The first equality can be taken as a definition of $\delta(x-x')$.

X. Matrix Representations of Functions and Operators

As we saw above, if the set of functions $\{f_i(x)\}$ is complete, any function of x can be expanded in terms of the $\{f_i(x)\}$

$$\begin{aligned}
g(x) &= \sum_i |f_i\rangle \langle f_i| g \rangle \\
&= \sum_i f_i(x) \int f_i^*(x') g(x') dx'
\end{aligned}$$

The column vector of numbers $a_i = \int f_i^*(x') g(x') dx'$ is called the representation of $g(x)$ in the $f_i(x)$ basis. Note that this vector may have an infinite number of components because there may be an infinite number of $f_i(x)$.

Suppose that the function $g(x)$ obeys some operator equation (e.g., an eigenvalue equation) such as

$$\frac{dg(x)}{dx} = g(x).$$

Then, if we express $g(x)$ in terms of the $\{f_i(x)\}$, the equation for $g(x)$ can be replaced by a corresponding matrix problem for the a_j vector coefficients:

$$\sum_i \frac{df_i(x)}{dx} \int f_i^*(x') g(x') dx' = \sum_i f_i(x) \int f_i^*(x') g(x') dx'$$

If we now multiply by $f_j^*(x)$ and integrate over x , we obtain

$$\int f_j^*(x) \frac{d}{dx} f_i(x) dx a_i = \int f_j^*(x) f_i(x) dx a_i.$$

If the $f_i(x)$ functions are orthonormal, this result simplifies to

$$\frac{d}{dx} a_i = a_j$$

where we have defined the matrix representation of $\frac{d}{dx}$ in the $\{f_i\}$ basis by

$$\frac{d}{dx} a_j = \sum_i \int f_i^*(x) \frac{d}{dx} f_i(x) dx a_i.$$

Remember that a_i is the representation of $g(x)$ in the $\{f_i\}$ basis. So the operator eigenvalue equation is equivalent to the matrix eigenvalue problem if the functions $\{f_i\}$ form a complete set.

Let us consider an example, that of the derivative operator in the orthonormal basis of Harmonic Oscillator functions. The fact that the solutions of the quantum Harmonic Oscillator, $\psi_n(x)$, are orthonormal and complete means that:

$$\int_{-\infty}^{\infty} \psi_n^*(x) \psi_m(x) dx = \delta_{nm}.$$

The lowest members of this set are given as

$$\psi_0(x) = \frac{1}{\sqrt{4}} e^{-\frac{x^2}{2}},$$

$$\psi_1(x) = \frac{1}{\sqrt{4}} \frac{1}{\sqrt{2}} x e^{-\frac{x^2}{2}},$$

$$\psi_2(x) = \frac{1}{\sqrt{4}} \frac{1}{\sqrt{8}} (4x^2 - 2) e^{-\frac{x^2}{2}}, \dots,$$

$$\psi_n(x) = A_n H_n(x) e^{-\frac{x^2}{2}}.$$

The derivatives of these functions are

$$\psi_0'(x) = \frac{1}{\sqrt{4}} (-x) e^{-\frac{x^2}{2}},$$

$$\psi_1'(x) = \frac{1}{\sqrt{4}} \frac{1}{\sqrt{2}} e^{-\frac{x^2}{2}} - x^2 e^{-\frac{x^2}{2}},$$

,

$$= \frac{1}{\sqrt{2}} (0 - 2) e^{-\frac{x^2}{2}}, \text{ etc.}$$

In general, one finds that

$$\frac{d}{dx} \psi_n(x) = \frac{1}{\sqrt{2}} \frac{1}{\sqrt{n}} \psi_{n-1}(x) - (n+1) \frac{1}{\sqrt{2}} \psi_{n+1}(x).$$

From this general result, it is clear that the matrix representation of the $\frac{d}{dx}$ operator is given by

$$\underline{D} = 2^{-\frac{1}{2}} \begin{pmatrix} 0 & 1 & 0 & 0 & \dots \\ -1 & 0 & \sqrt{2} & 0 & \dots \\ 0 & -\sqrt{2} & 0 & \sqrt{3} & \dots \\ 0 & 0 & -\sqrt{3} & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix} .$$

The matrix \underline{D} operates on the unit vectors $\mathbf{e}_0 = (1,0,0\dots)$, $\mathbf{e}_1 = (0,1,\dots)$ etc. just like $\frac{d}{dx}$ operates on $\psi_n(x)$, because these unit vectors are the representations of the $\psi_n(x)$ in the basis of the $\psi_n(x)$, and \underline{D} is the representation of $\frac{d}{dx}$ in the basis of the $\psi_n(x)$. Since any vector can be represented by a linear combination of the \mathbf{e}_i vectors, the matrix \underline{D} operates on any vector $(a_0, a_1, a_2\dots)$ just like $\frac{d}{dx}$ operates on any $f(x)$. Note that the matrix is not Hermitian; it is actually antisymmetric. However, if we multiply \underline{D} by $-i\hbar$ we obtain a Hermitian matrix that represents the operator $-i\hbar \frac{d}{dx}$, the momentum operator.

It is easy to see that we can form the matrix representation of any linear operator for any complete basis in any space. To do so, we act on each basis function with the operator and express the resulting function as a linear combination of the original basis functions. The coefficients that arise when we express the operator acting on the functions in terms of the original functions form the the matrix representation of the operator.

It is natural to ask what the eigenvalues and eigenfunctions of the matrix you form through this process mean. If your operator is the Hamiltonian operator, then the matrix eigenvectors and eigenvalues are the representations of the solutions to the Schrodinger equation in this basis. Forming the representation of an operator reduces the solution of the operator eigenvalue equation to a matrix eigenvalue equation.

XI. Complex Numbers, Fourier Series, Fourier Transforms, Basis Sets

One of the techniques to which chemists are frequently exposed is Fourier transforms. They are used in NMR and IR spectroscopy, quantum mechanics, and classical mechanics.

Just as we expand a function in terms of a complete set of basis function or a vector in terms of a complete set of vectors, the Fourier transform expresses a function $f(x)$ of a continuous variable x in terms of a set of orthonormal functions that are not discretely

labeled but which are labeled by a continuous "index" t . These functions are $(2\pi)^{-\frac{1}{2}} e^{-i t x}$, and the "coefficients" in the expansion

$$f(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{+\infty} e^{-i t x} f(t) dt ,$$

are called the Fourier transform $f(t)$ of $f(x)$.

The orthonormality of the $(\frac{1}{2}) e^{-i t}$ functions will be demonstrated explicitly later. Before doing so however, it is useful to review both complex numbers and basis sets.

A. Complex numbers

A complex number has a real part and an imaginary part and is usually denoted:

$$z = x + iy.$$

For the complex number z , x is the real part, y is the imaginary part, and $i = \sqrt{-1}$. This is expressed as $x = \text{Re}(z)$, $y = \text{Im}(z)$. For every complex number z , there is a related one called its complex conjugate $z^* = x - iy$.

Complex numbers can be thought of as points in a plane where x and y are the abscissa and ordinate, respectively. This point of view prompts us to introduce polar coordinates r and θ to describe complex numbers, with

$$x = r \cos \theta \quad r = \sqrt{x^2 + y^2}$$

or,

$$y = r \sin \theta \quad \theta = \tan^{-1} \frac{y}{x} + \pi \quad (\text{if } x < 0).$$

Another name for r is the norm of z which is denoted $|z|$. The angle θ is sometimes called the argument of z , $\arg(z)$, or the phase of z .

Complex numbers can be added, subtracted, multiplied and divided like real numbers. For example, the multiplication of z by z^* gives:

$$zz^* = (x + iy)(x - iy) = x^2 + ixy - ixy + y^2 = x^2 + y^2 = r^2$$

Thus $zz^* = |z|^2$ is a real number.

An identity due to Euler is important to know and is quite useful. It states that

$$e^{i\theta} = \cos \theta + i \sin \theta.$$

It can be proven by showing that both sides of the identity obey the same differential equation. Here we will only demonstrate its plausibility by Taylor series expanding both sides:

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots,$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \dots,$$

and

$$\cos x = 1 - \frac{x^2}{2} + \frac{x^4}{4!} + \dots$$

Therefore, the exponential $\exp(i x)$ becomes

$$\begin{aligned} e^{ix} &= 1 + i x - \frac{x^2}{2} + i^3 \frac{x^3}{3!} + i^4 \frac{x^4}{4!} + i^5 \frac{x^5}{5!} + \dots \\ &= 1 - \frac{x^2}{2} + \frac{x^4}{4!} + \dots + i \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} + \dots \right) \end{aligned}$$

The odd powers of i clearly combine to give the Sine function; the even powers give the Cosine function, so

$$e^{ix} = \cos x + i \sin x$$

is established. From this identity, it follows that $\cos x = \frac{1}{2}(e^{ix} + e^{-ix})$ and $\sin x = \frac{1}{2i}(e^{ix} - e^{-ix})$.

It is now possible to express the complex number z as

$$\begin{aligned} z &= x + iy \\ &= r \cos \theta + i r \sin \theta \\ &= r(\cos \theta + i \sin \theta) \\ &= r e^{i\theta} \end{aligned}$$

This form for complex numbers is extremely useful. For instance, it can be used to easily show that

$$z z^* = r e^{i\theta} (r e^{-i\theta}) = r^2 e^{i\theta - i\theta} = r^2.$$

B. Fourier Series

Now let us consider a function that is periodic in time with period T . Fourier's theorem states that any periodic function can be expressed in a Fourier series as a linear combination (infinite series) of Sines and Cosines whose frequencies are multiples of a fundamental frequency corresponding to the period:

$$f(t) = \sum_{n=0}^{\infty} a_n \cos(n \omega t) + \sum_{n=1}^{\infty} b_n \sin(n \omega t),$$

where $\omega = \frac{2\pi}{T}$. The Fourier expansion coefficients are given by projecting $f(t)$ along each of the orthogonal Sine and Cosine functions:

$$a_0 = \frac{1}{T} \int_0^T f(t) dt,$$

$$a_n = \frac{2}{T} \int_0^T f(t) \cos(n\omega t) dt,$$

$$b_n = \frac{2}{T} \int_0^T f(t) \sin(n\omega t) dt.$$

The term in the Fourier expansion associated with a_0 is a constant giving the average value (i.e., the zero-frequency or DC component) of the function. The terms with $n = 1$ contain the fundamental frequency and higher terms contain the n th harmonics or overtones of the fundamental frequency. The coefficients, a_n and b_n , give the amplitudes of each of these frequencies. Note that if $f(t)$ is an even function (i.e., if $f(t) = f(-t)$), $b_n = 0$ for $n = 1, 2, 3, \dots$ so the series only has Cosine terms. If $f(t)$ is an odd function (i.e. $f(t) = -f(-t)$), $a_n = 0$ for $n = 0, 1, 2, 3, \dots$ and the series only has Sine terms.

The Fourier series expresses a continuous function as an infinite series of numbers $\dots a_0, a_1, b_1, a_2, b_2, \dots$. We say that the set of coefficients is a representation of the function in the Fourier basis. The expansion in the $\cos n\omega t$ and $\sin n\omega t$ basis is useful because the basis functions are orthogonal when integrated over the interval 0 to T . An alternative set of functions is sometimes used to carry out the Fourier expansion; namely the $\frac{1}{\sqrt{T}} \exp(i n\omega t)$ functions for n from $-\infty$ to $+\infty$. Their orthogonality can be proven as follows:

$$\begin{aligned} \frac{1}{T} \int_0^T \exp(i n\omega t) \exp(-i m\omega t) dt &= \frac{1}{T} \int_0^T \exp(i(m-n)\omega t) dt = 1 \text{ if } m=n, \text{ and} \\ &= \frac{1}{T} (i(m-n)\omega)^{-1} (\exp(i(m-n)\omega T) - 1) = 0 \text{ if } m \neq n. \end{aligned}$$

Let us consider the Fourier representation of $f(t)$ in terms of the complex exponentials introduced above. For an arbitrary periodic function $f(t)$, we can write

$$f(t) = \sum_{n=-\infty}^{+\infty} c_n e^{i n\omega t}, \text{ where } c_n = \frac{1}{T} \int_0^T f(t) e^{-i n\omega t} dt.$$

This form of the Fourier series is entirely equivalent to the first form and the a_n and b_n can be obtained from the c_n and vice versa. For example, the c_n amplitudes are obtained by projecting $f(t)$ onto $\exp(in t)$ as:

$$c_n = \frac{1}{T} \int_0^T f(t)(\cos(n t) - i\sin(n t)) dt$$

$$= \frac{1}{T} \int_0^T f(t)\cos(n t)dt - \frac{1}{T} \int_0^T f(t)\sin(n t)dt ,$$

but these two integrals are easily recognized as the expansion in the other basis, so

$$c_n = \frac{1}{2} a_n - \frac{1}{2} i b_n = \frac{1}{2} (a_n - ib_n) .$$

By using complex exponential functions instead of trigonometric functions, we only have one family of basis functions, $e^{in t}$, but we have twice as many of them. In this form, if $f(t)$ is even, then the c_n are real, but if $f(t)$ is odd, then c_n are imaginary.

It is useful to consider some examples to learn this material. First, let $f(t)$ be the odd function $f(t) = \sin 3t$. Then, one period has elapsed when $3T = 2\pi$, so $T = \frac{2\pi}{3}$ and $\omega = 3$. For this function, the complex series Fourier expansion coefficients are given by

$$c_0 = \frac{3}{2} \int_0^{\frac{2\pi}{3}} \sin 3t dt = 0.$$

$$c_1 = \frac{3}{2} \int_0^{\frac{2\pi}{3}} \sin 3t e^{-i3t} dt = \frac{3}{2} \frac{(e^{i3t} - e^{-i3t})}{2i} e^{-i3t} dt$$

$$= \frac{1}{2i} \frac{3}{2} \int_0^{\frac{2\pi}{3}} (1 - e^{6it}) dt = \frac{3}{4i} \left[\frac{2\pi}{3} - 0 \right] = \frac{1}{2i} = -\frac{i}{2}$$

Because $\sin 3t$ is real, it is straightforward to see from the definition of c_n that $c_{-n} = c_n^*$, as a result of which $c_{-1} = c_1^* = +\frac{i}{2}$. The orthogonality of the $\exp(in t)$ functions can be used to show that all of the higher c_n coefficients vanish:

$$c_n = 0 \text{ for } n \neq \pm 1.$$

Hence we see that this simple periodic function has just two terms in its Fourier series. In terms of the Sine and Cosine expansion, one finds for this same $f(t) = \sin 3t$ that $a_n = 0$, $b_n = 0$ for $n \neq 1$, and $b_1 = 1$.

As another example, let $f(t)$ be

$$f(t) = t, \quad -\pi < t < \pi,$$

and make $f(t)$ periodic with period 2π (so $\omega = 1$). For this function,

$$c_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} t dt = 0$$

$$\begin{aligned} c_n &= \frac{1}{2\pi} \int_{-\pi}^{\pi} t e^{-int} dt = \frac{1}{2\pi n^2} [e^{-int}(1 + int)] \Big|_{-\pi}^{\pi} \\ &= \frac{1}{2\pi n^2} (e^{-in\pi} (1 + in\pi) - e^{in\pi} (1 - in\pi)) \\ &= \frac{1}{2\pi n^2} ((-1)^n (1 + in\pi) - (-1)^n (1 - in\pi)) \\ &= \frac{(-1)^n}{2\pi n^2} (2in\pi) = \frac{i}{n} (-1)^n \pi \neq 0. \end{aligned}$$

Note that since t is an odd function, c_n is imaginary, and as $n \rightarrow \infty$, c_n approaches zero.

C. Fourier Transforms

Let us now imagine that the period T becomes very long. This means that the fundamental frequency $\omega = \frac{2\pi}{T}$ becomes very small and the harmonic frequencies $n\omega$ are very close together. In the limit of an infinite period one has a non-periodic function (i.e., an arbitrary function). The frequencies needed to represent such a function become continuous, and hence the Fourier sum becomes an integral.

To make this generalization more concrete, we first take the earlier Fourier expression for $f(t)$ and multiply it by unity in the form $\frac{T}{2\pi}$ to obtain:

$$f(t) = \sum c_n e^{in\omega t} \frac{T}{2\pi}.$$

Now we replace n (which are frequencies infinitesimally close together for sufficiently long T) by the continuous index ω :

$$f(t) = \frac{1}{2} \sum_{n=-\infty}^{\infty} (c_n T) e^{i n \omega t} .$$

In the limit of long T , this sum becomes an integral because ω becomes infinitesimally small ($\omega \rightarrow d\omega$). As T grows, the frequency spacing between n and $n+1$, which is $\frac{2\pi}{T}$, as well as the frequency associated with a given n -value become smaller and smaller. As a result, the product $c_n T$ evolves into a continuous function of ω which we denote $c(\omega)$. Before, c_n depended on the continuous index n , and represented the contribution to the function $f(t)$ of waves with frequency n . The function $c(\omega)$ is the contribution per unit frequency to $f(t)$ from waves with frequency in the range ω to $\omega + d\omega$. In summary, the Fourier series expansion evolves into an integral that defines the Fourier transformation of $f(t)$:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} c(\omega) e^{-i \omega t} d\omega .$$

It is convenient to define a new function $f(\omega) = (2\pi)^{-\frac{1}{2}} c(\omega)$ where $f(\omega)$ is called the Fourier transform of $f(t)$. The Fourier transform $f(\omega)$ is a representation of $f(t)$ in another basis, that of the orthonormal set of oscillating functions $e^{i \omega t} (2\pi)^{-\frac{1}{2}}$:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\omega) e^{-i \omega t} d\omega .$$

Earlier, for Fourier series, we had the orthogonality relation among the Fourier functions:

$$\frac{1}{T} \int_0^T e^{i n \omega t} e^{-i m \omega t} dt = \delta_{nm} ,$$

but for the continuous variable ω , we have a different kind of orthogonality

$$\int_{-\infty}^{\infty} e^{i \omega_1 t} e^{-i \omega_2 t} dt = 2\pi \delta(\omega_1 - \omega_2) ,$$

where $\delta(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t} f(t) dt$.

The function $\delta(t - t_0)$, called the Dirac delta function, is the continuous analog to δ_{nm} . It is zero unless $t = t_0$. If $t = t_0$, $\delta(t - t_0)$ is infinite, but it is infinite in such a way that the area under the curve is precisely unity. Its most useful definition is that $\delta(t - t_0)$ is the function which, for arbitrary $f(t)$, the following identity holds:

$$\int_{-\infty}^{\infty} f(t) \delta(t - t_0) dt = f(t_0).$$

That is, integrating $\delta(t - t_0)$ times any function evaluated at t just gives back the value of the function at t_0 .

The Dirac delta function can be expressed in a number of convenient ways, including the following two:

$$\begin{aligned} \delta(t - t_0) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega(t - t_0)} d\omega \\ &= \lim_{a \rightarrow 0} \frac{1}{\sqrt{a}} \int_0^{\infty} e^{-\frac{a}{2}(t - t_0)^2} dt \end{aligned}$$

As an example of applying the Fourier transform method to a non-periodic function, consider the localized pulse

$$f(t) = \begin{cases} \frac{1}{T} & -\frac{T}{2} \leq t \leq \frac{T}{2} \\ 0 & t > \frac{T}{2} \end{cases}$$

For this function, the Fourier transform is

$$\begin{aligned} F(\omega) &= \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \\ &= \int_{-\frac{T}{2}}^{\frac{T}{2}} \frac{1}{T} e^{-i\omega t} dt \end{aligned}$$

$$\begin{aligned}
&= (2\pi)^{-\frac{1}{2}} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} (\cos t - i \sin t) dt \\
&= (2\pi)^{-\frac{1}{2}} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \cos t dt + 0 \\
&= (2\pi)^{-\frac{1}{2}} \frac{1}{T} (\sin t) \Big|_{-\frac{T}{2}}^{\frac{T}{2}} \\
&= \frac{2}{T}^{-\frac{1}{2}} \sin(\pi/2) / T.
\end{aligned}$$

Note that $f(\omega)$ has its maximum value of $(2\pi)^{-\frac{1}{2}}$ for $\omega = 0$ and that $f(\omega)$ falls slowly in magnitude to zero as ω increases in magnitude, while oscillating to positive and negative values. However, the primary maximum in $f(\omega)$ near zero-frequency has a width that is inversely proportional to T . This inverse relationship between the width (T) in t -space and the width $\frac{2}{T}$ in ω -space is an example of the uncertainty principle. In the present case it means that if you try to localize a wave in time, you must use a wide range of frequency components.

D. Fourier Series and Transforms in Space

The formalism we have developed for functions of time can also be used for functions of space variables or any other variables for that matter. If $f(x)$ is a periodic function of the coordinate x , with period (repeat distance) $\frac{2}{K}$, then it can be represented in terms of Fourier functions as follows:

$$f(x) = \sum_{n=-\infty}^{\infty} f_n e^{inKx}$$

where the coefficients are

$$f_n = \frac{K}{2} \int_0^{\frac{2}{K}} f(x) e^{-inKx} dx .$$

If $f(x)$ is a non-periodic function of the coordinate x , we write it as

$$f(x) = \int_{-\infty}^{\infty} f(k) e^{ikx} dk$$

and the Fourier transform is given as

$$f(k) = \int_{-\infty}^{\infty} f(x) e^{-ikx} dx$$

If f is a function of several spatial coordinates and/or time, one can Fourier transform (or express as Fourier series) simultaneously in as many variables as one wishes. You can even Fourier transform in some variables, expand in Fourier series in others, and not transform in another set of variables. It all depends on whether the functions are periodic or not, and whether you can solve the problem more easily after you have transformed it.

E. Comments

So far we have seen that a periodic function can be expanded in a discrete basis set of frequencies and a non-periodic function can be expanded in a continuous basis set of frequencies. The expansion process can be viewed as expressing a function in a different basis. These basis sets are the collections of solutions to a differential equation called the wave equation. These sets of solutions are useful because they are complete sets. Completeness means that any arbitrary function can be expressed exactly as a linear combination of these functions. Mathematically, completeness can be expressed as

$$1 = \int_{-\infty}^{\infty} \delta(x) dx$$

in the Fourier transform case, and

$$1 = \sum_n \delta_{n,n}$$

in the Fourier series case.

The only limitation on the function expressed is that it has to be a function that has the same boundary properties and depends on the same variables as the basis. You would not want to use Fourier series to express a function that is not periodic, nor would you want to express a three-dimensional vector using a two-dimensional or four-dimensional basis.

Besides the intrinsic usefulness of Fourier series and Fourier transforms for chemists (e.g., in FTIR spectroscopy), we have developed these ideas to illustrate a point that is important in quantum chemistry. Much of quantum chemistry is involved with basis sets and expansions. This has nothing in particular to do with quantum mechanics. Any time one is dealing with linear differential equations like those that govern light (e.g. spectroscopy) or matter (e.g. molecules), the solution can be written as linear combinations of complete sets of solutions.

XII. Spherical Coordinates

A. Definitions

The relationships among cartesian and spherical polar coordinates are given as follows:

$$z = r \cos \theta \quad r = \sqrt{x^2 + y^2 + z^2}$$

$$x = r \sin \theta \cos \phi \quad \theta = \cos^{-1} \frac{z}{\sqrt{x^2 + y^2 + z^2}}$$

$$y = r \sin \theta \sin \phi \quad \phi = \cos^{-1} \frac{x}{\sqrt{x^2 + y^2}}$$

The ranges of the polar variables are $0 < r < \infty$, $0 < \theta < \pi$, $0 < \phi < 2\pi$.

B. The Jacobian in Integrals

In performing integration over all space, it is necessary to convert the multiple integral from cartesian to spherical coordinates:

$$dx \, dy \, dz \, f(x,y,z) = dr \, d\theta \, d\phi \, f(r, \theta, \phi) \, J,$$

where J is the so-called Jacobian of the transformation. J is computed by forming the determinant of the three-by-three matrix consisting of the partial derivatives relating x, y, z to r, θ, ϕ :

$$J = \frac{\partial(x,y,z)}{\partial(r, \theta, \phi)}$$

$$= \begin{vmatrix} \sin \theta \cos \phi & \sin \theta \sin \phi & \cos \theta \\ r \cos \theta \cos \phi & r \cos \theta \sin \phi & -r \sin \theta \\ -r \sin \theta \cos \phi & r \sin \theta \sin \phi & 0 \end{vmatrix}$$

The determinant, J , can be expanded, for example, by the method of diagonals, giving four nonzero terms:

$$J = r^2 \sin^3 \theta \cos^2 \phi + r^2 \cos^2 \theta \sin \theta \cos^2 \phi$$

$$+ r^2 \sin^3 \theta \sin^2 \phi + r^2 \cos^2 \theta \sin \theta \sin^2 \phi$$

$$= r^2 \sin \theta (\sin^2 \theta (\sin^2 \phi + \cos^2 \phi) + \cos^2 \theta (\cos^2 \phi + \sin^2 \phi))$$

$$= r^2 \sin \theta \, dr \, d\theta \, d\phi$$

Hence in converting integrals from x, y, z to r, θ, ϕ , one writes as a short hand $dx dy dz = r^2 \sin \theta \, dr \, d\theta \, d\phi$.

C. Transforming Operators

In many applications, derivative operators need to be expressed in spherical coordinates. In converting from cartesian to spherical coordinate derivatives, the chain rule is employed as follows:

$$\begin{aligned} \frac{\partial}{\partial x} &= \frac{\partial r}{\partial x} \frac{\partial}{\partial r} + \frac{\partial \theta}{\partial x} \frac{\partial}{\partial \theta} + \frac{\partial \phi}{\partial x} \frac{\partial}{\partial \phi} \\ &= \sin \theta \cos \phi \frac{\partial}{\partial r} + \frac{\cos \theta \cos \phi}{r} \frac{\partial}{\partial \theta} - \frac{\sin \phi}{r \sin \theta} \frac{\partial}{\partial \phi} \end{aligned}$$

Likewise

$$\frac{\partial}{\partial y} = \sin \theta \sin \phi \frac{\partial}{\partial r} + \frac{\cos \theta \sin \phi}{r} \frac{\partial}{\partial \theta} + \frac{\cos \phi}{r \sin \theta} \frac{\partial}{\partial \phi},$$

and

$$\frac{\partial}{\partial z} = \cos \theta \frac{\partial}{\partial r} - \frac{\sin \theta}{r} \frac{\partial}{\partial \theta} + 0 \frac{\partial}{\partial \phi}$$

Now to obtain an expression for $\frac{\partial^2}{\partial x^2}$ and other second derivatives, one needs to take the following derivatives:

$$\begin{aligned} \frac{\partial^2}{\partial x^2} &= \sin \theta \cos \phi \frac{\partial}{\partial r} \sin \theta \cos \phi \frac{\partial}{\partial r} + \frac{\cos \theta \cos \phi}{r} \frac{\partial}{\partial \theta} \frac{\partial}{\partial \theta} - \frac{\sin \phi}{r \sin \theta} \frac{\partial}{\partial \phi} \frac{\partial}{\partial \phi} \\ &\quad + \frac{\cos \theta \cos \phi}{r} \frac{\partial}{\partial r} \sin \theta \cos \phi \frac{\partial}{\partial r} + \frac{\cos \theta \cos \phi}{r} \frac{\partial}{\partial \theta} \frac{\partial}{\partial \theta} - \frac{\sin \phi}{r \sin \theta} \frac{\partial}{\partial \phi} \frac{\partial}{\partial \phi} \\ &\quad - \frac{\sin \phi}{r \sin \theta} \frac{\partial}{\partial r} \sin \theta \cos \phi \frac{\partial}{\partial r} + \frac{\cos \theta \cos \phi}{r} \frac{\partial}{\partial \theta} \frac{\partial}{\partial \theta} - \frac{\sin \phi}{r \sin \theta} \frac{\partial}{\partial \phi} \frac{\partial}{\partial \phi} \\ &= \sin^2 \theta \cos^2 \phi \frac{\partial^2}{\partial r^2} + \frac{\cos^2 \theta \cos^2 \phi}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{\sin^2 \phi}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \end{aligned}$$

$$\begin{aligned}
& + \frac{2\sin\theta \cos\theta \cos^2\theta}{r} \frac{\partial^2}{\partial r^2} - \frac{2\cos\theta \sin\theta}{r} \frac{\partial^2}{\partial r^2} - \frac{2\cos\theta \cos\theta \sin\theta}{r^2 \sin^2\theta} \frac{\partial^2}{\partial \theta^2} \\
& + \frac{\cos^2\theta \cos^2\theta}{r} + \frac{\sin^2\theta}{r} \frac{\partial^2}{\partial r^2} + \frac{2\sin\theta \cos\theta \cos^2\theta}{r^2} + \frac{\sin^2\theta \cos\theta}{r^2 \sin^2\theta} \frac{\partial^2}{\partial \theta^2} \\
& + \frac{\cos\theta \sin\theta}{r^2} + \frac{\cos^2\theta \cos\theta \sin\theta}{r^2 \sin^2\theta} \frac{\partial^2}{\partial \theta^2}
\end{aligned}$$

Analogous steps can be performed for $\frac{\partial^2}{\partial y^2}$ and $\frac{\partial^2}{\partial z^2}$. Adding up the three contributions, one obtains:

$$\Delta = \frac{1}{r^2} \frac{\partial^2}{\partial r^2} + \frac{1}{r^2 \sin^2\theta} \frac{\partial^2}{\partial \theta^2} + \frac{1}{r^2 \sin^2\theta} \frac{\partial^2}{\partial \phi^2}$$

As can be seen by reading Appendix G, the terms involving angular derivatives in Δ is identical to $-\frac{L^2}{r^2}$, where L^2 is the square of the rotational angular momentum operator.

Although in this appendix, we choose to treat it as a collection of differential operators that gives rise to differential equations in r, θ, ϕ to be solved, there are more general tools for treating all such angular momentum operators. These tools are developed in detail in Appendix G and are used substantially in the next two chapters.

XIII. Separation of Variables

In solving differential equations such as the Schrödinger equation involving two or more variables (e.g., equations that depend on three spatial coordinates $x, y,$ and z or r, θ, ϕ , and t or that depend on time t and several spatial variables denoted \mathbf{r}), it is sometimes possible to reduce the solution of this one multi-variable equation to the solution of several equations each depending on fewer variables. A commonly used device for achieving this goal is the separation of variables technique.

This technique is not always successful, but can be used for the type of cases illustrated now. Consider a two-dimensional differential equation that is of second order and of the eigenvalue type:

$$A \frac{\partial^2}{\partial x^2} + B \frac{\partial^2}{\partial y^2} + C \frac{\partial}{\partial x} + D \frac{\partial}{\partial y} + E = 0$$

The solution $\psi(x,y)$ must be a function of x and y because the differential equation refers to ψ 's derivatives with respect to these variables.

The separations of variables device assumes that $\psi(x,y)$ can be written as a product of a function of x and a function of y :

$$(x,y) = (x) (y).$$

Inserting this ansatz into the above differential equation and then dividing by $(x) (y)$ produces:

$$A^{-1} \frac{2}{x^2} + B^{-1} \frac{2}{y^2} + C^{-1} \frac{1}{x} \frac{1}{y} + D = E.$$

The key observations to be made are:

A. If A is independent of y, $A^{-1} \frac{2}{x^2}$ must be independent of y.

B. If B is independent of x, $B^{-1} \frac{2}{y^2}$ must be independent of x.

C. If C vanishes and D does not depend on both x and y, then there are no "cross terms" in the above equations (i.e., terms that contain both x and y). For the sake of argument, let us take D to be dependent on x only for the remainder of this discussion; the case for which D depends on y only is handled in like manner.

Under circumstances for which all three of the above conditions are true, the left-hand side of the above second-order equation in two variables can be written as the sum of

$$A^{-1} \frac{2}{x^2} + D,$$

which is independent of y, and

$$B^{-1} \frac{2}{y^2}$$

which is independent of x.

The full equation states that the sum of these two pieces must equal the eigenvalue E, which is independent of both x and y.

Because E and $B^{-1} \frac{2}{y^2}$ are both independent of x, the quantity $A^{-1} \frac{2}{x^2} + D$ (as a whole) can not depend on x, since $E - B^{-1} \frac{2}{y^2}$ must equal $A^{-1} \frac{2}{x^2} + D$ for all values of x and y. That $A^{-1} \frac{2}{x^2} + D$ is independent of x and (by assumption) independent of y allows us to write:

$$A^{-1} \frac{2}{x^2} + D = \text{constant},$$

a constant.

Likewise, because E and $A^{-1} \frac{2}{x^2} + D$ are independent of y, the quantity $B^{-1} \frac{2}{y^2}$ (as a whole) can not depend on y, since $E - A^{-1} \frac{2}{x^2} - D$ must equal $B^{-1} \frac{2}{y^2}$ for all values of x and y. That $B^{-1} \frac{2}{y^2}$ is independent of y and (by assumption) independent of x allows us to write:

$$B^{-1} \frac{2}{y^2} = \text{constant},$$

another constant.

The net result is that we now have two first-order differential equations of the eigenvalue form:

$$A \frac{d^2}{dx^2} + D = \quad ,$$

and

$$B \frac{d^2}{dy^2} = \quad ' ,$$

and the solution of the original equation has been successfully subjected to separation of variables. The two eigenvalues λ and μ of the separated x- and y- equations must obey:

$$\lambda + \mu = E,$$

which follows by using the two separate (x and y) eigenvalue equations in the full two-dimensional equation that contains E as its eigenvalue.

In summary, when separations of variables can be used, it:

- A. Reduces one multidimensional differential equation to two or more lower-dimensional differential equations.
- B. Expresses the eigenvalue of the original equation as a sum of eigenvalues (whose values are determined via boundary conditions as usual) of the lower-dimensional problems.
- C. Expresses the solutions to the original equation as a product of solutions to the lower-dimensional equations (i.e., $\psi(x,y) = \psi(x) \psi(y)$, for the example considered above).