Mathematics for Quantum Chemistry

JAY MARTIN ANDERSON
Bryn Mawr College

W. A. Benjamin, Inc.
NEW YORK AMSTERDAM
1966

mechanics and its applications to molecular motion from physics. I have selected these four topics because of their relevance to modern quantum chemistry, especially in the application of quantum mechanics to molecular spectroscopy. This emphasis on molecular spectroscopy betrays my personal interest and excitement in this growing and popular field of endeavor; it also eliminates from the pages of this brief book a consideration of other topics which may be equally stimulating to my colleagues and to their students. Relativity, electricity, magnetism, and radiation physics were eliminated because they are generally better treated elsewhere and in greater depth than this work allows; similarly, group theory and differential equations, including approximate methods of solution, are left to other treatises.

This book attempts to lay down a central core of physical and mathematical background for quantum chemistry in general, but for molecular spectroscopy in particular. It assumes a knowledge of calculus through partial derivatives and multiple integration (usually about one and one-half years), a year of physics, and chemistry through a year of physical chemistry. This material has been used as the basis of a one-semester course at Bryn Mawr College entitled "Applied Mathematics for Chemists" for students with approximately the indicated background; this course immediately precedes the first course in quantum mechanics.

The author is indebted to Addison-Wesley Publishing Co. for permission to quote from their publications, and W. A. Benjamin, Inc. for their continued help and encouragement.

JAY MARTIN ANDERSON

Bryn Mawr, Pennsylvania
October 1965

Contents

1

Introduction

1-1 EIGENVALUE PROBLEMS IN QUANTUM MECHANICS

The mathematics and physics that are relevant to quantum chemistry are, almost without exception, oriented toward the solution of a particular kind of problem, the calculation of properties of a molecular system from the fundamental properties (charge, mass) of the particles composing the system. A good example of this is the calculation of the *energy* of the electrons in a molecule, using only the charge of the electron, Planck's constant, and so forth. The reader is probably already aware of the nature of the answer to this problem. There are a number of discrete values for the energy which the electrons in the molecule can assume up to a point, but higher values for the electronic energy occur in a continuous range. These energy values are shown qualitatively in Fig. 1-1. Quantum mechanics does provide the result that some physical quantities may take on only *some* values, not *all* values, as experiments indicated. The allowed values for a physical quantity are called *eigenvalues*, from the German for *characteristic values*. A particular physical quantity may assume an eigenvalue from a continuum, or perhaps from a finite or infinite discrete set of eigenvalues. The energy of an atom, for instance, may take on one of an

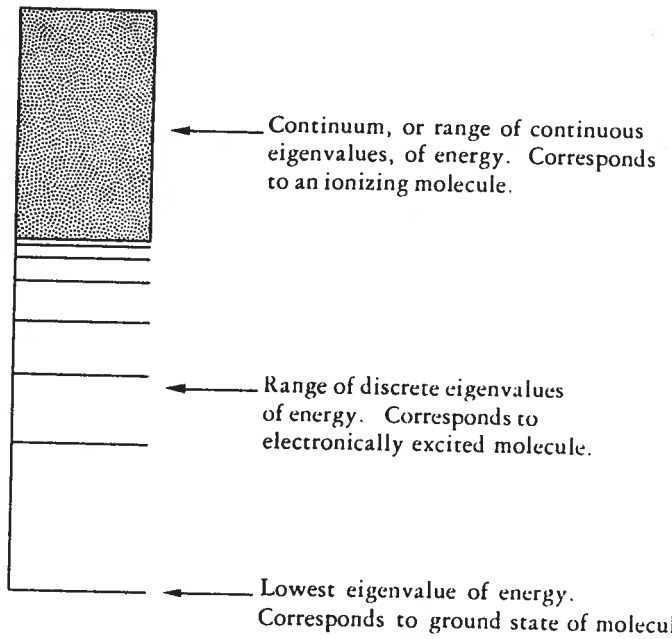


Figure 1–1 Eigenvalues of the energy of a molecule.

infinite number of discrete values, as well as values from a higher-lying range of continuous eigenvalues, called the *continuum*. More often than not, chemistry is concerned with the discrete eigenvalues of a quantity, rather than its continuum of eigenvalues.

The mathematical problem of finding the eigenvalues of a quantity is called an *eigenvalue problem;* it is usually cast in the form of an equation called an *eigenvalue equation*. An eigenvalue equation for a physical quantity Q has the deceptively simple appearance

$$\mathfrak{Q}f = qf \tag{1–1}$$

In this equation, f is a function, called the *eigenfunction* for the quantity Q, with the eigenvalue q. The element $\mathfrak{Q}$ is called an *operator*, and the statement $\mathfrak{Q}f$ tells us to change the function f into a new function, according to a set of instructions implicit in the definition of the operator $\mathfrak{Q}$. The eigenvalue equation, Eq. 1–1, then informs

us that, by applying these "instructions" of the operator $\mathfrak{Q}$ to f, we get merely a multiple, q, of the function f. The function $\mathfrak{Q}f$ differs from the function f by a multiplicative constant q. It may very well be the case that several eigenfunctions have the same eigenvalue; that is, $\mathfrak{Q}f_1 = qf_1$, $\mathfrak{Q}f_2 = qf_2$, and so forth. If this is the case, the eigenvalue q is said to be *degenerate*, and the number of eigenfunctions that have the same eigenvalue is called the *degree of degeneracy*.

Operators may simply be numbers or functions; for example, the operator $\mathfrak{X}$ may be defined by the instruction "multiply the operand function by x"; thus, $\mathfrak{X}\, x^2 = x^3$. On the other hand, operators may be more complex than just numbers or functions. For example, the student has already used the operator (although probably not by that name) Δ, which means, or is defined by the instructions, "evaluate the change in." For example, if we operate Δ on the thermodynamic function H, the enthalpy, we get a new function ΔH, the change in the enthalpy, $\Delta H = H_2 - H_1$. Another operator that is familiar is d/dx, meaning, "evaluate the derivative with respect to x."

It is the job of quantum mechanics to tell us how to form operators corresponding to the physical quantities which we wish to measure. Our task for the moment will be to learn how to solve the eigenvalue equations for such operators, and especially the vocabulary and concepts that are used to discuss the solutions. Quantum mechanics itself, however, grew up from two different points of view, which represent two analogous mathematical formulations of eigenvalue problems.

The first of these points of view is the *wave mechanics* of Schrödinger In wave mechanics, operators are differential expressions, such as the operator d/dx referred to above, and the eigenvalue equation then takes the form of a differential equation, and relies on the calculus for its solution. The second formulation is the *matrix mechanics* of Heisenberg, in which operators are represented by algebraic entities called *matrices;* instead of a function in the eigenvalue equation, the matrix operator operates on a *vector* ξ to transform ξ into a vector parallel to ξ, but q times as long:

$$\mathfrak{Q}\xi = q\xi \tag{1–2}$$

Equation 1–2 is the matrix-mechanical formulation of the eigenvalue

problem. Matrices and vectors are defined and discussed in detail in Chapter 3. As in Eq. 1–1, q is the eigenvalue of the quantity Q, ξ is the *eigenvector*, and $\mathcal{Q}$ is the operator represented as a matrix. The solution of this form of the eigenvalue problem relies on algebra.

These apparently different mathematical and physical approaches to quantum mechanical problems are really deeply interrelated; the work of Dirac shows the underlying equivalence of the two points of view, as well as of the corresponding mathematical techniques.

1–2 EIGENVALUE PROBLEMS IN CLASSICAL MECHANICS

We have briefly discussed the role of eigenvalue equations in quantum mechanics. But a number of problems of classical mechanics may also be expressed in a simple and meaningful way as eigenvalue problems. Among these are the problems of the vibrations and rotations of a mechanical system, such as a molecule. These physical problems are of importance to the chemist concerned with molecular motion and spectroscopy. In vibrations, the normal modes and frequencies of oscillation appear as eigenvectors and eigenvalues; in rotations, the principal axes and moments of inertia emerge from an eigenvalue problem. It should be noted, however, that a correct description of these systems on the molecular level nearly always requires quantum mechanics, not classical mechanics.

1–3 SCOPE OF THIS BOOK

With our course thus determined by the kinds of problems we wish to be able to set up, solve, and understand, we shall proceed first to a study of a certain class of functions germane to eigenfunction problems, then to a number of aspects of vector algebra and matrix algebra, finally to a synthesis of the two points of view of eigenvalue problems. We shall conclude with a study of classical mechanics to see how the vibrations of a mechanical system, such as a molecule, may be formulated as an eigenvalue problem. We shall also attempt to formulate Newtonian mechanics in such a way that the connection to quantum mechanics is clear.

Along the way, we shall learn some methods of solving eigenvalue problems, and take up applications of interest in chemistry. Our emphasis throughout will be primarily on concepts, secondarily on methods, and only lastly on the detailed proofs of the mathematical theorems. At the end of each chapter, a set of problems is given. Answers and hints for solution for many of the problems are found at the back of the book.

Problem

1. Find the eigenfunctions of the operator d/dx.

2

Orthogonal Functions

Two properties are, almost without exception, possessed by the eigenfunctions of operators corresponding to important physical quantities: *orthogonality* and *normality*. The purpose of this chapter is to develop these concepts in detail and to illustrate a number of their applications. Of primary usefulness is the idea of an *expansion in orthogonal functions*. As an example of this technique, we shall examine the *Fourier series* in some detail. We shall also learn how to construct orthogonal functions by the *Schmidt orthogonalization procedure*, and how orthogonal functions arise from the solution of particular differential equations. To illustrate the latter concepts, we shall investigate the properties of the *Legendre polynomials*, and briefly mention other of the important "special functions" which arise in quantum chemistry. A brief discussion of some of the elements of the calculus and of complex variables are given in the Appendix. The reader would be wise to check his familiarity with this material before advancing into the present chapter.

2-1 INTRODUCTORY CONCEPTS: ORTHOGONALITY AND NORMALIZATION

We may best begin our discussion of orthogonal functions by reviewing the concept of *function*. The concept of function has three essential ingredients. We agree first to define a function on a particular region of the number scale, say, from a to b. Second, we agree that there exists a variable (say, x) that can independently assume values in the region from a to b. Third, we agree by some prescribed rule that for any value of x there exists a *definite* value of y. Then we say that y is a function of x on the range $a \leq x \leq b$. This definition may be modified in a number of ways—so as to include more than one independent variable—but these three essential ingredients persist: an independent variable; a range on which the independent variable assumes its values; a dependent variable related to the independent variable by a prescribed rule.

The simplest way of notating the statement "y is a function of x" is to write $y = y(x)$. This notation is compact, yet may be misleading. The left side of the equation is simply the name of a variable—we do not know it is the dependent variable until we see the right side of the equation. The right side uses the letter y again, but here the symbol $y(\ )$ means something different than just the name of the variable. The meaning of $y(\ )$ is that y is a dependent variable whose value may be found by some prescribed rule from the quantity inside the parentheses. Left out of the notation $y = y(x)$ is the interval, or range, of the independent variable x for which the functional relationship is defined. This is not always of importance in elementary considerations of the idea of function, but it is of supreme importance to the notion of expansion of a function.

Hence, we introduce a definition.

Definition *Expansion interval* (or, simply, *interval*). **The expansion interval is the range of the independent variable assumed by the functions under consideration. This does not imply that the function may not be defined for other values of the independent variable; we just decline to consider those other values.**

The expansion interval is usually notated $[a, b]$, meaning that the independent variable x is allowed values on the range $a \leq x \leq b$.

We proceed now to four definitions in rapid succession.

Definition *Inner product.* **The inner product of the two (in general, complex-valued) functions f and g of a continuous variable on their expansion interval $[a, b]$ is**

$$\langle f \mid g \rangle = \int_a^b f(x)^* g(x)\, dx \quad . \tag{2-1}$$

The inner product of two functions is defined *on their expansion interval*. The inner product is notated by some authors (f, g), but this can easily be confused with the notation for two-dimensional coordinates or for an open interval. We shall use the notation $\langle f \mid g \rangle$. The order is quite important:

$$\langle g \mid f \rangle = \int g(x)^* f(x)\, dx = \left(\int f(x)^* g(x)\, dx\right)^* = \langle f \mid g \rangle^* \tag{2-2}$$

For real-valued functions, the order is not important. Equation 2-2 illustrates an important feature of the inner product that arises again and again: "turning around," or *transposing an inner product gives the complex conjugate of that inner product*. Constants may be removed at will from the inner product symbol: if b and c are (complex) numbers, $\langle bf \mid cg \rangle = b^* c \langle f \mid g \rangle$.

The inner product is a concept of no small significance. It has a geometrical analog, that of the *dot product* or *scalar product* of vectors that may already be familiar, which we shall discuss in Chapter 3.

In analogy to the geometrical property of perpendicularity of vectors, both functions and vectors afford the sweeping and general concept of *orthogonality*.

Definition Two functions, $f(x)$ and $g(x)$, are said to be *orthogonal* on the interval $[a, b]$ if their inner product on $[a, b]$ is zero:

$$\langle f \mid g \rangle = \int_a^b f^* g = 0 = \int_a^b g^* f = \langle g \mid f \rangle \tag{2-3}$$

If the inner product is to be zero, it does not matter which function "comes first" in the inner product, so the orthogonality of f and g may be expressed by either $\langle f \mid g \rangle = 0$ or $\langle g \mid f \rangle = 0$. The perpendicularity of two vectors is related to this definition of orthogonality: two vectors are perpendicular if their dot product is zero.

Definition The *norm* of a function on the interval $[a, b]$ is the inner product of the function with itself, and may be symbolized by N:

$$N(f) = \langle f \mid f \rangle = \int_a^b f^* f \tag{2-4}$$

The norm of a function is a real, positive quantity; it is analogous to the square of the length of a vector. That the norm is real and positive may be easily demonstrated by

$$f^* f = (\mathrm{Re}\, f - i\, \mathrm{Im}\, f)(\mathrm{Re}\, f + i\, \mathrm{Im}\, f) = (\mathrm{Re}\, f)^2 + (\mathrm{Im}\, f)^2 \tag{2-5}$$

which is positive definite. Then the integral of $f^* f$, which gives the norm of f, is also positive definite. The positiveness of the norm is of use to us at once.

Definition A function is said to be *normalized*[1] if its norm is one; that is, if $\langle f \mid f \rangle = 1$.

Since the norm of a function on a particular interval is always simply a positive real number, we can always form a multiple of a given function which is normalized. Suppose f has a norm N. Then the function $f/N^{1/2}$ will have a norm of one, since

$$\left\langle \frac{f}{N^{1/2}} \,\middle|\, \frac{f}{N^{1/2}} \right\rangle = \frac{1}{N}\langle f \mid f \rangle = \frac{N}{N} = 1 \tag{2-6}$$

The process of dividing a function by the square root of its norm is called *normalizing the function*, or, sometimes, normalizing the function to unity.

Let us use the five definitions we have introduced thus far in some examples. Suppose we consider functions defined on the interval $[-1, 1]$. As an example of the computation of an inner product, let us evaluate $\langle x \mid x^2 \rangle$.

$$\langle x \mid x^2 \rangle = \int_{-1}^{+1} x^* x^2\, dx = \int_{-1}^{+1} x^3\, dx = \frac{x^4}{4}\Big|_{-1}^{+1} = 0 \tag{2-7}$$

The computation of this simple inner product gives zero. We therefore may state that, on $[-1, 1]$, x and x^2 are orthogonal functions. Notice the importance of specifying the interval: on the interval $[0, 1]$, the inner product $\langle x \mid x^2 \rangle$ is

$$\langle x \mid x^2 \rangle = \int_0^1 x^3 = \frac{x^4}{4}\Big|_0^1 = \frac{1}{4} \tag{2-8}$$

[1] Normalization to unity is not the only possible normalization, but it is the most common, and will be used throughout this book.

and the functions are not orthogonal. The expansion interval must be specified before a statement about orthogonality can be made. The same is true for normality. On $[-1, 1]$, the function x has the norm

$$N(x) = \langle x \mid x \rangle = \int_{-1}^{+1} x^2 = \tfrac{2}{3} \tag{2-9}$$

but on the interval $[0, 1]$, the norm

$$N(x) = \langle x \mid x \rangle = \int_{0}^{1} x^2 = \tfrac{1}{3} \tag{2-10}$$

One very useful property of functions may be introduced at this point. Very often, the integrals which form inner products may be simplified by using symmetry properties of the functions. This symmetry may be expressed by two definitions.

Definition An *even function* is a function for which $f(x) = f(-x)$; an *odd function* is a function for which $f(x) = -f(-x)$.

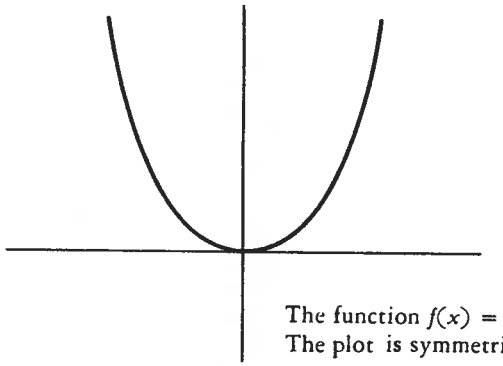
Evenness or oddness is easily pictured. Figure 2-1a shows a graph of the function $f(x) = x^2$, which is even, since $(x)^2 = (-x)^2$. Graphically speaking, the plot of $f(x)$ is symmetrical about the ordinate axis. Figure 2-1b shows the function $f(x) = x^3$, which is odd, since $(x)^3 = -(-x)^3$. The plot to the right is the negative of the plot to the left of the ordinate axis. The integrals of even or odd functions are especially simple *if the interval is symmetric*. The following theorem results.

Theorem The integral of an even function on a symmetric interval is twice the integral on the half-interval; the integral of an odd function on a symmetric interval is zero.

This theorem is illustrated graphically in Fig. 2-2. It may be proven by dividing the full symmetric interval into two half-intervals:

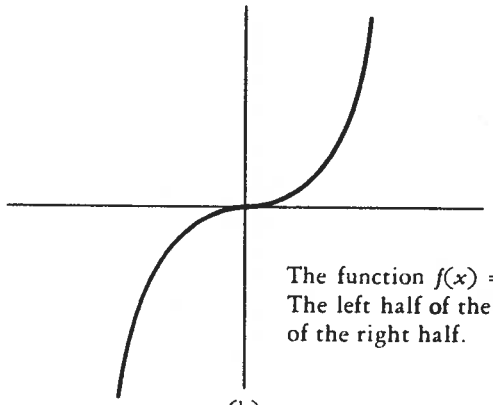
$$\int_{-a}^{a} (\text{even}) = \int_{-a}^{0} (\text{even}) + \int_{0}^{a} (\text{even}) \tag{2-11}$$

But, since an even function of x is the same as the even function of $-x$, we may replace the integral over the negative half-interval



The function $f(x) = x^2$, an even function. The plot is symmetric.

(a)



The function $f(x) = x^3$, an odd function. The left half of the plot is the negative of the right half.

(b)

Figure 2-1 Even and odd functions.

$[-a, 0]$ by the integral over the positive half-interval $[0, a]$ without changing the function:

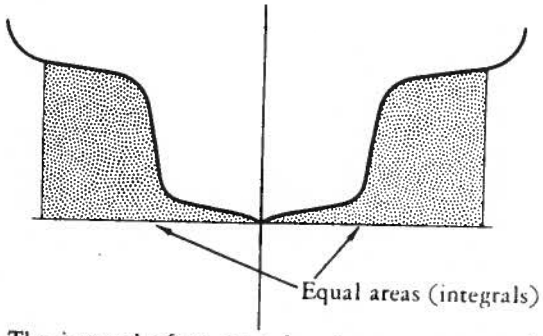
$$\int_{-a}^{a} (\text{even}) = \int_{0}^{a} (\text{even}) + \int_{0}^{a} (\text{even}) = 2\int_{0}^{a} (\text{even}) \tag{2-12}$$

which proves the first part of the theorem. The second part is as simple:

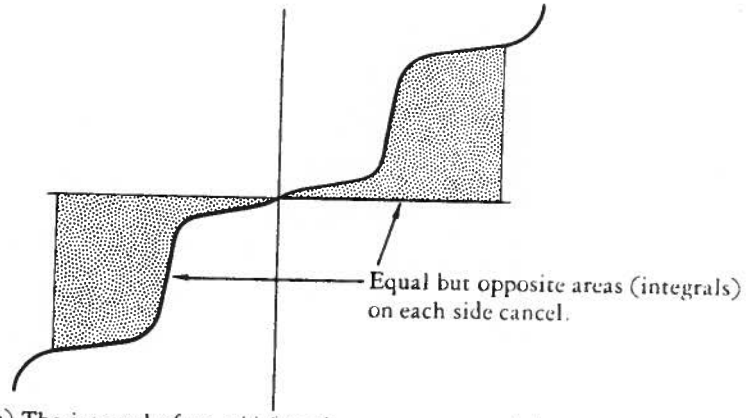
$$\int_{-a}^{a} (\text{odd}) = \int_{-a}^{0} (\text{odd}) + \int_{0}^{a} (\text{odd})$$

$$= -\int_{0}^{a} (\text{odd}) + \int_{0}^{a} (\text{odd}) = 0 \qquad (2\text{–}13)$$

For odd functions, the integral over the negative half-interval may be replaced by an integral over the positive half-interval if the sign of the function is changed. Equation 2–13 gives the second part of the theorem.



(a) The integral of an even function on a symmetric interval is twice the integral on the half-interval.



(b) The integral of an odd function on a symmetric interval is zero.

Figure 2–2 Integrals of even and odd functions.

Application of this theorem to the calculation of inner products on symmetric intervals gives another result.

Theorem On the symmetric interval $[-a, a]$, the inner product of an odd function with an odd function, or of an even function with an even function, is not zero, and may be calculated by twice the inner product over either half-interval $[-a, 0]$ or $[0, a]$; the inner product of an even function with an odd function is zero, no matter what the form of the function.

In the foregoing five definitions, we have considered only two arbitrary functions. However, the power and usefulness of the definitions established above really become apparent when we consider *sets of functions*. A set of functions is a collection of functions of the *same variable*, defined on the *same interval*, and *characterized by a rule* for figuring out all the functions in the set. For example, all the powers of x constitute a set of functions. We write this set, using braces, as $\{x^n\}$, indicating all these functions of x: $x^0 = 1$, $x^1 = x$, x^2, x^3, x^4, and so on. The general exponent n indicates the rule for figuring out each member of the set of functions. To be complete, we must specify the interval under consideration, and the values that n, often called the *index*, may take on, like this: "the set of functions $\{x^n\}$, on $[-1, 1]$, for $n = 0$ and all positive integers."

To round out our description of sets of functions which are useful in quantum chemistry, we introduce three new definitions.

Definition A *complete set of functions* $\{F_i\}$, is a set of functions such that any other function f may be expressed as a linear combination of members of the set $\{F_i\}$ on a prescribed expansion interval to whatever precision may be desired.[2]

If the set $\{F_i\}$ is complete, then we may expand f in terms of the functions F_i like this:

$$f(x) = a_1F_1(x) + a_2F_2(x) + \cdots + a_nF_n(x) + \cdots$$

$$= \sum_{n=1}^{\infty} a_nF_n(x) \qquad (2\text{–}14)$$

[2] The uniform continuity of the functions of the set $\{F_i\}$ and of the function f is tacitly assumed here, and throughout; this definition could be made more rigorous if such concepts were used, but this is not of importance in quantum chemistry.

Generally speaking, proving that a set of functions is complete is quite difficult, and for our purposes we shall consider only such sets of functions which are known to be complete, and not consider completeness proofs per se.

Definition An *orthogonal set of functions*, or a *set of orthogonal functions*, is a set of functions each of which is orthogonal on a prescribed interval to all other members of the set.

That is, the set $\{F_i\}$ is an orthogonal set if every member is orthogonal to every other member:

$$\langle F_j \mid F_k \rangle = 0 \qquad (2\text{–}15)$$

for all j and k such that $j \neq k$. Such a set of functions, which we shall consider in the third section of this chapter, is the set $\{\sin nx, \cos nx\}$, on $[-\pi, \pi]$, for n zero or positive integers. The proof of the orthogonality of these functions is one of the problems at the end of this chapter; notice that it involves three separate proofs:

$$\langle \sin nx \mid \sin mx \rangle = 0 \qquad n \neq m \qquad (2\text{–}16a)$$

$$\langle \cos nx \mid \cos mx \rangle = 0 \qquad n \neq m \qquad (2\text{–}16b)$$

$$\langle \sin nx \mid \cos nx \rangle = 0 \qquad \text{for all } n \qquad (2\text{–}16c)$$

Finally, we combine the definitions of orthogonality and normalization.

Definition An *orthonormal set of functions* is an orthogonal set of functions, each of which is normalized.

That is, the set $\{F_i\}$ is orthonormal if

$$\langle F_j \mid F_k \rangle = 0 \qquad \text{for all } j \neq k \qquad (2\text{–}17a)$$

and

$$\langle F_j \mid F_j \rangle = 1 \qquad \text{for all } j \qquad (2\text{–}17b)$$

The pair of equations 2–17a and 2–17b occurs so often in discussing orthonormal functions that a special symbol has been introduced to combine Eqs. 2–17a and 2–17b. The *Kronecker delta symbol*, or *Kronecker delta*, or *delta symbol* δ_{ij} has the meaning $\delta_{ij} = 0$ for $i \neq j$, $\delta_{ij} = 1$ for $i = j$. If we use the Kronecker delta symbol, the condition for orthonormality, Eqs. 2–17a and 2–17b may be simply expressed as

$$\langle F_j \mid F_k \rangle = \delta_{jk} \qquad \text{for all } j \text{ and } k \qquad (2\text{–}18)$$

In this chapter, orthonormal sets will be indicated by lower case Greek letters, such as $\{\phi_i\}$.

In this section we have defined a number of terms of importance, inner product, orthogonality, norm and normalization, completeness, and orthogonal and orthonormal sets of functions; we have also used the property of evenness or oddness of functions to simplify integrals over symmetric intervals.

2–2 EXPANSION IN TERMS OF ORTHONORMAL FUNCTIONS

In this section we shall learn how to expand a given function, on a prescribed interval, in terms of a set of orthonormal functions. Since the operations that occur in the calculation which follows occur often in discussions of orthonormal functions and later in discussions of orthonormal vectors, the calculation is set aside from the text.

$$f(x) = \sum_i a_i \phi_i(x) \qquad (2\text{–}19)$$

$$\phi_j(x)^* f(x) = \sum_i a_i \phi_j(x)^* \phi_i(x) \qquad (2\text{–}20)$$

$$\int \phi_j(x)^* f(x)\, dx = \sum_i a_i \int \phi_j(x)^* \phi_i(x)\, dx \qquad (2\text{–}21)$$

$$\langle \phi_j \mid f \rangle = \sum_i a_i \langle \phi_j \mid \phi_i \rangle \qquad (2\text{–}22)$$

$$\langle \phi_j \mid f \rangle = \sum_i a_i \delta_{ji} \qquad (2\text{–}23)$$

$$\langle \phi_j \mid f \rangle = a_j \qquad (2\text{–}24)$$

Equation 2–19 shows the expansion that we desire to use to express $f(x)$ on a particular expansion interval (not specified here) as a linear

combination of the members of the set of orthonormal functions $\{\phi_i\}$.

Equation 2–20 gives the result of multiplying each side of the equation by $\phi_j(x)^*$, the complex conjugate of some member of the set $\{\phi_i\}$. Equation 2–21 gives the result of integrating both sides of Eq. 2–20 over the expansion interval. However, the result of Eqs. 2–20 and 2–21 is simply the formation of the inner product of ϕ_j (on the left) with Eq. 2–19 (on the right).

The definition of orthonormality is used in Eq. 2–23 to replace $\langle\phi_j \mid \phi_i\rangle$ with the Kronecker delta δ_{ji}.

Lastly, the sum over i on the right side of Eq. 2–23 is evaluated. If this sum were written out, it would look like this:

$$\sum_i a_i\delta_{ji} = a_1\delta_{j1} + a_2\delta_{j2} + \cdots + a_j\delta_{jj} + \cdots + a_n\delta_{jn} + \cdots \tag{2–25}$$

All the delta symbols but one are identically zero. The only one that is not zero is $\delta_{jj} = 1$. However, then the sum gives

$$\sum_i a_i\delta_{ji} = a_j\delta_{jj} = a_j \cdot 1 = a_j \tag{2–26}$$

The use of the delta symbols, which in turn stems from the property of orthonormality, is the key to a simple evaluation of the coefficients a_i for the expansion in orthonormal functions. The evaluation of the sum in Eq. 2–23 follows a simple rule: *A sum involving the products of a Kronecker delta with other quantities "picks out" that term for which the subscripts of the Kronecker delta are identical;* or *only that term survives for which the subscripts are identical.*

Finally, Eq. 2–24 is a formula for calculating the expansion coefficients in an expansion of a function of a given interval in members of an orthonormal set, $a_j = \langle\phi_j \mid f\rangle$. The expansion coefficients may be complex numbers. The use of a letter subscript should not obscure the issue: if we needed a_1, we would evaluate $\langle\phi_1 \mid f\rangle$; if a_2, $\langle\phi_2 \mid f\rangle$; if a_{309}, $\langle\phi_{309} \mid f\rangle$, and so on. We shall work an example in depth in the following section.

We turn next to a question of practical importance. If an expansion in orthonormal functions is curtailed after a finite number of terms, what error is incurred? The answer to this question reveals a new property of the expansion coefficients: these coefficients minimize the error of a curtailed expansion. Denote the error after taking n terms by M_n. This error is evaluated[3] by

$$M_n = \int |f(x) - \sum_{j=1}^{n} b_j\phi_j|^2\, dx \tag{2–27}$$

that is, by the area under a plot of the square of the absolute value of the residual as a function of x. The integral is taken over the expansion interval. The meaning of M_n is illustrated graphically in Fig. 2–3. Since $c^*c = |c|^2$, we may write

$$\begin{aligned} M_n &= \int \Big[f^* - \sum_{j=1}^{n} b_j^*\phi_j^*\Big]\Big[f - \sum_{i=1}^{n} b_i\phi_i\Big] \\ &= \langle f \mid f\rangle - \sum_{j=1}^{n} b_j^*\langle\phi_j \mid f\rangle - \sum_{i=1}^{n} b_i\langle f \mid \phi_i\rangle \\ &\quad + \sum_{i,\,j=1}^{n} b_j^* b_i\langle\phi_j \mid \phi_i\rangle \end{aligned} \tag{2–28}$$

Notice that a different summation index is used in each factor in Eq. 2–28. Summation indices are often referred to as *dummy indices*, since their *name alone* confers no special meaning. However, they must often be distinguished *by name* with care. For example,

$$\Big(\sum_i c_i\phi_i\Big)\Big(\sum_i c_i\phi_i\Big)$$

might casually be written

$$\sum_i c_i^2\phi_i^2,$$

but this is not what the product connotes. If the sums are written out, $(c_1\phi_1 + c_2\phi_2 + \cdots)(c_1\phi_1 + c_2\phi_2 + \cdots)$ is equal to $c_1^2\phi_1^2 + c_1c_2\phi_1\phi_2 + c_2c_1\phi_2\phi_1 + c_2^2\phi_2^2 + \cdots$, which is quite different from

[3] This is the "least-squares" criterion for error. Others are also applicable. It should also be noted that the curtailed expansion will not be normalized if the total function is.

$$\sum_i c_i^2\phi_i^2 = c_1^2\phi_1^2 + c_2^2\phi_2^2 + \cdots.$$

To ensure that the proper answer—the one with the cross terms—is

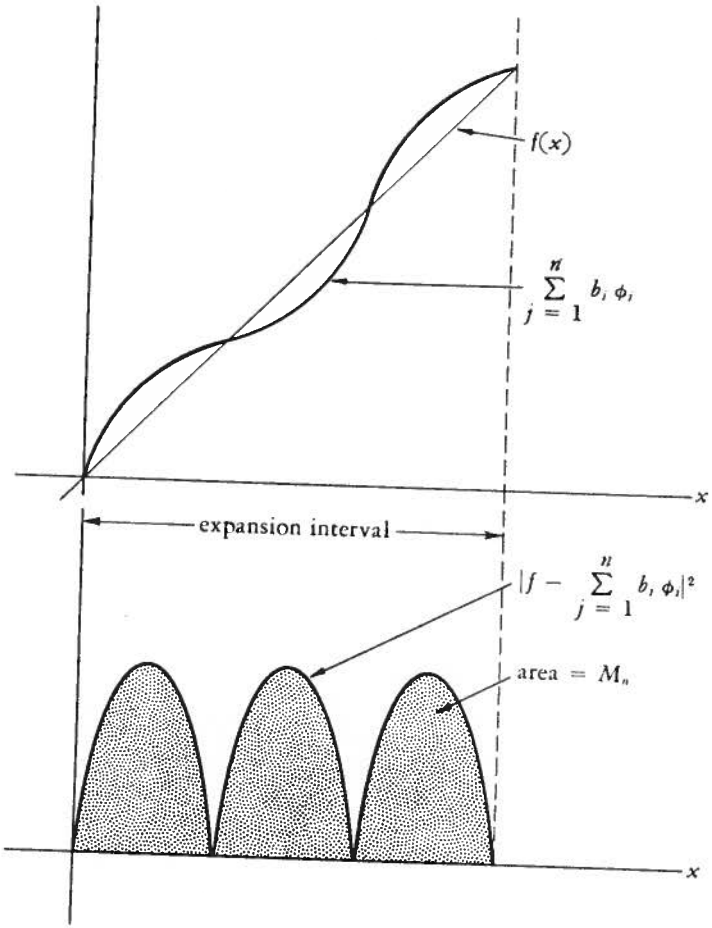


Figure 2–3 The error in a curtailed expansion in orthonormal functions is $M_n = \int |f - \sum_{j=1}^{n} b_j\phi_j|^2$.

written, we use the notation

$$\left(\sum_i c_i\phi_i\right)\left(\sum_j c_j\phi_j\right) = \sum_{i,j} c_ic_j\phi_i\phi_j.$$

In the last term of Eq. 2–28 we may substitute the Kronecker delta, because the set $\{\phi_i\}$ is orthonormal:

$$M_n = \langle f \mid f\rangle - \sum_{j=1}^{n} b_j^*\langle\phi_j \mid f\rangle - \sum_{i=1}^{n} b_i\langle f \mid \phi_i\rangle + \sum_{i,\ j=1}^{n} b_j^*b_i\delta_{ji}$$

$$= \langle f \mid f\rangle - \sum_{j=1}^{n} [b_j\langle f \mid \phi_j\rangle + b_j^*\langle\phi_j \mid f\rangle - b_j^*b_j] \qquad (2\text{–}29)$$

The sum over i involving only the delta symbol leaves just the sum over j, with the terms surviving for which $i = j$, and, since the name of the summation index is unimportant, each of the three terms has been expressed in the index j, and combined.

We would find the "best" set of coefficients for the expansion by minimizing the error M_n with respect to these coefficients, b_j. This involves two steps since the b_j have both a real and an imaginary part. This double-minimum problem may be expressed by requiring $\partial M_n/\partial b_j = \partial M_n/\partial(b_j^*) = 0$ for all j. Taking derivatives of Eq. 2–29, we obtain

$$\frac{\partial M_n}{\partial b_j} = 0 = -\langle f \mid \phi_j\rangle + b_j^* \qquad (2\text{–}30a)$$

$$\frac{\partial M_n}{\partial(b_j^*)} = 0 = -\langle\phi_j \mid f\rangle + b_j \qquad (2\text{–}30b)$$

These equations are complex conjugates of one another, and the "best" coefficients for the expansion, in the sense of minimizing the error, are given by

$$b_j = \langle\phi_j \mid f\rangle \qquad (2\text{–}31)$$

The coefficients given by Eq. 2–31 are, in fact, the coefficients that *must* be used in an expansion in orthonormal functions. Hence, $a_j = b_j = \langle\phi_j \mid f\rangle$ are not only the "right" coefficients (Eq. 2–24), but also the "best" (Eq. 2–31).

After substituting $a_j = b_j = \langle \phi_j \mid f \rangle$ into Eq. 2–29, we calculate the error to be

$$M_n = \langle f \mid f \rangle - \sum_{j=1}^{n} a_j a_j^* + a_j^* a_j - a_j^* a_j$$

$$= \langle f \mid f \rangle - \sum_{j=1}^{n} |a_j|^2 \qquad (2\text{–}32)$$

We conclude this section with an important relation concerning inner products, called the *expansion theorem*.

Theorem The inner product $\langle f \mid g \rangle$ may be expanded in terms of an orthonormal set $\{\phi_i\}$ as

$$\langle f \mid g \rangle = \sum_i \langle f \mid \phi_i \rangle \langle \phi_i \mid g \rangle \qquad (2\text{–}33)$$

To prove this theorem, expand f as $f = \sum_i a_i \phi_i$, and g as $g = \sum b_j \phi_j$. Then,

$$\langle f \mid g \rangle = \int f^* g = \sum_i \sum_j \int a_i^* \phi_i^* b_j \phi_j$$

$$= \sum_{ij} a_i^* b_j \langle \phi_i \mid \phi_j \rangle = \sum_{ij} a_i^* b_j \delta_{ij}.$$

However, $a_i^* = \langle \phi_i \mid f \rangle^* = \langle f \mid \phi_i \rangle$, and $b_i = \langle \phi_i \mid g \rangle$. Hence,

$$\langle f \mid g \rangle = \sum_i \langle f \mid \phi_i \rangle \langle \phi_i \mid g \rangle.$$

The structure $\mid \phi_i \rangle \langle \phi_i \mid$ which occurs in the expansion theorem will come to have more meaning later on in the book. For the time being, we may notice that in the expansion theorem we have but inserted $\mid \phi_i \rangle \langle \phi_i \mid$ between $\langle f \mid$ and $\mid g \rangle$ and summed over i. Such is indeed the case; this operation may be called "inserting a complete set of states." The content of the expansion theorem may be written $\sum_i \mid \phi_i \rangle \langle \phi_i \mid = 1$. The structure $\mid \phi_i \rangle \langle \phi_i \mid$ is not an inner product as

defined in Eq. 2–1; it is an element that has not heretofore entered our discussions. Actually, it is an operator. The sense in which it is an operator will become apparent later.

In this section, we have derived the formula for the coefficients in an expansion in terms of orthonormal functions; we have demonstrated that these coefficients minimize the error in a curtailed expansion, and we have derived a formula for the error that is incurred by curtailing an expansion. Finally, we have derived the expansion theorem for the inner product.

2 3 THE FOURIER SERIES

The Fourier series is an expansion of a function in the orthonormal functions which are proportional to $\{\sin mx, \cos mx\}$. We mentioned in the first section that these functions are orthogonal, but we derive now their norm on the interval $[-\pi, \pi]$.

$$N(\sin mx) = \langle \sin mx \mid \sin mx \rangle = \int_{-\pi}^{\pi} \sin^2 mx \, dx$$

$$= \frac{1}{m} \int_{-m\pi}^{m\pi} \sin^2 y \, dy = \pi \qquad (2\text{–}34)$$

Similarly,

$$N(\cos mx) = \langle \cos mx \mid \cos mx \rangle = \frac{1}{m} \int_{-m\pi}^{m\pi} \cos^2 y \, dy = \pi \qquad (2\text{–}35)$$

Equations 2–34 and 2–35 are valid for all values of m, except for $m = 0$, where each result yields the ambiguous $N = 0/0$. To clarify the $m = 0$ case, we must separately evaluate

$$N(\sin 0x) = \langle \sin 0x \mid \sin 0x \rangle = \int_{-\pi}^{\pi} 0 \, dx = 0 \qquad (2\text{–}36)$$

and

$$N(\cos 0x) = \langle \cos 0x \mid \cos 0x \rangle = \int_{-\pi}^{\pi} 1 \, dx = 2\pi \qquad (2\text{–}37)$$

All these results may be expressed by

$$\begin{aligned} N(\sin mx) &= \pi \qquad \text{for } m \neq 0 \\ &= 0 \qquad \text{for } m = 0 \\ N(\cos mx) &= \pi \qquad \text{for } m \neq 0 \\ &= 2\pi \qquad \text{for } m = 0 \end{aligned} \tag{2-38a}$$

or, using the Kronecker delta, by

$$\begin{aligned} N(\sin mx) &= \pi - \pi\delta_{m0} \\ N(\cos mx) &= \pi + \pi\delta_{m0} \end{aligned} \tag{2-38b}$$

Together with the orthogonality relations, we may then write all possible inner products of $\{\sin mx, \cos mx\}$ by

$$\begin{aligned} \langle \sin mx \mid \cos nx \rangle &= 0 \\ \langle \sin mx \mid \sin nx \rangle &= (1 - \delta_{m0})\pi\delta_{mn} \\ \langle \cos mx \mid \cos nx \rangle &= (1 + \delta_{m0})\pi\delta_{mn} \end{aligned} \tag{2-39}$$

for all m, n = any positive integer or zero. Since we shall be using our previous formula, Eq. 2-24, for calculating the expansion coefficients for an expansion in orthonormal functions, we must use, instead of the set of functions $\{\sin mx, \cos mx\}$, the set of functions $\{(2\pi)^{-1/2}, \pi^{-1/2}\sin mx, \pi^{-1/2}\cos mx\}$, for $m = 1, 2, \ldots$, for expansion on the interval $[-\pi, \pi]$.

For these orthonormal functions, the expansion coefficients will be

$$\begin{aligned} a_0 &= \langle (2\pi)^{-1/2} \mid f \rangle \\ a_m &= \langle \pi^{-1/2}\cos mx \mid f \rangle \\ b_m &= \langle \pi^{-1/2}\sin mx \mid f \rangle \end{aligned} \tag{2-40}$$

where the expansion is

$$f(x) = a_0(2\pi)^{-1/2} + \sum_{m=1}^{\infty} a_m(\pi^{-1/2}\cos mx) + \sum_{m=1}^{\infty} b_m(\pi^{-1/2}\sin mx) \tag{2-41}$$

This is not the usual form of the Fourier series, but this is an example of an expansion in orthonormal functions. Usually, the constants are removed from the terms by explicitly writing out the expansion coefficients:

$$f(x) = \frac{1}{2\pi}\langle 1 \mid f\rangle + \frac{1}{\pi}\sum_{m=1}^{\infty}[\langle \cos mx \mid f\rangle \cos mx + \langle \sin mx \mid f\rangle \sin mx] \tag{2-42}$$

This gives the final result

$$f(x) = \frac{c_0}{2} + \sum_{m=1}^{\infty} c_m \cos mx + \sum_{m=1}^{\infty} d_m \sin mx \tag{2-43}$$

$$c_m = \frac{1}{\pi}\langle \cos mx \mid f\rangle \tag{2-44a}$$

$$d_m = \frac{1}{\pi}\langle \sin mx \mid f\rangle \tag{2-44b}$$

on the interval $[-\pi, \pi]$. This is the form in which the Fourier series is usually written. Notice that the lead term is divided by two. This is because the norm of $\cos 0x$ is 2π, whereas for all other values of m, the norm of $\cos mx$ is π.

We can make two simple extensions of the Fourier series at once by using the property of evenness and oddness of functions. The sine function is odd; $\sin x = -\sin(-x)$. The cosine function is even; $\cos x = \cos(-x)$. This being the case, we may formulate these two rules.

(1) *The Fourier expansion on* $[-\pi, \pi]$ *of an odd function is made up only of sine terms:* $f(x) = \sum_{m=1}^{\infty} d_m \sin mx$. (2) *The Fourier expansion on* $[-\pi, \pi]$ *of an even function is made up only of cosine terms:* $f(x) = c_0/2 + \sum_{m=1}^{\infty} c_m \cos mx$. For, if f is odd, all inner products $\langle \cos mx \mid f\rangle \equiv 0$; and, if f is even, all inner products $\langle \sin mx \mid f\rangle \equiv 0$.

EXAMPLE

The expansion of $f(x) = x$ on $[-\pi, \pi]$. Since $f(x) = x$ is odd, only sine terms occur, and we may write

$$x = \sum_{m=1}^{\infty} d_m \sin mx \qquad m = 1, 2, \ldots \text{ on } [-\pi, \pi]$$

where

$$d_m = \frac{1}{\pi}\langle \sin mx \mid x\rangle$$

$$= \frac{1}{\pi}\int_{-\pi}^{\pi} x \sin mx\, dx = \frac{1}{m^2\pi}\int_{-m\pi}^{m\pi} y \sin y\, dy$$

$$= \frac{1}{m^2\pi}(-y\cos y + \sin y)\Big|_{-m\pi}^{m\pi}$$

$$= \frac{-m\pi\cos m\pi - m\pi\cos m\pi}{m^2\pi} = \frac{2(-1)^{m+1}}{m}$$

and the Fourier series is

$$x = \sum_{m=1}^{\infty} \frac{2(-1)^{m+1}}{m}\sin mx$$

$$= 2\sin x - \sin 2x + \tfrac{2}{3}\sin 3x - \cdots$$

The comparison of the Fourier series curtailed after three terms with the function itself is shown in Fig. 2–4. The mean-square error after taking three terms can be found using Eq. 2–32 as follows:

$$M_3 = \langle x \mid x\rangle - \sum_{j=1}^{3} |a_j|^2 \qquad (2\text{–}45)$$

According to Eqs. 2–40 and 2–44a, $a_j = \pi^{1/2}c_j$, so that

$$M_3 = \int_{-\pi}^{\pi} x^2\, dx - 4\pi - 1\pi - \frac{4}{9}\pi = \frac{2\pi^3}{3} - \frac{49\pi}{9} \simeq 6.8;$$

the square root of this, 2.6, may be compared with the area under the function $\pi^2 = 9.9$. Therefore, after three terms, a relative error of about 26% persists.

The use of Fourier series to describe functions has an important electronic analog. It is usually the case that one can generate sine and cosine functions electronically with ease. The generation of other functions can then be done by an electronic Fourier synthesis. Such a synthesis could be used to generate the so-called "sawtooth" wave shown in Fig. 2–5a. The sawtooth pattern is an endless series of plots of $f(x) = x$ vs. x on $[-\pi, \pi]$, which has the Fourier components derived in the example. Figure 2–5b shows a three-term approximation to the sawtooth wave. Such waveforms are used for sweeps, as in oscilloscopes and television.

The functions $\{\sin mx\}$ and $\{\cos mx\}$ are individually complete on either half-interval $[-\pi, 0]$ or $[0, \pi]$. These functions are also orthogonal on either half-interval, and all have the norm $\pi/2$, except $\cos 0x$, whose norm is π. Therefore, we can construct two distinct "half Fourier series":

$$f(x) = \frac{c_0}{2} + \sum_{m=1}^{\infty} c_m \cos mx \qquad (2\text{–}46a)$$

$$c_m = \frac{2}{\pi}\langle \cos mx \mid f\rangle \qquad (2\text{–}46b)$$

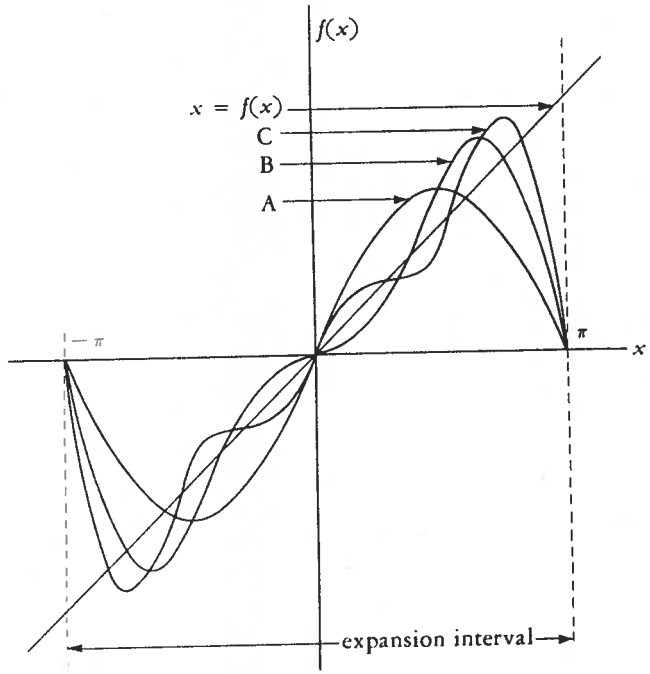


Figure 2–4 The function $f(x) = x$ on the interval $(-\pi, \pi)$, and the Fourier series for $f(x)$ after (A) one term, (B) two terms, and (C) three terms.

$$f(x) = \sum_{m=1}^{\infty} d_m \sin mx \qquad (2\text{–}47a)$$

$$d_m = \frac{2}{\pi} \langle \sin mx \mid f \rangle \qquad (2\text{–}47b)$$

each on either interval $[-\pi, 0]$ or $[0, \pi]$. Likewise, either the full or half Fourier series may be extended to any symmetric interval $[-a, a]$ or half of a symmetric interval $[-a, 0]$ or $[0, a]$ by a scale expansion, such as

$$f(x) = \frac{c_0}{2} + \sum_{m=1} \left(c_m \cos \frac{m\pi x}{a} + d_m \sin \frac{m\pi x}{a} \right) \qquad (2\text{–}48)$$

with $c_m = (1/a)\langle \cos(m\pi x/a) \mid f \rangle$, $d_m = (1/a)\langle \sin(m\pi x/a) \mid f \rangle$. A very important revision of the Fourier series for quantum mechanics is the formulation

$$f(x) = \sum_{m=-\infty}^{\infty} a_m e^{imx} \qquad \text{on } [-\pi, \pi] \qquad (2\text{–}49)$$

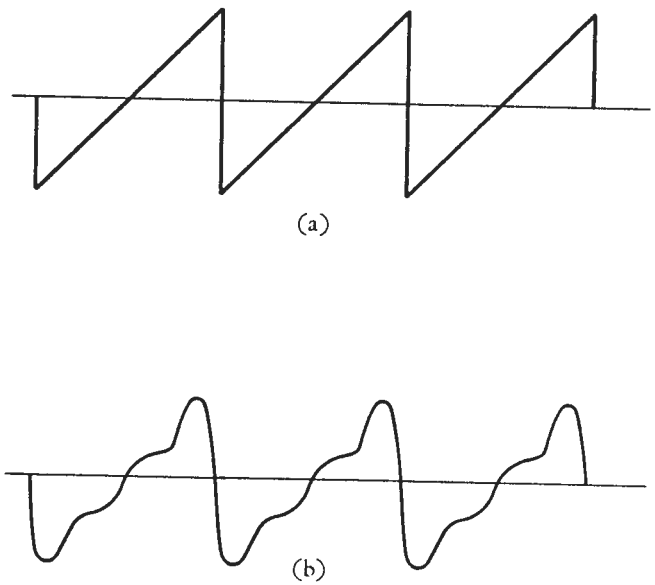


Figure 2–5 (a) The "sawtooth" wave, and (b) a three-term Fourier synthesis of the sawtooth wave.

For, if $\{\cos mx, \sin mx\}$ is complete for all positive values of m and zero, then $\{e^{imx} = \cos mx + i \sin mx\}$ is complete for all integral values—positive, negative, and zero—of m. To put it another way, including the complex conjugate of the functions $\{e^{imx}\}$ completes the set. These modifications of the basic Fourier series are subjects of a number of the problems at the end of the chapter.

In this section, we have exhibited a very useful expansion in orthonormal functions, the Fourier series (expansion in sines and cosines), and have commented on its usefulness and extension.

2–4 CONSTRUCTION OF ORTHONORMAL FUNCTIONS

Thus far, we have discussed in general the properties of orthonormal functions, their use in series expansions, and the particular example of the Fourier series. We have pointed out that a sufficiently precise expansion can be achieved only if the set of functions, orthonormal or not, is complete. In this section, we shall show how an orthonormal complete set of functions may be formed from a set of complete functions. Before we take up the actual formation, we must introduce one aspect of completeness which plays an important rôle in this discussion and also in the work that follows. We begin with a definition.

Definition A set of functions is said to be ***linearly independent*** if none of the functions can be expressed as a linear combination of the rest. Similarly, a set of functions is said to be ***linearly dependent*** if at least one of the functions may be related to one or more of the other functions by a linear equation.

This definition may be expressed mathematically by saying that a set of functions $\{F_i\}$ is linearly independent if the identity

$$\sum_i c_i F_i = c_1 F_1 + c_2 F_2 + \cdots = 0 \qquad (2\text{–}50)$$

can be solved only with all the $c_i = 0$; the set $\{F_i\}$ is linearly dependent if this equation can be solved with at least one of the $c_i \neq 0$.

Take, for example, these two sets of functions:

$$\begin{array}{ll} F_1 = 1 & G_1 = 1 \\ F_2 = x & G_2 = x \\ F_3 = 7x + 4 & G_3 = x^2 \\ F_4 = x^3 & G_4 = x^3 \end{array}$$

The set $\{G_i\}$ is linearly independent. It is not possible for any linear combination of those functions to add to zero, as in Eq. 2–50. To put it another way, there is no linear relation connecting those four functions: the function $G_3 = x^2$ is not a linear combination of 1, x, and x^3. The set $\{F_i\}$ is linearly dependent. Function F_3 is a linear combination of F_1, F_2, and F_4; $F_3 = 7F_2 + 4F_1$. That is, Eq. 2–50 can be solved with some of the $c_i \neq 0$; $-4F_1 - 7F_2 + 1F_3 + 0F_4 = 0$.

A set of complete functions will always contain a linearly independent subset. This is an important statement that is offered without proof; it is the first step in generating a complete set of orthonormal functions.

As an example, consider the powers of x on the interval $[-1, 1]$. Taylor's theorem essentially guarantees that the set $\{x^n\}$, $n = 0, 1, 2, \cdots$ is complete. We have just seen that these functions are also linearly independent. Consider the sets of functions below. The set $\{F_i'\}$ is complete; the linearly independent set $\{F_i\}$ was formed by removing the linearly dependent member of $\{F_i'\}$. The set $\{G_i\}$ is also linearly independent and complete; it differs from $\{F_i\}$ only in the order of arrangement.

$\{F_i'\}$	$\{F_i\}$	$\{G_i\}$
1	1	1
x	x	x^2
$2x$	$\cdots$	x^4
x^2	x^2	$\cdots$
x^3	x^3	x
x^4	x^4	x^3
$\cdots$	$\cdots$	$\cdots$

Construction of a complete set of orthonormal functions can always be accomplished from a complete, linearly independent set of functions. The procedure for this construction is known as the *Schmidt orthogonalization procedure*. Rather than give a complete account of the procedure, we shall illustrate the idea, state the general

conclusion, and then work an illustrative example. Suppose we have a complete, linearly independent set of functions $\{f_i\}$, defined on a prescribed interval. We wish to form an orthonormal set $\{\phi_i\}$ from the $\{f_i\}$.

Step 1. Let ϕ_0 be f_0 normalized. That is, let the first ϕ function be simply proportional to the first f function. The norm of f_0 is $N_0 = \langle f_0 \mid f_0 \rangle$, so that $\phi_0 = f_0/N_0^{1/2}$.

Step 2. Let ϕ_1 be a linear combination of ϕ_0 and f_1 which is orthogonal to ϕ_0, and normalized. This can be done in general since the set $\{f_i\}$ is linearly independent. That is,

$$\phi_1 = N_1^{-1/2}(c\phi_0 + f_1) \tag{2–51}$$

where

$$\langle \phi_0 \mid \phi_1 \rangle = 0 \tag{2–52}$$

The orthogonality condition gives

$$c\langle \phi_0 \mid \phi_0 \rangle + \langle \phi_0 \mid f_1 \rangle = 0 \tag{2–53a}$$

$$c = -\langle \phi_0 \mid f_1 \rangle \tag{2–53b}$$

Then,

$$\phi_1 = N_1^{-1/2}(f_1 - \langle \phi_0 \mid f_1 \rangle \phi_0) \tag{2–54}$$

and

$$\begin{aligned} N_1 &= \langle f_1 \mid f_1 \rangle - \langle \phi_0 \mid f_1 \rangle \langle f_1 \mid \phi_0 \rangle - \langle \phi_0 \mid f_1 \rangle \langle \phi_0 \mid f_1 \rangle + \langle \phi_0 \mid f_1 \rangle^2 \\ &= \langle f_1 \mid f_1 \rangle - |\langle \phi_0 \mid f_1 \rangle|^2 \end{aligned} \tag{2–55}$$

This process is then continued until the complete, orthonormal set is generated. The general term is

$$\phi_k = N_k^{-1/2}\left(f_k - \sum_{j=0}^{k-1} \langle \phi_j \mid f_k \rangle \phi_j\right) \tag{2–56}$$

with

$$N_k = \langle f_k \mid f_k \rangle - \sum_{j=0}^{k-1} |\langle \phi_j \mid f_k \rangle|^2 \tag{2–57}$$

It should be clear from this discussion that functions will not be normalizable if their norm is not finite. Not all functions on all

intervals are normalizable: the set $\{x^n\}$ on $(-\infty, \infty)$ is not, for example. The quantum mechanics of molecular structure is concerned almost exclusively with normalizable, or *square-integrable* functions.

As an example of the Schmidt process, we shall consider the functions $\{\exp[-x^2/2]x^n\}$ on the full line, $(-\infty, \infty)$. We begin by showing that these functions are normalizable.

$$\langle f_n \mid f_n \rangle = \int_{-\infty}^{\infty} \exp\left[\frac{-x^2}{2}\right] x^n \exp\left[\frac{-x^2}{2}\right] x^n \, dx$$
$$= \int_{-\infty}^{\infty} \exp[-x^2] x^{2n} \, dx \qquad (2\text{-}58)$$

Since the integrand is an even function, we may simplify the integral to

$$\langle f_n \mid f_n \rangle = 2 \int_0^{\infty} \exp[-x^2] x^{2n} \, dx = \int_0^{\infty} e^{-y} y^{(n-1/2)} \, dy \qquad (2\text{-}59)$$

which may successively be integrated by parts to give

$$\langle f_n \mid f_n \rangle = (n - \tfrac{1}{2})(n - \tfrac{3}{2}) \cdots (\tfrac{1}{2}) \int_0^{\infty} e^{-y} y^{-1/2} \, dy$$
$$= 2(n - \tfrac{1}{2})(n - \tfrac{3}{2}) \cdots (\tfrac{1}{2}) \int_0^{\infty} e^{-x^2} \, dx$$
$$= (2n - 1)(2n - 3) \cdots (3)(1)\pi^{1/2}/2^n \qquad (2\text{-}60)$$

where the known integral $\int_0^{\infty} e^{-x^2} \, dx = \frac{\pi^{1/2}}{2}$ has been substituted.

The norm N_n is finite for any n, so all functions in the set $\{\exp[-x^2/2]x^n\}$ are normalizable. It is the presence of the decaying exponential $\exp[-x^2/2]$ that renders the integral finite and the functions square integrable.

To construct an orthonormal set, we begin by normalizing $f_0 = \exp[-x^2/2]x^0 = \exp[-x^2/2]$. The norm is $\pi^{1/2}$, so we have $\phi_0 = \pi^{-1/4} \exp[-x^2/2]$.

The next function is found from Eqs. 2-54 and 2-55, which are

specific cases of Eqs. 2-56 and 2-57, respectively. In order to use these equations, we must first evaluate $\langle \phi_0 \mid f_1 \rangle$ and $\langle f_1 \mid f_1 \rangle$:

$$\langle \phi_0 \mid f_1 \rangle = \int_{-\infty}^{\infty} \exp\left[\frac{-x^2}{2}\right] \pi^{-1/4} \exp\left[\frac{-x^2}{2}\right] x \, dx = 0 \qquad (2\text{-}61)$$

because the integrand is odd, and

$$\langle f_1 \mid f_1 \rangle = \pi^{1/2}/2 \qquad (2\text{-}62)$$

from Eq. 2-60. Then

$$N_1 = \langle f_1 \mid f_1 \rangle - |\langle \phi_0 \mid f_1 \rangle|^2 = \pi^{1/2}/2 \qquad (2\text{-}63)$$

and

$$\phi_1 = N_1^{-1/2}(f_1 - \langle \phi_0 \mid f_1 \rangle \phi_0) = 2^{1/2}\pi^{-1/4} \exp\left[\frac{-x^2}{2}\right] x \qquad (2\text{-}64)$$

The reader should notice that all integrals of the form $\langle \phi_i \mid f_j \rangle$ where i is even and j odd, or vice versa, are zero for these sets of functions because the integrand is odd. This greatly simplifies the calculation.

The next function will be calculated according to Eqs. 2-56 and 2-57, as follows:

$$\phi_2 = N_2^{-1/2}(f_2 - \langle \phi_0 \mid f_2 \rangle \phi_0 - \langle \phi_1 \mid f_2 \rangle \phi_1)$$
$$= N_2^{-1/2}(f_2 - \langle \phi_0 \mid f_2 \rangle \phi_0) \qquad (2\text{-}65)$$

and

$$N_2 = \langle f_2 \mid f_2 \rangle - |\langle \phi_0 \mid f_2 \rangle|^2 - |\langle \phi_1 \mid f_2 \rangle|^2$$
$$= \langle f_2 \mid f_2 \rangle - |\langle \phi_0 \mid f_2 \rangle|^2 \qquad (2\text{-}66)$$

We need to evaluate $\langle f_2 \mid f_2 \rangle$ and $\langle \phi_0 \mid f_2 \rangle$. From Eq. 2-60,

$$\langle f_2 \mid f_2 \rangle = \tfrac{3}{4}\pi^{1/2} \qquad (2\text{-}67)$$

and

$$\langle \phi_0 \mid f_2 \rangle = \int_{-\infty}^{\infty} \pi^{-1/4} \exp\left[\frac{-x^2}{2}\right] \exp\left[\frac{-x^2}{2}\right] x^2 \, dx$$
$$= \pi^{-1/4} \int_{-\infty}^{\infty} x^2 \exp[-x^2] \, dx = \pi^{-1/4} \frac{\pi^{1/2}}{2}$$
$$= \frac{\pi^{1/4}}{2} \qquad (2\text{-}68)$$

Finally,

$$N_2 = \frac{3}{4}\pi^{1/2} - \frac{1}{4}\pi^{1/2} = \frac{\pi^{1/2}}{2} \tag{2-69}$$

$$\phi_2 = (\sqrt{2}\pi^{-1/4})\left\{\exp\left[\frac{-x^2}{2}\right]x^2 - \frac{\pi^{1/4}}{2}\pi^{-1/4}\exp\left[\frac{-x^2}{2}\right]\right\}$$

$$= \pi^{-1/4}\exp\left[\frac{-x^2}{2}\right]\frac{(2x^2-1)}{\sqrt{2}} \tag{2-70}$$

The set of orthonormal functions that would be generated in this way are all of the form $\pi^{-1/4}\,(1/\sqrt{2^n n!})\,H_n(x)\exp[-x^2/2]$, where $H_n(x)$ represents one of the polynomials

$$\begin{aligned} H_0(x) &= 1 \\ H_1(x) &= 2x \\ H_2(x) &= 4x^2 - 2 \\ &\cdots \end{aligned} \tag{2-71}$$

The even-numbered polynomials contain even powers of x only, and the odd-numbered polynomials contain odd powers of x only. These polynomials are proportional to the famous *Hermite polynomials*. The functions that we have constructed to be orthonormal on $(-\infty, \infty)$ from the set $\{\exp[-x^2/2]x^n\}$ are the solutions for the wave function of a quantum-mechanical harmonic oscillator. We have, of course, carried out our construction with no attention to any particular physical problem. That we accidentally—by this route—come across the solution to a physical problem emphasizes the significance and simplicity of the harmonic oscillator in quantum mechanics.

Thus, in this section, we have further developed the properties of functions with the concepts of linear independence and normalizability, and we have shown how a complete orthonormal set of functions can be constructed from a complete, linearly independent set. Lastly, we have begun crudely to relate the properties of orthonormal functions to the wave functions describing physical situations.

2–5 THE LEGENDRE POLYNOMIALS AND OTHER SPECIAL FUNCTIONS

In this section we examine a number of sets of orthonormal functions on various intervals. This could be a mammoth undertaking.

The relations among such functions are tabulated in larger treatises, and the proofs of these relations are seldom of importance in quantum chemistry. Therefore, this section is devoted to a detailed examination of one of the special functions, the Legendre polynomials, and a cursory tabulation of others, including Laguerre and Hermite functions.

The Legendre polynomials are of importance in a number of quantum-chemical problems. They are the basis for the wave functions for angular momentum, and therefore occur in problems involving spherical motion, such as that of the electron in a hydrogen atom or the rotations of a molecule. In this context, the Legendre polynomials are important in describing the angular dependence of one-electron atomic orbitals; this dependence, in turn, forms the basis of the geometry of chemical compounds.

There are a number of ways of forming these polynomials, three of which are quite general.

(a) Schmidt orthogonalization of a linearly independent complete set.

(b) Solution of a differential equation.

(c) Use of a generating function.

In addition, we shall describe a formula for the Legendre polynomials. We begin with the orthogonalization procedure.

Legendre polynomials by Schmidt orthogonalization. The Legendre polynomials result from applying the Schmidt procedure to the set $\{x^n\}$ on the interval $[-1, 1]$. The choice of the interval $[-1, 1]$ is specific to the Legendre polynomials. We begin by letting ϕ_0 be proportional to $x^0 = 1$, and normalized on the interval:

$$N_0 = \int_{-1}^{1} 1\,dx = 2 \tag{2-72}$$

$$\phi_0 = (\tfrac{1}{2})^{1/2} \tag{2-73}$$

Then, using Eqs. 2-56 and 2-57,

$$N_1 = \langle f_1 \mid f_1\rangle - |\langle\phi_0 \mid f_1\rangle|^2 = \langle f_1 \mid f_1\rangle = \int_{-1}^{1} x^2\,dx = \tfrac{2}{3} \tag{2-74}$$

$$\phi_1 = N_1^{-1/2}(f_1 - \langle\phi_0 \mid f_1\rangle\,\phi_0) = (\tfrac{3}{2})^{1/2}\,x \tag{2-75}$$

Here, by realizing the oddness of the integrand, we have set $\langle\phi_0 \mid f_1\rangle = 0$; and similarly, for all inner products where the sum of

the indices is odd. This behavior was also noticed for the Hermite polynomials in the previous section.

Proceeding, we find

$$N_2 = \langle f_2 \mid f_2 \rangle - |\langle \phi_0 \mid f_2 \rangle|^2 \tag{2-76}$$

$$\phi_2 = N_2^{-1/2}(f_2 - \langle \phi_0 \mid f_2 \rangle \phi_0) \tag{2-77}$$

so we need to calculate

$$\langle f_2 \mid f_2 \rangle = \int_{-1}^{1} x^4 \, dx = \tfrac{2}{5} \tag{2-78}$$

$$\langle \phi_0 \mid f_2 \rangle = \int_{-1}^{1} (\tfrac{1}{2})^{1/2} x^2 \, dx = (\tfrac{1}{2})^{1/2} \tfrac{2}{3} \tag{2-79}$$

whence

$$N_2 = \tfrac{2}{5} - \tfrac{1}{2} \cdot \tfrac{4}{9} = \tfrac{8}{45} \tag{2-80}$$

$$\phi_2 = (\tfrac{45}{8})^{1/2}(x^2 - \tfrac{1}{3}) = (\tfrac{5}{2})^{1/2}(\tfrac{3}{2}x^2 - \tfrac{1}{2}) \tag{2-81}$$

The functions that are appearing are

$$\begin{aligned} \phi_0 &= (\tfrac{1}{2})^{1/2} \cdot 1 \\ \phi_1 &= (\tfrac{3}{2})^{1/2} \cdot x \\ \phi_2 &= (\tfrac{5}{2})^{1/2} \cdot (\tfrac{3}{2}x^2 - \tfrac{1}{2}) \\ \phi_3 &= (\tfrac{7}{2})^{1/2} \cdot (\tfrac{5}{2}x^3 - \tfrac{3}{2}x) \end{aligned} \tag{2-82}$$

The polynomials at the right are the Legendre polynomials; the whole orthonormal set of functions is $\{[(2n + 1)/2]^{1/2} P_n(x)\}$, where $P_n(x)$ signifies the *Legendre polynomial of rank n*.

Solution of Legendre's differential equation. The Legendre polynomials are the solutions of the differential equation

$$\frac{d}{dx}\left[(1 - x^2)\frac{df}{dx}\right] + l(l + 1)f = 0 \tag{2-83}$$

where l is a positive integer or zero. That the polynomials that have been generated by orthogonalization of $\{x^n\}$ on $[-1, 1]$ are solutions of Eq. 2-83 can be verified by direct substitution. However, we may derive this result directly using a method that is a prototype for the solution of many differential equations.

We attempt a power series solution for f, $f = \sum_n a_n x^n$. Substituting this trial solution (often called *Ansatz*, from the German) into Eq. 2-83, we obtain

$$\frac{d}{dx}\left[(1 - x^2)\sum_n n a_n x^{n-1}\right] + \left[l(l + 1)\sum_n a_n x^n\right] = 0$$

$$\frac{d}{dx}\left[\sum_n n a_n x^{n-1} - \sum_n n a_n x^{n+1}\right] + \sum_n l(l + 1) a_n x^n = 0$$

$$\sum_n n(n - 1) a_n x^{n-2} - \sum_n n(n + 1) a_n x^n + \sum_n l(l + 1) a_n x^n = 0 \tag{2-84}$$

If we collect all similar powers of x, we obtain

$$\sum_n x^n[(n + 2)(n + 1)a_{n+2} - n(n + 1)a_n + l(l + 1)a_n] = 0 \tag{2-85}$$

but, for the series to be identically zero, each term must be identically zero, since the functions $\{x^n\}$ are linearly independent. This gives the *recursion relation* for the expansion coefficients,

$$a_{n+2} = \frac{[n(n + 1) - l(l + 1)]a_n}{(n + 2)(n + 1)} \tag{2-86}$$

Equation 2-86 gives a rule for finding every other coefficient starting with the first. That is, if $a_0 = 1$, then Eq. 2-86 gives $a_2 = -l(l + 1)/2$, $a_4 = [6 - l(l + 1)]/12 \cdot [-l(l + 1)/2]$, and so forth. The coefficients depend on the value of l. Suppose we continue on, finding all the even coefficients. If the *rank* of the equation, l, is even, eventually n will equal l for some term, $n(n + 1)$ will equal $l(l + 1)$, and the following coefficient, a_{n+2}, will equal zero. Then all following coefficients will also equal zero. In other words, if l is even, the even terms of the power series cut off at $n = l$. The odd terms, of course, keep on going. This infinite series of odd powers diverges at $x = \pm 1$, and is not of physical interest; the finite series of even powers gives

$l = 0$	$a_0 = 1$	$f_0 = 1$
$l = 2$	$a_0 = 1$	
	$a_2 = -3$	$f_2 = -3x^2 + 1$

and so forth, which are proportional to the Legendre polynomials derived by orthogonalization. It has become customary to choose a_0 such that the polynomial $f_n = 1$ at $x = +1$. Then

$$\begin{array}{lll} l = 0 & a_0 = 1 & P_0 = 1 \\ l = 2 & a_0 = -\frac{1}{2} & \\ & a_2 = \frac{3}{2} & P_2 = \frac{3}{2}x^2 - \frac{1}{2} \end{array}$$

and so on.

Similarly, if l is odd, the series in odd powers terminates at $n = l$, and the series in even powers is an infinite series, divergent at $x = \pm 1$. Again, we choose a_0 such that $f_n(+1) = 1$. Then

$$\begin{array}{lll} l = 1 & a_1 = 1 & P_1 = x \\ l = 3 & a_1 = -\frac{3}{2} & \\ & a_3 = \frac{5}{2} & P_3 = \frac{5}{2}x^3 - \frac{3}{2}x \end{array}$$

and so forth. The method of power series substitution is very powerful. The usual result is a recursion relation which may be interpreted, within the context of the problem, as we have done here.

The generating function for Legendre polynomials. A generating function for a set of functions is some function of two variables such that when this function is expanded as a power series in one of the variables, the coefficients are the set of functions in the other variable. The generating function for Legendre polynomials is

$$F(x, t) = (1 - 2xt + t^2)^{-1/2} = \sum_n P_n(x)t^n \qquad (2\text{-}87)$$

If the generating function is written as $[1 - (2xt - t^2)]^{-1/2}$, it may be expanded in the binomial series

$$F(x, t) = 1 + \frac{1}{2}\cdot\frac{1}{1!}(2xt - t^2) + \frac{1}{2}\cdot\frac{3}{2}\cdot\frac{1}{2!}(2xt - t^2)^2 + \frac{1}{2}\cdot\frac{3}{2}\cdot\frac{5}{2}\cdot\frac{1}{3!}(2xt - t^2)^3 + \cdots \qquad (2\text{-}88)$$

and then rearranged in powers of t:

$$F(x, t) = 1 + (x)(t) + \left(\frac{-1}{2} + \frac{3}{2}x^2\right)(t^2) + \left(\frac{-3}{2}x + \frac{5}{2}x^3\right)(t^3) + \cdots = \sum_n P_n(x)t^n \qquad (2\text{-}89)$$

Often many useful relations may be derived from a generating function; however, there are no general methods for constructing a generating function.

Rodrigues' formula for the Legendre polynomials. As a final example Of a formula for finding Legendre polynomials, we derive Rodrigues' formula,

$$P_l(x) = \frac{1}{2^l l!}\frac{d^l}{dx^l}(x^2 - 1)^l \qquad (2\text{-}90)$$

Consider the differential equation

$$(1 - x^2)\frac{dy}{dx} + 2lxy = 0 \qquad (2\text{-}91)$$

which can be solved by the elementary "separation of variables" technique, giving

$$\frac{dy}{y} = \frac{-2lx\,dx}{1 - x^2} \qquad (2\text{-}92)$$

whose integral is

$$y = (1 - x^2)^l \qquad (2\text{-}93)$$

If the original differential equation, Eq. 2-91, is now differentiated $(l + 1)$ times,[4] the result is

$$(1 - x^2)\frac{d^{l+2}y}{dx^{l+2}} + (l + 1)(-2x)\frac{d^{l+1}y}{dx^{l+1}} + \frac{l(l + 1)}{2}(-2)\frac{d^l y}{dx^l} + 2xl\frac{d^{l+1}y}{dx^{l+1}} + 2l(l + 1)\frac{d^l y}{dx^l} = 0 \qquad (2\text{-}94)$$

[4] In this derivation, use has been made of the expansion of the nth derivative of a product: $d^n/dx^n(uv) = (d^n u/dx^n)v + n(d^{n-1}u/dx^{n-1})(dv/dx) + \cdots$, a series with binomial coefficients.

If we now substitute into this equation $f = d^l y/dx^l$, we find that f is a solution of Legendre's differential equation:

$$(1 - x^2)\frac{d^2f}{dx^2} - 2x\frac{df}{dx} + l(l+1)f = 0$$

$$\frac{d}{dx}\left[(1 - x^2)^l \frac{df}{dx}\right] + l(l+1)f = 0 \qquad (2\text{-}95)$$

Therefore, $f = d^l y/dx^l = d^l/dx^l(1 - x^2)^l$ is a solution of Legendre's equation. However, the solutions of Legendre's equation are the Legendre polynomials. We have only to insure $f(+1) = 1$, as we required, to establish the relation desired. If the formula is expanded as a derivative of a product,[4] we obtain

$$\begin{aligned} f &= \frac{d^l}{dx^l}[(x+1)^l(x-1)^l] \\ &= [(x-1)^l]\left[\frac{d^l}{dx^l}(x+1)^l\right] \\ &\quad + l\left[\frac{d}{dx}(x-1)^l\right]\left[\frac{d^{l-1}}{dx^{l-1}}(x+1)^l\right] + \cdots \\ &\quad + l\left[\frac{d^{l-1}}{dx^{l-1}}(x-1)^l\right]\left[\frac{d}{dx}(x+1)^l\right] \\ &\quad + \left[\frac{d^l}{dx^l}(x-1)^l\right][(x+1)^l] \end{aligned} \qquad (2\text{-}96)$$

Evaluated at $x = 1$, all the terms but the last are zero, since each will contain powers of $(x - 1)$. The last term, however, does not contain powers of $(x - 1)$: $(d^l/dx^l)(x-1)^l = l!$ and $(x+1)^l = 2^l$. Therefore, $f(1) = 2^l l!$; to make $f(1) = 1$, and agree with the Legendre polynomials, we divide by the factor

$$P_l(x) = \frac{1}{2^l l!}\frac{d^l}{dx^l}(x^2 - 1)^l \qquad (2\text{-}97)$$

Having described four different methods for finding Legendre polynomials (and there are others, too), we shall now use these formulas to derive some of the salient properties of the polynomials.

Evaluation of $P_l(x)$ at $x = -1$, $x = 0$. We have restricted our definition of the polynomials by the boundary condition $P_l(+1) =$

$+1$, but it is also useful to know the values of $P_l(x)$ at the other boundary, $P_l(-1)$, and in the middle of the interval, $P_l(0)$. Rodrigues' formula is useful for obtaining $P_l(-1)$. From an expansion of $P_l(x) = (1/2^l l!)(d^l/dx^l)(1 - x^2)^l$ similar to that given in Eq. 2–96, it is clear that only the first term will be nonzero at $x = -1$, since all other terms involve powers of $(x + 1)$. That first term at $x = -1$ is $(-2)^l l!$. Then,

$$P_l(-1) = \frac{1}{2^l l!}(-2)^l l! = (-1)^l \qquad (2\text{-}98)$$

That is, $P_l(-1)$ is $+1$ or -1 as l is even or odd.

We may also use Rodrigues' formula to evaluate $P_l(0)$. The binomial expansion of $(x^2 - 1)^l$ gives

$$\begin{aligned} (x^2 - 1)^n &= x^{2n} + \frac{n(-1)}{1!}x^{2n-2} + \cdots \\ &\quad + \frac{(-1)^k n(n-1)(n-2)\cdots(n-k+1)}{k!}x^{2n-2k} \\ &\quad + \cdots + n(-1)^{n-1}x^2 + (-1)^n \end{aligned} \qquad (2\text{-}99)$$

For the polynomial P_n, we require the nth derivative of Eq. 2–99. All terms in the binomial expansion afford zero for the result of taking n derivatives and evaluating at $x = 0$, except a possible term in x^n. If there is a term in x^n, the nth derivative, at $x = 0$, will give $n!$. There will be a term in x^n only if n is even, so that there can be some value of k in Eq. 2–99 for which $2n - 2k = n$, namely $k = n/2$. From these considerations, we conclude

$$\frac{d^n}{dx^n}(x^2 - 1)^n\Big|_{x=0} = 0 \qquad (2\text{-}100a)$$

if n is odd, and

$$\frac{d^n}{dx^n}(x^2 - 1)^n\Big|_{x=0} = (-1)^{n/2}\frac{n(n-1)\cdots\left(n - \frac{n}{2} + 1\right)n!}{(n/2)!} \qquad (2\text{-}100b)$$

if n is even.

The nonzero result can be simplified. Let $2l = n$, and then

$$P_{2l}(0) = \left(\frac{1}{2^{2l}(2l)!}\right)(-1)^l\frac{(2l)(2l+1)\cdots(l+2)(l+1)(2l)!}{l!}$$

$$= \left(\frac{1}{2^{2l}}\right)(-1)^l$$
$$\times \frac{(2l)(2l-1)\cdots(l+2)(l+1)(l)(l-1)\cdots(2)(1)}{l!l!}$$
$$= \left(\frac{1}{2^{2l}}\right)(-1)^l \frac{(2l-1)(2l-3)\cdots(3)(1)(2^l)}{l!}$$
$$= (-1)^l \frac{(2l-1)!!}{2^l l!} \qquad (2\text{-}101)$$

where the symbol $(2l-1)!!$, read "$2l-1$ double factorial," indicates a product of every other integer, $n!! = n(n-2)(n-4)\cdots$. This result may also be derived using the generating function for the Legendre polynomials. For, setting $x = 0$ in $F(x, t) = [1 - 2xt + t^2]^{-1/2}$, we obtain

$$F(0, t) = \sum_n P_n(0)t^n = (1 + t^2)^{-1/2}$$
$$= 1 - \frac{1}{1!}\cdot\frac{1}{2}t^2 + \frac{1}{2!}\cdot\frac{1}{2}\cdot\frac{3}{2}t^4 - \cdots$$
$$= \sum_{n \text{ even}} \frac{(n-1)!!(-1)^{n/2}}{(2^{n/2})}\left(\frac{n}{2}\right)!t^n$$
$$= \sum_l \frac{(2l-1)!!(-1)^l}{2^l l! t^{2l}} \qquad (2\text{-}102)$$

A plot of the first four Legendre polynomials is shown in Fig. 2-6. The values of $P_l(\pm 1)$ and $P_l(0)$ are apparent.

Recursion relations among Legendre polynomials. As a further example of the use of the generating function, we derive two recursion relations interrelating the Legendre polynomials. By differentiating Eq. 2-87 with respect to t, we obtain

$$\frac{\partial F(x, t)}{\partial t} = (x - t)(1 - 2xt + t^2)^{-3/2} = \sum_n nt^{n-1}P_n(x) \qquad (2\text{-}103)$$

which may be rearranged to give

$$(x - t)(1 - 2xt + t^2)^{-1/2} = (1 - 2xt + t^2)\sum_n nt^{n-1}P_n(x) \qquad (2\text{-}104)$$

$$(x - t)\sum_n t^n P_n(x) = (1 - 2xt + t^2)\sum_n nt^{n-1}P_n(x)$$

Since the powers of t are linearly independent, we can equate coefficients of t^n:

$$xP_n - P_{n-1} = (n + 1)P_{n+1} - 2nxP_n + (n - 1)P_{n-1}$$

or

$$(n + 1)P_{n+1} - (2n + 1)xP_n + nP_{n-1} = 0 \qquad (2\text{-}105)$$

which interrelates the various Legendre polynomials. For instance, given P_1 and P_2, we may find P_3 by using $n = 2$ in Eq. 2-105:

$$P_3 = \frac{(2\cdot 2 + 1)xP_2 - 2P_1}{2 + 1} = \frac{3P_2 - 2P_1}{3}$$
$$= \frac{5x(\frac{3}{2}x^2 - \frac{1}{2}) - 2x}{3} = \frac{5}{2}x^2 - \frac{5}{6}x - \frac{2}{3}x$$
$$= \frac{5}{2}x^3 - \frac{3}{2}x \qquad (2\text{-}106)$$

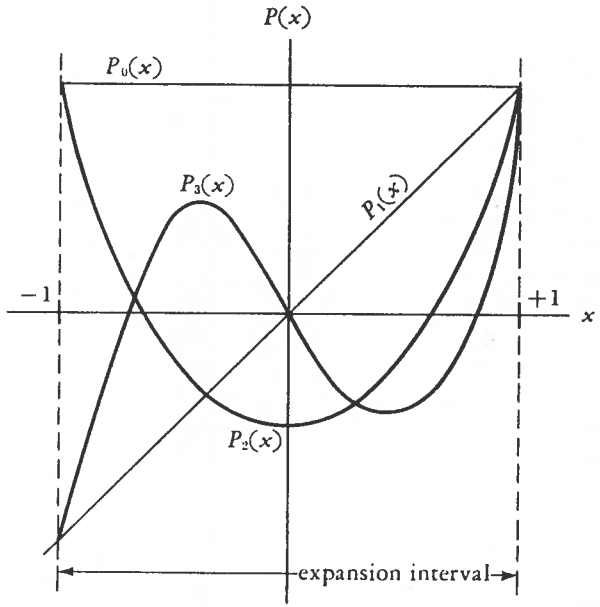


Figure 2-6 Legendre polynomials $P_n(x)$ for $n = 0, 1, 2, 3$.

This recursion relation can be used to generate higher Legendre polynomials.

Another relation, obtained by differentiating Eq. 2-87 with respect to x, is

$$\frac{\partial F(x,t)}{\partial x} = t(1 - 2xt + t^2)^{-3/2} = \sum_n P_n'(x)t^n \qquad (2\text{-}107)$$

which yields

$$t\sum_n P_n t^n = (1 - 2xt + t^2)P_n' t^n \qquad (2\text{-}108)$$

or, for t^n,

$$P_{n-1} = P_n' - 2xP_{n-1}' + P_{n-2}' \qquad (2\text{-}109)$$

This recursion relation relates the Legendre polynomials to their first derivatives. Many other useful relations may be derived by analogous manipulations of the generating function.

Spherical harmonics: an orthonormal set in two variables. Legendre's differential equation, Eq. 2-83, is a special case of the equation

$$\frac{d}{dx}\left[(1 - x^2)\frac{df}{dx}\right] + \left[l(l+1) - \frac{m^2}{1 - x^2}\right]f = 0 \qquad (2\text{-}110)$$

where m is an integer. If $m = 0$, Eq. 2-110 is identical to Eq. 2-83. The solutions to this equation are known as the associated Legendre functions, $P_l^m(x) = (-1)^m(1 - x^2)^{m/2}\, d^m P_l(x)/dx^m$. The associated Legendre functions are marked by two indices, m and l. The associated Legendre functions are also orthogonal on the interval $[-1, 1]$ in the sense $\langle P_{l'}^m \mid P_l^m \rangle = 0$. The inner product of two associated Legendre functions for different l but the same m is zero. The norm of the associated Legendre functions is

$$N(P_l^m) = (2)(l + m)!/(2l + 1)(l - m)!$$

The set of functions

$$\left\{\sqrt{\frac{(2l+1)(l-m)!}{2(l+m)!}}\, P_l^m(x)\right\}$$

defined on $[-1, 1]$, for a fixed value of m and $l = m, m + 1, \ldots$ are complete and orthonormal. The Legendre polynomials themselves are an example of this set, for, if $m = 0$, this set is identical to the set formed by Schmidt orthogonalization.

The associated Legendre functions may be rewritten as functions of an angle variable θ, on $0 \leq \theta \leq \pi$, by replacing x by $\cos\theta$. The functions $\{P_l^m(\theta)\}$ are therefore a complete orthogonal set of functions on the interval $[0, \pi]$ in the variable θ. However, we saw in Section 2-3 that the set $\{(2\pi)^{-1/2}\, e^{im\phi}\}$ is a complete orthonormal set of the interval $[0, 2\pi]$ in the variable ϕ. By multiplying each of the functions $[(2l + 1)(l - m)!/2(l + m)!]^{1/2} P_l^m(\cos\theta)$ by the members of this other complete, orthonormal set, $\{(2\pi)^{-1/2} e^{im\phi}\}$, we obtain a set of functions, complete and orthonormal in two variables, called the *spherical harmonics:*

$$Y_{l,m}(\theta, \phi) = \left[\frac{(2l+1)(l-m)!}{4\pi(l+m)!}\right]^{1/2} P_l^{|m|}(\cos\theta) e^{im\phi} \qquad (2\text{-}111)$$

with

$$Y_{l,-m}(\theta, \phi) = (-1)^m Y_{l,m}^* \qquad (2\text{-}112)$$

The index m is usually called the *order*, and the index l the *rank* of the spherical harmonics. The rank assumes the values $l = 0, 1, 2, \ldots$; the order m, the values $-l, (-l + 1), \ldots, 0, \ldots, (l - 1), l$.

We have now spent some time discussing the ins and outs of Legendre polynomials and the associated Legendre functions. Lest the student miss the forest for the trees, we now recapitulate the salient points of our discussion.

1. A set of polynomials can be formed by Schmidt orthogonalization from the complete set $\{x^n\}$ on $[-1, 1]$. Except for a normalization factor, these are the Legendre polynomials.

2. These same polynomials are solutions of the differential equation $(d/dx)[(1 - x^2)\, df/dx] + l(l + 1)f = 0$.

3. These same polynomials are the coefficients in a series expansion in powers of t of $F(x, t) = (1 - 2xt + t^2)^{-1/2}$.

4. These same polynomials are given by $(1/2^l l!)(d^l/dx^l)(1 - x^2)^l$.

5. A number of relations may be derived from points 3 and 4.

6. Functions $P_l^m(x) = (-1)^m (1 - x^2)^{m/2} (d^m P_l/dx^m)$, called the associated Legendre functions, are solutions of the differential equation $(d/dx)[(1 - x^2)(df/dx)] + [l(l + 1) - (m^2/(1 - x^2))]f = 0$. These functions are orthogonal for different l and the same m.

7. The functions of two variables (the angles θ and ϕ), $Y_{l,m}(\theta, \phi)$, formed by multiplying the normalized associated Legendre functions by the harmonic functions $e^{im\phi}/(2\pi)^{-1/2}$, are orthonormal and complete on the intervals $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$.

This section has attempted to present the two following ideas.

1. An illustration of the ways of forming the important special functions of mathematical physics: orthogonalization, power series solution of a differential equation, and generating function.

2. A record for the student's use—but by no means his memorization—of some of the important properties of a number of functions which occur in chemistry. The behavior of these functions, even in a qualitative way, gives useful information about chemical systems. For example, since $P_{2l+1}(0) = 0$, the p and f atomic orbitals have no electron density perpendicular to their axes. The student should endeavor to become as familiar with the Legendre (and other) functions as he is with sines and cosines, so that he may easily picture the behavior of chemical systems which these functions describe.

For the most part, these and other facts can be found elsewhere. Some students, hopefully, will come to enjoy the mathematical manipulations—investigating the properties of integrals and derivatives—but this is not usually the most important part of attacking a physical problem.

This section concludes with a summary of the occurrence of the special functions of use in quantum chemistry.

Special functions of quantum chemistry

Harmonic, $\{e^{imz}\}$.
Differential equation: $(d^2f/dx^2) + m^2f = 0$.
Orthogonalization: on $[0, 2\pi]$.
Normalization: multiply by $(2\pi)^{-1/2}$.
Occurrence: translational motion.

Legendre, $\{P_l(x)\}$.
Differential equation: $(1 - x^2)(d^2f/dx^2) - 2x\,df/dx + l(l + 1)f = 0$.
Orthogonalization: on $[-1, 1]$ from $\{x^n\}$.
Normalization: $[(2l + 1)/2]^{1/2}$.
Generating function: $(1 - 2xt + t^2)^{-1/2} = \sum_n P_n(x)t^n$.
Occurrence: angular motion.

Associated Legendre, $\{P_l^m(x)\}$.
Differential equation: $(1 - x^2)(d^2f/dx^2) - 2x(df/dx) + [l(l + 1) - m^2/(1 - x^2)]f = 0$.
Orthogonalization: for same m, different l, on $[-1, 1]$.
Normalization: $[(2l + 1)(l - m)!/2(l + m)!]^{1/2}$.
Occurrence: angular motion.

Laguerre, $\{L_n(x)\}$.
Differential equation:
$$x(d^2f/dx^2) + (df/dx) - (\tfrac{1}{2} + x/4 + n)f = 0.$$
Orthogonalization: on $[0, \infty)$, from $\{e^{-x/2}x^n\}$.
Normalization: $1/n!$.
Generating function: $\exp[-xt/(1 - t)]/(1 - t) = \sum_n L_n(x)t^n$.
Occurrence: radial motion.

Hermite, $\{H_n(x)\}$.
Differential equation: $(d^2f/dx^2) - 2x(df/dx) + 2nf = 0$.
Orthogonalization: on $(-\infty, \infty)$, from $\{\exp[-x^2/2]x^n\}$.
Normalization: $[1/2^n n!\pi^{1/2}]^{1/2}$.
Generating function: $\exp[2tx - t^2] = \sum_n H_n(x)t^n/n!$.
Occurrence: harmonic oscillator.

Problems

1. Show that the set of functions $\{\sin mx, \cos nx\}$, for positive, integral values of m and n, are orthogonal on the interval $[-\pi, \pi]$.

2. Classify the following functions as even, odd, or neither even nor odd. If the function is neither even nor odd, decompose the function into the sum of an even function and an odd function. (a) x^2; (b) x^3; (c) $x \sin x$; (d) $x^3 \cos nx$ (n integral); (e) x^4; (f) $\log[(1 + x)/(1 - x)]$; (g) e^x; (h) e^{ix}.

3. Show that the sets of functions $\{\sin mx\}$, $\{\cos mx\}$ are orthogonal on

the interval $[0, \pi]$. Find, on this interval, the norms of these functions. Calculate the expansion coefficients for an expansion in these sets of functions.

4. Expand the function $\pi x - x^2$ on the interval $[-\pi, \pi]$; on the interval $[0, \pi]$ in sines; on the interval $[0, \pi]$ in cosines. Which expansion converges most rapidly on $[0, \pi]$?

5. Expand the function $\sin^2 x$ in a sine series and a cosine series on $[0, \pi]$. Which expansion converges most rapidly? Compare the situation with that of Problem 4.

6. Using the information and results of Problems 4 and 5 and the expansion theorem, calculate the inner product $\langle \sin^2 x \mid \pi x - x^2 \rangle$ on $[0, \pi]$.

7. Show that the functions $\{e^{imx}\}$ are orthogonal on the interval $[0, 2\pi]$. Find the norm of these functions. Calculate the expansion coefficients for an expansion in these functions. Compare these coefficients with the usual expansion coefficients in a Fourier series.

8. A *square wave* is defined by $f(x) = a$ for $0 < x < \pi$, $f(x) = -a$ for $-\pi < x < 0$. Carry out a Fourier series expansion for the square wave.

9. A *triangular wave* is defined by $f(x) = x + \pi$ on $-\pi \leq x \leq 0$, $f(x) = \pi - x$ on $0 \leq x \leq \pi$. Carry out a Fourier series expansion for the triangular wave.

10. Show that the functions $\{\exp[2ni\pi x/(b-a)]\}$ are orthogonal on the interval $[a, b]$.

11. A Taylor series about zero is an expansion in powers of x. Is the Taylor series an expansion in orthogonal functions? Show how different sets of orthonormal functions may be constructed from the powers of x used in the Taylor series.

12. The functions $\{\sin^n x, \cos^n x\}$, for $n = 0, 1, 2, \ldots$ are not an orthonormal set. Show that the orthonormal set which can be constructed from this set on the interval $[-\pi, \pi]$ are just the functions of Problem 1.

13. Use the differential equation, Eq. 2–110, to prove that the associated Legendre functions are, in fact, orthogonal as stated.

14. Verify the orthogonalization of the Laguerre polynomials as stated in the list of special functions; verify their normalization.

15. Verify the normalization of the Hermite polynomials as stated in the list of special functions.

16. Expand the function e^x in Legendre, Hermite, and Laguerre polynomials, and compare the results. Notice that the expansion intervals are quite different.

17. The discussion of the Schmidt procedure was somewhat arbitrary in the choice of the expansion interval. With a number of trial examples, show

what effect changing the expansion interval has on the orthogonal functions.

18. In spectroscopy, *selection rules* govern what transitions between energy levels are allowed. For so-called *electric dipole transitions*, the intensity of a spectroscopic absorption is proportional to $|\langle \psi_1 \mid x\psi_2 \rangle|^2$, which is called the *transition dipole moment*. Calculate the transition dipole moment for the following transitions of the harmonic oscillator, where $\psi_n(x) = \exp[-x^2/2]H_n(x)$.

$$\psi_1 \rightarrow \psi_2 \qquad \psi_2 \rightarrow \psi_3$$
$$\psi_1 \rightarrow \psi_3 \qquad \psi_3 \rightarrow \psi_4$$
$$\psi_1 \rightarrow \psi_4$$

Do you see a pattern emerging? Can you formulate and prove the general selection rules for the harmonic oscillator?

3

Linear Algebra

We mentioned briefly in Chapter 1 that quantum mechanics may be looked upon from two points of view. The first, Schrödinger's wave mechanics, casts eigenvalue equations in the form of differential equations of functions of one or more variables. With this in mind, we studied in depth the concept of orthogonal functions in Chapter 2 in order to establish the kind of language needed to discuss the Schrödinger wave functions. We concluded by solving one such differential equation, Legendre's differential equation, and related the solution to the established concepts and behavior of orthonormal functions.

The second point of view from which quantum mechanics can be studied is Heisenberg's matrix mechanics. In this chapter we shall be concerned with building up the vocabulary of algebra that is used in studying matrix mechanics. We shall also introduce in a more complete fashion the entities called operators, which were mentioned in Chapter 1, and derive some traditional results of interest and importance in quantum chemistry.

3–1 INTRODUCTION

With no further ado, we begin with a barrage of definitions.

Definition An n-dimensional *vector* α is an n-tuple, or list of n numbers: $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$. The n numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$ are

called the *components* of the vector; they are arranged in a prespecified order; they may be real or complex.

We shall denote vectors by Greek letters. An example is a conventional physical vector in three-dimensional space represented by three numbers that denote the x, y, and z components, respectively.

Definition A *vector space* is a set of vectors with two properties:
(a) A unique, commutative sum is defined by $\alpha^1 + \alpha^2 = \alpha^2 + \alpha^1 = (\alpha_1{}^1 + \alpha_1{}^2, \alpha_2{}^1 + \alpha_2{}^2, \ldots, \alpha_n{}^1 + \alpha_n{}^2)$.
(b) Any vector α and any scalar c define a unique product $c\alpha = (c\alpha_1, c\alpha_2, \ldots, c\alpha_n)$ which is both distributive [i.e., $c(\alpha + \beta) = c\alpha + c\beta$] and associative [i.e., $(c_1c_2)\alpha = c_1(c_2\alpha)$].

The student should realize that these properties are familiar in the case of three-dimensional physical vectors. We denote members of a *set* of vectors with a superscript, $\{\alpha^i\}$, but components of a single vector with a subscript.

Definition A *Euclidean vector space* is a vector space for which the components of the vectors are real, and such that with any two vectors we may associate a real *inner product* $\langle\alpha \mid \beta\rangle = \alpha_1\beta_1 + \alpha_2\beta_2 + \cdots + \alpha_n\beta_n$, that is symmetric $[\langle\alpha \mid \beta\rangle = \langle\beta \mid \alpha\rangle]$, bilinear $[\langle\alpha + \beta \mid \gamma\rangle = \langle\alpha \mid \gamma\rangle + \langle\beta \mid \gamma\rangle]$, and positive $[\langle\alpha \mid \alpha\rangle \geq 0]$.

This definition of inner product uses the same symbol, $\langle\ \mid\ \rangle$, that was used for inner products of functions.

A slightly different, and more general, definition of inner product is involved in the definition of a slightly different, and more general, vector space.

Definition A *Hermitian vector space* is a vector space for which the components of the vectors may be complex, and such that with any two vectors α and β, we may associate a complex inner product $\langle\alpha \mid \beta\rangle = \alpha_1{}^*\beta_1 + \alpha_2{}^*\beta_2 + \cdots + \alpha_n{}^*\beta_n$, which is Hermitian $[\langle\alpha \mid \beta\rangle = \langle\beta \mid \alpha\rangle^*]$, bilinear $[\langle\alpha + \beta \mid \gamma\rangle = \langle\alpha \mid \gamma\rangle + \langle\beta \mid \gamma\rangle]$, and positive $[\langle\alpha \mid \alpha\rangle \geq 0]$.

Two comparisons should be made. First, compare the definition of inner product that appears here

$$\langle\alpha \mid \beta\rangle = \sum_i \alpha_i^*\beta_i \tag{3-1}$$

with that of Chapter 2 (Eq. 2-1)

$$\langle f \mid g \rangle = \int f(x)^* g(x)\, dx$$

The difference is that the vector inner product is defined by a sum over a discrete index, whereas the function inner product is defined by an integral over a continuous variable. This difference—the exchange of a discrete index for a continuous variable—is much of what distinguishes the Heisenberg from the Schrödinger picture.

Again, notice the difference between Euclidean and Hermitian vector spaces: the inner product in a Euclidean vector space is real and symmetric; in a Hermitian vector space the inner product is complex and Hermitian. This property of inner products in Hermitian vector spaces should recall the analogous property of the inner product of two functions: *Transposing an inner product gives the complex conjugate of that inner product.*

Definition Two vectors are said to be *orthogonal* if their inner product is zero.

This definition is strictly parallel to the definition of orthogonality of functions; it makes the concept of orthogonality one of sweeping generality, and establishes a connection to the familiar property of perpendicular vectors.

Definition Again, by analogy with the material on functions which has gone before, we define the *norm* of a vector by $N(\alpha) = \langle \alpha \mid \alpha \rangle = |\alpha|^2$.

The norm corresponds to the length of a vector in physical three-dimensional space. Two final definitions continue the parallel structure of the algebra of vector spaces with the calculus of functions.

Definition A vector is said to be *normalized* if its norm is one.

Definition An *orthonormal set of vectors* is a set of vectors, each of which is normalized, and each of which is orthogonal to every other vector in the set.

As an illustration of these definitions, we prove Schwarz's inequality.

Theorem *Schwarz's inequality:* $|\langle \alpha \mid \beta \rangle| \leq |\alpha| \cdot |\beta|$.

If either α or β is zero, the equality sign obtains, and the theorem is trivial. If neither α nor β is zero, we construct the theorem with the aid of two arbitrary scalars, c and d. By the positiveness property of inner products,

$$0 \leq \langle c\alpha + d\beta \mid c\alpha + d\beta \rangle \tag{3-2}$$

and, by bilinearity,

$$\begin{aligned} 0 &\leq c^* \langle \alpha \mid c\alpha + d\beta \rangle + d^* \langle \beta \mid c\alpha + d\beta \rangle \\ &\leq c^* c \langle \alpha \mid \alpha \rangle + c^* d \langle \alpha \mid \beta \rangle + d^* c \langle \beta \mid \alpha \rangle + d^* d \langle \beta \mid \beta \rangle \end{aligned} \tag{3-3}$$

This equation is true for any value of c and d; therefore, it must be true for the particular values $c = -\langle \alpha \mid \beta \rangle$, and $d = \langle \alpha \mid \alpha \rangle$. With this, Eq. 3-3 yields

$$\begin{aligned} 0 &\leq c^* c d - c^* d c - d^* c c^* + d^* d \langle \beta \mid \beta \rangle \\ &\leq d[-c^* c + d^* \langle \beta \mid \beta \rangle] \end{aligned} \tag{3-4}$$

so that

$$0 \leq -|\langle \alpha \mid \beta \rangle|^2 + \langle \alpha \mid \alpha \rangle \langle \beta \mid \beta \rangle \tag{3-5}$$

or

$$|\langle \alpha \mid \beta \rangle|^2 \leq |\alpha|^2 |\beta|^2 \tag{3-6}$$

$$|\langle \alpha \mid \beta \rangle| \leq |\alpha||\beta| \tag{3-7}$$

Having established definitions of inner product and norm which apply to both functions and vectors, we may apply to vectors results which were studied in detail for functions. We proceed to a consideration of sets of vectors; our results will, to a large extent, parallel the results obtained in our earlier study of sets of functions.

The concept of completeness of a set of functions has an analog in terms of vector spaces.

Definition A vector space is *spanned* by a set of vectors $\{\alpha^i\}$ if any vector in the vector space can be expressed as a linear combination of the set $\{\alpha^i\}$.

This definition should be easily pictured in terms of the usual three-dimensional Euclidean vector space (which we shall call 3D space hereafter). As an example, the vectors (1, 0, 0), (0, 1, 0), (0, 0, 1) *span* 3D space. These are just the unit vectors in the three coordinate directions, and it is a familiar fact that any vector may

be expressed as a linear combination of the three unit vectors. This definition prompts a more rigorous interpretation of the idea of dimension.

Definition The *dimension* of a vector space is the minimum number of vectors required to span the vector space.

In the example given above, no two vectors are sufficient to span 3D space, but four vectors, say, (1, 0, 0), (0, 1, 0), (0, 0, 1), and (1, 1, 1) are redundant. Any one of them could be omitted and the remaining three would span 3D space. Of course, not all sets of three vectors span 3D space. The vectors (1, 0, 0), (0, 1, 0), (1, 1, 0), although three (the dimension) in number, do not span 3D space. Hopefully, the reader can see why. These vectors are linearly dependent, and so one is redundant. The concept of linear independence thus makes an appearance again in our consideration of vector spaces.

Definition A set of vectors $\alpha^1, \alpha^2, \ldots, \alpha^n$ is *linearly independent* if the equation

$$c_1\alpha^1 + c_2\alpha^2 + \cdots + c_n\alpha^n = \sum_i c_i\alpha^i = 0 \qquad (3\text{–}8)$$

requires all $c_i = 0$. The set $\{\alpha_i\}$ is *linearly dependent* if Eq. 3–8 can be solved with some $c_i \neq 0$.

These properties of vector spaces allow the convenient concept of a basis.

Definition A *basis* of (or for) a vector space is *some* linearly independent set of vectors that spans the vector space; an *orthonormal basis* of a vector space is some orthonormal set which spans the space.

These definitions should recall the process used to form an orthonormal set of functions from a linearly independent set, the Schmidt orthogonalization procedure. This procedure may be used for vectors with no modification. If the linearly independent set is notated $\{\alpha^i\}$ and the orthonormal set $\{\phi^i\}$, the pertinent formulas are

$$\phi^k = N_k^{-1/2}(\alpha^k - \sum_{j=0}^{k-1} \langle\phi^j \mid \alpha^k\rangle\phi^j) \qquad (3\text{–}9)$$

$$N_k = \langle\alpha^k \mid \alpha^k\rangle - \sum_{j=0}^{k-1} |\langle\phi^j \mid \alpha^k\rangle|^2 \qquad (3\text{–}10)$$

With this background, in almost all respects parallel to that of orthonormal functions, it should be clear that all the algebraic "machinery" is now available for the expansion of vectors in terms of an orthonormal set of vectors. Precisely the same derivation that was set aside from the text in Chapter 2, Eqs. 2–19 through 2–24, applies to the vector problem. We have, therefore, the expansion of an arbitrary vector ξ in the set $\{\phi^i\}$, according to

$$\xi = \sum_i c_i\phi^i \qquad (3\text{–}11)$$

with the expansion coefficients (perhaps complex)

$$c_i = \langle\phi^i \mid \xi\rangle \qquad (3\text{–}12)$$

We may also borrow from Chapter 2 the expansion theorem for inner products,

$$\langle\xi \mid \eta\rangle = \sum_i \langle\xi \mid \phi^i\rangle\langle\phi^i \mid \eta\rangle \qquad (3\text{–}13)$$

where the set $\{\phi^i\}$ is orthonormal.

One departure—a useful one—from the parallel study of vector spaces and functions is that for finite-dimensional vector spaces we may carry out straightforward tests to determine whether or not a set of vectors is linearly independent. The reader may recall that this question was generally bypassed offhand in talking about sets of functions, and we relied there more on intuition than on thorough reasoning.

For sets of vectors, however, we may carefully make an analysis to determine if the set (finite) is linearly independent. This analysis makes a fitting conclusion to this section and an introduction to the next, since the mathematical element called a *matrix* appears.

In considering linear independence, we want to try to find some c_i's, not zero, that satisfy Eq. 3–8. In order to make this search something better than trial and error, a simple procedure has been developed.

Line up the vectors in rows. Suppose there are three vectors α, β, γ, each of which has four components. Our lineup looks like this:

$$\begin{matrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 \\ \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 \end{matrix}$$

Begin with the bottom row. We can surely find some multiple of α_1, which, when added to γ_1, gives 0. In fact, that multiple is $-\gamma_1/\alpha_1$. Multiply all the α_i's (the first row) by this constant, and add to the bottom row. Our lineup now looks like this:

$$\begin{matrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 \\ 0 & \gamma_2 - \dfrac{\gamma_1\alpha_2}{\alpha_1} & \gamma_3 - \dfrac{\gamma_1\alpha_3}{\alpha_1} & \gamma_4 - \dfrac{\gamma_1\alpha_4}{\alpha_1} \end{matrix}$$

In the same way, we can "annihilate" β_1 by finding a multiple of α_1 that when added to β_1 gives 0. By continuing this process (next annihilating γ_2), we tend toward a lineup that looks like this:

$$\begin{matrix} \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & 0 & \times & \times \end{matrix}$$

A lineup like this, with no row all zeros when the lower left corner is all zeros, indicates that no linear combination of the α, β, γ vectors adds to zero, and the vectors are linearly independent. If, however, the lineup were to come out

$$\begin{matrix} \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 \end{matrix}$$

then there was some linear combination of the α, β, and γ vectors which added to zero (final last row), and the vectors would be linearly dependent.

This idea can now be expressed in formal terms.

Definition A rectangular array of scalars (real or complex) is called a *matrix*. The scalar located in the ith row and jth column is called the i, j *elements* of the matrix. A matrix with n rows and m columns is designated an $n \times m$ matrix.

Definition The *elementary row operations* which may be performed on a matrix are

1. Multiplying a row by a constant.
2. Adding two rows together.
3. Exchanging two rows.

Definition The *diagonal* (or, *principal diagonal*) of a matrix A consists of the elements $\{a_{ii}\}$, that is, $a_{11}, a_{22}, a_{33}, \ldots$, and so forth. These elements are called the *diagonal elements*.

Definition A matrix, all of whose elements below and to the left of the diagonal are zero, is called a *triangular matrix*.

Definition Matrices which may be formed from one another by elementary row operations are said to be *row equivalent*.

Theorem A set of vectors is said to be linearly independent if the components of these vectors may be made the rows of a matrix which is row equivalent to a triangular matrix containing no all-zero rows.

This theorem expresses the method that we have outlined for testing for linear independence. The theorem and the method will be clearer with two examples.

EXAMPLE 1

Are the vectors $(1, -1, 3)$, $(2, -4, 1)$, $(0, 3, 2)$ linearly independent? We form the matrix (contained in parentheses) and perform elementary row operations on this matrix.

$$\begin{pmatrix} 1 & -1 & 3 \\ 2 & -4 & 1 \\ 0 & 3 & 2 \end{pmatrix}$$

The 3, 1 element is already zero. Force the 2, 1 element to be zero by adding to the second row -2 times the first row.

$$\xrightarrow[-2 \times [1\text{st row}]]{[2\text{nd row}]} \begin{pmatrix} 1 & -1 & 3 \\ 0 & -2 & -5 \\ 0 & 3 & 2 \end{pmatrix}$$

For simplicity, multiply the second row by $-\frac{1}{2}$.

$$\xrightarrow{[2\text{nd row}] \times (-\frac{1}{2})} \begin{pmatrix} 1 & -1 & 3 \\ 0 & 1 & \frac{5}{2} \\ 0 & 3 & 2 \end{pmatrix}$$

Force the 3, 2 element to be zero by multiplying the second row by -3 and adding to the third row.

$$\xrightarrow[-3 \times [2\text{nd row}]]{[3\text{rd row}]} \begin{pmatrix} 1 & -1 & 3 \\ 0 & 1 & \frac{5}{2} \\ 0 & 0 & -\frac{11}{2} \end{pmatrix}$$

The matrix is now in triangular form. All rows contain at least one nonzero element; therefore, the vectors are linearly independent.

EXAMPLE 2

Are the vectors $(1, 2i, 1 + i)$, $(4, 6 - i, 7)$, and $(-2, -6 + 5i, -5 + 2i)$ linearly independent? We again form the matrix and systematically perform elementary row operations.

$$\begin{pmatrix} 1 & 2i & 1+i \\ 4 & 6-i & 7 \\ -2 & -6+5i & -5+2i \end{pmatrix}$$

$$\xrightarrow{[3\text{rd}] + 2 \times [1\text{st}]} \begin{pmatrix} 1 & 2i & 1+i \\ 4 & 6-i & 7 \\ 0 & -6+9i & -3+4i \end{pmatrix}$$

$$\xrightarrow{[2\text{nd}] - 4 \times [1\text{st}]} \begin{pmatrix} 1 & 2i & 1+i \\ 0 & 6-9i & 3-4i \\ 0 & -6+9i & -3+4i \end{pmatrix}$$

$$\xrightarrow{[3\text{rd}] + [2\text{nd}]} \begin{pmatrix} 1 & 2i & 1+i \\ 0 & 6-9i & 3-4i \\ 0 & 0 & 0 \end{pmatrix}$$

The matrix is now in triangular form. One row is all zeros; therefore, the vectors are linearly dependent. We can go back and find the constants that satisfy Eq. 3–8:

$$2\alpha^1 + \alpha^3 + (-4\alpha^1 + \alpha^2) = 0$$
$$-2\alpha^1 + \alpha^2 + \alpha^3 = 0$$

In this section we have developed the properties of vectors in a way parallel to those developed for functions. The Schmidt orthogonalization procedure, expansion in an orthonormal basis set, and the expansion theorem are results which may be borrowed from Chapter 2. A simple test for linear independence has been demonstrated. This test introduces algebraic elements called matrices, and some of their properties and parts, such as elements, diagonals, rows and columns, and row equivalence.

3–2 MATRICES, DETERMINANTS, AND LINEAR EQUATIONS

Matrices. The definition of a matrix was introduced in the previous section as a formal aid to the test for linear independence of vectors. However, matrices have more applications than to serve as aids in this one kind of problem. From the point of view of quantum mechanics, matrices will be used to represent operators, as was hinted at in Chapter 1. This use of matrices will be studied in detail in Section 4 of this chapter, and prompts our consideration of more of the properties of matrices at this time.

Definition **Two matrices may be added if they have the same dimensions. The *sum* of an $n \times m$ matrix A and an $n \times m$ matrix B is an $n \times m$ matrix C whose elements are given by the formula**

$$c_{ij} = a_{ij} + b_{ij} \tag{3–14}$$

That is, matrices are added by adding their elements. We shall notate matrices by capital letters, and their elements by lower case letters.

EXAMPLE

$$\begin{pmatrix} 1 & 3 & 2 & 5 \\ 0 & 7 & 9 & 4 \\ 6 & -2 & 5 & 1 \end{pmatrix} + \begin{pmatrix} 4 & -3 & 1 & -6 \\ 0 & 0 & -2 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 5 & 0 & 3 & -1 \\ 0 & 7 & 7 & 5 \\ 7 & -3 & 6 & 0 \end{pmatrix}$$

The definition of addition is not one which surprises students. The definition of multiplication, however, does not resemble the usual idea of multiplying two numbers.

Definition **Two matrices may be multiplied if the multiplier has the same number of columns as the multiplicand has rows. The *product* of an $n \times m$ matrix A and an $m \times p$ matrix B is an $n \times p$ matrix C whose elements are given by the formula**

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj} \tag{3–15}$$

Notice the rule which restricts the matrices which can be multiplied.

$$\begin{pmatrix} n \times m \end{pmatrix} \cdot \begin{pmatrix} m \times p \end{pmatrix} = \begin{pmatrix} n \times p \end{pmatrix}$$

In particular, square matrices of the same dimension may always be multiplied.

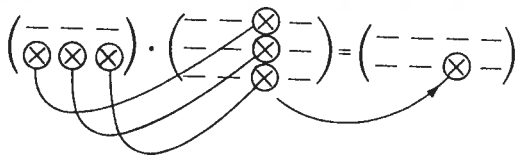
$$\begin{pmatrix} n \times n \end{pmatrix} \cdot \begin{pmatrix} n \times n \end{pmatrix} = \begin{pmatrix} n \times n \end{pmatrix}$$

Since we shall be representing operators by square matrices, their properties are of particular importance to us.

Compare the way in which a product of two matrices is formed to the way in which a sum is formed. Matrices are added by adding their elements, but matrices are not multiplied simply by multiplying their elements. Although we shall have many occasions to use Eq. 3–15 per se, nevertheless, it is useful to see what that equation means in terms of the actual operations. Suppose we wish to compute the product of a 2×3 and a 3×4 matrix. The result is a 2×4 matrix. The 2, 3 element would be given by

$$c_{23} = \sum_{k=1}^{3} a_{2k} b_{k3} = a_{21} b_{13} + a_{22} b_{23} + a_{23} b_{33} \qquad (3\text{–}16)$$

A diagram might help visualize what elements these are.



Often people (even professionals) will multiply matrices by running their left forefinger across the row of the multiplier (left) matrix,

and their right forefinger down the column of the multiplicand (right) matrix.

EXAMPLE 1

As an example, find the product shown below:

$$\begin{pmatrix} 3 & 1 & 2 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \end{pmatrix}$$

By Eq. 3–15 we compute the eight matrix elements:

$$\begin{aligned} c_{11} &= 3 \cdot 1 + 1 \cdot 0 + 2 \cdot 1 = 5 & \qquad c_{21} &= 1 \cdot 1 + 2 \cdot 0 + 3 \cdot 1 = 4 \\ c_{12} &= 3 \cdot 0 + 1 \cdot 1 + 2 \cdot 0 = 1 & \qquad c_{22} &= 1 \cdot 0 + 2 \cdot 1 + 3 \cdot 0 = 2 \\ c_{13} &= 3 \cdot 1 + 1 \cdot 0 + 2 \cdot 1 = 5 & \qquad c_{23} &= 1 \cdot 1 + 2 \cdot 0 + 3 \cdot 1 = 4 \\ c_{14} &= 3 \cdot 0 + 1 \cdot 1 + 2 \cdot 0 = 1 & \qquad c_{24} &= 1 \cdot 0 + 2 \cdot 1 + 3 \cdot 0 = 2 \end{aligned}$$

The product is

$$\begin{pmatrix} 5 & 1 & 5 & 1 \\ 4 & 2 & 4 & 2 \end{pmatrix}$$

The student should verify these elements, and also confirm the "forefinger method" for multiplying the two matrices shown.

"L method"

EXAMPLE 2

Work out the product of two square matrices shown below.

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 4 & 3 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 8 & 5 \\ 20 & 13 \end{pmatrix}$$

Another peculiar property of matrix multiplication is that it is noncommutative. The multiplication of scalars, and the addition of scalars and matrices are all commutative operations: $ab = ba$, $a + b = b + a$, $A + B = B + A$. Matrix multiplication is, in general, not commutative: $AB \neq BA$. We can be concerned with commutation only for square matrices, since nonsquare matrices cannot be multiplied in both directions. In Example 2 above,

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 4 & 3 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 8 & 5 \\ 20 & 13 \end{pmatrix} \neq \begin{pmatrix} 13 & 20 \\ 5 & 8 \end{pmatrix} = \begin{pmatrix} 4 & 3 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

We continue our discussion of the elementary arithmetic of matrices by mentioning two matrices which are the analogs of the numbers zero and one. The *zero*, or *null matrix*, 0, is a matrix, all of whose elements are zero. Analogous to the number zero, the

zero matrix is the matrix that, when added to any matrix (of the same dimensions), gives the same matrix for the sum:

$$A + 0 = A \tag{3-17}$$

The *identity, or unit matrix*, E (denoted by some I or 1), is the (square) matrix whose diagonal elements are all ones, and whose off-diagonal elements are all zeros. This definition may be made very concise by the use of the Kronecker delta. We define

$$e_{ii} = 1 \tag{3-18a}$$

for all i and

$$e_{ij} = 0 \tag{3-18b}$$

for all $i \neq j$. This is equivalent to defining

$$e_{ij} = \delta_{ij} \tag{3-18c}$$

The identity matrix, defined only for square matrices, has a property analogous to the number one, namely, the product of E with any matrix (square and of the same dimension as E) gives that matrix as product: $EA = A$.

For practice in using the general formula, Eq. 3-15, we prove that E commutes with all matrices (of the same dimension). The i, j element of the product EA is

$$(EA)_{ij} = \sum_k e_{ik}a_{kj} = \sum_k \delta_{ik}a_{kj} = a_{ij} \tag{3-19}$$

but

$$(AE)_{ij} = \sum_k a_{ik}e_{kj} = \sum_k a_{ik}\delta_{kj} = a_{ij} = (E.1)_{ij} \tag{3-20}$$

That there is a matrix with properties analogous to the number one prompts us to ask whether there is an analog to the reciprocal (and therefore to division) for matrices.

Definition The *inverse* of a square matrix A, denoted A^{-1}, is that matrix for which $A^{-1}A = AA^{-1} = E$. *Inverse* is defined only for square matrices.

We shall discuss two ways of finding the inverse of a matrix. The first of these is quite similar to the test for linear independence introduced in the previous section. The second is defined in terms of determinants, and will be discussed somewhat later. The idea of row equivalence is the clue to the first method.

Theorem A square matrix A has an inverse if and only if its rows are linearly independent. The inverse may be found by the following procedure. Perform elementary row operations on A to form E. Then, perform these same row operations in the same order on E to form A^{-1}.

Notice that a matrix, even a nonzero matrix, need not have an inverse.

As the theorem in the previous section implied, A is row equivalent to E only if the rows of A are linearly independent. The proof of this theorem is left to the interested reader to find in one of the many excellent texts on algebra. We shall illustrate the theorem by an example.

EXAMPLE

To find the inverse of

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

we begin by performing elementary row operations on A to form E:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \xrightarrow[3 \times [1st]]{[2nd] -} \begin{pmatrix} 1 & 2 \\ 0 & -2 \end{pmatrix} \xrightarrow[-2]{[2nd] \div} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \xrightarrow[2 \times [2nd]]{[1st] -} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = E$$

We then perform these same operations on E:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \xrightarrow[3 \times [1st]]{[2nd] -} \begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix} \xrightarrow[\div -2]{[2nd]} \begin{pmatrix} 1 & 0 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix} \xrightarrow[2 \times [2nd]]{[1st] -} \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix} = A^{-1}$$

As a check, multiply the answer by A:

$$\begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

This technique is cumbersome and prone to error; the student will find the technique using determinants to be somewhat safer. We conclude our discussion of matrix algebra with one further definition.

Definition The *transpose* of a square matrix A, denoted A', has elements $a'_{ij} = a_{ji}$. It can be visualized by flipping the matrix A about its diagonal.

The transpose of a matrix plays an important role in quantum mechanics. Multiplication of transposes follows a simple rule,

$$(AB)' = B'A' \tag{3-21}$$

which is also followed by inverses when they exist,

$$(AB)^{-1} = B^{-1}A^{-1} \tag{3-22}$$

Determinants. The formal definition of a determinant is sufficiently complicated that we begin with a practical example of the use of determinants. One of the important applications of determinants is in the solution of simultaneous linear equations. As an example, consider this set of equations:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{21}x_1 + a_{22}x_2 &= b_2 \end{aligned} \tag{3-23}$$

Writing the equations in this form does, of course, prejudice us toward a matrix point of view which we shall develop later. For the present, we solve Eqs. 3-23 by substitution. We find from the second equation that $x_2 = (b_2 - a_{21}x_1)/a_{22}$, and, substituting this result into the first equation, we obtain the solution

$$x_1 = \frac{a_{22}b_1 - a_{12}b_2}{a_{11}a_{22} - a_{12}a_{21}} \tag{3-24}$$

The denominator of this expression contains all four of the coefficients in the simultaneous equations 3-23. In fact, this denominator is the *determinant* of a matrix which we shall call the *coefficient matrix*.

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{3-25a}$$

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \equiv a_{11}a_{22} - a_{12}a_{21} \tag{3-25b}$$

The numerator of Eq. 3-24 may also be written as a determinant, so that the results are

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} \qquad x_2 = \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} \tag{3-26}$$

This result is a special case of a theorem that we shall discuss later called *Cramer's rule*. The 2×2 determinant has the simple definition given in Eq. 3-25b. Determinants corresponding to large square matrices have a definition that is more complex, and not immediately generalizable from Eq. 3-25b.

Definition A *permutation of n integers*, P, is a way of ordering the n integers; there are $n!$ such permutations. The *sign of a permutation*, sign P, is positive if an even number of interchanges of two indices achieves the permutation, and negative if an odd number of interchanges achieves the permutation.

EXAMPLE

For the integers (1, 2, 3), an odd permutation would be (2, 1, 3); an even permutation would be (2, 3, 1).

Definition The *determinant* of a square matrix A, denoted $|A|$ or $\det A$, is a number, real or complex, such that

$$|A| = \sum_{\text{all perm. } P}^{n!} (\text{sign } P) a_{1,P1}a_{2,P2}\cdots a_{n,Pn} \tag{3-27}$$

EXAMPLE

As an example of this formula for finding determinants, consider a 3×3 example. There are $3! = 6$ terms. Evaluate

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

Term	Permutation	Interchanges	Sign P
$a_{11}a_{22}a_{33}$	1 2 3	0	+
$a_{11}a_{23}a_{32}$	1 3 2	1	−
$a_{12}a_{21}a_{33}$	2 1 3	1	−
$a_{12}a_{23}a_{31}$	2 3 1	2	+
$a_{13}a_{22}a_{31}$	3 2 1	1	−
$a_{13}a_{21}a_{32}$	3 1 2	2	+

Putting these terms together, with the correct signs, we get

$$|A| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \tag{3-28}$$

This is a cumbersome procedure at best, and there is a simpler way. We may get a clue to this simpler way by grouping together all the terms which contain a_{11}, a_{12}, and a_{13}, respectively.

$$|A| = a_{11}(a_{22}a_{33} - a_{23}a_{32}) + a_{12}(a_{23}a_{31} - a_{21}a_{33}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \tag{3-29a}$$

$$= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \tag{3-29b}$$

Eq. 3-29a illustrates the *expansion of a determinant in cofactors;* Eq. 3-29b, the *expansion of a determinant in minors.*

Definition The *cofactor*, A_{ij}, of an element a_{ij} in an $n \times n$ square matrix A is a number that can be represented by an $(n-1) \times (n-1)$ determinant, such that the determinant of A is given by

$$|A| = \sum_j a_{ij}A_{ij}$$

for any and all i. The *minor*, M_{ij}, of an element a_{ij} in an $n \times n$ square matrix A is a number, represented by an $(n-1) \times (n-1)$ determinant found by striking the ith row and jth column from A. The minor and the cofactor are related by at most a change of sign: $A_{ij} = M_{ij}(-1)^{i+j}$.

EXAMPLE

As an example, consider the 2, 3 minor and cofactor for this 4×4 matrix:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$$

strike out 2nd row

strike out 3rd column

$$M_{23} = \begin{vmatrix} a_{11} & a_{12} & a_{14} \\ a_{31} & a_{32} & a_{34} \\ a_{41} & a_{42} & a_{44} \end{vmatrix}$$

$$A_{23} = -\begin{vmatrix} a_{11} & a_{12} & a_{14} \\ a_{31} & a_{32} & a_{34} \\ a_{41} & a_{42} & a_{44} \end{vmatrix} = (-1)^{2+3}M_{23}$$

Theorem The determinant of an $n \times n$ matrix A may be expanded in cofactors or minors:

$$\det A = \sum_j a_{ij}A_{ij} = \sum_j a_{ij}(-1)^{i+j}M_{ij} \tag{3-30}$$

for any i.

EXAMPLE

Evaluate the 3×3 determinant shown below by expansion in minors and by direct use of Eq. 3-28.

$$\begin{vmatrix} 1 & 0 & 2 \\ 3 & 4 & -1 \\ -2 & 1 & 1 \end{vmatrix} = 1\begin{vmatrix} 4 & -1 \\ 1 & 1 \end{vmatrix} - 0\begin{vmatrix} 3 & -1 \\ -2 & 1 \end{vmatrix} + 2\begin{vmatrix} 3 & 4 \\ -2 & 1 \end{vmatrix} = 5 - 0 + 22 = 27$$

$$= -3\begin{vmatrix} 0 & 2 \\ 1 & 1 \end{vmatrix} + 4\begin{vmatrix} 1 & 2 \\ -2 & 1 \end{vmatrix} - (-1)\begin{vmatrix} 1 & 0 \\ -2 & 1 \end{vmatrix} = 6 + 20 + 1 = 27$$

$$= -2\begin{vmatrix} 0 & 2 \\ 4 & -1 \end{vmatrix} - 1\begin{vmatrix} 1 & 2 \\ 3 & -1 \end{vmatrix} + 1\begin{vmatrix} 1 & 0 \\ 3 & 4 \end{vmatrix} = 16 + 7 + 4 = 27$$

$$= (1)(4)(1) + (0)(-1)(-2) + (2)(3)(1) - (1)(-1)(1) - (0)(3)(1) - (2)(4)(-2) = 4 + 0 + 6 + 1 - 0 + 16 = 27$$

The first equation in the example is an expansion in minors of the first row; the second, of the second row; the third, of the third row. The last equation is a direct evaluation, using Eq. 3-28, the special case of Eq. 3-27.

Expansion by minors is especially helpful if there are zeros in the determinant, as in the first equation of the previous example.

The elementary row operations that we have used to test for linear independence in vectors may also be used to aid in the evaluation of determinants.

Theorem Multiplying a row of a square matrix by a constant multiplies the determinant of that matrix by that constant.

If we expand A in cofactors of the row in question,

$$|A| = \sum_j a_{ij}A_{ij} \tag{3-31}$$

Multiplying the ith row by a constant c gives a new matrix B, whose determinant is

$$|B| = \sum_j b_{ij}B_{ij} = \sum_j ca_{ij}A_{ij} = c\sum_j a_{ij}A_{ij} = c|A| \quad (3\text{-}32)$$

Theorem The exchange of two rows of a matrix changes the sign of the determinant of that matrix.

This result can be proved by relying on the general definition of a determinant. Without going into details, all the terms in the expansion are the same for the original matrix and the matrix with the exchanged rows. The signs of the terms will be different throughout because one additional exchange of indices has been made. This theorem has an immediate corollary.

Corollary A square matrix with two identical rows has a zero determinant.

This is so because exchanging the identical rows gives a determinant of opposite sign, but exchanging the identical rows must also give the same determinant. Hence $|A| = -|A|$, which is true only if $|A| = 0$. Finally, the result of performing the last elementary row operation is obtained.

Theorem The addition of two rows of a square matrix leaves the determinant unchanged.

We may prove this by expanding the determinant of B, where the sum of the ith and kth rows of A forms the ith row of B. Use cofactors of the ith row of B.

$$|B| = \sum_j b_{ij}B_{ij} = \sum_j (a_{ij} + a_{kj})A_{ij} = \sum_j a_{ij}A_{ij} + \sum_j a_{kj}A_{ij} \quad (3\text{-}33)$$

The first term is the expansion of $|A|$ in cofactors of the ith row of A. The second term looks like an expansion in cofactors—it would be, if the matrix A had identical ith and kth rows. Hence the second term represents a cofactor expansion of a determinant with two identical rows; that determinant is zero. Therefore, $|B| = |A|$.

By way of recapitulation, the accompanying table gives the results of the elementary row operations on the value of a determinant.

Elementary row operation	Effect on value of determinant
1. Multiply row by constant.	Multiply determinant by constant.
2. Exchange rows.	Change sign or determinant.
3. Add rows.	No change.

We conclude this discussion of elementary properties of determinants with two further results of importance.

Theorem The determinant of the product of two square matrices is the product of their determinants: $|AB| = |A| \cdot |B|$.

Theorem The determinant of a square matrix is equal to the determinant of the transpose of that matrix: $|A| = |A'|$.

The proof of the latter theorem relies on an examination of the general definition of a determinant. All the terms are alike, and all the signs match as well. This theorem allows us to rephrase all our results about matrices in a column language rather than a row language. These statements are given in the following list, which also serves to sum up our knowledge of determinants.

Row-language statements	Analogous column-language statements
1. Expansion of a determinant by row cofactors or minors: $\|A\| = \sum_i a_{ij}A_{ij} = \sum_i a_{ij}M_{ij}(-1)^{i+j}$.	Expansion of a determinant by column cofactors or minors: $\|A\| = \sum_j a_{ij}A_{ij} = \sum_j a_{ij}M_{ij}(-1)^{i+j}$.
2. Multiply row by constant, multiply determinant by constant.	Multiply column by constant, multiply determinant by constant.
3. Interchange rows, change sign of determinant.	Interchange columns, change sign of determinant.
4. Matrix with two identical rows has zero determinant.	Matrix with two identical columns has zero determinant.
5. Adding two rows together leaves determinant unchanged.	Adding two columns together leaves determinant unchanged.

With this background in the properties of determinants, we may approach two problems in a simpler way. Both the problem of testing for linear independence of vectors (as rows of a matrix) and the problem of matrix inversion have been solved using row equivalence. Each of these procedures is cumbersome, especially for large matrices. A straightforward application of the properties of determinants affords a more compact solution to both of these problems.

Theorem **A square matrix A has an inverse if $|A| \neq 0$; the i, j element of the inverse matrix A^{-1} is**

$$(A^{-1})_{ij} = \frac{A'_{ij}}{|A|} = \frac{A_{ji}}{|A|} \tag{3-34}$$

Expanding $|A|$ in cofactors, we have

$$a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} = |A| = \sum_j a_{ij}A_{ij} \tag{3-35}$$

but, using the same principle as we did in Eq. 3-33 to set the second term equal to zero, we have also

$$\sum_j a_{kj}A_{ij} = 0 \tag{3-36}$$

for $i \neq k$. These may be combined as

$$\sum_j a_{kj}\frac{A_{ij}}{|A|} = \delta_{ki} \tag{3-37}$$

where δ_{ik} is an element of the unit matrix E. The left side of Eq. 3-37 looks almost like a matrix product. We may make it be such by writing, for the cofactor A_{ij}, the cofactor A'_{ji} from the transposed matrix. Then the element $A'_{ji}/|A| = a^{-1}_{ji}$, since

$$\sum_j a_{kj}a^{-1}_{ji} = \delta_{ki}.$$

Notice that these elements exist only for nonzero $|A|$. Here, the symbol a^{-1}_{ij} is not used to represent $1/a_{ij}$, but to represent the i, j element of the matrix A^{-1}. This affords a more straightforward path to matrix inversion. As an example, let us do the previous example again.

EXAMPLE

Find the inverse of

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

By the formula given above,

$$a^{-1}_{11} = \frac{A'_{11}}{|A|} = \frac{4}{-2} = -2$$

$$a^{-1}_{12} = \frac{A'_{12}}{|A|} = \frac{-2}{-2} = 1$$

$$a^{-1}_{21} = \frac{A'_{21}}{|A|} = \frac{-3}{-2} = \frac{3}{2}$$

$$a^{-1}_{22} = \frac{A'_{22}}{|A|} = \frac{1}{-2} = \frac{-1}{2}$$

hence

$$A^{-1} = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}$$

as was found previously.

Finally, we may combine our result on linear independence and matrix inversion (derived by row-equivalence arguments) with our result on nonzero determinants and matrix inversion to yield the following result relating linear independence to nonzero determinants.

Theorem **The rows of a square matrix are linearly independent if and only if the determinant of the matrix is nonzero.**

Simultaneous Linear Equations. The methods of matrix and determinant algebra may be applied with force to the solution of simultaneous linear equations. The results will be stated without proof, and a number of examples will illustrate the methods.

By way of introduction to this problem, note that the general set of simultaneous linear equations may be represented as a single matrix equation. A set of m equations in n unknowns may be represented either by

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \quad\quad \vdots \quad\quad\quad\quad \vdots \quad\quad \vdots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{3-38}$$

or by the matrix equation

$$AX = B \tag{3-39}$$

where A is an $m \times n$ matrix of coefficients, X an $n \times 1$ matrix of unknowns, and B an $m \times 1$ matrix of constants:

$$\begin{pmatrix} a_{11} & a_{12} \cdots a_{1n} \\ a_{21} & a_{22} \cdots a_{2n} \\ \vdots & \vdots \quad \vdots \\ a_{m1} & a_{m2} \cdots a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \tag{3-40}$$

We shall define at the outset two concepts.

Definition The *augmented matrix* for a set of m simultaneous linear equations in n unknowns is an $m \times (n + 1)$ matrix formed by appending the $m \times 1$ matrix of constants on the right of the $n \times m$ matrix of coefficients. If, for example, A and B are the matrices in Eq. 3–40, then the augmented matrix $^*A$ is

$$^*A = \begin{pmatrix} a_{11} & a_{12} \cdots a_{1n} & b_1 \\ a_{21} & a_{22} \cdots a_{2n} & b_2 \\ \vdots & \vdots \quad \vdots & \vdots \\ a_{m1} & a_{m2} \cdots a_{mn} & b_m \end{pmatrix} \tag{3-41}$$

Definition A set of simultaneous linear equations is called *homogeneous* if the matrix B is zero.

The results, offered almost entirely without proof, are expressed in the following theorems.

Theorem A set of n simultaneous linear equations in n unknowns has a solution if the coefficient determinant is not zero. The solution may be constructed by matrix inversion, or by *Cramer's rule*.

If $AX = B$, then $A^{-1}AX = A^{-1}B$. So the entire set of solutions X may be generated if A^{-1} is known. However, A^{-1} exists only if $|A| \neq 0$. The result is

$$X = A^{-1}B \tag{3-42}$$

In particular,

$$x_i = \sum_j (A^{-1})_{ij} b_j$$

$$= \sum_j \frac{A_{ji}}{|A|} b_j = \sum_j \frac{b_j A_{ji}}{|A|} \tag{3-43}$$

Equation 3–43 involves a sum,

$$\sum_j b_j A_{ji}$$

which is the column cofactor expansion of a determinant whose ith column is the constant matrix B. However, this is Cramer's rule in its familiar form:

$$x_i = \frac{\begin{vmatrix} a_{11} \cdots b_1 \cdots a_{1n} \\ a_{21} \cdots b_2 \cdots a_{2n} \\ \vdots \quad \vdots \quad \vdots \\ a_{n1} \cdots b_n \cdots a_{nn} \end{vmatrix}}{|A|} \tag{3-44}$$

Theorem A set of m simultaneous linear equations in n unknowns, $AX = B$, has a solution if the number of linearly independent rows, r, in A is the same as the number of linearly independent rows in $^*A$. If such is the case, r of the unknowns may be expressed in terms of the remaining $n - r$, which may be given arbitrary values.

Theorem A set of m simultaneous, linear, homogeneous equations in n unknowns always has the trivial solution $X = 0$. There will be a nonzero (nontrivial) solution if the coefficient matrix, A has fewer linearly independent rows than there are unknowns. Again, r of the unknowns may be expressed in terms of the remaining $n - r$, which may be assigned arbitrary values.

We conclude this section with a number of examples of these theorems.

EXAMPLE 1

Two linear equations in two unknowns.

$$4x + 4y = 2$$
$$8x - 2y = 4$$

The Cramer's rule solution is

$$x = \frac{\begin{vmatrix} 2 & 4 \\ 4 & -2 \end{vmatrix}}{\begin{vmatrix} 4 & 4 \\ 8 & -2 \end{vmatrix}} = \frac{-20}{-40} = \frac{1}{2}$$

$$y = \frac{\begin{vmatrix} 4 & 2 \\ 8 & -4 \end{vmatrix}}{\begin{vmatrix} 4 & 4 \\ 8 & -2 \end{vmatrix}} = \frac{0}{-40} = 0$$

The matrix solution can be found from the rule that, if $AX = C$, $X = A^{-1}C$. The inverse of the coefficient matrix is the matrix

$$\begin{pmatrix} \frac{1}{20} & \frac{1}{10} \\ \frac{1}{8} & -\frac{1}{10} \end{pmatrix}$$

Then the solution to the set of equations is

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{20} & \frac{1}{10} \\ \frac{1}{8} & -\frac{1}{10} \end{pmatrix} \begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} \frac{2}{20} + \frac{4}{10} \\ \frac{2}{8} - \frac{4}{10} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix}$$

This matrix equation says $x = \frac{1}{2}$, $y = 0$.

EXAMPLE 2

Three equations in two unknowns.

$$\begin{aligned} 4x + 4y &= 2 \\ 8x - 2y &= 4 \\ x + y &= 1 \end{aligned}$$

The matrix of coefficients,

$$\begin{pmatrix} 4 & 4 \\ 8 & -2 \\ 1 & 1 \end{pmatrix}$$

is row equivalent to

$$\begin{pmatrix} 1 & 1 \\ 0 & 10 \\ 0 & 0 \end{pmatrix}$$

and has two linearly independent rows. The augmented matrix,

$$\begin{pmatrix} 4 & 4 & 2 \\ 8 & -2 & 4 \\ 1 & 1 & 1 \end{pmatrix}$$

is row equivalent to

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

and has three linearly independent rows. There is, therefore, no solution to this set of equations.

EXAMPLE 3

Three equations in two unknowns.

$$\begin{aligned} 4x + 4y &= 2 \\ 8x - 2y &= 4 \\ 3x + \frac{y}{2} &= \frac{3}{2} \end{aligned}$$

The matrix of coefficients,

$$\begin{pmatrix} 4 & 4 \\ 8 & -2 \\ 3 & \frac{1}{2} \end{pmatrix}$$

is row equivalent to

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

and has two linearly independent rows. The augmented matrix

$$\begin{pmatrix} 4 & 4 & 2 \\ 8 & -2 & 4 \\ 3 & \frac{1}{2} & \frac{3}{2} \end{pmatrix}$$

is row equivalent to

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

and also has two linearly independent rows. This set of equations does have a solution. The solution can be obtained from any matrix row equivalent to the augmented matrix. Using this last matrix as a guide, write the *equivalent* matrix equation

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix}$$

The solution to this equation can be found by the usual technique (see Example 1), and is $x = \frac{1}{2}, y = 0$, which is the same as the solution to Example 1. This could also be expressed by saying that, since the equations of this example are linearly dependent, and since two of these equations are identical to the equations of Example 1, the solution to this set must also be identical to the solution to Example 1.

EXAMPLE 4

Two equations in three unknowns.

$$\begin{aligned} x + y + z &= 2 \\ x - y - z &= 1 \end{aligned}$$

The coefficient matrix,

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix}$$

has two linearly independent rows, since it is row equivalent to the matrix

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & -2 & -2 \end{pmatrix}$$

The augmented matrix,

$$\begin{pmatrix} 1 & 1 & 1 & 2 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

which is row equivalent to

$$\begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -2 & -2 & -1 \end{pmatrix}$$

also has two linearly independent rows; therefore, this set of equations has a solution. As in the previous example, this set of equations can be solved by appealing to a matrix which is row equivalent to the augmented matrix, and writing

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & -2 & -2 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x + y + z \\ -2y - 2z \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

whence, $x = 2 - y - z$ and $y + z = \frac{1}{2}$. The solution can be expressed $x = \frac{3}{2}, z = \frac{1}{2} - y$. There are an infinite number of solutions. That is, there is

one solution for every possible value of y. The general solution can be checked by substitution into the equations themselves, which gives an identity in each case. One solution might be $x = \frac{3}{2}, y = 0, z = \frac{1}{2}$; another might be $x = \frac{3}{2}, y = 1, z = -\frac{1}{2}$, and so forth.

EXAMPLE 5

Two equations in three unknowns.

$$\begin{aligned} x + y + z &= 2 \\ x + y + z &= 1 \end{aligned}$$

This set of equations has no solutions, which is obvious by inspection. The result can be proved by examining the coefficient matrix,

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

which is row equivalent to

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

and therefore has one linearly independent row. On the other hand, the augmented matrix,

$$\begin{pmatrix} 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

is row equivalent to

$$\begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

and therefore has two linearly independent rows. Since the coefficient matrix and the augmented matrix have different numbers of linearly independent rows, there is no solution to this set of equations.

EXAMPLE 6

Three homogeneous equations in three unknowns.

$$\begin{aligned} 3x + 4y + z &= 0 \\ 2x + 6y + 4z &= 0 \\ x - y + z &= 0 \end{aligned}$$

The coefficient matrix,

$$\begin{pmatrix} 3 & 4 & 1 \\ 2 & 6 & 4 \\ 1 & -1 & 1 \end{pmatrix}$$

has a determinant equal to 30, and three linearly independent rows, and therefore there is no solution to this set of homogeneous equations, except the trivial solution $x = y = z = 0$.

EXAMPLE 7

Three homogeneous equations in three unknowns.

$$\begin{aligned} x + y + z &= 0 \\ x - y - z &= 0 \\ x + 3y + 3z &= 0 \end{aligned}$$

The coefficient matrix has a determinant

$$\begin{vmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & 3 & 3 \end{vmatrix} = 0$$

and there is a solution to this set of equations. To find the solution, rely again on a matrix which is row equivalent to the augmented matrix. The augmented matrix is

$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & -1 & -1 & 0 \\ 1 & 3 & 3 & 0 \end{pmatrix}$$

which is row eqiuvalent to

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

and has only two linearly independent rows. The solution of a matrix equation written from this augmented matrix is $x + y + z = 0$, and $y + z = 0$. That is to say, a general solution is $x = 0$, $y = -z$. There are any number of particular solutions, as before. One is the ever present trivial solution $x = y = z = 0$; another might be $x = 0$, $y = 1$, $z = -1$.

3–3 LINEAR TRANSFORMATIONS

In defining a function in Chapter 2, we emphasized the single value that the function delivered if presented with a single value of the independent variable on some specified interval. A transformation is a generalization of this concept.

Definition Let there exist n independent variables x_i $(i = 1, 2, \ldots, n)$, each defined on a specified interval, such as $a_1 \leq x_1 \leq b_1$, $a_2 \leq x_2 \leq b_2$, and so forth. If there then exist m dependent variables, y_i, each of which is a single-valued function of the n independent variables x_i, we say that there exists a *transformation* which carries a region in n-dimensional space (or n space) into m-dimensional space (or m space). We write $y_i = T(x_i)$, and refer to the set of values y_i as the *image* of the set of values x_i under T. We shall represent transformations with capital letters.

To continue the algebraic theme established in the first sections of this chapter, we restrict the discussion at once to linear transformations.

Definition A transformation A is linear if

1. $A(x_i + x_j) = A(x_i) + A(x_j)$
2. $A(cx_i) = cA(x_i)$

where c is a scalar.

This definition contains two aspects of linearity which should be familiar in retrospect. These two aspects necessarily imply that a linear transformation has the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= y_1 \\ \vdots \quad\quad \vdots \quad\quad & \quad\quad \vdots \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= y_m \end{aligned} \tag{3–45}$$

which, expressed in matrix language, is

$$AX = Y \tag{3–46}$$

where A is an $m \times n$ matrix, X an $n \times 1$ matrix, and Y an $m \times 1$ matrix. Having discovered that matrices may be used to represent linear transformations, we may call upon our past development of matrix algebra to help understand the properties of these transformations. The introduction of the term *rank* will simplify the discussion.

Definition The *rank of a matrix* is the largest number of linearly independent rows in the matrix; alternatively, the rank of a matrix is the dimension of the largest nonzero determinant which can be formed from the matrix elements. The *rank of a linear transformation* is equal to the rank of the matrix which represents the linear transformation.

The relationship between nonzero determinants and linear independence establishes the alternative definition of rank. We shall now present a general result and then illustrate it with a specific example.

Theorem **Suppose a linear transformation of n space into m space has rank r. The image in m space will be an r-dimensional subregion which is described by a linear equation in the m coordinates of m space. In particular, the transformation will map n space into all of m space if the transformation has rank $r = m$.**

By way of proof, consider the $m \times n$ matrix which represents a linear transformation. If only r of the rows of this matrix are linearly independent, then $m - r$ of the rows are dependent. This implies that $m - r$ of the coordinates in m space are linearly dependent. As a consequence, the image of n space in m space is on an r-dimensional subregion whose equation is linear in the m coordinates.

EXAMPLE

Consider the transformation

$$\left.\begin{aligned} x + y + z &= u \\ x + y + z &= v \end{aligned}\right\} T$$

The rank of T is the rank of the matrix

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

which has one linearly independent row; the rank is therefore one. In fact, since the rows of the matrix are the same, $u = v$. Hence the image of XYZ space on UV space is the line $u = v$, a one-dimensional subregion in UV space.

In quantum mechanics, where, as the student will learn, a premium is placed on the properties of orthogonality and normality of functions or vectors, particular attention is paid to the kinds of transformations which retain these properties. In a Euclidean vector space, this kind of transformation is called *orthogonal;* in a Hermitian vector space, the analogous transformation is called *unitary*.

Definition **An *orthogonal transformation* preserves the length (normalization) and orthogonality of vectors in a Euclidean vector space; a *unitary transformation* preserves orthogonality and normalization of vectors in a Hermitian vector space.**

This definition is basic, but rather sterile unless we interpret it in terms of the structure of the transformation—the relations between the matrix elements of the matrix representing the transformation. We may examine this structure by using this very definition. Suppose there is a complete, orthonormal set of vectors $\{\phi^i\}$. We have already established that an arbitrary vector α may be expanded in terms of the vectors $\{\phi^i\}$ according to

$$\alpha = \sum_i a_i \phi^i \tag{3-47}$$

as described in Eq. 3-11 and the discussion there. Now suppose we require that the expansion of α in terms of the ϕ^i's preserve length and orthogonality. That is, suppose α is a member of a second complete orthonormal set $\{\psi^i\}$, $\alpha = \psi^j$, say. Then

$$\psi^j = \sum_i a_{ji} \phi^i \tag{3-48}$$

The coefficient a_i in Eq. 3-47 has now been modified with a second index j indicating which of the vectors in the set $\{\psi^j\}$ is being considered. The inverse expansion is also possible (this is to be proved in one of the problems), so that

$$\phi^i = \sum_k b_{ik} \psi^k \tag{3-49}$$

Combining these equations, we find

$$\psi^j = \sum_i a_{ji} \phi^i = \sum_i a_{ji} \sum_k b_{ik} \psi^k = \sum_k \left(\sum_i a_{ji} b_{ik} \right) \psi^k \tag{3-50}$$

The only way for Eq. 3-50 to be true, since the ψ^i's are linearly independent, is for $\sum_i a_{ji} b_{ik} = 1$ for $j = k$, and $\sum_i a_{ji} b_{ik} = 0$ for $j \neq k$. This then gives

$$\sum_i a_{ji} b_{ik} = \delta_{jk} \tag{3-51}$$

a statement that, if expressed not in the language of expansion coefficients but in the language of matrices, becomes

$$AB = E \tag{3-52}$$

or

$$A = B^{-1} \tag{3-53}$$

This should not be surprising. The transformation from the basis set $\{\phi^i\}$ to the basis set $\{\psi^i\}$ must be representable by a square matrix with nonzero determinant; hence, its inverse must exist. It is only reasonable that the inverse transformation be directly related to the reverse expansion of vectors.

We have not yet drawn upon the property that the orthonormality of the vectors be preserved under this transformation. Equation 3–53 is a result of requiring that completeness be preserved. For orthonormality of the ψ^i's, we have, in a Euclidean vector space,

$$\begin{aligned} \langle \psi^i \mid \psi^j \rangle = \delta_{ij} &= \sum_k \sum_l \langle \phi^k b_{ik} \mid \phi^l b_{jl} \rangle \\ &= \sum_{kl} \langle \phi^k \mid \phi^l \rangle b_{ik} b_{jl} = \sum_{kl} \delta_{kl} b_{ik} b_{jl} \\ &= \sum_k b_{ik} b_{jk} \end{aligned} \tag{3-54}$$

Equation 3–54 is almost a matrix product, but not quite. If we again invoke the definition of a transpose, we derive

$$\delta_{ij} = \sum_k b_{ik} b'_{kj} \tag{3-55}$$

which implies that

$$BB' = E \tag{3-56}$$

or

$$B^{-1} = B' \tag{3-57}$$

We have thereby proven the following results.

Theorem **(a) The inverse of an orthogonal matrix is the transpose of that orthogonal matrix.**
(b) The inverse of a unitary matrix is the transpose complex conjugate of that matrix.
(c) The rows of orthogonal or unitary matrices are orthonormal vectors.
(d) The columns of orthogonal or unitary matrices are orthonormal vectors.

Part (b) can be proved by deriving Eqs. 3–54 through 3–57 again for Hermitian vectors by supplying complex conjugation at the appropriate places. Part (c) should be clear from Eq. 3–54, and (d) is analogous. The last aspect of the structure of orthogonal and unitary transformations is a statement about the determinants of their representative matrices. We know, for orthogonal transformations, that $BB' = E$; hence,

$$|B|\,|B'| = |E| = 1 \tag{3-58}$$

Since $|B'| = |B|$,

$$|B| = \pm 1 \tag{3-59}$$

Together with the analogous result for unitary transformations, we have proven the following theorem:

Theorem **The determinant of an orthogonal matrix is either +1 or −1; the determinant of a unitary matrix has modulus 1.**

A particular, and useful, geometrical application of orthogonal transformations is rotation in two and three dimensions. We begin by considering a rotation in two dimensions. If a rotation through the angle α is carried out, the point p will be transformed into the point p'. The coordinates of p' are

$$\begin{aligned} x' &= x \cos\alpha + y \sin\alpha \\ y' &= -x \sin\alpha + y \cos\alpha \end{aligned} \tag{3-60}$$

wrong!

which is a set of linear equations. In matrix language, the rotational transformation may be represented by

$$R_{\text{point}} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \tag{3-61}$$

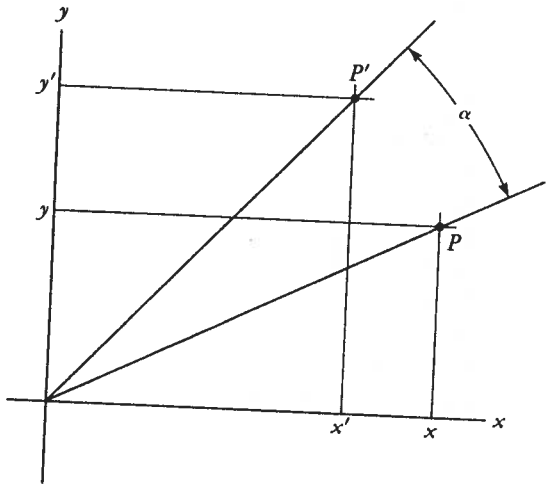


Figure 3–1 Rotation in two dimensions.

The student should notice that in this process the coordinate axes of Fig. 3–1 stay fixed but the point moves. The problem can also be described by a rotation of the coordinate axes in the opposite sense, that is, through an angle of $-\alpha$:

$$R_{\text{axis}} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \tag{3-62}$$

Both these points of view occur, and the important point to remember is that a rotational transformation of a point has the opposite sense of a rotation of the coordinate system.

With the matrix representative of a two-dimensional rotational transformation in mind, we may now examine rotations in three dimensions. There are many possible ways to consider a three-dimensional rotation, but all these ways have two principles in common. First, three angles are, in general, necessary to describe the rotation of a three-dimensional object. Playing with an object, preferably unsymmetrical, will convince the reader of this. Second, the customary way of describing the rotation is by the following sequence.

(a) Rotation about one of the coordinate axes by an angle ϕ.

(b) Rotation about the new position of another one of the coordinate axes by an angle θ.

(c) Rotation about the new position of the original axis by an angle ψ.

This sequence of rotations is called a rotation through *Eulerian angles* (after the mathematician Euler). Modern writers have made different choices for the sequence, and the student should carefully note a particular writer's convention. In agreement with a number of well-known textbooks, we shall make the following choice, illustrated in Fig. 3–2.

(a) Rotation through an angle ϕ about the z axis to yield a new $x'y'z$ axis system (matrix representative A).

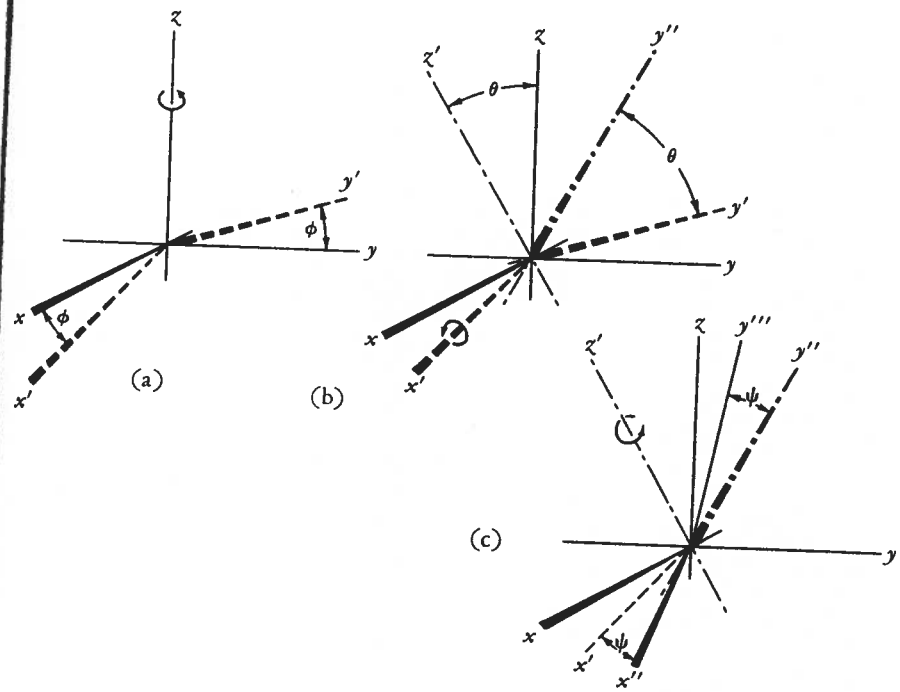


Figure 3–2 Rotation in three dimensions through Eulerian angles: (A) rotation about z by ϕ; (B) rotation about new x by θ; (C) rotation about new z by ψ.

(b) Rotation through an angle θ about the x' axis to yield a new $x'y''z'$ axis system (matrix representative B).

(c) Rotation through an angle ψ about the z' axis to yield a new, and final, $x''y'''z'$ axis system (matrix representative C).

This sequence of operations represents a rotation of the coordinate system. The matrix representatives of each separate transformation can be found by analogy with Eq. 3–62, and are

$$A = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3–63a}$$

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} \tag{3–63b}$$

$$C = \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3–63c}$$

The over-all effect of these three separate rotations gives the general three-dimensional rotation matrix, R,

$$R = CBA = \begin{pmatrix} \cos\psi\cos\phi - \cos\theta\sin\phi\sin\psi & \cos\psi\sin\phi + \cos\theta\cos\phi\sin\psi & \sin\psi\sin\theta \\ -\sin\psi\cos\phi - \cos\theta\sin\phi\cos\psi & -\sin\psi\sin\phi + \cos\theta\cos\phi\cos\psi & \cos\psi\sin\theta \\ \sin\theta\sin\phi & -\sin\theta\cos\phi & \cos\theta \end{pmatrix} \tag{3–64}$$

which, although complicated, plays an important role in the consideration of molecular rotation.

In conclusion of this section, and by way of offering a partial summary of the past two sections, the following lists show the interrelation between statements about transformations, matrices, determinants, and linear independence. The student should endeavor to realize the underlying unity of the subject, and to use it.

3–4 LINEAR OPERATORS

In this section we shall take a major step forward in realizing the application of the algebra of vector spaces to quantum mechanics.

Equivalence of statements: $n \times n$ matrices

Case 1. *Determinant is not zero.*

1. Linear independence of a set of vectors.
2. Row equivalence to unit matrix.
3. Matrix has an inverse.
4. Matrix has rank r equal to dimension n.
5. Linear transformation is a one-to-one mapping of n space onto n space.
6. Determinant of matrix is not zero.
7. Set of n simultaneous linear equations in n unknowns has a solution.

Case 2. *Determinant is zero.*

1. Linear dependence of a set of vectors.
2. Row equivalence to triangular matrix with at least one all-zero row.
3. Matrix has no inverse.
4. Matrix has rank r less than dimension n.
5. Linear transformation is a many-to-one mapping of n space onto r-dimensional subregion of n space.
6. Determinant of matrix is zero.
7. Set of n homogeneous simultaneous linear equations in n unknowns has a solution.

Equivalence of statements: $m \times n$ matrices with rank r

1. Linear independence of any r of the m vectors; dependence of $m - r$ of the vectors on the r independent vectors.
2. Row equivalence to triangular matrix with $m - r$ all-zero rows.
3. Linear transformation is mapping of n space onto r-dimensional subregion of m space.
4. Set of simultaneous linear equations has solution if rank of augmented matrix is also r.

Early in Chapter 1 we had mentioned the role of eigenvalue equations in quantum mechanics and seen the form which such equations take, such as Eq. 1–2. Although the description of specific operators will be left to a work on quantum mechanics per se, the general results can be set forth here. As has been our custom, we begin with a number of definitions.

Definition An *operator* is a set of instructions, defined for some vector space, for changing one vector belonging to the space into another vector belonging to the space. Thus we shall write $\mathcal{A}\xi = \eta$ to mean that the result of applying a particular set of instructions, embodied in the definition of the operator $\mathcal{A}$, to the vector ξ is to form a new vector η. Operators will be notated in script letters.

We might well inquire into the difference between this definition of an operator and the definition of a transformation in the previous section. In the final analysis, there is none. The word *operator* is more often used in the context of quantum mechanics to represent particular physical quantities, whereas the word *transformation* is used to represent changes in a coordinate system. However, their definitions are identical in form, and furthermore, the definition of a linear operator is analogous to that of a linear transformation.

Definition A *linear operator* $\mathcal{A}$ obeys these equations:
(a) $\mathcal{A}(c\xi) = c\mathcal{A}\xi$, where c is a constant (perhaps complex).
(b) $\mathcal{A}(\xi + \eta) = \mathcal{A}\xi + \mathcal{A}\eta$, where ξ and η are both vectors.

The definition of *operator* as a "set of instructions" leaves us with no way of representing operators. How may we represent linear operators? The result of operating an operator $\mathcal{A}$ on any vector can be found from the result of operating the operator $\mathcal{A}$ on the basis vectors. Suppose, for example, we know that

$$\mathcal{A}\phi^i = \sum_j A_{ij}\phi^j \tag{3-65}$$

where the set $\{\phi^j\}$ is a complete, orthonormal basis set for the vector space in question. Any vectors may be expanded in terms of the ϕ^j vectors:

$$\xi = \sum_i c_i\phi^i \qquad \eta = \sum_j d_j\phi^j \tag{3-66}$$

We may then use the operator definition, $\mathcal{A}\xi = \eta$, to relate the numbers $\{c_i\}$ and $\{d_i\}$:

$$\mathcal{A}\xi = \mathcal{A}\sum_i c_i\phi^i = \sum_{ij} c_i A_{ij}\phi^j = \eta = \sum_j d_j\phi^j \tag{3-67}$$

so that

$$d_j = \sum_i A_{ij}c_i \tag{3-68}$$

We have now only to discover what the numbers $\{A_{ij}\}$ are. This is easy if the basis set $\{\phi^i\}$ is orthonormal, for then

$$\langle\phi^j \mid \mathcal{A}\phi^i\rangle = \sum_k \langle\phi^j \mid A_{ik}\phi^k\rangle = \sum_k A_{ik}\langle\phi^j \mid \phi^k\rangle = A_{ij} \tag{3-69}$$

We often see the left side of Eq. 3-69 written with an extra vertical bar, $\langle\phi^j \mid \mathcal{A} \mid \phi^i\rangle$. This extra vertical bar adds nothing new, but only serves to call attention to the operator $\mathcal{A}$ in the center. An inner product like $\langle\phi^j \mid \mathcal{A} \mid \phi^i\rangle$ is often called a *matrix element*. It will be far more convenient to use these elements as we go further. For example, Eq. 3-68 can now be written

$$d_j = \sum_i \langle\phi^j|\mathcal{A}|\phi^i\rangle c_i \tag{3-70}$$

which has the form of a matrix product,

$$\begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix} = \begin{pmatrix} \langle\phi^1|\mathcal{A}|\phi^1\rangle & \cdots & \langle\phi^1|\mathcal{A}|\phi^n\rangle \\ \langle\phi^2|\mathcal{A}|\phi^1\rangle & \cdots & \langle\phi^2|\mathcal{A}|\phi^n\rangle \\ \vdots & & \vdots \\ \langle\phi^n|\mathcal{A}|\phi^1\rangle & \cdots & \langle\phi^n|\mathcal{A}|\phi^n\rangle \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \tag{3-71}$$

This being the case, we shall do away with the expansion coefficients A_{ij} (small capital letter), and replace them with the usual matrix elements $a_{ji} = \langle\phi^j \mid \mathcal{A} \mid \phi^i\rangle$:

$$d_j = \sum_i A_{ij}c_i = \sum_i a_{ji}c_i \tag{3-72}$$

Here we have made use of a column matrix (an $n \times 1$ matrix) to represent a vector. An inner product may be formed from two vectors in matrix notation if the left vector is written as a $1 \times n$ row matrix and the right vector as a $n \times 1$ column matrix. Their product is, of course, a 1×1 matrix, or a scalar:

$$\langle\xi \mid \eta\rangle = [c_1 \; c_2^* \cdots c_n^*] \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix} = [\;] \tag{3-73}$$

In our case, $\eta = \mathfrak{A}\xi$. Hence,

$$\langle \xi \mid \eta \rangle = \langle \xi \mid \mathfrak{A} \mid \xi \rangle = [c_1^* \cdots c_n^*] \begin{pmatrix} a_{11} \cdots a_{1n} \\ \vdots \quad \vdots \\ a_{n1} \cdots a_{nn} \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \tag{3-74}$$

Two features of importance have appeared so far. The use of a *square matrix to represent a linear operator* and the use of a *column matrix to represent a vector*. The particular matrices used were predicated on the choice of the basis set $\{\phi^i\}$, since the matrix elements were defined by $a_{ji} = \langle \phi^j \mid \mathfrak{A} \mid \phi^i \rangle$. Of course, *any* basis set would work. There are, therefore, any number of matrices which represent $\mathfrak{A}$, each matrix representative arising from a different basis. Because of this, it would, in a strict sense, be incorrect to say that the matrix A whose elements are a_{ij} is the *same* as the operator $\mathfrak{A}$; accordingly, we say that the matrix *represents* $\mathfrak{A}$, or *is a representative of* the operator $\mathfrak{A}$ in the basis $\{\phi^i\}$. Similarly, we would not be strictly correct to say that the column vector

$$\begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix}$$

is the *same* as η, but that this column vector *is a representative of* or *represents* η.

We should then investigate the relations between two matrix representatives of a given operator. Let A^ϕ be the representative of $\mathfrak{A}$ in the $\{\phi^i\}$ basis, and A^ψ be the representative of $\mathfrak{A}$ in the $\{\psi^i\}$ basis. The matrix elements are $a^\phi_{ij} = \langle \phi^i \mid \mathfrak{A} \mid \phi^j \rangle$ and $a^\psi_{ij} = \langle \psi^i \mid \mathfrak{A} \mid \psi^j \rangle$. Finally, let the two bases (which are orthonormal) be related by a unitary transformation,

$$\psi^i = \sum_k u_{ik} \phi^k,$$

and, in reverse,

$$\phi^i = \sum_j u^*_{ji} \psi^j = \sum_j u'^*_{ij} \psi^j.$$

The relation between the matrix elements a^ϕ_{ij} and a^ψ_{ij} is then

$$\begin{aligned} a^\psi_{ij} &= \langle \psi^i \mid \mathfrak{A} \mid \psi^j \rangle = \sum_k u^*_{ik} \langle \phi^k \mid \mathfrak{A} \mid \psi^j \rangle \\ &= \sum_{kl} u^*_{ik} u_{jl} \langle \phi^k \mid \mathfrak{A} \mid \phi^l \rangle = \sum_{kl} u^*_{ik} u_{jl} a^\phi_{kl} \\ &= \sum_{kl} u'^{-1}_{ik} a^\phi_{kl} u'_{lj} = [U'^{-1} A^\phi U']_{ij} \end{aligned} \tag{3-75}$$

The structure $A^\psi = U'^{-1} A^\phi U'$ occurs often in algebraic equations. We say that A is obtained from B by means of a *similarity transformation* if $A = S^{-1} B S$. If S is unitary, as in the present case ($S = U'$), the transformation is called a *unitary transformation;* if S is orthogonal, an *orthogonal transformation*. From this we get the following result.

Theorem The matrix representative of $\mathfrak{A}$ in the ψ basis may be found by a similarity transformation of the representative of $\mathfrak{A}$ in the ϕ basis; this similarity transformation is the transpose of the transformation connecting the bases.

We have established the relations between linear operators and their matrix representatives; now it is only natural to extend to operators the features of matrix algebra, the matrix product, the inverse, and the transpose. A few further features are of importance, and are contained in the following definitions.

Definitions The *commutator* of two operators $\mathfrak{A}$ and $\mathfrak{B}$ is the operator $\mathfrak{A}\mathfrak{B} - \mathfrak{B}\mathfrak{A}$, and is designated $[\mathfrak{A}, \mathfrak{B}]$.

The *adjoint* of an operator $\mathfrak{A}$, denoted $\mathfrak{A}^\dagger$, is that operator whose matrix elements are related to those of $\mathfrak{A}$ by

$$\langle \phi^i \mid \mathfrak{A}^\dagger \mid \phi^j \rangle = \langle \mathfrak{A} \phi^i \mid \phi^j \rangle.$$

The operator adjoint to $\mathfrak{A}$ is represented by a matrix which is related by a transposition and a complex conjugation to the representative of $\mathfrak{A}$ itself, since

$$\langle \phi^i \mid \mathfrak{A}^\dagger \mid \phi^j \rangle = \langle \mathfrak{A} \phi^i \mid \phi^j \rangle = \langle \phi^j \mid \mathfrak{A} \mid \phi^i \rangle^* \tag{3-76}$$

The dagger symbol is used to signify an adjoint.

Definition The *trace* of an operator is the sum of the diagonal matrix elements of any representative of that operator: $\text{tr } \mathcal{A} = \sum_i a_{ii}$. (In German references, *trace* is called *Spur* and is abbreviated Sp.)

As an example of some of the principles that we have been discussing, let us consider a typical operator and its representatives in some bases.

Definition A *projection operator* $\mathcal{P}_\epsilon$, which gives the projection of some vector in the direction of a unit vector ϵ, is defined by $\mathcal{P}_\epsilon \xi = \langle \epsilon \mid \xi \rangle \epsilon$.

The projection operator gives the component of a vector in a particular direction. We may very simply solve for the eigenvalue of $\mathcal{P}_\epsilon$ using only operator methods:

$$\mathcal{P}_\epsilon{}^2 \xi = \mathcal{P}_\epsilon(\mathcal{P}_\epsilon \xi) = \mathcal{P}_\epsilon(\langle \epsilon \mid \xi \rangle \epsilon) = \langle \epsilon \mid \xi \rangle \mathcal{P}_\epsilon \epsilon = \langle \epsilon \mid \xi \rangle \epsilon \qquad (3\text{-}77)$$

Hence, $\mathcal{P}_\epsilon{}^2 = \mathcal{P}_\epsilon$. This is expressed by saying that the operator $\mathcal{P}_\epsilon$ is *idempotent*. Then, $(\mathcal{P}_\epsilon{}^2 - \mathcal{P}_\epsilon) = 0 = \mathcal{P}_\epsilon(\mathcal{P}_\epsilon - 1) = 0$, and either $\mathcal{P}_\epsilon = 0$ or $\mathcal{P}_\epsilon = 1$. We already know the possible eigenvalues for $\mathcal{P}_\epsilon$ —one or zero—without even setting down the matrix representative of $\mathcal{P}_\epsilon$. Now let us see what a matrix representative of $\mathcal{P}_\epsilon$ looks like.

EXAMPLE

The projection operator in a two-dimensional Euclidean vector space, with the projection direction vector $\epsilon = (1/\sqrt{2}, 1/\sqrt{2})$. Choose as the basis the conventional Cartesian basis $\phi^1 = (1, 0)$, $\phi^2 = (0, 1)$. Then, by direct evaluation,

$$p^{\phi}_{11} = \langle \phi^1 \mid \mathcal{P} \mid \phi^1 \rangle = \left\langle (1,0) \middle| \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \cdot \left(\frac{1}{\sqrt{2}} \right) \right\rangle = \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} = \frac{1}{2}$$

$$p^{\phi}_{12} = \tfrac{1}{2}$$

$$p^{\phi}_{21} = \tfrac{1}{2}$$

$$p^{\phi}_{22} = \tfrac{1}{2}$$

The matrix P^{ϕ} is

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Consider now the basis $\psi^1 = (1/2, \sqrt{3}/2)$, $\psi^2 = (\sqrt{3}/2, -1/2)$. A calculation similar to the above gives the elements p^{ψ}_{ij}, and the matrix

$$P^{\psi} = \begin{pmatrix} \frac{1}{2} + \frac{\sqrt{3}}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} - \frac{\sqrt{3}}{4} \end{pmatrix}$$

We must now confirm the transformation theorem, $P^{\psi} = U'^{-1} P^{\phi} U'$. Since

$$\psi^1 = \frac{1}{2}(1, 0) + \frac{\sqrt{3}}{2}(0, 1) = \frac{1}{2}\phi^1 + \frac{\sqrt{3}}{2}\phi^2$$

$$\psi^2 = \frac{\sqrt{3}}{2}(1, 0) - \frac{1}{2}(0, 1) = \frac{\sqrt{3}}{2}\phi^1 - \frac{1}{2}\phi^2$$

we have

$$U = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{-1}{2} \end{pmatrix} \qquad U' = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{-1}{2} \end{pmatrix} \qquad U'^{-1} = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{-1}{2} \end{pmatrix}$$

Then, by direct multiplication

$$\begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{-1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{-1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} + \frac{\sqrt{3}}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} - \frac{\sqrt{3}}{4} \end{pmatrix} = P^{\psi}$$

We will soon find that in the basis $x^1 = (1/\sqrt{2}, 1/\sqrt{2})$, $x^2 = (1/\sqrt{2}, -1/\sqrt{2})$ $\mathcal{P}$ takes the form

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

With these three examples we may (somewhat facetiously) illustrate the concept of matrix representation with Fig. 3–3.

We turn finally to the problem of finding the eigenvalues and eigenvectors of linear operators. Not all linear operators obey an eigenvalue equation, but, from the point of view of quantum me-

chanics, two classes of operators which do obey an eigenvalue equation are of great importance. These are *Hermitian operators* and *unitary operators*. We shall summarize the results first, and then offer the proofs.

Definition An eigenvalue is said to be *degenerate* if the eigenvalue satisfies an eigenvalue-eigenvector equation for more than one eigenvector; the number of eigenvectors with the same eigenvalue is called the *degree of degeneracy* of the eigenvalue.

Definition A *Hermitian operator* is a linear operator which is self-adjoint; that is, $\mathcal{H} = \mathcal{H}^\dagger$, or $h_{ij} = h_{ji}^*$. The diagonal elements of a Hermitian matrix are real.

Theorem A Hermitian operator in an n-dimensional vector space has n *distinct eigenvectors* and n *real eigenvalues*. If the eigenvalues are nondegenerate, the eigenvectors are mutually orthogonal, and, with suitable normalization, form an orthonormal set. Even if some of the eigenvalues are degenerate, an orthonormal set can be constructed from the distinct eigenvectors.

Definition A *unitary operator* is a linear operator whose adjoint is its inverse: $\mathcal{U}^{-1} = \mathcal{U}^\dagger$.

Theorem A unitary operator in an n-dimensional vector space has n distinct eigenvectors (which may, as above, be made to constitute an orthonormal set), and n eigenvalues, all of which have unit modulus.

We begin by proving the Hermitian operator theorem. If $\mathcal{H}\phi^i = h_i\phi^i$, then $\langle\phi^i|\mathcal{H}|\phi^i\rangle = h_i\langle\phi^i|\phi^i\rangle$. Since $\mathcal{H}$ is Hermitian, $\langle\phi^i|\mathcal{H}|\phi^i\rangle = \langle\phi^i|\mathcal{H}|\phi^i\rangle^*$; and $h_i = h_i^*$, so h_i is real. For two different eigenvectors,

$$\mathcal{H}\phi^i = h_i\phi^i \tag{3-78a}$$

$$\mathcal{H}\phi^j = h_j\phi^j \tag{3-78b}$$

From Eq. 3-78a we get $\langle\phi^j|\mathcal{H}|\phi^i\rangle = h_i\langle\phi^j|\phi^i\rangle$; from Eq. 3-78b, $\langle\mathcal{H}\phi^j|\phi^i\rangle = h_j^*\langle\phi^j|\phi^i\rangle$. However, $\langle\mathcal{H}\phi^j|\phi^i\rangle = \langle\phi^i|\mathcal{H}|\phi^j\rangle^* = \langle\phi^j|\mathcal{H}|\phi^i\rangle = h_i\langle\phi^j|\phi^i\rangle$. Then, $(h_i - h_j^*)\langle\phi^j|\phi^i\rangle = 0$. Since h_j is real, if $h_i \neq h_j$, $\langle\phi^j|\phi^i\rangle = 0$ and the eigenvectors are orthogonal.

The proof of the theorem for unitary operators is very similar. If $\mathcal{U}\psi^i = u_i\psi^i$, then $\mathcal{U}^\dagger\mathcal{U}\psi^i = \psi^i = \mathcal{U}^\dagger u_i\psi^i = u_i\mathcal{U}^\dagger\psi^i$. Then $\mathcal{U}^\dagger\psi^i = (1/u_i)\psi^i$, or $\mathcal{U}^\dagger$ has the same eigenvectors as $\mathcal{U}$, but with reciprocal eigenvalues. Then $\langle\psi^i|\mathcal{U}^\dagger|\psi^i\rangle = u_i^*\langle\psi^i|\psi^i\rangle$ and $\langle\mathcal{U}^\dagger\psi^i|\psi^i\rangle = (1/u_i^*)\langle\psi^i|\psi^i\rangle$. However, $\langle\mathcal{U}^\dagger\psi^i|\psi^i\rangle$ also equals $\langle\psi^i|\mathcal{U}|\psi^i\rangle$. Hence, $1/u_i^* = u_i$, and $u_iu_i^* = 1$, or the eigenvalues u_i have modulus one. The orthogonality of the eigenvectors may be proved in a way analogous to the proof for Hermitian operators. The particular problems which arise for degenerate eigenvalues are discussed at the end of this section.

Neither of these proofs shows why Hermitian or unitary operators

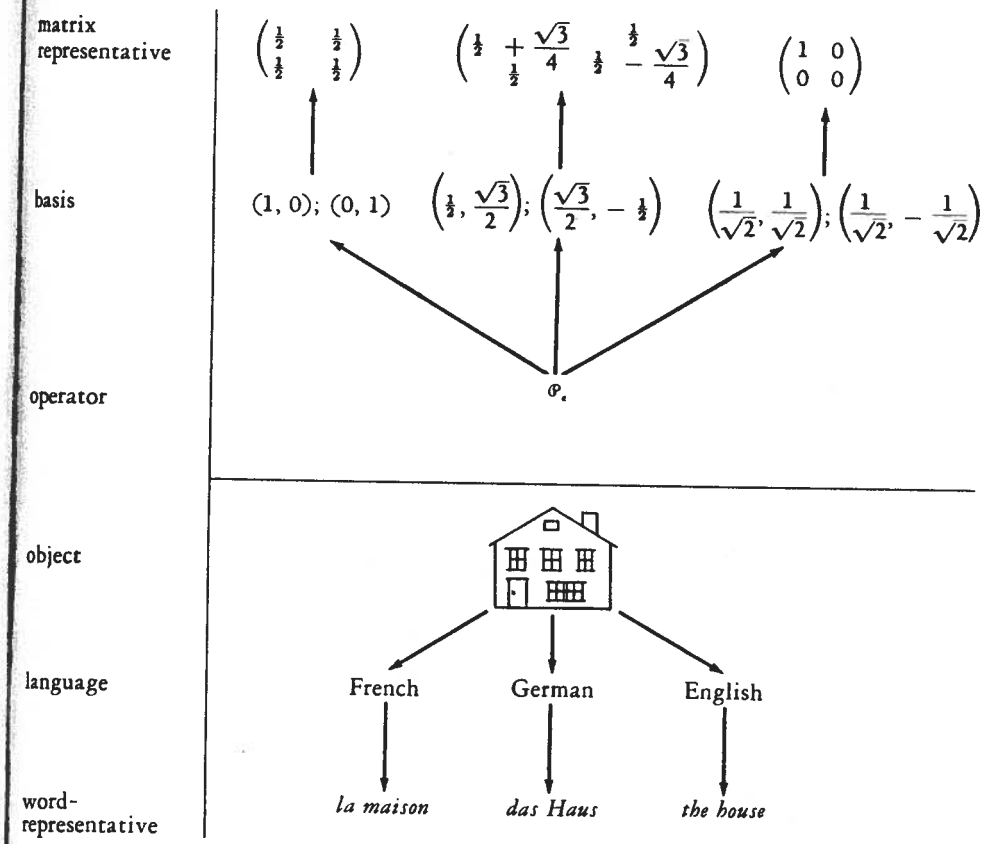


Figure 3-3 The operator $\mathcal{P}_\epsilon$, where $\epsilon = (1/\sqrt{2}, 1/\sqrt{2})$ and its matrix representatives in three bases; by analogy, the object *house*, and its word representatives in three languages.

must obey an eigenvalue equation, and this is offered without proof at the present. However, each of these theorems has a corollary for Euclidean vector spaces.

Corollary Corresponding to the theorem for Hermitian operators in a Hermitian vector space there is the analogous theorem for *symmetric operators* ($S_{ij} = S_{ji}$) in a Euclidean vector space; corresponding to the theorem for unitary operators in a Hermitian vector space there is the analogous theorem for *orthogonal operators* ($\mathcal{R}' = \mathcal{R}^{-1}$) in a Euclidean vector space.

All these theorems and corollaries have dealt with the *existence* of eigenvalues and eigenvectors, but nothing so far has shown us how to *find* these eigenvalues. We can undertake a study of the solution of eigenvalue equations in general terms from two points of view, the *secular equation* and the *similarity transformation*.

Consider the eigenvalue equation, Eq. 1–2, in matrix form:

$$\mathcal{Q}\phi = q\phi \tag{3-79}$$

This equation is a one-line statement of a number of linear equations in the components of the vector ϕ:

$$\begin{aligned} \mathcal{Q}_{11}\phi_1 + \mathcal{Q}_{12}\phi_2 + \cdots + \mathcal{Q}_{1n}\phi_n &= q\phi_1 \\ \vdots \qquad\quad \vdots \qquad\quad & \vdots \\ \mathcal{Q}_{n1}\phi_1 + \mathcal{Q}_{n2}\phi_2 + \cdots + \mathcal{Q}_{nn}\phi_n &= q\phi_n \end{aligned} \tag{3-80}$$

These, in turn, may be rearranged to give a set of simultaneous linear homogeneous equations in the n unknowns ϕ_i (the components of ϕ).

$$\begin{aligned} (\mathcal{Q}_{11} - q)\phi_1 + \quad \mathcal{Q}_{12}\phi_2 \quad + \cdots + \quad \mathcal{Q}_{1n}\phi_n \quad &= 0 \\ \mathcal{Q}_{21}\phi_1 \quad + (\mathcal{Q}_{22} - q)\phi_2 + \cdots + \quad \mathcal{Q}_{2n}\phi_n \quad &= 0 \\ \vdots \qquad\quad \vdots \qquad\quad & \vdots \\ \mathcal{Q}_{n1}\phi_1 \quad + \quad \mathcal{Q}_{n2}\phi_2 \quad + \cdots + (\mathcal{Q}_{nn} - q)\phi_n &= 0 \end{aligned} \tag{3-81}$$

We have already learned that such a set of equations has a nontrivial solution only if the determinant of the coefficients vanishes. Hence, we obtain the nth degree equation for q, called the *secular equation:*

$$\begin{vmatrix} (\mathcal{Q}_{11} - q) & \mathcal{Q}_{12} & \cdots & \mathcal{Q}_{1n} \\ \mathcal{Q}_{21} & (\mathcal{Q}_{22} - q) & \cdots & \mathcal{Q}_{2n} \\ \vdots & \vdots & & \vdots \\ \mathcal{Q}_{n1} & \mathcal{Q}_{n2} & \cdots & (\mathcal{Q}_{nn} - q) \end{vmatrix} = 0 \tag{3-82}$$

Secular equations are commonplace in quantum mechanics. Think a minute about what Eq. 3-82 says. If we expanded Eq. 3-82 we would get a polynomial of degree n in the unknown q. Setting aside (for the moment) the practical question of solving that equation,

notice that one significant feature appears: there are n roots to the equation. The theorems we proved above stated that there would be n eigenvectors and n eigenvalues, and now we see why. It is because the eigenvalues are the solutions of an nth degree equation. Of course, as we saw, some of these roots (the eigenvalues) may be equal, but we shall defer this problem.

Very well, we now have n eigenvalues. However, the set of equations 3–81 can then be solved for the components ϕ_i for the eigenvector, *for each of the n eigenvalues*. Hence, Eq. 3-81 gives, ultimately, n eigenvectors, one eigenvector for each eigenvalue. To conclude the calculation, we shall require that the eigenvectors be normalized.

By way of recapitulation, we solve a matrix eigenvalue equation by the following steps.

1. Set up and solve the secular equation; obtain eigenvalues.
2. Use eigenvalues (now known) in eigenvalue equation to get eigenvectors.
3. Normalize eigenvectors.

We may look at the eigenvalue problem from a second point of view as well. Imagine that we know the eigenvalues and eigenvectors of a matrix:

$$\begin{aligned} \mathcal{Q}\phi^1 &= q_1\phi^1 \\ \mathcal{Q}\phi^2 &= q_2\phi^2 \\ \vdots \qquad & \vdots \\ \mathcal{Q}\phi^n &= q_n\phi^n \end{aligned} \tag{3-83}$$

We could line up all the column vectors side by side and make an $n \times n$ matrix of them, such as

$$\Phi = \begin{pmatrix} \phi_1{}^1 & \phi_1{}^2 \cdots \phi_1{}^n \\ \phi_2{}^1 & \phi_2{}^2 \cdots \phi_2{}^n \\ \vdots & \vdots \qquad \vdots \\ \phi_n{}^1 & \phi_n{}^2 \cdots \phi_n{}^n \end{pmatrix} \tag{3-84}$$

The eigenvectors ϕ^i form the columns of the matrix Φ. The effect of operating $\mathcal{Q}$ on Φ is simply to generate a matrix whose columns are $q_i\phi^i$:

$$\mathcal{Q}\Phi = \begin{pmatrix} q_1\phi_1{}^1 \cdots q_n\phi_1{}^n \\ q_1\phi_2{}^1 \cdots q_n\phi_2{}^n \\ \vdots \qquad\quad \vdots \\ q_1\phi_n{}^1 \cdots q_n\phi_n{}^n \end{pmatrix} = \begin{pmatrix} \phi_1{}^1 \cdots \phi_1{}^n \\ \phi_2{}^1 \cdots \phi_2{}^n \\ \vdots \qquad\quad \vdots \\ \phi_n{}^1 \quad\; \phi_n{}^n \end{pmatrix} \begin{pmatrix} q_1 & & \bigcirc \\ & q_2 & \\ \bigcirc & & \ddots \\ & & & q_n \end{pmatrix} \tag{3-85}$$

$$= \Phi\hat{q}$$

where we have used the symbol $\hat{q}$ to represent the matrix which is composed of the n eigenvalues q_i on the diagonal, and zeros everywhere else. Multiplying each side of Eq. 3-85 by Φ^{-1} gives

$$\Phi^{-1}Q\Phi = \hat{q} \tag{3-86}$$

This is a statement that *the similarity transformation of Q by the matrix composed of its eigenvectors taken columnwise gives a diagonal matrix whose entries are the eigenvalues of Q.*

Hence, if we can find a way of transforming Q to a diagonal matrix, called *diagonalizing the matrix Q*, the columns of that similarity transformation are the eigenvectors and the entries in the diagonal matrix are the eigenvalues. This procedure is readily transcribed for digital computation, and often forms the basis for computer solution of eigenvalue problems. Since, as we have proven, the vectors ϕ^i are orthonormal, the transformation $\Phi^{-1}Q\Phi$ is a unitary transformation. As an example of these ideas, we find the eigenvalues of the P_ϵ matrix discussed previously.

EXAMPLE

Consider the operator $\mathcal{P}_\epsilon$, where $\epsilon = (1/\sqrt{2}, 1/\sqrt{2})$ in the usual Cartesian basis. The matrix representative of this projection operator in this basis is

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

as found before. To find the eigenvalues of this operator, we solve the secular equation.

$$\begin{vmatrix} \frac{1}{2} - p & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} - p \end{vmatrix} = 0$$

which gives the quadratic

$$\tfrac{1}{4} - p + p^2 - \tfrac{1}{4} = 0$$

which has the roots

$$p^2 - p = 0$$
$$p_1 = 1 \qquad p_2 = 0$$

These roots, which are the eigenvalues of $\mathcal{P}_\epsilon$, are just what we had predicted before from operator properties alone. We now find the eigenvector ϕ^1 corresponding to the eigenvalue $p_1 = 1$:

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}\begin{pmatrix} \phi_1{}^1 \\ \phi_2{}^1 \end{pmatrix} = \begin{pmatrix} \phi_1{}^1 \\ \phi_2{}^1 \end{pmatrix}$$

which becomes

$$\tfrac{1}{2}\phi_1{}^1 + \tfrac{1}{2}\phi_2{}^1 = \phi_1{}^1$$
$$\tfrac{1}{2}\phi_1{}^1 + \tfrac{1}{2}\phi_2{}^1 = \phi_2{}^1$$

or

$$\phi_2{}^1 = \phi_1{}^1$$
$$\phi_1{}^1 = \phi_2{}^1$$

Note that the eigenvector-component equations are identical. To fully determine the components, we require normalization, which gives $\phi_1{}^1 = \phi_2{}^1 = 1/\sqrt{2}$, or $\phi^1 = (1/\sqrt{2}, 1/\sqrt{2})$. The eigenvector ϕ^2 corresponding to the eigenvalue $p_2 = 0$ is found by an identical process to be $\phi^2 = (1/\sqrt{2}, -1/\sqrt{2})$.

Stop and think about these eigenvectors: the eigenvector with eigenvalue one is *exactly the same* as the projection direction ϵ; that is, a vector in the projection direction projects to a vector of the same length. On the other hand, the eigenvector ϕ^2 with eigenvalue zero is perpendicular to the projection direction ϵ; a vector perpendicular to the projection direction projects to a point (zero length).

We may also confirm the similarity transformation $\Phi^{-1}P_\epsilon\Phi$ which diagonalizes P_ϵ by direct substitution:

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

There is another technique for finding the eigenvectors which is somewhat simpler. The components of an eigenvector obey Eq. 3-81. Each line of Eq. 3-81 looks like the expansion of a determinant by cofactors. For, if $\phi_1 = [Q - \hat{q}]_{11}$, $\phi_2 = [Q - \hat{q}]_{12}$, and so on, then the first line of Eq. 3-81 gives

$$\begin{aligned}(Q_{11} - q)\phi_1 + Q_{12}\phi_2 + \cdots + Q_{1n}\phi_n \\ = (Q_{11} - q)[Q - \hat{q}]_{11} + Q_{12}[Q - \hat{q}]_{12} + \cdots + Q_{1n}[Q - \hat{q}]_{1n} \\ = |Q - \hat{q}| = 0\end{aligned} \tag{3-87}$$

where the numbers Q_{ij} are the matrix elements of Q, and the numbers

q are the eigenvalues of Q. Hence the cofactor of the secular determinant for *any* row, ith column, gives a number proportional to ϕ_i.

EXAMPLE

In the case of the operator $\mathcal{P}_1$ given above, the secular determinant for the eigenvalue $p_1 = 1$ is

$$\begin{vmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = 0$$

and the eigenvector ϕ^1 has components proportional to $(-\frac{1}{2}, -\frac{1}{2})$, if we choose the cofactors of the first row; ϕ^1 has components proportional to $(-\frac{1}{2}, -\frac{1}{2})$ from cofactors of the second row; each of these gives $\phi^1 = (1/\sqrt{2}, 1/\sqrt{2})$ with suitable normalization.

The student might well inquire what conditions must prevail in order that two operators have the *same* set of eigenvectors. This question is of great importance in quantum mechanics, because it tells us when two operators can both assume definite eigenvalues in the same single, stationary quantum state. Our question could equally well be phrased by asking when two matrices can be diagonalized by the same similarity transformation. The result is beautifully simple.

Theorem **Two Hermitian operators can have the same set of eigenvectors (eigenfunctions) if, and only if, they commute.**

Let us prove first that if $\mathfrak{A}$ and $\mathfrak{B}$ commute, then they have the same eigenvectors. Suppose $\mathfrak{A}$ has eigenvectors ψ^i; then $\mathfrak{A}\psi^i = a_i\psi^i$. However, because $\mathfrak{A}$ and $\mathfrak{B}$ commute, $\mathfrak{A}\mathfrak{B}\psi^i = \mathfrak{B}\mathfrak{A}\psi^i$. Then,

$$\mathfrak{B}\mathfrak{A}\psi^i = \mathfrak{A}\mathfrak{B}\psi^i = \mathfrak{B}a_i\psi^i = a_i\mathfrak{B}\psi^i \tag{3-88}$$

Equation 3–88 shows that the vector $(\mathfrak{B}\psi^i)$ is also an eigenvector of $\mathfrak{A}$ with the eigenvalue a_i. This can only be true if $(\mathfrak{B}\psi^i)$ is a multiple of ψ^i; hence,

$$\mathfrak{B}\psi^i = b_i\psi^i \tag{3-89}$$

and ψ^i is also an eigenvector of $\mathfrak{B}$. The student may detect that this argument fails if the eigenvalue a_i is degenerate; such cases are considered at the end of this section.

To prove the converse—that if $\mathfrak{A}$ and $\mathfrak{B}$ have the same set of eigenvectors, then they commute—we begin by showing how an operator's effect may be restated in terms of projection operators. A vector ξ may be expanded in a complete orthonormal set ϕ^i:

$$\xi = \sum_i \langle\phi^i \mid \xi\rangle\phi^i = \sum_i \mathcal{P}_{\phi^i}\xi \tag{3-90}$$

That is, the effect of *expanding* a vector in a basis is equivalent to adding up the *projections* of that vector along the basis vectors. Suppose the eigenvectors of $\mathfrak{A}$ are the set $\{\phi^i\}$. Then, because of the foregoing relation, we may write

$$\mathfrak{A}\xi = \mathfrak{A}\sum_i \langle\phi^i \mid \xi\rangle\phi^i = \sum_i \langle\phi^i \mid \xi\rangle a_i\phi^i = \sum_i a_i\mathcal{P}_{\phi^i} \tag{3-91}$$

which is an equation stating that the effect of operating $\mathfrak{A}$ on a given vector is to generate the vector sum of the given vector projected along an eigenvector times the eigenvalue. If we represent $\mathfrak{A}$ as

$$\mathfrak{A} = \sum_i a_i\mathcal{P}_{\phi^i} \tag{3-92}$$

and $\mathfrak{B}$, having the same set of eigenvectors $\{\phi^i\}$, as

$$\mathfrak{B} = \sum_j b_j\mathcal{P}_{\phi^j} \tag{3-93}$$

then it is not difficult to show that $\mathfrak{A}$ and $\mathfrak{B}$ must commute, since

$$\left[\sum_i a_i\mathcal{P}_{\phi^i}, \sum_j b_j\mathcal{P}_{\phi^j}\right] = 0.$$

(The student should work out this commutator to his satisfaction.)

The final topic of this section is the discussion of how the properties of the eigenvalues and eigenvectors of linear operators must be modified in the case of degenerate eigenvalues. We consider first how to form an orthonormal set of eigenvectors from the eigenvectors of Hermitian (or unitary, or symmetric, or orthogonal) operators with a degenerate eigenvalue.

Suppose it happens that two eigenvectors, ξ_1 and ξ_2 have the same eigenvalue q. These vectors need not be orthogonal. Our theorems provide only that the eigenvectors corresponding to different eigen-

values are orthogonal. If $\mathfrak{Q}\xi_1 = q\xi_1$ and $\mathfrak{Q}\xi_2 = q\xi_2$, then any linear combination of ξ_1 and ξ_2 also obeys the eigenvalue equation $\mathfrak{Q}(\xi_1 + c\xi_2) = q(\xi_1 + c\xi_2)$, where c is a constant. We want the vector $\xi_1 + c\xi_2$ (*which is an eigenvector*) to be normalized and orthogonal to ξ_1. Then we will have two vectors, ξ_1 and $\xi_1 + c\xi_2$, that are orthonormal. This is precisely the problem which confronted us in the Schmidt orthogonalization procedure. Suppose ξ_1 is already normalized. Then, requiring

$$\langle \xi_1 \mid \xi_1 + c\xi_2 \rangle = 0 \tag{3-94}$$

gives

$$c = \frac{-\langle \xi_1 \mid \xi_1 \rangle}{\langle \xi_1 \mid \xi_2 \rangle} = -\frac{1}{\langle \xi_1 \mid \xi_2 \rangle} \tag{3-95}$$

This equation is analogous to Eq. 3-9; the procedure can be repeated indefinitely as indicated by the degree of degeneracy of the eigenvalue q.

Degeneracy also affects our discussion of the simultaneous eigenvectors of commuting operators. We had pointed out that if $\mathfrak{A}$ and $\mathfrak{B}$ are Hermitian and commute, and that if $\mathfrak{A}$ has the eigenvectors $\{\psi^i\}$ and eigenvalues $\{a_i\}$, then

$$\mathfrak{A}(\mathfrak{B}\psi^i) = a_i(\mathfrak{B}\psi^i). \tag{3-88}$$

If a_i is a nondegenerate eigenvalue, then it must be the case that $\mathfrak{B}\psi^i$ is a multiple of ψ^i, as we had noted. If a_i is degenerate, however, $\mathfrak{B}\psi^i$ is, in general, a linear combination of all the eigenvectors belonging to the eigenvalue a_i. Suppose these are labeled by a second index, k, running from 1 to n (we say a_i is n-fold degenerate). Then

$$\mathfrak{B}\psi^i = \sum_k b_{ik}\psi^{ik} \tag{3-96}$$

The coefficients b_{ik} form an n-dimensional submatrix of the matrix representative of $\mathfrak{B}$ that is not diagonal. To put it another way, if the matrix A is a diagonal, the matrix B is diagonal wherever A is nondegenerate. Where A is degenerate and B is not diagonal, we may find the eigenvalues of $\mathfrak{B}$ by diagonalizing the smaller submatrix.

This section has introduced a number of ideas which are central to a concise formulation of quantum mechanics. By way of conclusion, a summary of these concepts is presented.

1. Linear operators may be represented by matrices, once a basis set is specified. The matrix representative of the operator $\mathfrak{A}$ is that matrix whose elements are $a_{ij} = \langle \phi^i \mid \mathfrak{A}\phi^j \rangle$, where the basis $\{\phi^i\}$ is specified.
2. The effect of operating $\mathfrak{A}$ on a vector ξ may be simulated by multiplying the matrix representative of $\mathfrak{A}$ by a column matrix representative of ξ (both in the same basis).
3. Change of basis is effected by a similarity transformation.
4. Hermitian operators have real eigenvalues and orthonormal eigenvectors.
5. Unitary operators have eigenvalues of modulus one and orthonormal eigenvectors.
6. Eigenvalue equations may be solved by (a) setting up and solving the secular equation, substituting back into the eigenvalue equation, and normalizing; (b) solving the secular equation, using cofactors of the secular determinant to find the eigenvectors, and normalizing; or (c) finding the similarity transformation which diagonalizes the matrix.
7. Two Hermitian operators have the same set of eigenvectors if and only if they commute.

Problems

1. If the following vectors are linearly independent, form an orthonormal set from them; if they are linearly dependent, exhibit this linear dependence.
(a) (1, 1, 2), (0, −1, 0), (−1, 0, 1).
(b) (1, 2, −1, 0), (0, 3, 4, 1), (1, 1, 1, 1), (2, 0, −4, 0).
(c) (i, 1, 2), ($2i + 1$, −1, $3i$), (4, $5i$, $6 + i$).
(d) (0, 0, 2, 0, 0), (1, 1, 1, 1, 1), (1, 3, 1, 2, 2), (0, 0, 1, 2, 2), (1, 0, 1, 0, 1).
2. Are a set of mutually orthogonal vectors linearly independent? Why or why not?
3. Show that, in a vector space, $c\alpha = 0$ implies either $c = 0$ or $\alpha = 0$.
4. Form a multiplication table and compute the inner products of all vectors in Problem 1, parts (a), (b), and (c).
5. Show that matrix multiplication is associative.

6. Let $A = \begin{pmatrix} 1 & 2 & -1 \\ 3 & 0 & 2 \\ 4 & 5 & 0 \end{pmatrix}$ $B = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 1 & 3 \end{pmatrix}$

Find AB and BA. Do A and B commute? Find A^{-1} and B^{-1}. Verify that $(AB)' = B'A'$, $(AB)^{-1} = B^{-1}A^{-1}$.

7. Evaluate the determinant

$$\begin{vmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & 1 \end{vmatrix}$$

by expansion in cofactors; by direct evaluation.

8. A determinant has zeros below the main diagonal and nonzero elements on and above the main diagonal. Show that the value of this determinant is simply the product of the diagonal elements.

9. Solve the following sets of simultaneous linear equations where possible. If a solution is impossible, comment.

(a)
$$\begin{aligned} 2x - 3y + 5z &= 0 \\ x - y - 2z &= 2 \\ 5x - z &= -1 \end{aligned}$$

(b)
$$\begin{aligned} 2x - y + 3z - w &= 0 \\ 4x - 2y - z + 3w &= 0 \\ 2x - y - 4z + 4w &= 0 \\ 10x - 5y - 6z + 10w &= 0 \end{aligned}$$

(c)
$$\begin{aligned} 2x - y + 3z &= 1 \\ 4x - 2y - z &= -3 \\ 2x - y - 4z &= -4 \\ 10x - 5y - 6z &= -10 \end{aligned}$$

(d)
$$\begin{aligned} 4x + 2y + z &= 11 \\ x - y - z &= -4 \\ x + y + z &= 6 \end{aligned}$$

10. What orthogonal transformation carries the Cartesian basis $(\hat{x}, \hat{y}, \hat{z})$ into the spherical basis $(\hat{r}, \hat{\theta}, \hat{\phi})$?

11. A linear transformation maps the xy plane onto the uv plane according to

$$L = \begin{pmatrix} 2 & -1 \\ -3 & 0 \end{pmatrix}$$

Find the image of the points (1, 2), (−2, 1), (1, 0), (0, 1) under L.

12. Compute the rank of each of these transformations and comment.

(a)
$$\begin{aligned} u &= x + 2y - 3z \\ v &= 2x - y + 4z \\ w &= 3x + y + z \end{aligned}$$

(b)
$$\begin{aligned} u &= y - z \\ v &= 3x - y + 3z \\ w &= x + z \end{aligned}$$

13. In a two-dimensional Hermitian space, two bases are related by

$$\psi^1 = \frac{1}{\sqrt{2}}(\phi^1 + i\phi^2) \qquad \psi^2 = \frac{1}{\sqrt{2}}(\phi^1 - i\phi^2)$$

(a) Determine the unitary matrix U which transforms the ϕ vectors into the ψ vectors.

(b) If in the ϕ representation operator $\mathfrak{A}$ has the matrix representative

$$A^\phi = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

what will be its representative A^ψ in the ψ representation?

(c) Show that a unitary or orthogonal transformation always has an inverse.

14. Show that the matrix

$$A = \begin{pmatrix} a & b \\ b & b \end{pmatrix}$$

a symmetric, real matrix is transformed into the diagonal matrix

$$B = \begin{pmatrix} c & 0 \\ 0 & d \end{pmatrix}$$

by the similarity transformation $B = T^{-1}AT$, where

$$T = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Derive the value of θ for which this diagonalization transformation works, and evaluate c and d.

15. Prove the following properties of Hermitian operators.
(a) Any matrix representative of a Hermitian operator has a real determinant.
(b) The inverse of a Hermitian operator is Hermitian.
(c) The product of two Hermitian operators is Hermitian if and only if they commute.

16. Prove the following properties of the *trace*.
(a) $\text{tr } A^\dagger = (\text{tr } A)^*$
(b) $\text{tr}(aA) = a \text{ tr } A$
(c) $\text{tr}(A + B) = \text{tr } A + \text{tr } B$
(d) $\text{tr}(AB) = \text{tr}(BA)$
(e) $\text{tr}(A)$ is independent of the basis in which $\mathfrak{A}$ is represented by A.

17. Show that a unitary operator may always be written in the following forms.
(a) The form $\mathfrak{U} = \mathfrak{A} + i\mathfrak{B}$, where $\mathfrak{A}$ and $\mathfrak{B}$ are Hermitian, and $[\mathfrak{A}, \mathfrak{B}] = 0$.
(b) The form $\mathfrak{U} = e^{i\mathfrak{A}}$, where $\mathfrak{A}$ is Hermitian.

18. Prove the following properties of orthogonal and unitary operators.
(a) The product of two orthogonal operators is orthogonal.
(b) If $\mathfrak{A}$ is symmetric, and $\mathfrak{U}$ orthogonal, then $\mathfrak{U}^{-1}\mathfrak{A}\mathfrak{U}$ is symmetric.
(c) The product of two unitary operators is unitary.
(d) If $\mathfrak{A}$ is Hermitian and $\mathfrak{U}$ unitary, then $\mathfrak{U}^{-1}\mathfrak{A}\mathfrak{U}$ is Hermitian.

19. Verify the matrix representing a rotation in 3D space through Eulerian angles given in Eq. 3–64.

20. What happens in the determination of eigenvectors if the eigenvalues are degenerate?

21. Find the eigenvalues and eigenvectors of the operator $\mathfrak{K}$, whose matrix representative in 3D space is

$$K = \begin{pmatrix} 7 & -3 & -\sqrt{2} \\ -3 & 7 & \sqrt{2} \\ -\sqrt{2} & \sqrt{2} & 10 \end{pmatrix}$$

Check your result by applying an appropriate similarity transformation to K to diagonalize K.

22. Show that the eigenvectors of a unitary operator are orthogonal.

23. Find the eigenvalues and eigenvectors of

$$\begin{pmatrix} 1 & i & 0 \\ -i & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

4

Classical Mechanics

Our purpose in this chapter is twofold. We shall first formulate classical mechanics in such a way that the connection between classical and quantum mechanics is clear.

Our second purpose is to learn to handle two mechanical problems of importance in quantum chemistry, vibrational and rotational motion of molecules. These two kinds of molecular motion lie at the center of infrared and microwave spectroscopy.

We begin with a section of review and basic definitions, proceed from there to Lagrange's and Hamilton's equations to point up the connection with quantum mechanics, and conclude with applications to molecular motion.

4–1 INTRODUCTION AND THE CONSERVATION LAWS

In considering classical mechanics, we shall freely use vector quantities. The world in which we, and our experiments, live is a three-dimensional Euclidean vector space, so that we may quickly summarize the properties of vectors in this space as follows:

(a) A vector **V** will be specified by three components, all real, usually in the Cartesian basis,