MATHEMATICAL METHODS FOR PHYSICISTS

SIXTH EDITION

George B. Arfken
Miami University
Oxford, OH

Hans J. Weber
University of Virginia
Charlottesville, VA

ELSEVIER
ACADEMIC
PRESS

Amsterdam Boston Heidelberg London New York Oxford
Paris San Diego San Francisco Singapore Sydney Tokyo

CHAPTER 3

DETERMINANTS AND MATRICES

3.1 DETERMINANTS

We begin the study of matrices by solving linear equations that will lead us to determinants and matrices. The concept of *determinant* and the notation were introduced by the renowned German mathematician and philosopher Gottfried Wilhelm von Leibniz.

Homogeneous Linear Equations

One of the major applications of determinants is in the establishment of a condition for the existence of a nontrivial solution for a set of linear homogeneous algebraic equations. Suppose we have three unknowns x_1, x_2, x_3 (or n equations with n unknowns):

$$\begin{aligned} a_1x_1 + a_2x_2 + a_3x_3 &= 0, \\ b_1x_1 + b_2x_2 + b_3x_3 &= 0, \\ c_1x_1 + c_2x_2 + c_3x_3 &= 0. \end{aligned} \tag{3.1}$$

The problem is to determine under what conditions there is any solution, apart from the trivial one $x_1 = 0,\ x_2 = 0,\ x_3 = 0$. If we use vector notation $\mathbf{x} = (x_1, x_2, x_3)$ for the solution and three rows $\mathbf{a} = (a_1, a_2, a_3),\ \mathbf{b} = (b_1, b_2, b_3),\ \mathbf{c} = (c_1, c_2, c_3)$ of coefficients, then the three equations, Eqs. (3.1), become

$$\mathbf{a} \cdot \mathbf{x} = 0, \qquad \mathbf{b} \cdot \mathbf{x} = 0, \qquad \mathbf{c} \cdot \mathbf{x} = 0. \tag{3.2}$$

These three vector equations have the **geometrical** interpretation that $\mathbf{x}$ is orthogonal to $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$. If the volume spanned by $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ given by the determinant (or triple scalar

product, see Eq. (1.50) of Section 1.5)

$$D_3 = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \det(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \tag{3.3}$$

is not zero, then there is only the trivial solution $\mathbf{x} = 0$.

Conversely, if the aforementioned determinant of coefficients vanishes, then one of the row vectors is a linear combination of the other two. Let us assume that $\mathbf{c}$ lies in the plane spanned by $\mathbf{a}$ and $\mathbf{b}$, that is, that the third equation is a linear combination of the first two and not independent. Then $\mathbf{x}$ is orthogonal to that plane so that $\mathbf{x} \sim \mathbf{a} \times \mathbf{b}$. Since homogeneous equations can be multiplied by arbitrary numbers, only ratios of the x_i are relevant, for which we then obtain ratios of 2×2 determinants

$$\begin{aligned} \frac{x_1}{x_3} &= \frac{a_2 b_3 - a_3 b_2}{a_1 b_2 - a_2 b_1} \\ \frac{x_2}{x_3} &= -\frac{a_1 b_3 - a_3 b_1}{a_1 b_2 - a_2 b_1} \end{aligned} \tag{3.4}$$

from the components of the cross product $\mathbf{a} \times \mathbf{b}$, provided $x_3 \sim a_1 b_2 - a_2 b_1 \neq 0$. This is **Cramer's rule** for three homogeneous linear equations.

Inhomogeneous Linear Equations

The simplest case of two equations with two unknowns,

$$a_1 x_1 + a_2 x_2 = a_3, \qquad b_1 x_1 + b_2 x_2 = b_3, \tag{3.5}$$

can be reduced to the previous case by imbedding it in three-dimensional space with a solution vector $\mathbf{x} = (x_1, x_2, -1)$ and row vectors $\mathbf{a} = (a_1, a_2, a_3)$, $\mathbf{b} = (b_1, b_2, b_3)$. As before, Eqs. (3.5) in vector notation, $\mathbf{a} \cdot \mathbf{x} = 0$ and $\mathbf{b} \cdot \mathbf{x} = 0$, imply that $\mathbf{x} \sim \mathbf{a} \times \mathbf{b}$, so the analog of Eqs. (3.4) holds. For this to apply, though, the third component of $\mathbf{a} \times \mathbf{b}$ must not be zero, that is, $a_1 b_2 - a_2 b_1 \neq 0$, because the third component of $\mathbf{x}$ is $-1 \neq 0$. This yields the x_i as

$$x_1 = \frac{a_3 b_2 - b_3 a_2}{a_1 b_2 - a_2 b_1} = \frac{\begin{vmatrix} a_3 & a_2 \\ b_3 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}}, \tag{3.6a}$$

$$x_2 = \frac{a_1 b_3 - a_3 b_1}{a_1 b_2 - a_2 b_1} = \frac{\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}}{\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}}. \tag{3.6b}$$

The determinant in the numerator of $x_1 (x_2)$ is obtained from the determinant of the coefficients $\left|\begin{smallmatrix} a_1 & a_2 \\ b_2 & b_2 \end{smallmatrix}\right|$ by replacing the first (second) column vector by the vector $\binom{a_3}{b_3}$ of the inhomogeneous side of Eq. (3.5). This is **Cramer's rule** for a set of two inhomogeneous linear equations with two unknowns.

These solutions of linear equations in terms of determinants can be generalized to n dimensions. The determinant is a square array

$$D_n = \begin{vmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \\ c_1 & c_2 & \cdots & c_n \\ \cdot & \cdot & \cdots & \cdot \end{vmatrix} \tag{3.7}$$

of numbers (or functions), the coefficients of n linear equations in our case here. The number n of columns (and of rows) in the array is sometimes called the **order** of the determinant. The generalization of the expansion in Eq. (1.48) of the triple scalar product (of row vectors of three linear equations) leads to the following value of the determinant D_n in n dimensions,

$$D_n = \sum_{i,j,k,\ldots} \varepsilon_{ijk\ldots} a_i b_j c_k \cdots, \tag{3.8}$$

where $\varepsilon_{ijk\ldots}$, analogous to the Levi-Civita symbol of Section 2.9, is $+1$ for even permutations[1] $(ijk\cdots)$ of $(123\cdots n)$, -1 for odd permutations, and zero if any index is repeated.

Specifically, for the third-order determinant D_3 of Eq. (3.3), Eq. (3.8) leads to

$$D_3 = +a_1b_2c_3 - a_1b_3c_2 - a_2b_1c_3 + a_2b_3c_1 + a_3b_1c_2 - a_3b_2c_1. \tag{3.9}$$

The third-order determinant, then, is this particular linear combination of products. Each product contains one and only one element from each row and from each column. Each product is added if the columns (indices) represent an even permutation of (123) and subtracted if we have an odd permutation. Equation (3.3) may be considered shorthand notation for Eq. (3.9). The number of terms in the sum (Eq. (3.8)) is 24 for a fourth-order determinant, $n!$ for an nth-order determinant. Because of the appearance of the negative signs in Eq. (3.9) (and possibly in the individual elements as well), there may be considerable cancellation. It is quite possible that a determinant of large elements will have a very small value.

Several useful properties of the nth-order determinants follow from Eq. (3.8). Again, to be specific, Eq. (3.9) for third-order determinants is used to illustrate these properties.

Laplacian Development by Minors

Equation (3.9) may be written

$$\begin{aligned} D_3 &= a_1(b_2c_3 - b_3c_2) - a_2(b_1c_3 - b_3c_1) + a_3(b_1c_2 - b_2c_1) \\ &= a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix}. \end{aligned} \tag{3.10}$$

In general, the nth-order determinant may be expanded as a linear combination of the products of the elements of any row (or any column) and the $(n-1)$th-order determinants

[1] In a linear sequence $abcd\cdots$, any single, simple transposition of adjacent elements yields an **odd** permutation of the original sequence: $abcd \to bacd$. Two such transpositions yield an even permutation. In general, an odd number of such interchanges of adjacent elements results in an odd permutation; an even number of such transpositions yields an even permutation.

formed by striking out the row and column of the original determinant in which the element appears. This reduced array (2×2 in this specific example) is called a **minor**. If the element is in the ith row and the jth column, the sign associated with the product is $(-1)^{i+j}$. The minor with this sign is called the **cofactor**. If M_{ij} is used to designate the minor formed by omitting the ith row and the jth column and C_{ij} is the corresponding cofactor, Eq. (3.10) becomes

$$D_3 = \sum_{j=1}^{3} (-1)^{j+1} a_j M_{1j} = \sum_{j=1}^{3} a_j C_{1j}. \tag{3.11}$$

In this case, expanding along the first row, we have $i = 1$ and the summation over j, the columns.

This Laplace expansion may be used to advantage in the evaluation of high-order determinants in which a lot of the elements are zero. For example, to find the value of the determinant

$$D = \begin{vmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{vmatrix}, \tag{3.12}$$

we expand across the top row to obtain

$$D = (-1)^{1+2} \cdot (1) \begin{vmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{vmatrix}. \tag{3.13}$$

Again, expanding across the top row, we get

$$D = (-1) \cdot (-1)^{1+1} \cdot (-1) \begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix} = 1. \tag{3.14}$$

(This determinant D (Eq. (3.12)) is formed from one of the Dirac matrices appearing in Dirac's relativistic electron theory in Section 3.4.)

Antisymmetry

The determinant changes sign if any two rows are interchanged or if any two columns are interchanged. This follows from the even–odd character of the Levi-Civita ε in Eq. (3.8) or explicitly from the form of Eqs. (3.9) and (3.10).[2]

This property was used in Section 2.9 to develop a totally antisymmetric linear combination. It is also frequently used in quantum mechanics in the construction of a many-particle wave function that, in accordance with the Pauli exclusion principle, will be antisymmetric under the interchange of any two identical spin $\frac{1}{2}$ particles (electrons, protons, neutrons, etc.).

[2]The sign reversal is reasonably obvious for the interchange of two adjacent rows (or columns), this clearly being an odd permutation. Show that the interchange of **any** two rows is still an odd permutation.

- As a special case of antisymmetry, any determinant with two rows equal or two columns equal equals zero.
- If each element in a row or each element in a column is zero, the determinant is equal to zero.
- If each element in a row or each element in a column is multiplied by a constant, the determinant is multiplied by that constant.
- The value of a determinant is unchanged if a multiple of one row is added (column by column) to another row or if a multiple of one column is added (row by row) to another column.[3]

We have

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 + ka_2 & a_2 & a_3 \\ b_1 + kb_2 & b_2 & b_3 \\ c_1 + kc_2 & c_2 & c_3 \end{vmatrix}. \tag{3.15}$$

Using the Laplace development on the right-hand side, we obtain

$$\begin{vmatrix} a_1 + ka_2 & a_2 & a_3 \\ b_1 + kb_2 & b_2 & b_3 \\ c_1 + kc_2 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} + k \begin{vmatrix} a_2 & a_2 & a_3 \\ b_2 & b_2 & b_3 \\ c_2 & c_2 & c_3 \end{vmatrix}, \tag{3.16}$$

then by the property of antisymmetry the second determinant on the right-hand side of Eq. (3.16) vanishes, verifying Eq. (3.15).

As a special case, a determinant is equal to zero if any two rows are proportional or any two columns are proportional.

Some useful relations involving determinants or matrices appear in Exercises of Sections 3.2 and 3.4.

Returning to the homogeneous Eqs. (3.1) and multiplying the determinant of the coefficients by x_1, then adding x_2 times the second column and x_3 times the third column, we can directly establish the condition for the presence of a nontrivial solution for Eqs. (3.1):

$$\begin{aligned} x_1 \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} &= \begin{vmatrix} a_1 x_1 & a_2 & a_3 \\ b_1 x_1 & b_2 & b_3 \\ c_1 x_1 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 x_1 + a_2 x_2 + a_3 x_3 & a_2 & a_3 \\ b_1 x_1 + b_2 x_2 + b_3 x_3 & b_2 & b_3 \\ c_1 x_1 + c_2 x_2 + c_3 x_3 & c_2 & c_3 \end{vmatrix} \\ &= \begin{vmatrix} 0 & a_2 & a_3 \\ 0 & b_2 & b_3 \\ 0 & c_2 & c_3 \end{vmatrix} = 0. \end{aligned} \tag{3.17}$$

Therefore x_1 (and x_2 and x_3) must be zero **unless the determinant of the coefficients vanishes**. Conversely (see text below Eq. (3.3)), we can show that if the determinant of the coefficients vanishes, a nontrivial solution does indeed exist. This is used in Section 9.6 to establish the linear dependence or independence of a set of functions.

[3]This derives from the geometric meaning of the determinant as the volume of the parallelepiped spanned by its column vectors. Pulling it to the side without changing its height leaves the volume unchanged.

If our linear equations are **inhomogeneous**, that is, as in Eqs. (3.5) if the zeros on the right-hand side of Eqs. (3.1) are replaced by a_4, b_4, and c_4, respectively, then from Eq. (3.17) we obtain, instead,

$$x_1 = \frac{\begin{vmatrix} a_4 & a_2 & a_3 \\ b_4 & b_2 & b_3 \\ c_4 & c_2 & c_3 \end{vmatrix}}{\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}}, \tag{3.18}$$

which generalizes Eq. (3.6a) to $n = 3$ dimensions, etc. If the determinant of the coefficients vanishes, the inhomogeneous set of equations has no solution — unless the numerators also vanish. In this case solutions may exist but they are not unique (see Exercise 3.1.3 for a specific example).

For numerical work, this determinant solution, Eq. (3.18), is exceedingly unwieldy. The determinant may involve large numbers with alternate signs, and in the subtraction of two large numbers the relative error may soar to a point that makes the result worthless. Also, although the determinant method is illustrated here with three equations and three unknowns, we might easily have 200 equations with 200 unknowns, which, involving up to 200! terms in each determinant, pose a challenge even to high-speed computers. There must be a better way.

In fact, there are better ways. One of the best is a straightforward process often called **Gauss elimination**. To illustrate this technique, consider the following set of equations.

Example 3.1.1 GAUSS ELIMINATION

Solve

$$\begin{aligned} 3x + 2y + z &= 11 \\ 2x + 3y + z &= 13 \\ x + y + 4z &= 12. \end{aligned} \tag{3.19}$$

The determinant of the inhomogeneous linear equations (3.19) is 18, so a solution exists.

For convenience and for the optimum numerical accuracy, the equations are rearranged so that the largest coefficients run along the main diagonal (upper left to lower right). This has already been done in the preceding set.

The Gauss technique is to use the first equation to eliminate the first unknown, x, from the remaining equations. Then the (new) second equation is used to eliminate y from the last equation. In general, we work down through the set of equations, and then, with one unknown determined, we work back up to solve for each of the other unknowns in succession.

Dividing each row by its initial coefficient, we see that Eqs. (3.19) become

$$\begin{aligned} x + \tfrac{2}{3}y + \tfrac{1}{3}z &= \tfrac{11}{3} \\ x + \tfrac{3}{2}y + \tfrac{1}{2}z &= \tfrac{13}{2} \\ x + y + 4z &= 12. \end{aligned} \tag{3.20}$$

Now, using the first equation, we eliminate x from the second and third equations:

$$\begin{aligned} x + \tfrac{2}{3}y + \tfrac{1}{3}z &= \tfrac{11}{3} \\ \tfrac{5}{6}y + \tfrac{1}{6}z &= \tfrac{17}{6} \\ \tfrac{1}{3}y + \tfrac{11}{3}z &= \tfrac{25}{3} \end{aligned} \tag{3.21}$$

and

$$\begin{aligned} x + \tfrac{2}{3}y + \tfrac{1}{3}z &= \tfrac{11}{3} \\ y + \tfrac{1}{5}z &= \tfrac{17}{5} \\ y + 11z &= 25. \end{aligned} \tag{3.22}$$

Repeating the technique, we use the new second equation to eliminate y from the third equation:

$$\begin{aligned} x + \tfrac{2}{3}y + \tfrac{1}{3}z &= \tfrac{11}{3} \\ y + \tfrac{1}{5}z &= \tfrac{17}{5} \\ 54z &= 108, \end{aligned} \tag{3.23}$$

or

$$z = 2.$$

Finally, working back up, we get

$$y + \tfrac{1}{5} \times 2 = \tfrac{17}{5},$$

or

$$y = 3.$$

Then with z and y determined,

$$x + \tfrac{2}{3} \times 3 + \tfrac{1}{3} \times 2 = \tfrac{11}{3},$$

and

$$x = 1.$$

The technique may not seem so elegant as Eq. (3.18), but it is well adapted to computers and is far faster than the time spent with determinants.

This Gauss technique may be used to convert a determinant into triangular form:

$$D = \begin{vmatrix} a_1 & b_1 & c_1 \\ 0 & b_2 & c_2 \\ 0 & 0 & c_3 \end{vmatrix}$$

for a third-order determinant whose elements are not to be confused with those in Eq. (3.3). In this form $D = a_1 b_2 c_3$. For an nth-order determinant the evaluation of the triangular form requires only $n - 1$ multiplications, compared with the $n!$ required for the general case.

A variation of this progressive elimination is known as Gauss–Jordan elimination. We start as with the preceding Gauss elimination, but each new equation considered is used to eliminate a variable from **all** the other equations, not just those below it. If we had used this Gauss–Jordan elimination, Eq. (3.23) would become

$$\begin{aligned} x + \tfrac{1}{5}z &= \tfrac{7}{5} \\ y + \tfrac{1}{5}z &= \tfrac{17}{5} \\ z &= 2, \end{aligned} \tag{3.24}$$

using the second equation of Eqs. (3.22) to eliminate y from both the first and third equations. Then the third equation of Eqs. (3.24) is used to eliminate z from the first and second, giving

$$\begin{aligned} x &= 1 \\ y &= 3 \\ z &= 2. \end{aligned} \tag{3.25}$$

We return to this Gauss–Jordan technique in Section 3.2 for inverting matrices.

Another technique suitable for computer use is the Gauss–Seidel iteration technique. Each technique has its advantages and disadvantages. The Gauss and Gauss–Jordan methods may have accuracy problems for large determinants. This is also a problem for matrix inversion (Section 3.2). The Gauss–Seidel method, as an iterative method, may have convergence problems. The IBM Scientific Subroutine Package (SSP) uses Gauss and Gauss–Jordan techniques. The Gauss–Seidel iterative method and the Gauss and Gauss–Jordan elimination methods are discussed in considerable detail by Ralston and Wilf and also by Pennington.[4] Computer codes in FORTRAN and other programming languages and extensive literature for the Gauss–Jordan elimination and others are also given by Press *et al.*[5] ■

Linear Dependence of Vectors

Two nonzero two-dimensional vectors

$$\mathbf{a}_1 = \begin{pmatrix} a_{11} \\ a_{12} \end{pmatrix} \neq 0, \qquad \mathbf{a}_2 = \begin{pmatrix} a_{21} \\ a_{22} \end{pmatrix} \neq 0$$

are defined to be **linearly dependent** if two numbers x_1, x_2 can be found that are not both zero so that the linear relation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = 0$ holds. They are **linearly independent** if $x_1 = 0 = x_2$ is the only solution of this linear relation. Writing it in Cartesian components, we obtain two homogeneous linear equations

$$a_{11}x_1 + a_{21}x_2 = 0, \qquad a_{12}x_1 + a_{22}x_2 = 0$$

[4] A. Ralston and H. Wilf, eds., *Mathematical Methods for Digital Computers*. New York: Wiley (1960); R. H. Pennington, *Introductory Computer Methods and Numerical Analysis*. New York: Macmillan (1970).

[5] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*, 2nd ed. Cambridge, UK: Cambridge University Press (1992), Chapter 2.

from which we extract the following criterion for linear independence of two vectors using Cramer's rule. **If $\mathbf{a}_1$, $\mathbf{a}_2$ span a nonzero area**, that is, **their determinant** $\begin{vmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{vmatrix} \neq 0$, **then the set of homogeneous linear equations has only the solution** $x_1 = 0 = x_2$. **If the determinant is zero, then there is a nontrivial solution** x_1, x_2, and **our vectors are linearly dependent**. In particular, the unit vectors in the x- and y-directions are linearly independent, the linear relation $x_1\hat{x}_1 + x_2\hat{x}_2 = \binom{x_1}{x_2} = \binom{0}{0}$ having only the trivial solution $x_1 = 0 = x_2$.

Three or more vectors in two-dimensional space are always linearly dependent. Thus, the maximum number of linearly independent vectors in two-dimensional space is 2. For example, given $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, the linear relation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 = 0$ always has nontrivial solutions. If one of the vectors is zero, linear dependence is obvious because the coefficient of the zero vector may be chosen to be nonzero and that of the others as zero. So we assume all of them as nonzero. If $\mathbf{a}_1$ and $\mathbf{a}_2$ are linearly independent, we write the linear relation

$$a_{11}x_1 + a_{21}x_2 = -a_{31}x_3, \qquad a_{12}x_1 + a_{22}x_2 = -a_{32}x_3,$$

as a set of two inhomogeneous linear equations and apply Cramer's rule. Since the determinant is nonzero, we can find a nontrivial solution x_1, x_2 for any nonzero x_3. This argument goes through for any pair of linearly independent vectors. If all pairs are linearly dependent, any of these linear relations is a linear relation among the three vectors, and we are finished. If there are more than three vectors, we pick any three of them and apply the foregoing reasoning and put the coefficients of the other vectors, $x_j = 0$, in the linear relation.

- Mutually orthogonal vectors are linearly independent.

Assume a linear relation $\sum_i c_i\mathbf{v}_i = 0$. Dotting $\mathbf{v}_j$ into this using $\mathbf{v}_j \cdot \mathbf{v}_i = 0$ for $j \neq i$, we obtain $c_j\mathbf{v}_j \cdot \mathbf{v}_j = 0$, so every $c_j = 0$ because $\mathbf{v}_j^2 \neq 0$.

It is straightforward to extend these theorems to n or more vectors in n-dimensional Euclidean space. Thus, **the maximum number of linearly independent vectors in n-dimensional space is n**. The coordinate unit vectors are linearly independent because they span a nonzero parallelepiped in n-dimensional space and their determinant is unity.

Gram–Schmidt Procedure

In an n-dimensional vector space with an inner (or scalar) product, we can always construct an orthonormal basis of n vectors $\mathbf{w}_i$ with $\mathbf{w}_i \cdot \mathbf{w}_j = \delta_{ij}$ starting from n linearly independent vectors $\mathbf{v}_i$, $i = 0, 1, \ldots, n-1$.

We start by normalizing $\mathbf{v}_0$ to unity, defining $\mathbf{w}_0 = \frac{\mathbf{v}_0}{\sqrt{\mathbf{v}_0^2}}$. Then we project $\mathbf{v}_0$ from $\mathbf{v}_1$, forming $\mathbf{u}_1 = \mathbf{v}_1 + a_{10}\mathbf{w}_0$, with the admixture coefficient a_{10} chosen so that $\mathbf{v}_0 \cdot \mathbf{u}_1 = 0$. Dotting $\mathbf{v}_0$ into $\mathbf{u}_1$ yields $a_{10} = -\frac{\mathbf{v}_0 \cdot \mathbf{v}_1}{\sqrt{\mathbf{v}_0^2}} = -\mathbf{v}_1 \cdot \mathbf{w}_0$. Again, we normalize $\mathbf{u}_1$ defining $\mathbf{w}_1 = \frac{\mathbf{u}_1}{\sqrt{\mathbf{u}_1^2}}$. Here, $\mathbf{u}_1^2 \neq 0$ because $\mathbf{v}_0$, $\mathbf{v}_1$ are linearly independent. This first step generalizes to

$$\mathbf{u}_j = \mathbf{v}_j + a_{j0}\mathbf{w}_0 + a_{j1}\mathbf{w}_1 + \cdots + a_{jj-1}\mathbf{w}_{j-1},$$

with coefficients $a_{ji} = -\mathbf{v}_j \cdot \mathbf{w}_i$. Normalizing $\mathbf{w}_j = \frac{\mathbf{u}_j}{\sqrt{\mathbf{u}_j^2}}$ completes our construction.

It will be noticed that although this Gram–Schmidt procedure is one possible way of constructing an orthogonal or orthonormal set, the vectors $\mathbf{w}_i$ are not unique. There is an infinite number of possible orthonormal sets.

As an illustration of the freedom involved, consider two (nonparallel) vectors $\mathbf{A}$ and $\mathbf{B}$ in the xy-plane. We may normalize $\mathbf{A}$ to unit magnitude and then form $\mathbf{B}' = a\mathbf{A} + \mathbf{B}$ so that $\mathbf{B}'$ is perpendicular to $\mathbf{A}$. By normalizing $\mathbf{B}'$ we have completed the Gram–Schmidt orthogonalization for two vectors. But any two perpendicular unit vectors, such as $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$, could have been chosen as our orthonormal set. Again, with an infinite number of possible rotations of $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ about the z-axis, we have an infinite number of possible orthonormal sets.

Example 3.1.2 VECTORS BY GRAM–SCHMIDT ORTHOGONALIZATION

To illustrate the method, we consider two vectors

$$\mathbf{v}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{v}_1 = \begin{pmatrix} 1 \\ -2 \end{pmatrix},$$

which are neither orthogonal nor normalized. Normalizing the first vector $\mathbf{w}_0 = \mathbf{v}_0/\sqrt{2}$, we then construct $\mathbf{u}_1 = \mathbf{v}_1 + a_{10}\mathbf{w}_0$ so as to be orthogonal to $\mathbf{v}_0$. This yields

$$\mathbf{u}_1 \cdot \mathbf{v}_0 = 0 = \mathbf{v}_1 \cdot \mathbf{v}_0 + \frac{a_{10}}{\sqrt{2}}\mathbf{v}_0^2 = -1 + a_{10}\sqrt{2},$$

so the adjustable admixture coefficient $a_{10} = 1/\sqrt{2}$. As a result,

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ -2 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{3}{2}\begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

so the second orthonormal vector becomes

$$\mathbf{w}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

We check that $\mathbf{w}_0 \cdot \mathbf{w}_1 = 0$. The two vectors $\mathbf{w}_0$, $\mathbf{w}_1$ form an orthonormal set of vectors, a basis of two-dimensional Euclidean space. ■

Exercises

3.1.1 Evaluate the following determinants:

$$\text{(a)} \begin{vmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{vmatrix}, \qquad \text{(b)} \begin{vmatrix} 1 & 2 & 0 \\ 3 & 1 & 2 \\ 0 & 3 & 1 \end{vmatrix}, \qquad \text{(c)} \frac{1}{\sqrt{2}} \begin{vmatrix} 0 & \sqrt{3} & 0 & 0 \\ \sqrt{3} & 0 & 2 & 0 \\ 0 & 2 & 0 & \sqrt{3} \\ 0 & 0 & \sqrt{3} & 0 \end{vmatrix}.$$

3.1.2 Test the set of linear homogeneous equations

$$x + 3y + 3z = 0, \qquad x - y + z = 0, \qquad 2x + y + 3z = 0$$

to see if it possesses a nontrivial solution, and find one.

3.1.3 Given the pair of equations

$$x + 2y = 3, \qquad 2x + 4y = 6,$$

(a) Show that the determinant of the coefficients vanishes.
(b) Show that the numerator determinants (Eq. (3.18)) also vanish.
(c) Find at least two solutions.

3.1.4 Express the **components** of $\mathbf{A} \times \mathbf{B}$ as 2×2 determinants. Then show that the dot product $\mathbf{A} \cdot (\mathbf{A} \times \mathbf{B})$ yields a Laplacian expansion of a 3×3 determinant. Finally, note that two rows of the 3×3 determinant are identical and hence that $\mathbf{A} \cdot (\mathbf{A} \times \mathbf{B}) = 0$.

3.1.5 If C_{ij} is the cofactor of element a_{ij} (formed by striking out the ith row and jth column and including a sign $(-1)^{i+j}$), show that

(a) $\sum_i a_{ij} C_{ij} = \sum_i a_{ji} C_{ji} = |A|$, where $|A|$ is the determinant with the elements a_{ij},
(b) $\sum_i a_{ij} C_{ik} = \sum_i a_{ji} C_{ki} = 0, \; j \neq k$.

3.1.6 A determinant with all elements of order unity may be surprisingly small. The Hilbert determinant $H_{ij} = (i + j - 1)^{-1}, \; i, j = 1, 2, \ldots, n$ is notorious for its small values.

(a) Calculate the value of the Hilbert determinants of order n for $n = 1, 2$, and 3.
(b) If an appropriate subroutine is available, find the Hilbert determinants of order n for $n = 4, 5$, and 6.

ANS.

n	$\text{Det}(H_n)$
1	1.
2	8.33333×10^{-2}
3	4.62963×10^{-4}
4	1.65344×10^{-7}
5	3.74930×10^{-12}
6	5.36730×10^{-18}

3.1.7 Solve the following set of linear simultaneous equations. Give the results to five decimal places.

$$\begin{aligned}
1.0x_1 + 0.9x_2 + 0.8x_3 + 0.4x_4 + 0.1x_5 &= 1.0 \\
0.9x_1 + 1.0x_2 + 0.8x_3 + 0.5x_4 + 0.2x_5 + 0.1x_6 &= 0.9 \\
0.8x_1 + 0.8x_2 + 1.0x_3 + 0.7x_4 + 0.4x_5 + 0.2x_6 &= 0.8 \\
0.4x_1 + 0.5x_2 + 0.7x_3 + 1.0x_4 + 0.6x_5 + 0.3x_6 &= 0.7 \\
0.1x_1 + 0.2x_2 + 0.4x_3 + 0.6x_4 + 1.0x_5 + 0.5x_6 &= 0.6 \\
0.1x_2 + 0.2x_3 + 0.3x_4 + 0.5x_5 + 1.0x_6 &= 0.5.
\end{aligned}$$

Note. These equations may also be solved by matrix inversion, Section 3.2.

3.1.8 Solve the linear equations $\mathbf{a}\cdot\mathbf{x}=c$, $\mathbf{a}\times\mathbf{x}+\mathbf{b}=0$ for $\mathbf{x}=(x_1,x_2,x_3)$ with constant vectors $\mathbf{a}\neq 0$, $\mathbf{b}$ and constant c.

ANS. $\mathbf{x}=\frac{c}{a^2}\mathbf{a}+(\mathbf{a}\times\mathbf{b})/a^2$.

3.1.9 Solve the linear equations $\mathbf{a}\cdot\mathbf{x}=d$, $\mathbf{b}\cdot\mathbf{x}=e$, $\mathbf{c}\cdot\mathbf{x}=f$, for $\mathbf{x}=(x_1,x_2,x_3)$ with constant vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ and constants d, e, f such that $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}\neq 0$.

ANS. $[(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}]\mathbf{x}=d(\mathbf{b}\times\mathbf{c})+e(\mathbf{c}\times\mathbf{a})+f(\mathbf{a}\times\mathbf{b})$.

3.1.10 Express in vector form the solution (x_1,x_2,x_3) of $\mathbf{a}x_1+\mathbf{b}x_2+\mathbf{c}x_3+\mathbf{d}=0$ with constant vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ so that $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}\neq 0$.

3.2 MATRICES

Matrix analysis belongs to linear algebra because matrices are linear operators or maps such as rotations. Suppose, for instance, we rotate the Cartesian coordinates of a two-dimensional space, as in Section 1.2, so that, in vector notation,

$$\begin{pmatrix}x_1'\\x_2'\end{pmatrix}=\begin{pmatrix}x_1\cos\varphi+x_2\sin\varphi\\-x_2\sin\varphi+x_2\cos\varphi\end{pmatrix}=\begin{pmatrix}\sum_j a_{1j}x_j\\\sum_j a_{2j}x_j\end{pmatrix}. \tag{3.26}$$

We label the array of elements $\begin{pmatrix}a_{11}&a_{12}\\a_{21}&a_{22}\end{pmatrix}$ a 2×2 matrix A consisting of two rows and two columns and consider the vectors x, x' as 2×1 matrices. **We take the summation of products in Eq. (3.26) as a definition of matrix multiplication involving the scalar product of each row vector of A with the column vector x.** Thus, in matrix notation Eq. (3.26) becomes

$$x'=\mathsf{A}x. \tag{3.27}$$

To extend this definition of multiplication of a matrix times a column vector to the product of two 2×2 matrices, let the coordinate rotation be followed by a second rotation given by matrix B such that

$$x''=\mathsf{B}x'. \tag{3.28}$$

In component form,

$$x_i''=\sum_j b_{ij}x_j'=\sum_j b_{ij}\sum_k a_{jk}x_k=\sum_k\left(\sum_j b_{ij}a_{jk}\right)x_k. \tag{3.29}$$

The summation over j is matrix multiplication defining a matrix $\mathsf{C}=\mathsf{BA}$ such that

$$x_i''=\sum_k c_{ik}x_k, \tag{3.30}$$

or $x''=\mathsf{C}x$ in matrix notation. Again, this definition involves the scalar products of row vectors of B with column vectors of A. This definition of matrix multiplication generalizes to $m\times n$ matrices and is found useful; indeed, **this usefulness is the justification for its existence.** The geometrical interpretation is that the matrix product of the two matrices BA is the rotation that carries the unprimed system directly into the double-primed coordinate

system. Before passing to formal definitions, the your should note that operator A is described by its effect on the coordinates or basis vectors. The matrix elements a_{ij} constitute a **representation** of the operator, a representation that depends on the choice of a basis.

The special case where a matrix has one column and n rows is called a column vector, $|x\rangle$, with components x_i, $i = 1, 2, \ldots, n$. If A is an $n \times n$ matrix, $|x\rangle$ an n-component column vector, $A|x\rangle$ is defined as in Eqs. (3.27) and (3.26). Similarly, if a matrix has one row and n columns, it is called a row vector, $\langle x|$ with components x_i, $i = 1, 2, \ldots, n$. Clearly, $\langle x|$ results from $|x\rangle$ by interchanging rows and columns, a matrix operation called **transposition**, and transposition for any matrix A, $\tilde{\mathsf{A}}$ is called[6] "A transpose" with matrix elements $(\tilde{\mathsf{A}})_{ik} = A_{ki}$. Transposing a product of matrices AB reverses the order and gives $\tilde{\mathsf{B}}\tilde{\mathsf{A}}$; similarly $\mathsf{A}|x\rangle$ transpose is $\langle x|\mathsf{A}$. The scalar product takes the form $\langle x|y\rangle = \sum_i x_i y_i$ (x_i^* in a complex vector space). This **Dirac bra-ket notation** is used in quantum mechanics extensively and in Chapter 10 and here subsequently.

More abstractly, we can define the **dual space $\tilde{V}$ of linear functionals** F on a vector space V, where each linear functional F of $\tilde{V}$ assigns a number $F(\mathbf{v})$ so that

$$F(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1F(\mathbf{v}_1) + c_2F(\mathbf{v}_2)$$

for any vectors $\mathbf{v}_1$, $\mathbf{v}_2$ from our vector space V and numbers c_1, c_2. If we define the sum of two functionals by linearity as

$$(F_1 + F_2)(\mathbf{v}) = F_1(\mathbf{v}) + F_2(\mathbf{v}),$$

then $\tilde{V}$ is a linear space by construction.

Riesz' theorem says that there is a one-to-one correspondence between linear functionals F in $\tilde{V}$ and vectors $\mathbf{f}$ in a vector space V that has an inner (or scalar) product $\langle \mathbf{f}|\mathbf{v}\rangle$ defined for any pair of vectors $\mathbf{f}$, $\mathbf{v}$.

The proof relies on the scalar product by defining a linear functional F for any vector $\mathbf{f}$ of V as $F(\mathbf{v}) = \langle \mathbf{f}|\mathbf{v}\rangle$ for any $\mathbf{v}$ of V. The linearity of the scalar product in $\mathbf{f}$ shows that these functionals form a vector space (contained in $\tilde{V}$ necessarily). Note that a linear functional is completely specified when it is defined for every vector $\mathbf{v}$ of a given vector space.

On the other hand, starting from any nontrivial linear functional F of $\tilde{V}$ we now construct a unique vector $\mathbf{f}$ of V so that $F(\mathbf{v}) = \mathbf{f} \cdot \mathbf{v}$ is given by an inner product. We start from an orthonormal basis $\mathbf{w}_i$ of vectors in V using the Gram–Schmidt procedure (see Section 3.2). Take any vector $\mathbf{v}$ from V and expand it as $\mathbf{v} = \sum_i \mathbf{w}_i \cdot \mathbf{v}\mathbf{w}_i$. Then the linear functional $F(\mathbf{v}) = \sum_i \mathbf{w}_i \cdot \mathbf{v}F(\mathbf{w}_i)$ is well defined on V. If we define the specific vector $\mathbf{f} = \sum_i F(\mathbf{w}_i)\mathbf{w}_i$, then its inner product with an arbitrary vector $\mathbf{v}$ is given by $\langle \mathbf{f}|\mathbf{v}\rangle = \mathbf{f} \cdot \mathbf{v} = \sum_i F(\mathbf{w}_i)\mathbf{w}_i \cdot \mathbf{v} = F(\mathbf{v})$, which proves Riesz' theorem.

Basic Definitions

A matrix is defined as a square or rectangular array of numbers or functions that obeys certain laws. This is a perfectly logical extension of familiar mathematical concepts. In arithmetic we deal with single numbers. In the theory of complex variables (Chapter 6) we deal with ordered pairs of numbers, $(1, 2) = 1 + 2i$, in which the ordering is important. We

[6] Some texts (including ours sometimes) denote A transpose by A^T.

now consider numbers (or functions) ordered in a square or rectangular array. For convenience in later work the numbers are distinguished by two subscripts, the first indicating the row (horizontal) and the second indicating the column (vertical) in which the number appears. For instance, a_{13} is the matrix element in the first row, third column. Hence, if A is a matrix with m rows and n columns,

$$\mathsf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}. \tag{3.31}$$

Perhaps the most important fact to note is that the elements a_{ij} are not combined with one another. A matrix is not a determinant. It is an ordered array of numbers, not a single number.

The matrix A, so far just an array of numbers, has the properties we assign to it. Literally, this means constructing a new form of mathematics. We define that matrices A, B, and C, with elements a_{ij}, b_{ij}, and c_{ij}, respectively, combine according to the following rules.

Rank

Looking back at the homogeneous linear Eqs. (3.1), we note that the matrix of coefficients, A, is made up of three row vectors that each represent one linear equation of the set. If their triple scalar product is not zero, than they span a nonzero volume and are linearly independent, and the homogeneous linear equations have only the trivial solution. In this case the matrix is said to have **rank** 3. In n dimensions the volume represented by the triple scalar product becomes the determinant, det(A), for a square matrix. If $\det(\mathsf{A}) \neq 0$, the $n \times n$ matrix A has **rank** n. The case of Eqs. (3.1), where the vector **c** lies in the plane spanned by **a** and **b**, corresponds to rank 2 of the matrix of coefficients, because only two of its row vectors (**a**, **b** corresponding to two equations) are independent. In general, the **rank r of a matrix is the maximal number of linearly independent row or column vectors** it has, with $0 \leq r \leq n$.

Equality

Matrix A = Matrix B if and only if $a_{ij} = b_{ij}$ for all values of i and j. This, of course, requires that A and B each be $m \times n$ arrays (m rows, n columns).

Addition, Subtraction

$\mathsf{A} \pm \mathsf{B} = \mathsf{C}$ if and only if $a_{ij} \pm b_{ij} = c_{ij}$ for all values of i and j, the elements combining according to the laws of ordinary algebra (or arithmetic if they are simple numbers). This means that $\mathsf{A} + \mathsf{B} = \mathsf{B} + \mathsf{A}$, commutation. Also, an associative law is satisfied $(\mathsf{A} + \mathsf{B}) + \mathsf{C} = \mathsf{A} + (\mathsf{B} + \mathsf{C})$. If all elements are zero, the matrix, called the **null matrix**, is denoted by O. For all A,

$$\mathsf{A} + \mathsf{O} = \mathsf{O} + \mathsf{A} = \mathsf{A},$$

with

$$\mathsf{O} = \begin{pmatrix} 0 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}. \tag{3.32}$$

Such $m \times n$ matrices form a linear space with respect to addition and subtraction.

Multiplication (by a Scalar)

The multiplication of matrix A by the scalar quantity α is defined as

$$\alpha\mathsf{A} = (\alpha\mathsf{A}), \tag{3.33}$$

in which the elements of $\alpha\mathsf{A}$ are αa_{ij}; that is, each element of matrix A is multiplied by the scalar factor. This is in striking contrast to the behavior of determinants in which the factor α multiplies only one column or one row and not every element of the entire determinant. A consequence of this scalar multiplication is that

$$\alpha\mathsf{A} = \mathsf{A}\alpha, \qquad \text{commutation.}$$

If A is a square matrix, then

$$\det(\alpha\mathsf{A}) = \alpha^n \det(\mathsf{A}).$$

Matrix Multiplication, Inner Product

$$\boxed{\mathsf{AB} = \mathsf{C}} \qquad \text{if and only if}^7 \qquad \boxed{c_{ij} = \sum_k a_{ik} b_{kj}.} \tag{3.34}$$

The ij element of C is formed as a scalar product of the ith row of A with the jth column of B (which demands that A have the same number of columns (n) as B has rows). The dummy index k takes on all values $1, 2, \ldots, n$ in succession; that is,

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} \tag{3.35}$$

for $n = 3$. Obviously, the dummy index k may be replaced by any other symbol that is not already in use without altering Eq. (3.34). Perhaps the situation may be clarified by stating that Eq. (3.34) defines the method of combining certain matrices. This method of combination, to give it a label, is called **matrix multiplication**. To illustrate, consider two (so-called Pauli) matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \text{and} \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.36}$$

[7] Some authors follow the summation convention here (compare Section 2.6).

The $_{11}$ element of the product, $(\sigma_1\sigma_3)_{11}$ is given by the sum of the products of elements of the first **row** of σ_1 with the corresponding elements of the first **column** of σ_3:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \to 0\cdot 1 + 1\cdot 0 = 0.$$

Continuing, we have

$$\sigma_1\sigma_3 = \begin{pmatrix} 0\cdot 1+1\cdot 0 & 0\cdot 0+1\cdot(-1) \\ 1\cdot 1+0\cdot 0 & 1\cdot 0+0\cdot(-1) \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{3.37}$$

Here

$$(\sigma_1\sigma_3)_{ij} = \sigma_{1_{i1}}\sigma_{3_{1j}} + \sigma_{1_{i2}}\sigma_{3_{2j}}.$$

Direct application of the definition of matrix multiplication shows that

$$\sigma_3\sigma_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{3.38}$$

and by Eq. (3.37)

$$\sigma_3\sigma_1 = -\sigma_1\sigma_3. \tag{3.39}$$

Except in special cases, matrix multiplication is not commutative:[8]

$$\mathsf{AB} \neq \mathsf{BA}. \tag{3.40}$$

However, from the definition of matrix multiplication we can show[9] that an associative law holds, $(\mathsf{AB})\mathsf{C} = \mathsf{A}(\mathsf{BC})$. There is also a distributive law, $\mathsf{A}(\mathsf{B}+\mathsf{C}) = \mathsf{AB}+\mathsf{AC}$.

The unit matrix 1 has elements δ_{ij}, Kronecker delta, and the property that $1\mathsf{A} = \mathsf{A}1 = \mathsf{A}$ for all A,

$$1 = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}. \tag{3.41}$$

It should be noted that it is possible for the product of two matrices to be the null matrix without either one being the null matrix. For example, if

$$\mathsf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathsf{B} = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix},$$

$\mathsf{AB} = \mathsf{O}$. This differs from the multiplication of real or complex numbers, which form a **field**, whereas the additive and multiplicative structure of matrices is called a **ring** by mathematicians. See also Exercise 3.2.6(a), from which it is evident that, if $\mathsf{AB} = 0$, at

[8] Commutation or the lack of it is conveniently described by the commutator bracket symbol, $[\mathsf{A}, \mathsf{B}] = \mathsf{AB} - \mathsf{BA}$. Equation (3.40) becomes $[\mathsf{A}, \mathsf{B}] \neq 0$.

[9] Note that the basic definitions of equality, addition, and multiplication are given in terms of the matrix elements, the a_{ij}. All our matrix operations can be carried out in terms of the matrix elements. However, we can also treat a matrix as a single algebraic operator, as in Eq. (3.40). Matrix elements and single operators each have their advantages, as will be seen in the following section. We shall use both approaches.

least one of the matrices must have a zero determinant (that is, be singular as defined after Eq. (3.50) in this section).

If A is an $n \times n$ matrix with determinant $|\mathsf{A}| \neq 0$, then it has a unique inverse A^{-1} satisfying $\mathsf{A}\mathsf{A}^{-1} = \mathsf{A}^{-1}\mathsf{A} = 1$. If B is also an $n \times n$ matrix with inverse B^{-1}, then the product AB has the inverse

$$(\mathsf{AB})^{-1} = \mathsf{B}^{-1}\mathsf{A}^{-1} \tag{3.42}$$

because $\mathsf{ABB}^{-1}\mathsf{A}^{-1} = 1 = \mathsf{B}^{-1}\mathsf{A}^{-1}\mathsf{AB}$ (see also Exercises 3.2.31 and 3.2.32).

The **product theorem**, which says that the determinant of the product, $|\mathsf{AB}|$, of two $n \times n$ matrices A and B is equal to the product of the determinants, $|\mathsf{A}||\mathsf{B}|$, links matrices with determinants. To prove this, consider the n column vectors $\mathbf{c}_k = (\sum_j a_{ij}b_{jk}, i = 1, 2, \ldots, n)$ of the product matrix $\mathsf{C} = \mathsf{AB}$ for $k = 1, 2, \ldots, n$. Each $\mathbf{c}_k = \sum_{j_k} b_{j_k k}\mathbf{a}_{j_k}$ is a sum of n column vectors $\mathbf{a}_{j_k} = (a_{ij_k}, i = 1, 2, \ldots, n)$. Note that we are now using a different product summation index j_k for each column $\mathbf{c}_k$. Since any determinant $D(b_1\mathbf{a}_1 + b_2\mathbf{a}_2) = b_1 D(\mathbf{a}_1) + b_2 D(\mathbf{a}_2)$ is linear in its column vectors, we can pull out the summation sign in front of the determinant from each column vector in C together with the common column factor $b_{j_k k}$ so that

$$|\mathsf{C}| = \sum_{j_k's} b_{j_1 1}b_{j_2 2}\cdots b_{j_n n}\det(\mathbf{a}_{j_1}\mathbf{a}_{j_2}, \ldots, \mathbf{a}_{j_n}). \tag{3.43}$$

If we rearrange the column vectors $\mathbf{a}_{j_k}$ of the determinant factor in Eq. (3.43) in the proper order, then we can pull the common factor $\det(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) = |\mathsf{A}|$ in front of the n summation signs in Eq. (3.43). These column permutations generate just the right sign $\varepsilon_{j_1 j_2 \cdots j_n}$ to produce in Eq. (3.43) the expression in Eq. (3.8) for $|\mathsf{B}|$ so

$$|\mathsf{C}| = |\mathsf{A}| \sum_{j_k's} \varepsilon_{j_1 j_2 \cdots j_n} b_{j_1 1}b_{j_2 2}\cdots b_{j_n n} = |\mathsf{A}||\mathsf{B}|, \tag{3.44}$$

which proves the product theorem.

Direct Product

A second procedure for multiplying matrices, known as the **direct** tensor or Kronecker **product**, follows. If A is an $m \times m$ matrix and B is an $n \times n$ matrix, then the direct product is

$$\mathsf{A} \otimes \mathsf{B} = \mathsf{C}. \tag{3.45}$$

C is an $mn \times mn$ matrix with elements

$$C_{\alpha\beta} = A_{ij}B_{kl}, \tag{3.46}$$

with

$$\alpha = m(i-1) + k, \qquad \beta = n(j-1) + l.$$

For instance, if A and B are both 2×2 matrices,

$$\mathsf{A} \otimes \mathsf{B} = \begin{pmatrix} a_{11}\mathsf{B} & a_{12}\mathsf{B} \\ a_{21}\mathsf{B} & a_{22}\mathsf{B} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{pmatrix}. \tag{3.47}$$

The direct product is associative but not commutative. As an example of the direct product, the Dirac matrices of Section 3.4 may be developed as direct products of the Pauli matrices and the unit matrix. Other examples appear in the construction of groups (see Chapter 4) and in vector or Hilbert space in quantum theory.

Example 3.2.1 DIRECT PRODUCT OF VECTORS

The direct product of two two-dimensional vectors is a four-component vector,

$$\begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \otimes \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 y_0 \\ x_0 y_1 \\ x_1 y_0 \\ x_1 y_1 \end{pmatrix};$$

while the direct product of three such vectors,

$$\begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \otimes \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} \otimes \begin{pmatrix} z_0 \\ z_1 \end{pmatrix} = \begin{pmatrix} x_0 y_0 z_0 \\ x_0 y_0 z_1 \\ x_0 y_1 z_0 \\ x_0 y_1 z_1 \\ x_1 y_0 z_0 \\ x_1 y_0 z_1 \\ x_1 y_1 z_0 \\ x_1 y_1 z_1 \end{pmatrix},$$

is a $(2^3 = 8)$-dimensional vector. ■

Diagonal Matrices

An important special type of matrix is the square matrix in which all the nondiagonal elements are zero. Specifically, if a 3×3 matrix A is diagonal, then

$$\mathsf{A} = \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}.$$

A physical interpretation of such diagonal matrices and the method of reducing matrices to this diagonal form are considered in Section 3.5. Here we simply note a significant property of diagonal matrices — multiplication of diagonal matrices is commutative,

$$\mathsf{AB} = \mathsf{BA}, \qquad \text{if A and B are each diagonal.}$$

Multiplication by a diagonal matrix $[d_1, d_2, \ldots, d_n]$ that has only nonzero elements in the diagonal is particularly simple:

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2\cdot 3 & 2\cdot 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 6 & 8 \end{pmatrix};$$

while the opposite order gives

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2\cdot 2 \\ 3 & 2\cdot 4 \end{pmatrix} = \begin{pmatrix} 1 & 4 \\ 3 & 8 \end{pmatrix}.$$

Thus, **a diagonal matrix does not commute with another matrix unless both are diagonal, or the diagonal matrix is proportional to the unit matrix.** This is borne out by the more general form

$$[d_1, d_2, \ldots, d_n]\mathsf{A} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdot \\ 0 & 0 & \cdots & d_n \end{pmatrix}\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

$$= \begin{pmatrix} d_1a_{11} & d_1a_{12} & \cdots & d_1a_{1n} \\ d_2a_{21} & d_2a_{22} & \cdots & d_2a_{2n} \\ \cdots & \cdots & \cdots & \cdot \\ d_na_{n1} & d_na_{n2} & \cdots & d_na_{nn} \end{pmatrix},$$

whereas

$$\mathsf{A}[d_1, d_2, \ldots, d_n] = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdot \\ 0 & 0 & \cdots & d_n \end{pmatrix}$$

$$= \begin{pmatrix} d_1a_{11} & d_2a_{12} & \cdots & d_na_{1n} \\ d_1a_{21} & d_2a_{22} & \cdots & d_na_{2n} \\ \cdots & \cdots & \cdots & \cdot \\ d_1a_{n1} & d_2a_{n2} & \cdots & d_na_{nn} \end{pmatrix}.$$

Here we have denoted by $[d_1, \ldots, d_n]$ a diagonal matrix with diagonal elements $d_1, \ldots, d_n$. In the special case of multiplying two diagonal matrices, we simply multiply the corresponding diagonal matrix elements, which obviously is commutative.

Trace

In any square matrix the sum of the diagonal elements is called the **trace**.

Clearly the trace is a linear operation:

$$\operatorname{trace}(\mathsf{A} - \mathsf{B}) = \operatorname{trace}(\mathsf{A}) - \operatorname{trace}(\mathsf{B}).$$

One of its interesting and useful properties is that the trace of a product of two matrices A and B is independent of the order of multiplication:

$$\begin{aligned}\text{trace}(\mathsf{AB}) &= \sum_i (\mathsf{AB})_{ii} = \sum_i \sum_j a_{ij} b_{ji} \\ &= \sum_j \sum_i b_{ji} a_{ij} = \sum_j (\mathsf{BA})_{jj} \\ &= \text{trace}(\mathsf{BA}).\end{aligned} \tag{3.48}$$

This holds even though $\mathsf{AB} \neq \mathsf{BA}$. Equation (3.48) means that the trace of any commutator $[\mathsf{A}, \mathsf{B}] = \mathsf{AB} - \mathsf{BA}$ is zero. From Eq. (3.48) we obtain

$$\text{trace}(\mathsf{ABC}) = \text{trace}(\mathsf{BCA}) = \text{trace}(\mathsf{CAB}),$$

which shows that the trace is invariant under cyclic permutation of the matrices in a product.

For a real symmetric or a complex Hermitian matrix (see Section 3.4) the trace is the sum, and the determinant the product, of its eigenvalues, and both are coefficients of the characteristic polynomial. In Exercise 3.4.23 the operation of taking the trace selects one term out of a sum of 16 terms. The trace will serve a similar function relative to matrices as orthogonality serves for vectors and functions.

In terms of tensors (Section 2.7) the trace is a contraction and, like the contracted second-rank tensor, is a scalar (invariant).

Matrices are used extensively to represent the elements of groups (compare Exercise 3.2.7 and Chapter 4). The trace of the matrix representing the group element is known in group theory as the **character**. The reason for the special name and special attention is that, the trace or character remains invariant under similarity transformations (compare Exercise 3.3.9).

Matrix Inversion

At the beginning of this section matrix A is introduced as the representation of an operator that (linearly) transforms the coordinate axes. A rotation would be one example of such a linear transformation. Now we look for the inverse transformation A^{-1} that will restore the original coordinate axes. This means, as either a matrix or an operator equation,[10]

$$\mathsf{A}\mathsf{A}^{-1} = \mathsf{A}^{-1}\mathsf{A} = 1. \tag{3.49}$$

With $(\mathsf{A}^{-1})_{ij} \equiv a_{ij}^{(-1)}$,

$$a_{ij}^{(-1)} \equiv \frac{C_{ji}}{|\mathsf{A}|}, \tag{3.50}$$

[10]Here and throughout this chapter our matrices have finite rank. If A is an infinite-rank matrix ($n \times n$ with $n \to \infty$), then life is more difficult. For A^{-1} to be the inverse we must demand that both

$$\mathsf{A}\mathsf{A}^{-1} = 1 \quad \text{and} \quad \mathsf{A}^{-1}\mathsf{A} = 1.$$

one relation no longer implies the other.

with C_{ji} the cofactor (see discussion preceding Eq. (3.11)) of a_{ij} and the assumption that the determinant of A, $|\mathsf{A}| \neq 0$. If it is zero, we label A singular. No inverse exists.

There is a wide variety of alternative techniques. One of the best and most commonly used is the Gauss–Jordan matrix inversion technique. The theory is based on the results of Exercises 3.2.34 and 3.2.35, which show that there exist matrices M_L such that the product $\mathsf{M}_L\mathsf{A}$ will be A but with

a. one row multiplied by a constant, or

b. one row replaced by the original row minus a multiple of another row, or

c. rows interchanged.

Other matrices M_R operating on the right ($\mathsf{A}\mathsf{M}_R$) can carry out the same operations on the **columns** of A.

This means that the matrix rows and columns may be altered (by matrix multiplication) as though we were dealing with determinants, so we can apply the Gauss–Jordan elimination techniques of Section 3.1 to the matrix elements. Hence there exists a matrix M_L (or M_R) such that[11]

$$\mathsf{M}_L\mathsf{A} = 1. \tag{3.51}$$

Then $\mathsf{M}_L = \mathsf{A}^{-1}$. We determine M_L by carrying out the identical elimination operations on the unit matrix. Then

$$\mathsf{M}_L 1 = \mathsf{M}_L. \tag{3.52}$$

To clarify this, we consider a specific example.

Example 3.2.2 GAUSS–JORDAN MATRIX INVERSION

We want to invert the matrix

$$\mathsf{A} = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{pmatrix}. \tag{3.53}$$

For convenience we write A and 1 side by side and carry out the identical operations on each:

$$\begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.54}$$

To be systematic, we multiply each row to get $a_{k1} = 1$,

$$\begin{pmatrix} 1 & \frac{2}{3} & \frac{1}{3} \\ 1 & \frac{3}{2} & \frac{1}{2} \\ 1 & 1 & 4 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.55}$$

[11] Remember that $\det(\mathsf{A}) \neq 0$.

Subtracting the first row from the second and third rows, we obtain

$$\begin{pmatrix} 1 & \frac{2}{3} & \frac{1}{3} \\ 0 & \frac{5}{6} & \frac{1}{6} \\ 0 & \frac{1}{3} & \frac{11}{3} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ -\frac{1}{3} & \frac{1}{2} & 0 \\ -\frac{1}{3} & 0 & 1 \end{pmatrix}. \tag{3.56}$$

Then we divide the second row (of **both** matrices) by $\frac{5}{6}$ and subtract $\frac{2}{3}$ times it from the first row and $\frac{1}{3}$ times it from the third row. The results for both matrices are

$$\begin{pmatrix} 1 & 0 & \frac{1}{5} \\ 0 & 1 & \frac{1}{5} \\ 0 & 0 & \frac{18}{5} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \frac{3}{5} & -\frac{2}{5} & 0 \\ -\frac{2}{5} & \frac{3}{5} & 0 \\ -\frac{1}{5} & -\frac{1}{5} & 1 \end{pmatrix}. \tag{3.57}$$

We divide the third row (of **both** matrices) by $\frac{18}{5}$. Then as the last step $\frac{1}{5}$ times the third row is subtracted from each of the first two rows (of both matrices). Our final pair is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathsf{A}^{-1} = \begin{pmatrix} \frac{11}{18} & -\frac{7}{18} & -\frac{1}{18} \\ -\frac{7}{18} & \frac{11}{18} & -\frac{1}{18} \\ -\frac{1}{18} & -\frac{1}{18} & \frac{5}{18} \end{pmatrix}. \tag{3.58}$$

The check is to multiply the original A by the calculated A^{-1} to see if we really do get the unit matrix 1. ■

As with the Gauss–Jordan solution of simultaneous linear algebraic equations, this technique is well adapted to computers. Indeed, this Gauss–Jordan matrix inversion technique will probably be available in the program library as a subroutine (see Sections 2.3 and 2.4 of Press *et al.*, loc. cit.).

For **matrices of special form, the inverse matrix can be given in closed form**. For example, for

$$\mathsf{A} = \begin{pmatrix} a & b & c \\ b & d & b \\ c & b & e \end{pmatrix}, \tag{3.59}$$

the inverse matrix has a similar but slightly more general form,

$$\mathsf{A}^{-1} = \begin{pmatrix} \alpha & \beta_1 & \gamma \\ \beta_1 & \delta & \beta_2 \\ \gamma & \beta_2 & \epsilon \end{pmatrix}, \tag{3.60}$$

with matrix elements given by

$$D\alpha = ed - b^2, \qquad D\gamma = -(cd - b^2), \qquad D\beta_1 = (c-e)b, \qquad D\beta_2 = (c-a)b,$$

$$D\delta = ae - c^2, \qquad D\epsilon = ad - b^2, \qquad D = b^2(2c - a - e) + d(ae - c^2),$$

where $D = \det(\mathsf{A})$ is the determinant of the matrix A. If $e = a$ in A, then the inverse matrix A^{-1} also simplifies to

$$\beta_1 = \beta_2, \qquad \epsilon = \alpha, \qquad D = (a^2 - c^2)d + 2(c-a)b^2.$$

As a check, let us work out the 11-matrix element of the product $\mathsf{A}\mathsf{A}^{-1} = 1$. We find

$$a\alpha + b\beta_1 + c\gamma = \frac{1}{D}\left[a\left(ed - b^2\right) + b^2(c - e) - c\left(cd - b^2\right)\right]$$
$$= \frac{1}{D}\left(-ab^2 + aed + 2b^2c - b^2e - c^2d\right) = \frac{D}{D} = 1.$$

Similarly we check that the 12-matrix element vanishes,

$$a\beta_1 + b\delta + c\beta_2 = \frac{1}{D}\left[ab(c - e) + b\left(ae - c^2\right) + cb(c - a)\right] = 0,$$

and so on.

Note though that we cannot always find an inverse of A^{-1} by solving for the matrix elements $a, b, \ldots$ of A, because not every inverse matrix A^{-1} of the form in Eq. (3.60) has a corresponding A of the special form in Eq. (3.59), as Example 3.2.2 clearly shows.

Matrices are square or rectangular arrays of numbers that define linear transformations, such as rotations of a coordinate system. As such, they are linear operators. Square matrices may be inverted when their determinant is nonzero. When a matrix defines a system of linear equations, the inverse matrix solves it. Matrices with the same number of rows and columns may be added and subtracted. They form what mathematicians call a ring with a unit and a zero matrix. Matrices are also useful for representing group operations and operators in Hilbert spaces.

Exercises

3.2.1 Show that matrix multiplication is associative, $(\mathsf{A}\mathsf{B})\mathsf{C} = \mathsf{A}(\mathsf{B}\mathsf{C})$.

3.2.2 Show that

$$(\mathsf{A} + \mathsf{B})(\mathsf{A} - \mathsf{B}) = \mathsf{A}^2 - \mathsf{B}^2$$

if and only if A and B commute,

$$[\mathsf{A}, \mathsf{B}] = 0.$$

3.2.3 Show that matrix A is a **linear operator** by showing that

$$\mathsf{A}(c_1\mathbf{r}_1 + c_2\mathbf{r}_2) = c_1\mathsf{A}\mathbf{r}_1 + c_2\mathsf{A}\mathbf{r}_2.$$

It can be shown that an $n \times n$ matrix is the **most general** linear operator in an n-dimensional vector space. This means that every linear operator in this n-dimensional vector space is equivalent to a matrix.

3.2.4 (a) Complex numbers, $a + ib$, with a and b real, may be represented by (or are isomorphic with) 2×2 matrices:

$$a + ib \leftrightarrow \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Show that this matrix representation is valid for (i) addition and (ii) multiplication.

(b) Find the matrix corresponding to $(a + ib)^{-1}$.

3.2.5 If A is an $n \times n$ matrix, show that

$$\det(-\mathsf{A}) = (-1)^n \det \mathsf{A}.$$

3.2.6 (a) The matrix equation $\mathsf{A}^2 = 0$ does not imply $\mathsf{A} = 0$. Show that the most general 2×2 matrix whose square is zero may be written as

$$\begin{pmatrix} ab & b^2 \\ -a^2 & -ab \end{pmatrix},$$

where a and b are real or complex numbers.

(b) If $\mathsf{C} = \mathsf{A} + \mathsf{B}$, in general

$$\det \mathsf{C} \neq \det \mathsf{A} + \det \mathsf{B}.$$

Construct a specific numerical example to illustrate this inequality.

3.2.7 Given the three matrices

$$\mathsf{A} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \mathsf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \mathsf{C} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix},$$

find all possible products of A, B, and C, two at a time, including squares. Express your answers in terms of A, B, and C, and 1, the unit matrix. These three matrices, together with the unit matrix, form a representation of a mathematical group, the **vierergruppe** (see Chapter 4).

3.2.8 Given

$$\mathsf{K} = \begin{pmatrix} 0 & 0 & i \\ -i & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix},$$

show that

$$\mathsf{K}^n = \mathsf{KKK} \cdots (n \text{ factors}) = 1$$

(with the proper choice of n, $n \neq 0$).

3.2.9 Verify the **Jacobi identity**,

$$[\mathsf{A}, [\mathsf{B}, \mathsf{C}]] = [\mathsf{B}, [\mathsf{A}, \mathsf{C}]] - [\mathsf{C}, [\mathsf{A}, \mathsf{B}]].$$

This is useful in matrix descriptions of elementary particles (see Eq. (4.16)). As a mnemonic aid, the you might note that the Jacobi identity has the same form as the *BAC–CAB* rule of Section 1.5.

3.2.10 Show that the matrices

$$\mathsf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathsf{B} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathsf{C} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

satisfy the commutation relations

$$[\mathsf{A}, \mathsf{B}] = \mathsf{C}, \qquad [\mathsf{A}, \mathsf{C}] = 0, \qquad \text{and} \qquad [\mathsf{B}, \mathsf{C}] = 0.$$

3.2.11 Let

$$\mathsf{i} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \qquad \mathsf{j} = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

and

$$\mathsf{k} = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}.$$

Show that

(a) $\mathsf{i}^2 = \mathsf{j}^2 = \mathsf{k}^2 = -1$, where 1 is the unit matrix.
(b) $\mathsf{ij} = -\mathsf{ji} = \mathsf{k}$,
$\mathsf{jk} = -\mathsf{kj} = \mathsf{i}$,
$\mathsf{ki} = -\mathsf{ik} = \mathsf{j}$.

These three matrices (i, j, and k) plus the unit matrix 1 form a basis for **quaternions**. An alternate basis is provided by the four 2×2 matrices, $i\sigma_1$, $i\sigma_2$, $-i\sigma_3$, and 1, where the σ are the Pauli spin matrices of Exercise 3.2.13.

3.2.12 A matrix with elements $a_{ij} = 0$ for $j < i$ may be called upper right triangular. The elements in the lower left (below and to the left of the main diagonal) vanish. Examples are the matrices in Chapters 12 and 13, Exercise 13.1.21, relating power series and eigenfunction expansions.
Show that the product of two upper right triangular matrices is an upper right triangular matrix.

3.2.13 The three Pauli spin matrices are

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \text{and} \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Show that

(a) $(\sigma_i)^2 = 1_2$,
(b) $\sigma_j \sigma_k = i\sigma_l$, $(j, k, l) = (1, 2, 3), (2, 3, 1), (3, 1, 2)$ (cyclic permutation),
(c) $\sigma_i \sigma_j + \sigma_j \sigma_i = 2\delta_{ij} 1_2$; 1_2 is the 2×2 unit matrix.

These matrices were used by Pauli in the nonrelativistic theory of electron spin.

3.2.14 Using the Pauli σ_i of Exercise 3.2.13, show that

$$(\boldsymbol{\sigma} \cdot \mathbf{a})(\boldsymbol{\sigma} \cdot \mathbf{b}) = \mathbf{a} \cdot \mathbf{b}\, 1_2 + i\boldsymbol{\sigma} \cdot (\mathbf{a} \times \mathbf{b}).$$

Here

$$\boldsymbol{\sigma} \equiv \hat{\mathbf{x}}\sigma_1 + \hat{\mathbf{y}}\sigma_2 + \hat{\mathbf{z}}\sigma_3,$$

$\mathbf{a}$ and $\mathbf{b}$ are ordinary vectors, and 1_2 is the 2×2 unit matrix.

3.2.15 One description of spin 1 particles uses the matrices

$$\mathsf{M}_x = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \mathsf{M}_y = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix},$$

and

$$\mathsf{M}_z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Show that

(a) $[\mathsf{M}_x, \mathsf{M}_y] = i\mathsf{M}_z$, and so on[12] (cyclic permutation of indices). Using the Levi-Civita symbol of Section 2.9, we may write

$$[\mathsf{M}_p, \mathsf{M}_q] = i\varepsilon_{pqr}\mathsf{M}_r.$$

(b) $\mathsf{M}^2 \equiv \mathsf{M}_x^2 + \mathsf{M}_y^2 + \mathsf{M}_z^2 = 2\,1_3$, where 1_3 is the 3×3 unit matrix.

(c) $[\mathsf{M}^2, \mathsf{M}_i] = 0$,
$[\mathsf{M}_z, \mathsf{L}^+] = \mathsf{L}^+$,
$[\mathsf{L}^+, \mathsf{L}^-] = 2\mathsf{M}_z$,
where
$\mathsf{L}^+ \equiv \mathsf{M}_x + i\mathsf{M}_y$,
$\mathsf{L}^- \equiv \mathsf{M}_x - i\mathsf{M}_y$.

3.2.16 Repeat Exercise 3.2.15 using an alternate representation,

$$\mathsf{M}_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \qquad \mathsf{M}_y = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix},$$

and

$$\mathsf{M}_z = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

In Chapter 4 these matrices appear as the **generators** of the rotation group.

3.2.17 Show that the matrix–vector equation

$$\left(\mathsf{M} \cdot \nabla + 1_3 \frac{1}{c}\frac{\partial}{\partial t}\right)\psi = 0$$

reproduces Maxwell's equations in vacuum. Here ψ is a column vector with components $\psi_j = B_j - iE_j/c$, $j = x, y, z$. M is a vector whose elements are the angular momentum **matrices** of Exercise 3.2.16. Note that $\varepsilon_0\mu_0 = 1/c^2$, 1_3 is the 3×3 unit matrix.

[12] $[\mathsf{A}, \mathsf{B}] = \mathsf{AB} - \mathsf{BA}$.

From Exercise 3.2.15(b),

$$\mathsf{M}^2\psi = 2\psi.$$

A comparison with the Dirac relativistic electron equation suggests that the "particle" of electromagnetic radiation, the photon, has zero rest mass and a spin of 1 (in units of h).

3.2.18 Repeat Exercise 3.2.15, using the matrices for a spin of 3/2,

$$\mathsf{M}_x = \frac{1}{2}\begin{pmatrix} 0 & \sqrt{3} & 0 & 0 \\ \sqrt{3} & 0 & 2 & 0 \\ 0 & 2 & 0 & \sqrt{3} \\ 0 & 0 & \sqrt{3} & 0 \end{pmatrix}, \qquad \mathsf{M}_y = \frac{i}{2}\begin{pmatrix} 0 & -\sqrt{3} & 0 & 0 \\ \sqrt{3} & 0 & -2 & 0 \\ 0 & 2 & 0 & -\sqrt{3} \\ 0 & 0 & \sqrt{3} & 0 \end{pmatrix},$$

and

$$\mathsf{M}_z = \frac{1}{2}\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -3 \end{pmatrix}.$$

3.2.19 An operator P commutes with J_x and J_y, the x and y components of an angular momentum operator. Show that P commutes with the third component of angular momentum, that is, that

$$[\mathsf{P}, \mathsf{J}_z] = 0.$$

Hint. The angular momentum components must satisfy the commutation relation of Exercise 3.2.15(a).

3.2.20 The L^+ and L^- matrices of Exercise 3.2.15 are *ladder operators* (see Chapter 4): L^+ operating on a system of spin projection m will raise the spin projection to $m+1$ if m is below its maximum. L^+ operating on $m_{\max}$ yields zero. L^- reduces the spin projection in unit steps in a similar fashion. Dividing by $\sqrt{2}$, we have

$$\mathsf{L}^+ = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathsf{L}^- = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Show that

$$\mathsf{L}^+|-1\rangle = |0\rangle, \mathsf{L}^-|-1\rangle = \text{null column vector},$$
$$\mathsf{L}^+|0\rangle = |1\rangle, \mathsf{L}^-|0\rangle = |-1\rangle,$$
$$\mathsf{L}^+|1\rangle = \text{null column vector}, \ \mathsf{L}^-|1\rangle = |0\rangle,$$

where

$$|-1\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \qquad |0\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad \text{and} \qquad |1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

represent states of spin projection -1, 0, and 1, respectively.
Note. Differential operator analogs of these ladder operators appear in Exercise 12.6.7.

3.2.21 Vectors **A** and **B** are related by the tensor T,

$$\mathbf{B} = \mathsf{T}\mathbf{A}.$$

Given **A** and **B**, show that there is **no unique solution** for the components of T. This is why vector division **B**/**A** is undefined (apart from the special case of **A** and **B** parallel and T then a scalar).

3.2.22 We might ask for a vector $\mathbf{A}^{-1}$, an inverse of a given vector **A** in the sense that

$$\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{A}^{-1} \cdot \mathbf{A} = 1.$$

Show that this relation does not suffice to define $\mathbf{A}^{-1}$ uniquely; **A** would then have an infinite number of inverses.

3.2.23 If A is a diagonal matrix, with all diagonal elements different, and A and B commute, show that B is diagonal.

3.2.24 If A and B are diagonal, show that A and B commute.

3.2.25 Show that trace(ABC) = trace(CBA) if any two of the three matrices commute.

3.2.26 Angular momentum matrices satisfy a commutation relation

$$[\mathsf{M}_j, \mathsf{M}_k] = i\mathsf{M}_l, \qquad j, k, l \text{ cyclic}.$$

Show that the trace of each angular momentum matrix vanishes.

3.2.27 (a) The operator trace replaces a matrix A by its trace; that is,

$$\text{trace}(\mathsf{A}) = \sum_i a_{ii}.$$

Show that trace is a **linear** operator.

(b) The operator det replaces a matrix A by its determinant; that is,

$$\det(\mathsf{A}) = \text{determinant of } \mathsf{A}.$$

Show that det is **not** a linear operator.

3.2.28 A and B anticommute: AB = −AB. Also, $\mathsf{A}^2 = 1$, $\mathsf{B}^2 = 1$. Show that trace(A) = trace(B) = 0.
Note. The Pauli and Dirac (Section 3.4) matrices are specific examples.

3.2.29 With $|x\rangle$ an N-dimensional column vector and $\langle y|$ an N-dimensional row vector, show that

$$\text{trace}\big(|x\rangle\langle y|\big) = \langle y|x\rangle.$$

Note. $|x\rangle\langle y|$ means direct product of column vector $|x\rangle$ with row vector $\langle y|$. The result is a square $N \times N$ matrix.

3.2.30 (a) If two nonsingular matrices anticommute, show that the trace of each one is zero. (*Nonsingular* means that the determinant of the matrix nonzero.)

(b) For the conditions of part (a) to hold, A and B must be $n \times n$ matrices with n **even**. Show that if n is **odd**, a contradiction results.

3.2.31 If a matrix has an inverse, show that the inverse is unique.

3.2.32 If A^{-1} has elements

$$(\mathsf{A}^{-1})_{ij} = a_{ij}^{(-1)} = \frac{C_{ji}}{|\mathsf{A}|},$$

where C_{ji} is the jith cofactor of $|\mathsf{A}|$, show that

$$\mathsf{A}^{-1}\mathsf{A} = 1.$$

Hence A^{-1} is the inverse of A (if $|\mathsf{A}| \neq 0$).

3.2.33 Show that $\det \mathsf{A}^{-1} = (\det \mathsf{A})^{-1}$.
Hint. Apply the product theorem of Section 3.2.
Note. If $\det \mathsf{A}$ is zero, then A has no inverse. A is singular.

3.2.34 Find the matrices M_L such that the product $\mathsf{M}_L\mathsf{A}$ will be A but with:

(a) the ith row multiplied by a constant k ($a_{ij} \to ka_{ij}$, $j = 1, 2, 3, \ldots$);
(b) the ith row replaced by the original ith row minus a multiple of the mth row ($a_{ij} \to a_{ij} - Ka_{mj}$, $i = 1, 2, 3, \ldots$);
(c) the ith and mth rows interchanged ($a_{ij} \to a_{mj}$, $a_{mj} \to a_{ij}$, $j = 1, 2, 3, \ldots$).

3.2.35 Find the matrices M_R such that the product $\mathsf{A}\mathsf{M}_R$ will be A but with:

(a) the ith column multiplied by a constant k ($a_{ji} \to ka_{ji}$, $j = 1, 2, 3, \ldots$);
(b) the ith column replaced by the original ith column minus a multiple of the mth column ($a_{ji} \to a_{ji} - ka_{jm}$, $j = 1, 2, 3, \ldots$);
(c) the ith and mth columns interchanged ($a_{ji} \to a_{jm}$, $a_{jm} \to a_{ji}$, $j = 1, 2, 3, \ldots$).

3.2.36 Find the inverse of

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

3.2.37 (a) Rewrite Eq. (2.4) of Chapter 2 (and the corresponding equations for dy and dz) as a single matrix equation

$$|dx_k\rangle = \mathsf{J}|dq_j\rangle.$$

J is a matrix of derivatives, the **Jacobian** matrix. Show that

$$\langle dx_k|dx_k\rangle = \langle dq_i|\mathsf{G}|dq_j\rangle,$$

with the metric (matrix) G having elements g_{ij} given by Eq. (2.6).

(b) Show that

$$\det(\mathsf{J})\, dq_1\, dq_2\, dq_3 = dx\, dy\, dz,$$

with $\det(\mathsf{J})$ the usual Jacobian.

3.2.38 Matrices are far too useful to remain the exclusive property of physicists. They may appear wherever there are linear relations. For instance, in a study of population movement the initial fraction of a fixed population in each of n areas (or industries or religions, etc.) is represented by an n-component column vector $\mathbf{P}$. The movement of people from one area to another in a given time is described by an $n \times n$ (stochastic) matrix T. Here T_{ij} is the fraction of the population in the jth area that moves to the ith area. (Those not moving are covered by $i = j$.) With $\mathbf{P}$ describing the initial population distribution, the final population distribution is given by the matrix equation $\mathsf{T}\mathbf{P} = \mathbf{Q}$.
From its definition, $\sum_{i=1}^{n} P_i = 1$.

(a) Show that conservation of people requires that

$$\sum_{i=1}^{n} T_{ij} = 1, \quad j = 1, 2, \ldots, n.$$

(b) Prove that

$$\sum_{i=1}^{n} Q_i = 1$$

continues the conservation of people.

3.2.39 Given a 6×6 matrix A with elements $a_{ij} = 0.5^{|i-j|}$, $i = 0, 1, 2, \ldots, 5$; $i = 0, 1, 2, \ldots, 5$, find A^{-1}. List its matrix elements to five decimal places.

ANS. $$\mathsf{A}^{-1} = \frac{1}{3}\begin{pmatrix} 4 & -2 & 0 & 0 & 0 & 0 \\ -2 & 5 & -2 & 0 & 0 & 0 \\ 0 & -2 & 5 & -2 & 0 & 0 \\ 0 & 0 & -2 & 5 & -2 & 0 \\ 0 & 0 & 0 & -2 & 5 & -2 \\ 0 & 0 & 0 & 0 & -2 & 4 \end{pmatrix}.$$

3.2.40 Exercise 3.1.7 may be written in matrix form:

$$\mathsf{A}\mathbf{X} = \mathbf{C}.$$

Find A^{-1} and calculate $\mathbf{X}$ as $\mathsf{A}^{-1}\mathbf{C}$.

3.2.41 (a) Write a **subroutine** that will multiply **complex** matrices. Assume that the complex matrices are in a general rectangular form.
(b) Test your subroutine by multiplying pairs of the Dirac 4×4 matrices, Section 3.4.

3.2.42 (a) Write a subroutine that will call the complex matrix multiplication subroutine of Exercise 3.2.41 and will calculate the commutator bracket of two complex matrices.
(b) Test your complex commutator bracket subroutine with the matrices of Exercise 3.2.16.

3.2.43 *Interpolating polynomial* is the name given to the $(n-1)$-degree polynomial determined by (and passing through) n points, (x_i, y_i) with all the x_i distinct. This interpolating polynomial forms a basis for numerical quadratures.

(a) Show that the requirement that an $(n-1)$-degree polynomial in x pass through each of the n points (x_i, y_i) with all x_i distinct leads to n simultaneous equations of the form

$$\sum_{j=0}^{n-1} a_j x_i^j = y_i, \qquad i = 1, 2, \ldots, n.$$

(b) Write a computer program that will read in n data points and return the n coefficients a_j. Use a subroutine to solve the simultaneous equations if such a subroutine is available.

(c) Rewrite the set of simultaneous equations as a matrix equation

$$\mathbf{XA} = \mathbf{Y}.$$

(d) Repeat the computer calculation of part (b), but this time solve for vector $\mathbf{A}$ by inverting matrix X (again, using a subroutine).

3.2.44 A calculation of the values of electrostatic potential inside a cylinder leads to

$$V(0.0) = 52.640 \quad V(0.6) = 25.844$$
$$V(0.2) = 48.292 \quad V(0.8) = 12.648$$
$$V(0.4) = 38.270 \quad V(1.0) = 0.0.$$

The problem is to determine the values of the argument for which $V = 10, 20, 30, 40,$ and 50. Express $V(x)$ as a series $\sum_{n=0}^{5} a_{2n} x^{2n}$. (Symmetry requirements in the original problem require that $V(x)$ be an even function of x.) Determine the coefficients a_{2n}. With $V(x)$ now a known function of x, find the root of $V(x) - 10 = 0$, $0 \le x \le 1$. Repeat for $V(x) - 20$, and so on.

ANS. $a_0 = 52.640,$
$a_2 = -117.676,$
$V(0.6851) = 20.$

3.3 ORTHOGONAL MATRICES

Ordinary three-dimensional space may be described with the Cartesian coordinates (x_1, x_2, x_3). We consider a second set of Cartesian coordinates (x_1', x_2', x_3'), whose origin and handedness coincides with that of the first set but whose orientation is different (Fig. 3.1). We can say that the primed coordinate **axes** have been **rotated** relative to the initial, unprimed coordinate axes. Since this rotation is a **linear** operation, we expect a matrix equation relating the primed basis to the unprimed basis.

This section repeats portions of Chapters 1 and 2 in a slightly different context and with a different emphasis. Previously, attention was focused on the vector or tensor. In the case of the tensor, transformation properties were strongly stressed and were critical. Here emphasis is placed on the description of the coordinate rotation itself—the matrix. Transformation properties, the behavior of the matrix when the basis is changed, appear at the end of this section. Sections 3.4 and 3.5 continue with transformation properties in complex vector spaces.

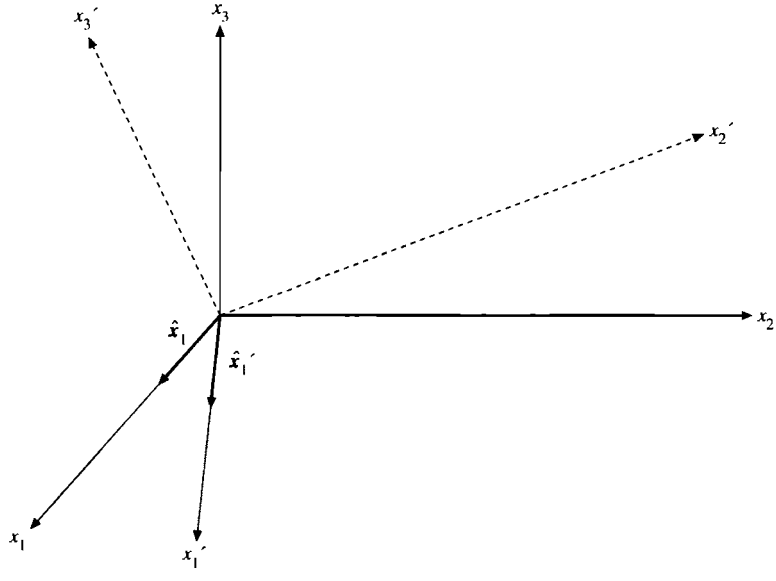

FIGURE 3.1 Cartesian coordinate systems.

Direction Cosines

A unit vector along the x_1'-axis ($\hat{\mathbf{x}}_1'$) may be resolved into components along the x_1-, x_2-, and x_3-axes by the usual projection technique:

$$\hat{\mathbf{x}}_1' = \hat{\mathbf{x}}_1 \cos(x_1', x_1) + \hat{\mathbf{x}}_2 \cos(x_1', x_2) + \hat{\mathbf{x}}_3 \cos(x_1', x_3). \tag{3.61}$$

Equation (3.61) is a specific example of the linear relations discussed at the beginning of Section 3.2.

For convenience these cosines, which are the direction cosines, are labeled

$$\begin{aligned} \cos(x_1', x_1) &= \hat{\mathbf{x}}_1' \cdot \hat{\mathbf{x}}_1 = a_{11}, \\ \cos(x_1', x_2) &= \hat{\mathbf{x}}_1' \cdot \hat{\mathbf{x}}_2 = a_{12}, \\ \cos(x_1', x_3) &= \hat{\mathbf{x}}_1' \cdot \hat{\mathbf{x}}_3 = a_{13}. \end{aligned} \tag{3.62a}$$

Continuing, we have

$$\begin{aligned} \cos(x_2', x_1) &= \hat{\mathbf{x}}_2' \cdot \hat{\mathbf{x}}_1 = a_{21}, \\ \cos(x_2', x_2) &= \hat{\mathbf{x}}_2' \cdot \hat{\mathbf{x}}_2 = a_{22}, \end{aligned} \tag{3.62b}$$

and so on, where $a_{21} \neq a_{12}$ in general. Now, Eq. (3.62) may be rewritten

$$\hat{\mathbf{x}}_1' = \hat{\mathbf{x}}_1 a_{11} + \hat{\mathbf{x}}_2 a_{12} + \hat{\mathbf{x}}_3 a_{13}, \tag{3.62c}$$

and also

$$\begin{aligned} \hat{\mathbf{x}}_2' &= \hat{\mathbf{x}}_1 a_{21} + \hat{\mathbf{x}}_2 a_{22} + \hat{\mathbf{x}}_3 a_{23}, \\ \hat{\mathbf{x}}_3' &= \hat{\mathbf{x}}_1 a_{31} + \hat{\mathbf{x}}_2 a_{32} + \hat{\mathbf{x}}_3 a_{33}. \end{aligned} \tag{3.62d}$$

We may also go the other way by resolving $\hat{\mathbf{x}}_1$, $\hat{\mathbf{x}}_2$, and $\hat{\mathbf{x}}_3$ into components in the primed system. Then

$$\begin{aligned}\hat{\mathbf{x}}_1 &= \hat{\mathbf{x}}_1' a_{11} + \hat{\mathbf{x}}_2' a_{21} + \hat{\mathbf{x}}_3' a_{31},\\ \hat{\mathbf{x}}_2 &= \hat{\mathbf{x}}_1' a_{12} + \hat{\mathbf{x}}_2' a_{22} + \hat{\mathbf{x}}_3' a_{32},\\ \hat{\mathbf{x}}_3 &= \hat{\mathbf{x}}_1' a_{13} + \hat{\mathbf{x}}_2' a_{23} + \hat{\mathbf{x}}_3' a_{33}.\end{aligned} \tag{3.63}$$

Associating $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_1'$ with the subscript 1, $\hat{\mathbf{x}}_2$ and $\hat{\mathbf{x}}_2'$ with the subscript 2, $\hat{\mathbf{x}}_3$ and $\hat{\mathbf{x}}_3'$ with the subscript 3, we see that in each case the first subscript of a_{ij} refers to the primed unit vector $(\hat{\mathbf{x}}_1', \hat{\mathbf{x}}_2', \hat{\mathbf{x}}_3')$, whereas the second subscript refers to the unprimed unit vector $(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \hat{\mathbf{x}}_3)$.

Applications to Vectors

If we consider a vector whose components are functions of the position in space, then

$$\begin{aligned}\mathbf{V}(x_1, x_2, x_3) &= \hat{\mathbf{x}}_1 V_1 + \hat{\mathbf{x}}_2 V_2 + \hat{\mathbf{x}}_3 V_3,\\ \mathbf{V}'(x_1', x_2', x_3') &= \hat{\mathbf{x}}_1' V_1' + \hat{\mathbf{x}}_2' V_2' + \hat{\mathbf{x}}_3' V_3',\end{aligned} \tag{3.64}$$

since the point may be given both by the coordinates (x_1, x_2, x_3) and by the coordinates (x_1', x_2', x_3'). Note that $\mathbf{V}$ and $\mathbf{V}'$ are geometrically the same vector (but with different components). The coordinate axes are being rotated; the vector stays fixed. Using Eqs. (3.62) to eliminate $\hat{\mathbf{x}}_1$, $\hat{\mathbf{x}}_2$, and $\hat{\mathbf{x}}_3$, we may separate Eq. (3.64) into three scalar equations,

$$\begin{aligned}V_1' &= a_{11} V_1 + a_{12} V_2 + a_{13} V_3,\\ V_2' &= a_{21} V_1 + a_{22} V_2 + a_{23} V_3,\\ V_3' &= a_{31} V_1 + a_{32} V_2 + a_{33} V_3.\end{aligned} \tag{3.65}$$

In particular, these relations will hold for the coordinates of a point (x_1, x_2, x_3) and (x_1', x_2', x_3'), giving

$$\begin{aligned}x_1' &= a_{11} x_1 + a_{12} x_2 + a_{13} x_3,\\ x_2' &= a_{21} x_1 + a_{22} x_2 + a_{23} x_3,\\ x_3' &= a_{31} x_1 + a_{32} x_2 + a_{33} x_3,\end{aligned} \tag{3.66}$$

and similarly for the primed coordinates. In this notation the set of three equations (3.66) may be written as

$$x_i' = \sum_{j=1}^{3} a_{ij} x_j, \tag{3.67}$$

where i takes on the values 1, 2, and 3 and the result is three **separate** equations.

Now let us set aside these results and try a different approach to the same problem. We consider two coordinate systems (x_1, x_2, x_3) and (x_1', x_2', x_3') with a common origin and one point (x_1, x_2, x_3) in the unprimed system, (x_1', x_2', x_3') in the primed system. Note the usual ambiguity. The same symbol x denotes both the coordinate axis and a particular

distance along that axis. Since our system is linear, x_i' must be a linear combination of the x_i. Let

$$x_i' = \sum_{j=1}^{3} a_{ij} x_j. \tag{3.68}$$

The a_{ij} may be identified as the direction cosines. This identification is carried out for the two-dimensional case later.

If we have two sets of quantities (V_1, V_2, V_3) in the unprimed system and (V_1', V_2', V_3') in the primed system, related in the same way as the coordinates of a point in the two different systems (Eq. (3.68)),

$$V_i' = \sum_{j=1}^{3} a_{ij} V_j, \tag{3.69}$$

then, as in Section 1.2, the quantities (V_1, V_2, V_3) are defined as the components of a vector that stays fixed while the coordinates rotate; that is, a vector is defined in terms of transformation properties of its components under a rotation of the coordinate axes. In a sense the coordinates of a point have been taken as a prototype vector. The power and usefulness of this definition became apparent in Chapter 2, in which it was extended to define pseudovectors and tensors.

From Eq. (3.67) we can derive interesting information about the a_{ij} that describe the orientation of coordinate system (x_1', x_2', x_3') relative to the system (x_1, x_2, x_3). The length from the origin to the point is the same in both systems. Squaring, for convenience,[13]

$$\sum_i x_i^2 = \sum_i x_i'^2 = \sum_i \left(\sum_j a_{ij} x_j \right) \left(\sum_k a_{ik} x_k \right)$$

$$= \sum_{j,k} x_j x_k \sum_i a_{ij} a_{ik}. \tag{3.70}$$

This can be true for all points if and only if

$$\sum_i a_{ij} a_{ik} = \delta_{jk}, \qquad j, k = 1, 2, 3. \tag{3.71}$$

Note that Eq. (3.71) is equivalent to the matrix equation (3.83); see also Eqs. (3.87a) to (3.87d).

Verification of Eq. (3.71), if needed, may be obtained by returning to Eq. (3.70) and setting $\mathbf{r} = (x_1, x_2, x_3) = (1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 0)$, and so on to evaluate the nine relations given by Eq. (3.71). This process is valid, since Eq. (3.70) must hold for all $\mathbf{r}$ for a given set of a_{ij}. Equation (3.71), a consequence of requiring that the length remain constant (invariant) under rotation of the coordinate system, is called the **orthogonality condition**. The a_{ij}, written as a matrix A subject to Eq. (3.71), form an orthogonal matrix, a first definition of an orthogonal matrix. Note that Eq. (3.71) is **not** matrix multiplication. Rather, it is interpreted later as a scalar product of two columns of A.

[13] Note that **two** independent indices j and k are used.

In matrix notation Eq. (3.67) becomes

$$|x'\rangle = \mathsf{A}|x\rangle. \tag{3.72}$$

Orthogonality Conditions — Two-Dimensional Case

A better understanding of the a_{ij} and the orthogonality condition may be gained by considering rotation in two dimensions in detail. (This can be thought of as a three-dimensional system with the x_1-, x_2-axes rotated about x_3.) From Fig. 3.2,

$$\begin{aligned} x_1' &= x_1\cos\varphi + x_2\sin\varphi, \\ x_2' &= -x_1\sin\varphi + x_2\cos\varphi. \end{aligned} \tag{3.73}$$

Therefore by Eq. (3.72)

$$\mathsf{A} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}. \tag{3.74}$$

Notice that A reduces to the unit matrix for $\varphi = 0$. Zero angle rotation means nothing has changed. It is clear from Fig. 3.2 that

$$\begin{aligned} a_{11} &= \cos\varphi = \cos(x_1', x_1), \\ a_{12} &= \sin\varphi = \cos\left(\tfrac{\pi}{2} - \varphi\right) = \cos(x_1', x_2), \end{aligned} \tag{3.75}$$

and so on, thus identifying the matrix elements a_{ij} as the direction cosines. Equation (3.71), the orthogonality condition, becomes

$$\begin{aligned} \sin^2\varphi + \cos^2\varphi &= 1, \\ \sin\varphi\cos\varphi - \sin\varphi\cos\varphi &= 0. \end{aligned} \tag{3.76}$$

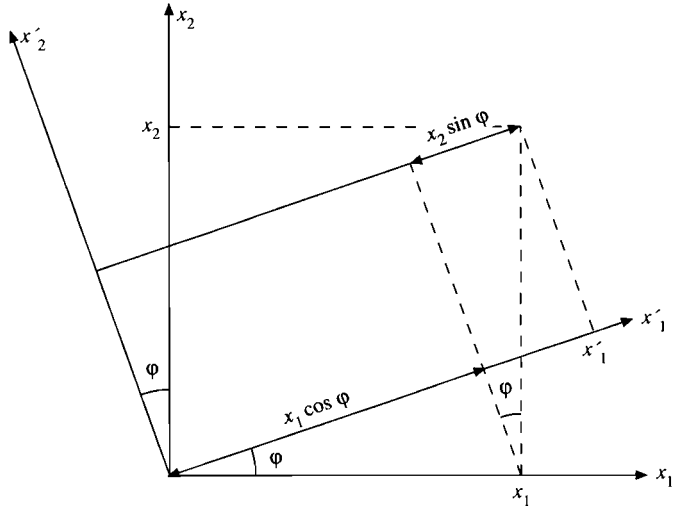

FIGURE 3.2 Rotation of coordinates.

The extension to three dimensions (rotation of the coordinates through an angle φ counterclockwise about x_3) is simply

$$\mathsf{A} = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.77}$$

The $a_{33} = 1$ expresses the fact that $x_3' = x_3$, since the rotation has been about the x_3-axis. The zeros guarantee that x_1' and x_2' do not depend on x_3 and that x_3' does not depend on x_1 and x_2.

Inverse Matrix, A^{-1}

Returning to the general transformation matrix A, the inverse matrix A^{-1} is defined such that

$$|x\rangle = \mathsf{A}^{-1}|x'\rangle. \tag{3.78}$$

That is, A^{-1} describes the reverse of the rotation given by A and returns the coordinate system to its original position. Symbolically, Eqs. (3.72) and (3.78) combine to give

$$|x\rangle = \mathsf{A}^{-1}\mathsf{A}|x\rangle, \tag{3.79}$$

and since $|x\rangle$ is arbitrary,

$$\mathsf{A}^{-1}\mathsf{A} = 1, \tag{3.80}$$

the unit matrix. Similarly,

$$\mathsf{A}\mathsf{A}^{-1} = 1, \tag{3.81}$$

using Eqs. (3.72) and (3.78) and eliminating $|x\rangle$ instead of $|x'\rangle$.

Transpose Matrix, $\tilde{\mathsf{A}}$

We can determine the elements of our postulated inverse matrix A^{-1} by employing the orthogonality condition. Equation (3.71), the orthogonality condition, does not conform to our definition of matrix multiplication, but it can be put in the required form by **defining** a new matrix $\tilde{\mathsf{A}}$ such that

$$\boxed{\tilde{a}_{ji} = a_{ij}.} \tag{3.82}$$

Equation (3.71) becomes

$$\boxed{\tilde{\mathsf{A}}\mathsf{A} = 1.} \tag{3.83}$$

This is a restatement of the orthogonality condition and may be taken as the constraint defining an orthogonal matrix, a second definition of an orthogonal matrix. Multiplying Eq. (3.83) by A^{-1} from the right and using Eq. (3.81), we have

$$\boxed{\tilde{\mathsf{A}} = \mathsf{A}^{-1},} \tag{3.84}$$

a third definition of an orthogonal matrix. This important result, that the inverse equals the transpose, holds only for orthogonal matrices and indeed may be taken as a further restatement of the orthogonality condition.

Multiplying Eq. (3.84) by A from the left, we obtain

$$\mathsf{A}\tilde{\mathsf{A}} = 1 \tag{3.85}$$

or

$$\boxed{\sum_i a_{ji}a_{ki} = \delta_{jk},} \tag{3.86}$$

which is still another form of the orthogonality condition.

Summarizing, the orthogonality condition may be stated in several equivalent ways:

$$\sum_i a_{ij}a_{ik} = \delta_{jk}, \tag{3.87a}$$

$$\sum_i a_{ji}a_{ki} = \delta_{jk}, \tag{3.87b}$$

$$\tilde{\mathsf{A}}\mathsf{A} = \mathsf{A}\tilde{\mathsf{A}} = 1, \tag{3.87c}$$

$$\tilde{\mathsf{A}} = \mathsf{A}^{-1}. \tag{3.87d}$$

Any one of these relations is a necessary and a sufficient condition for A to be orthogonal.

It is now possible to see and understand why the term **orthogonal** is appropriate for these matrices. We have the general form

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

a matrix of direction cosines in which a_{ij} is the cosine of the angle between x_i' and x_j. Therefore a_{11}, a_{12}, a_{13} are the direction cosines of x_1' relative to x_1, x_2, x_3. These three elements of A **define** a unit length along x_1', that is, a unit vector $\hat{\mathbf{x}}_1'$,

$$\hat{\mathbf{x}}_1' = \hat{\mathbf{x}}_1 a_{11} + \hat{\mathbf{x}}_2 a_{12} + \hat{\mathbf{x}}_3 a_{13}.$$

The orthogonality relation (Eq. (3.86)) is simply a statement that the unit vectors $\hat{\mathbf{x}}_1'$, $\hat{\mathbf{x}}_2'$, and $\hat{\mathbf{x}}_3'$ are mutually perpendicular, or orthogonal. Our orthogonal transformation matrix A transforms one orthogonal coordinate system into a second orthogonal coordinate system by rotation and/or reflection.

As an example of the use of matrices, the unit vectors in spherical polar coordinates may be written as

$$\begin{pmatrix} \hat{\mathbf{r}} \\ \hat{\boldsymbol{\theta}} \\ \hat{\boldsymbol{\varphi}} \end{pmatrix} = \mathsf{C} \begin{pmatrix} \hat{\mathbf{x}} \\ \hat{\mathbf{y}} \\ \hat{\mathbf{z}} \end{pmatrix}, \tag{3.88}$$

where C is given in Exercise 2.5.1. This is equivalent to Eqs. (3.62) with $\mathbf{x}_1'$, $\mathbf{x}_2'$, and $\mathbf{x}_3'$ replaced by $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\varphi}}$. From the preceding analysis C is orthogonal. Therefore the inverse relation becomes

$$\begin{pmatrix} \hat{\mathbf{x}} \\ \hat{\mathbf{y}} \\ \hat{\mathbf{z}} \end{pmatrix} = \mathsf{C}^{-1} \begin{pmatrix} \hat{\mathbf{r}} \\ \hat{\boldsymbol{\theta}} \\ \hat{\boldsymbol{\varphi}} \end{pmatrix} = \tilde{\mathsf{C}} \begin{pmatrix} \hat{\mathbf{r}} \\ \hat{\boldsymbol{\theta}} \\ \hat{\boldsymbol{\varphi}} \end{pmatrix}, \tag{3.89}$$

and Exercise 2.5.5 is solved by inspection. Similar applications of matrix inverses appear in connection with the transformation of a power series into a series of orthogonal functions (Gram–Schmidt orthogonalization in Section 10.3) and the numerical solution of integral equations.

Euler Angles

Our transformation matrix A contains nine direction cosines. Clearly, only three of these are independent, Eq. (3.71) providing six constraints. Equivalently, we may say that two parameters (θ and φ in spherical polar coordinates) are required to fix the axis of rotation. Then one additional parameter describes the amount of rotation about the specified axis. (In the Lagrangian formulation of mechanics (Section 17.3) it is necessary to describe A by using some set of three independent parameters rather than the redundant direction cosines.) The usual choice of parameters is the Euler angles.[14]

The goal is to describe the orientation of a final rotated system (x_1''', x_2''', x_3''') relative to some initial coordinate system (x_1, x_2, x_3). The final system is developed in three steps, with each step involving one rotation described by one Euler angle (Fig. 3.3):

1. The coordinates are rotated about the x_3-axis through an angle α counterclockwise into new axes denoted by x_1'-, x_2'-, x_3'. (The x_3- and x_3'-axes coincide.)

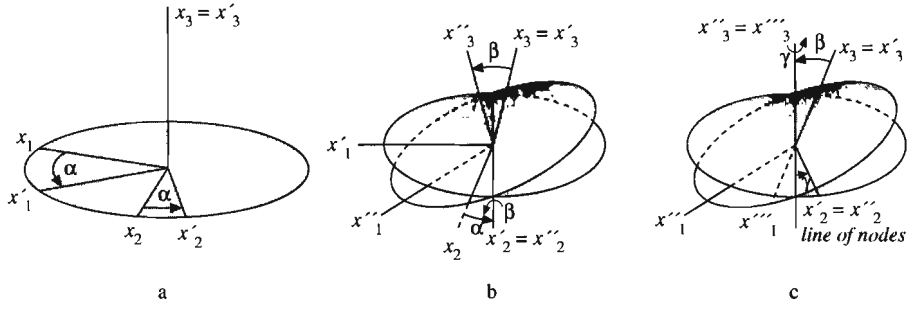

FIGURE 3.3 (a) Rotation about x_3 through angle α; (b) rotation about x_2' through angle β; (c) rotation about x_3'' through angle γ.

[14]There are almost as many definitions of the Euler angles as there are authors. Here we follow the choice generally made by workers in the area of group theory and the quantum theory of angular momentum (compare Sections 4.3, 4.4).

2. The coordinates are rotated about the x_2'-axis[15] through an angle β counterclockwise into new axes denoted by x_1''-, x_2''-, x_3''. (The x_2'- and x_2''-axes coincide.)
3. The third and final rotation is through an angle γ counterclockwise about the x_3''-axis, yielding the x_1''', x_2''', x_3''' system. (The x_3''- and x_3'''-axes coincide.)

The three matrices describing these rotations are

$$\mathsf{R}_z(\alpha) = \begin{pmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{3.90}$$

exactly like Eq. (3.77),

$$\mathsf{R}_y(\beta) = \begin{pmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{pmatrix} \tag{3.91}$$

and

$$\mathsf{R}_z(\gamma) = \begin{pmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.92}$$

The total rotation is described by the triple matrix product,

$$\mathsf{A}(\alpha, \beta, \gamma) = \mathsf{R}_z(\gamma)\mathsf{R}_y(\beta)\mathsf{R}_z(\alpha). \tag{3.93}$$

Note the order: $\mathsf{R}_z(\alpha)$ operates first, then $\mathsf{R}_y(\beta)$, and finally $\mathsf{R}_z(\gamma)$. Direct multiplication gives

$$\mathsf{A}(\alpha, \beta, \gamma) = \begin{pmatrix} \cos\gamma\cos\beta\cos\alpha - \sin\gamma\sin\alpha & \cos\gamma\cos\beta\sin\alpha + \sin\gamma\cos\alpha & -\cos\gamma\sin\beta \\ -\sin\gamma\cos\beta\cos\alpha - \cos\gamma\sin\alpha & -\sin\gamma\cos\beta\sin\alpha + \cos\gamma\cos\alpha & \sin\gamma\sin\beta \\ \sin\beta\cos\alpha & \sin\beta\sin\alpha & \cos\beta \end{pmatrix} \tag{3.94}$$

Equating $\mathsf{A}(a_{ij})$ with $\mathsf{A}(\alpha, \beta, \gamma)$, element by element, yields the direction cosines in terms of the three Euler angles. We could use this Euler angle identification to verify the direction cosine identities, Eq. (1.46) of Section 1.4, but the approach of Exercise 3.3.3 is much more elegant.

Symmetry Properties

Our matrix description leads to the rotation group $\mathsf{SO}(3)$ in three-dimensional space $\mathbb{R}^3$, and the Euler angle description of rotations forms a basis for developing the rotation group in Chapter 4. Rotations may also be described by the unitary group $\mathsf{SU}(2)$ in two-dimensional space $\mathbb{C}^2$ over the complex numbers. The concept of groups such as $\mathsf{SU}(2)$ and its generalizations and group theoretical techniques are often encountered in modern

[15] Some authors choose this second rotation to be about the x_1'-axis.

particle physics, where symmetry properties play an important role. The SU(2) group is also considered in Chapter 4. The power and flexibility of matrices pushed quaternions into obscurity early in the 20th century.[16]

It will be noted that matrices have been handled in two ways in the foregoing discussion: by their components and as single entities. Each technique has its own advantages and both are useful.

The transpose matrix is useful in a discussion of symmetry properties. If

$$\mathsf{A} = \tilde{\mathsf{A}}, \qquad a_{ij} = a_{ji}, \tag{3.95}$$

the matrix is called **symmetric**, whereas if

$$\mathsf{A} = -\tilde{\mathsf{A}}, \qquad a_{ij} = -a_{ji}, \tag{3.96}$$

it is called **antisymmetric** or **skewsymmetric**. The diagonal elements vanish. It is easy to show that any (square) matrix may be written as the sum of a symmetric matrix and an antisymmetric matrix. Consider the identity

$$\mathsf{A} = \tfrac{1}{2}[\mathsf{A} + \tilde{\mathsf{A}}] + \tfrac{1}{2}[\mathsf{A} - \tilde{\mathsf{A}}]. \tag{3.97}$$

$[\mathsf{A} + \tilde{\mathsf{A}}]$ is **clearly symmetric**, whereas $[\mathsf{A} - \tilde{\mathsf{A}}]$ is **clearly antisymmetric**. This is the matrix analog of Eq. (2.75), Chapter 2, for tensors. Similarly, a function may be broken up into its even and odd parts.

So far we have interpreted the orthogonal matrix as rotating the coordinate system. This changes the components of a fixed vector (not rotating with the coordinates) (Fig. 1.6, Chapter 1). However, an orthogonal matrix A may be interpreted equally well as a rotation of the **vector** in the **opposite** direction (Fig. 3.4).

These two possibilities, (1) rotating the vector keeping the coordinates fixed and (2) rotating the coordinates (in the opposite sense) keeping the vector fixed, have a direct analogy in quantum theory. Rotation (a time transformation) of the state vector gives the Schrödinger picture. Rotation of the basis keeping the state vector fixed yields the Heisenberg picture.

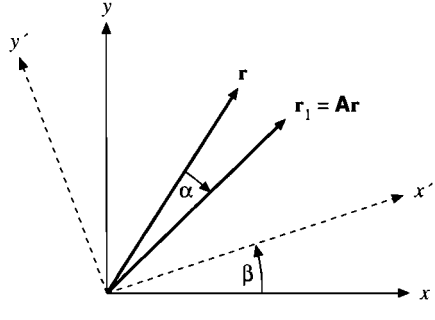

FIGURE 3.4 Fixed coordinates—rotated vector.

[16] R. J. Stephenson, Development of vector analysis from quaternions. *Am. J. Phys.* **34**: 194 (1966).

Suppose we interpret matrix A as rotating a **vector** $\mathbf{r}$ into the position shown by $\mathbf{r}_1$; that is, in a particular coordinate system we have a relation

$$\mathbf{r}_1 = \mathsf{A}\mathbf{r}. \tag{3.98}$$

Now let us rotate the **coordinates** by applying matrix B, which rotates (x, y, z) into (x', y', z'),

$$\begin{aligned}\mathbf{r}_1' &= \mathsf{B}\mathbf{r}_1 = \mathsf{B}\mathsf{A}\mathbf{r} = (\mathsf{A}\mathbf{r})' = \mathsf{B}\mathsf{A}(\mathsf{B}^{-1}\mathsf{B})\mathbf{r} \\ &= (\mathsf{B}\mathsf{A}\mathsf{B}^{-1})\mathsf{B}\mathbf{r} = (\mathsf{B}\mathsf{A}\mathsf{B}^{-1})\mathbf{r}'.\end{aligned} \tag{3.99}$$

$\mathsf{B}\mathbf{r}_1$ is just $\mathbf{r}_1$ in the new coordinate system, with a similar interpretation holding for $\mathsf{B}\mathbf{r}$. Hence **in this new system** $(\mathsf{B}\mathbf{r})$ is rotated into position $(\mathsf{B}\mathbf{r}_1)$ by the matrix $\mathsf{B}\mathsf{A}\mathsf{B}^{-1}$:

$$\begin{array}{ccccc} \mathsf{B}\mathbf{r}_1 & = & (\mathsf{B}\mathsf{A}\mathsf{B}^{-1}) & & \mathsf{B}\mathbf{r} \\ \downarrow & & \downarrow & & \downarrow \\ \mathbf{r}_1' & = & \mathsf{A}' & & \mathbf{r}' \end{array}$$

In the new system the coordinates have been rotated by matrix B; A has the form A', in which

$$\mathsf{A}' = \mathsf{B}\mathsf{A}\mathsf{B}^{-1}. \tag{3.100}$$

A' operates in the x', y', z' space as A operates in the x, y, z space.

The transformation defined by Eq. (3.100) with B any matrix, not necessarily orthogonal, is known as a **similarity transformation**. In component form Eq. (3.100) becomes

$$a_{ij}' = \sum_{k,l} b_{ik} a_{kl} (\mathsf{B}^{-1})_{lj}. \tag{3.101}$$

Now, if B is orthogonal,

$$(\mathsf{B}^{-1})_{lj} = (\tilde{\mathsf{B}})_{lj} = b_{jl}, \tag{3.102}$$

and we have

$$a_{ij}' = \sum_{k,l} b_{ik} b_{jl} a_{kl}. \tag{3.103}$$

It may be helpful to think of A again as an operator, possibly as rotating coordinate axes, relating angular momentum and angular velocity of a rotating solid (Section 3.5). Matrix A is the representation in a given coordinate system — or basis. But there are directions associated with A — crystal axes, symmetry axes in the rotating solid, and so on — so that the representation A depends on the basis. The similarity transformation shows just how the representation changes with a change of basis.

Relation to Tensors

Comparing Eq. (3.103) with the equations of Section 2.6, we see that it is the definition of a tensor of second rank. Hence a matrix that transforms by an **orthogonal** similarity transformation is, by definition, a tensor. Clearly, then, any **orthogonal** matrix A, interpreted as rotating a vector (Eq. (3.98)), may be called a tensor. If, however, we consider the orthogonal matrix as a collection of fixed direction cosines, giving the new orientation of a coordinate system, there is no tensor property involved.

The symmetry and antisymmetry properties defined earlier are preserved under **orthogonal** similarity transformations. Let A be a symmetric matrix, $\mathsf{A} = \tilde{\mathsf{A}}$, and

$$\mathsf{A}' = \mathsf{B}\mathsf{A}\mathsf{B}^{-1}. \tag{3.104}$$

Now,

$$\tilde{\mathsf{A}}' = \tilde{\mathsf{B}}^{-1}\tilde{\mathsf{A}}\tilde{\mathsf{B}} = \mathsf{B}\tilde{\mathsf{A}}\mathsf{B}^{-1}, \tag{3.105}$$

since B is orthogonal. But $\mathsf{A} = \tilde{\mathsf{A}}$. Therefore

$$\tilde{\mathsf{A}}' = \mathsf{B}\mathsf{A}\mathsf{B}^{-1} = \mathsf{A}', \tag{3.106}$$

showing that the property of symmetry is invariant under an orthogonal similarity transformation. In general, symmetry is **not** preserved under a nonorthogonal similarity transformation.

Exercises

Note. Assume all matrix elements are real.

3.3.1 Show that the product of two orthogonal matrices is orthogonal.
Note. This is a key step in showing that all $n \times n$ orthogonal matrices form a group (Section 4.1).

3.3.2 If A is orthogonal, show that its determinant $= \pm 1$.

3.3.3 If A is orthogonal and $\det \mathsf{A} = +1$, show that $(\det \mathsf{A})a_{ij} = C_{ij}$, where C_{ij} is the **cofactor** of a_{ij}. This yields the identities of Eq. (1.46), used in Section 1.4 to show that a cross product of vectors (in three-space) is itself a vector.
Hint. Note Exercise 3.2.32.

3.3.4 Another set of Euler rotations in common use is

(1) a rotation about the x_3-axis through an angle φ, counterclockwise,
(2) a rotation about the x_1'-axis through an angle θ, counterclockwise,
(3) a rotation about the x_3''-axis through an angle ψ, counterclockwise.

If

$$\begin{aligned} \alpha &= \varphi - \pi/2 & \varphi &= \alpha + \pi/2 \\ \beta &= \theta & \theta &= \beta \\ \gamma &= \psi + \pi/2 & \psi &= \gamma - \pi/2, \end{aligned}$$

show that the final systems are identical.

3.3.5 Suppose the Earth is moved (rotated) so that the north pole goes to 30° north, 20° west (original latitude and longitude system) and the 10° west meridian points due south.

(a) What are the Euler angles describing this rotation?
(b) Find the corresponding direction cosines.

$$\text{ANS. (b) } \mathsf{A} = \begin{pmatrix} 0.9551 & -0.2552 & -0.1504 \\ 0.0052 & 0.5221 & -0.8529 \\ 0.2962 & 0.8138 & 0.5000 \end{pmatrix}.$$

3.3.6 Verify that the Euler angle rotation matrix, Eq. (3.94), is invariant under the transformation

$$\alpha \to \alpha + \pi, \qquad \beta \to -\beta, \qquad \gamma \to \gamma - \pi.$$

3.3.7 Show that the Euler angle rotation matrix $\mathsf{A}(\alpha, \beta, \gamma)$ satisfies the following relations:

(a) $\mathsf{A}^{-1}(\alpha, \beta, \gamma) = \tilde{\mathsf{A}}(\alpha, \beta, \gamma)$,
(b) $\mathsf{A}^{-1}(\alpha, \beta, \gamma) = \mathsf{A}(-\gamma, -\beta, -\alpha)$.

3.3.8 Show that the trace of the product of a symmetric and an antisymmetric matrix is zero.

3.3.9 Show that the trace of a matrix remains invariant under similarity transformations.

3.3.10 Show that the determinant of a matrix remains invariant under similarity transformations.
Note. Exercises (3.3.9) and (3.3.10) show that the trace and the determinant are independent of the Cartesian coordinates. They are characteristics of the matrix (operator) itself.

3.3.11 Show that the property of antisymmetry is invariant under orthogonal similarity transformations.

3.3.12 A is 2×2 and orthogonal. Find the most general form of

$$\mathsf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Compare with two-dimensional rotation.

3.3.13 $|\mathbf{x}\rangle$ and $|\mathbf{y}\rangle$ are column vectors. Under an orthogonal transformation S, $|\mathbf{x}'\rangle = \mathsf{S}|\mathbf{x}\rangle$, $|\mathbf{y}'\rangle = \mathsf{S}|\mathbf{y}\rangle$. Show that the scalar product $\langle \mathbf{x} \mid \mathbf{y} \rangle$ is invariant under this orthogonal transformation.
Note. This is equivalent to the invariance of the dot product of two vectors, Section 1.3.

3.3.14 Show that the sum of the squares of the elements of a matrix remains invariant under orthogonal similarity transformations.

3.3.15 As a generalization of Exercise 3.3.14, show that

$$\sum_{jk} S_{jk} T_{jk} = \sum_{l,m} S'_{lm} T'_{lm},$$

where the primed and unprimed elements are related by an orthogonal similarity transformation. This result is useful in deriving invariants in electromagnetic theory (compare Section 4.6).
Note. This product $M_{jk} = \sum S_{jk}T_{jk}$ is sometimes called a *Hadamard product.* In the framework of tensor analysis, Chapter 2, this exercise becomes a double contraction of two second-rank tensors and therefore is clearly a scalar (invariant).

3.3.16 A rotation $\varphi_1 + \varphi_2$ about the z-axis is carried out as two successive rotations φ_1 and φ_2, each about the z-axis. Use the matrix representation of the rotations to derive the trigonometric identities

$$\cos(\varphi_1 + \varphi_2) = \cos\varphi_1 \cos\varphi_2 - \sin\varphi_1 \sin\varphi_2,$$
$$\sin(\varphi_1 + \varphi_2) = \sin\varphi_1 \cos\varphi_2 + \cos\varphi_1 \sin\varphi_2.$$

3.3.17 A column vector $\mathbf{V}$ has components V_1 and V_2 in an initial (unprimed) system. Calculate V_1' and V_2' for a

(a) rotation of the coordinates through an angle of θ **counterclockwise**,
(b) rotation of the vector through an angle of θ **clockwise**.

The results for parts (a) and (b) should be identical.

3.3.18 Write a subroutine that will test whether a real $N \times N$ matrix is symmetric. Symmetry may be defined as

$$0 \leq |a_{ij} - a_{ji}| \leq \varepsilon,$$

where ε is some small tolerance (which allows for truncation error, and so on in the computer).

3.4 Hermitian Matrices, Unitary Matrices

Definitions

Thus far it has generally been assumed that our linear vector space is a real space and that the matrix elements (the representations of the linear operators) are real. For many calculations in classical physics, real matrix elements will suffice. However, in quantum mechanics complex variables are unavoidable because of the form of the basic commutation relations (or the form of the time-dependent Schrödinger equation). With this in mind, we generalize to the case of complex matrix elements. To handle these elements, let us define, or label, some new properties.

1. Complex conjugate, A^*, formed by taking the complex conjugate ($i \to -i$) of each element, where $i = \sqrt{-1}$.
2. Adjoint, $\mathsf{A}^\dagger$, formed by transposing A^*,

$$\boxed{\mathsf{A}^\dagger = \widetilde{\mathsf{A}^*} = \tilde{\mathsf{A}}^*.} \tag{3.107}$$

3. Hermitian matrix: The matrix A is labeled **Hermitian** (or **self-adjoint**) if

$$\mathsf{A} = \mathsf{A}^\dagger. \tag{3.108}$$

If A is real, then $\mathsf{A}^\dagger = \tilde{\mathsf{A}}$ and real Hermitian matrices are real symmetric matrices. In quantum mechanics (or matrix mechanics) matrices are usually constructed to be Hermitian, or unitary.

4. Unitary matrix: Matrix U is labeled **unitary** if

$$\boxed{\mathsf{U}^\dagger = \mathsf{U}^{-1}.} \tag{3.109}$$

If U is real, then $\mathsf{U}^{-1} = \tilde{\mathsf{U}}$, so real unitary matrices are orthogonal matrices. This represents a generalization of the concept of orthogonal matrix (compare Eq. (3.84)).

5. $(\mathsf{AB})^* = \mathsf{A}^*\mathsf{B}^*$, $(\mathsf{AB})^\dagger = \mathsf{B}^\dagger\mathsf{A}^\dagger$.

If the matrix elements are complex, the physicist is almost always concerned with Hermitian and unitary matrices. Unitary matrices are especially important in quantum mechanics because they leave the length of a (complex) vector unchanged — analogous to the operation of an orthogonal matrix on a real vector. It is for this reason that the S matrix of scattering theory is a unitary matrix. One important exception to this interest in unitary matrices is the group of Lorentz matrices, Chapter 4. Using Minkowski space, we see that these matrices are not unitary.

In a complex n-dimensional linear space the square of the length of a point $\tilde{x} = x^T(x_1, x_2, \ldots, x_n)$, or the square of its distance from the origin 0, is defined as $x^\dagger x = \sum x_i^* x_i = \sum |x_i|^2$. If a coordinate transformation $y = \mathsf{U}x$ leaves the distance unchanged, then $x^\dagger x = y^\dagger y = (\mathsf{U}x)^\dagger \mathsf{U}x = x^\dagger \mathsf{U}^\dagger \mathsf{U}x$. Since x is arbitrary it follows that $\mathsf{U}^\dagger\mathsf{U} = 1_n$; that is, U is a unitary $n \times n$ matrix. If $x' = \mathsf{A}x$ is a linear map, then its matrix in the new coordinates becomes the unitary (analog of a similarity) transformation

$$\mathsf{A}' = \mathsf{U}\mathsf{A}\mathsf{U}^\dagger, \tag{3.110}$$

because $\mathsf{U}x' = y' = \mathsf{U}\mathsf{A}x = \mathsf{U}\mathsf{A}\mathsf{U}^{-1}y = \mathsf{U}\mathsf{A}\mathsf{U}^\dagger y$.

Pauli and Dirac Matrices

The set of three 2×2 Pauli matrices $\boldsymbol{\sigma}$,

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{3.111}$$

were introduced by W. Pauli to describe a particle of spin $1/2$ in nonrelativistic quantum mechanics. It can readily be shown that (compare Exercises 3.2.13 and 3.2.14) the Pauli $\boldsymbol{\sigma}$ satisfy

$$\sigma_i\sigma_j + \sigma_j\sigma_i = 2\delta_{ij}1_2, \qquad \text{anticommutation} \tag{3.112}$$

$$\sigma_i\sigma_j = i\sigma_k, \qquad i, j, k \text{ a cyclic permutation of } 1, 2, 3 \tag{3.113}$$

$$(\sigma_i)^2 = 1_2, \tag{3.114}$$

where 1_2 is the 2×2 unit matrix. Thus, the vector $\boldsymbol{\sigma}/2$ satisfies the same commutation relations,

$$[\sigma_i, \sigma_j] \equiv \sigma_i \sigma_j - \sigma_j \sigma_i = 2i\varepsilon_{ijk}\sigma_k, \tag{3.115}$$

as the orbital angular momentum $\mathbf{L}$ ($\mathbf{L} \times \mathbf{L} = i\mathbf{L}$, see Exercise 2.5.15 and the SO(3) and SU(2) groups in Chapter 4).

The three Pauli matrices $\boldsymbol{\sigma}$ and the unit matrix form a complete set, so any Hermitian 2×2 matrix M may be expanded as

$$\mathsf{M} = m_0 1_2 + m_1\sigma_1 + m_2\sigma_2 + m_3\sigma_3 = m_0 + \mathbf{m} \cdot \boldsymbol{\sigma}, \tag{3.116}$$

where the m_i form a constant vector $\mathbf{m}$. Using $(\sigma_i)^2 = 1_2$ and $\text{trace}(\sigma_i) = 0$ we obtain from Eq. (3.116) the expansion coefficients m_i by forming traces,

$$2m_0 = \text{trace}(\mathsf{M}), \qquad 2m_i = \text{trace}(\mathsf{M}\sigma_i), \qquad i = 1, 2, 3. \tag{3.117}$$

Adding and multiplying such 2×2 matrices we generate the Pauli algebra.[17] Note that $\text{trace}(\sigma_i) = 0$ for $i = 1, 2, 3$.

In 1927 P. A. M. Dirac extended this formalism to fast-moving particles of spin $\frac{1}{2}$, such as electrons (and neutrinos). To include special relativity he started from Einstein's energy, $E^2 = \mathbf{p}^2c^2 + m^2c^4$, instead of the nonrelativistic kinetic and potential energy, $E = \mathbf{p}^2/2m + V$. The key to the Dirac equation is to factorize

$$E^2 - \mathbf{p}^2c^2 = E^2 - (c\boldsymbol{\sigma} \cdot \mathbf{p})^2 = (E - c\boldsymbol{\sigma} \cdot \mathbf{p})(E + c\boldsymbol{\sigma} \cdot \mathbf{p}) = m^2c^4 \tag{3.118}$$

using the 2×2 matrix identity

$$(\boldsymbol{\sigma} \cdot \mathbf{p})^2 = \mathbf{p}^2 1_2. \tag{3.119}$$

The 2×2 unit matrix 1_2 is not written explicitly in Eq. (3.118), and Eq. (3.119) follows from Exercise 3.2.14 for $\mathbf{a} = \mathbf{b} = \mathbf{p}$. Equivalently, we can introduce two matrices γ' and γ to factorize $E^2 - \mathbf{p}^2c^2$ directly:

$$\begin{aligned}
&\left[E\gamma' \otimes 1_2 - c(\gamma \otimes \boldsymbol{\sigma}) \cdot \mathbf{p}\right]^2 \\
&\quad = E^2\gamma'^2 \otimes 1_2 + c^2\gamma^2 \otimes (\boldsymbol{\sigma} \cdot \mathbf{p})^2 - Ec(\gamma'\gamma + \gamma\gamma') \otimes \boldsymbol{\sigma} \cdot \mathbf{p} \\
&\quad = E^2 - \mathbf{p}^2c^2 = m^2c^4.
\end{aligned} \tag{3.119'}$$

For Eq. (3.119′) to hold, the conditions

$$\gamma'^2 = 1 = -\gamma^2, \qquad \gamma'\gamma + \gamma\gamma' = 0 \tag{3.120}$$

must be satisfied. Thus, the matrices γ' and γ anticommute, just like the three Pauli matrices; therefore they cannot be real or complex numbers. Because the conditions (3.120) can be met by 2×2 matrices, we have written direct product signs (see Example 3.2.1) in Eq. (3.119′) because γ', γ are multiplied by 1_2, $\boldsymbol{\sigma}$ matrices, respectively, with

$$\gamma' = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \gamma = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{3.121}$$

[17] For its geometrical significance, see W. E. Baylis, J. Huschilt, and Jiansu Wei, *Am. J. Phys.* **60**: 788 (1992).

The direct-product 4×4 matrices in Eq. (3.119′) are the four conventional Dirac γ-matrices,

$$\gamma^0 = \gamma' \otimes 1_2 = \begin{pmatrix} 1_2 & 0 \\ 0 & -1_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix},$$

$$\gamma^1 = \gamma \otimes \sigma_1 = \begin{pmatrix} 0 & \sigma_1 \\ -\sigma_1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix},$$

$$\gamma^3 = \gamma \otimes \sigma_3 = \begin{pmatrix} 0 & \sigma_3 \\ -\sigma_3 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \tag{3.122}$$

and similarly for $\gamma^2 = \gamma \otimes \sigma_2$. In vector notation $\boldsymbol{\gamma} = \gamma \otimes \boldsymbol{\sigma}$ is a vector with three components, each a 4×4 matrix, a generalization of the vector of Pauli matrices to a vector of 4×4 matrices. The four matrices γ^i are the components of the four-vector $\gamma^\mu = (\gamma^0, \gamma^1, \gamma^2, \gamma^3)$. If we recognize in Eq. (1.119′)

$$E\gamma' \otimes 1_2 - c(\gamma \otimes \boldsymbol{\sigma}) \cdot \mathbf{p} = \gamma^\mu p_\mu = \gamma \cdot p = (\gamma_0, \boldsymbol{\gamma}) \cdot (E, c\mathbf{p}) \tag{3.123}$$

as a scalar product of two four-vectors γ^μ and p^μ (see Lorentz group in Chapter 4), then Eq. (3.119′) with $p^2 = p \cdot p = E^2 - \mathbf{p}^2c^2$ may be regarded as a four-vector generalization of Eq. (3.119).

Summarizing the relativistic treatment of a spin $1/2$ *particle, it leads to* 4×4 *matrices, while the spin* $1/2$ *of a nonrelativistic particle is described by the* 2×2 *Pauli matrices* $\boldsymbol{\sigma}$.

By analogy with the Pauli algebra, we can form products of the basic γ^μ matrices and linear combinations of them and the unit matrix $1 = 1_4$, thereby generating a 16-dimensional (so-called **Clifford**[18]) algebra. A basis (with convenient Lorentz transformation properties, see Chapter 4) is given (in 2×2 matrix notation of Eq. (3.122)) by

$$1_4, \gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} 0 & 1_2 \\ 1_2 & 0 \end{pmatrix}, \qquad \gamma^\mu, \gamma^5\gamma^\mu, \sigma^{\mu\nu} = i(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu)/2. \tag{3.124}$$

The γ-matrices anticommute; that is, their symmetric combinations

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}1_4, \tag{3.125}$$

where $g^{00} = 1 = -g^{11} = -g^{22} = -g^{33}$, and $g^{\mu\nu} = 0$ for $\mu \neq \nu$, are zero or proportional to the 4×4 unit matrix 1_4, while the six antisymmetric combinations in Eq. (3.124) give new basis elements that transform like a tensor under Lorentz transformations (see Chapter 4). Any 4×4 matrix can be expanded in terms of these 16 elements, and the expansion coefficients are given by forming traces similar to the 2×2 case in Eq. (3.117) us-

[18] D. Hestenes and G. Sobczyk, *loc.cit.*; D. Hestenes, *Am. J. Phys.* **39**: 1013 (1971); and *J. Math. Phys.* **16**: 556 (1975).

ing $\text{trace}(1_4) = 4$, $\text{trace}(\gamma_5) = 0$, $\text{trace}(\gamma^\mu) = 0 = \text{trace}(\gamma_5\gamma^\mu)$, $\text{trace}(\sigma^{\mu\nu}) = 0$ for $\mu, \nu = 0, 1, 2, 3$ (see Exercise 3.4.23). In Chapter 4 we show that γ_5 is odd under parity, so $\gamma_5\gamma^\mu$ transform like an axial vector that has even parity.

The spin algebra generated by the Pauli matrices is just a matrix representation of the four-dimensional Clifford algebra, while Hestenes and coworkers (loc. cit.) have developed in their **geometric calculus** a representation-free (that is, "coordinate-free") algebra that contains complex numbers, vectors, the quaternion subalgebra, and generalized cross products as directed areas (called *bivectors*). This algebraic-geometric framework is tailored to nonrelativistic quantum mechanics, where spinors acquire geometric aspects and the Gauss and Stokes theorems appear as components of a unified theorem. Their geometric algebra corresponding to the 16-dimensional Clifford algebra of Dirac γ-matrices is the appropriate coordinate-free framework for relativistic quantum mechanics and electrodynamics.

The discussion of orthogonal matrices in Section 3.3 and unitary matrices in this section is only a beginning. Further extensions are of vital concern in "elementary" particle physics. With the Pauli and Dirac matrices, we can develop **spinor** wave functions for electrons, protons, and other (relativistic) spin $\frac{1}{2}$ particles. The coordinate system rotations lead to $\mathbf{D}^j(\alpha, \beta, \gamma)$, the rotation group usually represented by matrices in which the elements are functions of the Euler angles describing the rotation. The special unitary group SU(3) (composed of 3×3 unitary matrices with determinant $+1$) has been used with considerable success to describe mesons and baryons involved in the strong interactions, a gauge theory that is now called *quantum chromodynamics*. These extensions are considered further in Chapter 4.

Exercises

3.4.1 Show that

$$\det(\mathsf{A}^*) = (\det \mathsf{A})^* = \det\left(\mathsf{A}^\dagger\right).$$

3.4.2 Three angular momentum matrices satisfy the basic commutation relation

$$[\mathsf{J}_x, \mathsf{J}_y] = i\mathsf{J}_z$$

(and cyclic permutation of indices). If two of the matrices have real elements, show that the elements of the third must be pure imaginary.

3.4.3 Show that $(\mathsf{AB})^\dagger = \mathsf{B}^\dagger\mathsf{A}^\dagger$.

3.4.4 A matrix $\mathsf{C} = \mathsf{S}^\dagger\mathsf{S}$. Show that the trace is positive definite unless S is the null matrix, in which case trace $(\mathsf{C}) = 0$.

3.4.5 If A and B are Hermitian matrices, show that $(\mathsf{AB} + \mathsf{BA})$ and $i(\mathsf{AB} - \mathsf{BA})$ are also Hermitian.

3.4.6 The matrix C is **not** Hermitian. Show that then $\mathsf{C} + \mathsf{C}^\dagger$ and $i(\mathsf{C} - \mathsf{C}^\dagger)$ are Hermitian. This means that a non-Hermitian matrix may be resolved into two Hermitian parts,

$$\mathsf{C} = \frac{1}{2}\left(\mathsf{C} + \mathsf{C}^\dagger\right) + \frac{1}{2i}i\left(\mathsf{C} - \mathsf{C}^\dagger\right).$$

This decomposition of a matrix into two Hermitian matrix parts parallels the decomposition of a complex number z into $x + iy$, where $x = (z + z^*)/2$ and $y = (z - z^*)/2i$.

3.4.7 A and B are two noncommuting Hermitian matrices:

$$\mathsf{AB} - \mathsf{BA} = i\mathsf{C}.$$

Prove that C is Hermitian.

3.4.8 Show that a Hermitian matrix remains Hermitian under unitary similarity transformations.

3.4.9 Two matrices A and B are each Hermitian. Find a necessary and sufficient condition for their product AB to be Hermitian.

ANS. $[\mathsf{A}, \mathsf{B}] = 0$.

3.4.10 Show that the reciprocal (that is, inverse) of a unitary matrix is unitary.

3.4.11 A particular similarity transformation yields

$$\mathsf{A}' = \mathsf{UAU}^{-1},$$
$$\mathsf{A}'^{\dagger} = \mathsf{UA}^{\dagger}\mathsf{U}^{-1}.$$

If the adjoint relationship is preserved ($\mathsf{A}^{\dagger\prime} = \mathsf{A}'^{\dagger}$) and $\det \mathsf{U} = 1$, show that U must be unitary.

3.4.12 Two matrices U and H are related by

$$\mathsf{U} = e^{ia\mathsf{H}},$$

with a real. (The exponential function is defined by a Maclaurin expansion. This will be done in Section 5.6.)

(a) If H is Hermitian, show that U is unitary.
(b) If U is unitary, show that H is Hermitian. (H is independent of a.)

Note. With H the Hamiltonian,

$$\psi(x,t) = \mathsf{U}(x,t)\psi(x,0) = \exp(-it\mathsf{H}/\hbar)\psi(x,0)$$

is a solution of the time-dependent Schrödinger equation. $\mathsf{U}(x,t) = \exp(-it\mathsf{H}/\hbar)$ is the "evolution operator."

3.4.13 An operator $T(t+\varepsilon, t)$ describes the change in the wave function from t to $t+\varepsilon$. For ε real and small enough so that ε^2 may be neglected,

$$T(t+\varepsilon, t) = 1 - \frac{i}{\hbar}\varepsilon\mathsf{H}(t).$$

(a) If T is unitary, show that H is Hermitian.
(b) If H is Hermitian, show that T is unitary.

Note. When $\mathsf{H}(t)$ is independent of time, this relation may be put in exponential form — Exercise 3.4.12.

3.4.14 Show that an alternate form,

$$T(t+\varepsilon,t)=\frac{1-i\varepsilon \mathsf{H}(t)/2\hbar}{1+i\varepsilon \mathsf{H}(t)/2\hbar},$$

agrees with the T of part (a) of Exercise 3.4.13, neglecting ε^2, and is exactly unitary (for H Hermitian).

3.4.15 Prove that the direct product of two unitary matrices is unitary.

3.4.16 Show that γ_5 anticommutes with all four γ^μ.

3.4.17 Use the four-dimensional Levi-Civita symbol $\varepsilon_{\lambda\mu\nu\rho}$ with $\varepsilon_{0123}=-1$ (generalizing Eqs. (2.93) in Section 2.9 to four dimensions) and show that (i) $2\gamma_5\sigma_{\mu\nu}=-i\varepsilon_{\mu\nu\alpha\beta}\sigma^{\alpha\beta}$ using the summation convention of Section 2.6 and (ii) $\gamma_\lambda\gamma_\mu\gamma_\nu=g_{\lambda\mu}\gamma_\nu-g_{\lambda\nu}\gamma_\mu+g_{\mu\nu}\gamma_\lambda+i\varepsilon_{\lambda\mu\nu\rho}\gamma^\rho\gamma_5$. Define $\gamma_\mu=g_{\mu\nu}\gamma^\nu$ using $g^{\mu\nu}=g_{\mu\nu}$ to raise and lower indices.

3.4.18 Evaluate the following traces: (see Eq. (3.123) for the notation)

(i) $\text{trace}(\gamma\cdot a\gamma\cdot b)=4a\cdot b$,
(ii) $\text{trace}(\gamma\cdot a\gamma\cdot b\gamma\cdot c)=0$,
(iii) $\text{trace}(\gamma\cdot a\gamma\cdot b\gamma\cdot c\gamma\cdot d)=4(a\cdot bc\cdot d-a\cdot cb\cdot d+a\cdot db\cdot c)$,
(iv) $\text{trace}(\gamma_5\gamma\cdot a\gamma\cdot b\gamma\cdot c\gamma\cdot d)=4i\varepsilon_{\alpha\beta\mu\nu}a^\alpha b^\beta c^\mu d^\nu$.

3.4.19 Show that (i) $\gamma_\mu\gamma^\alpha\gamma^\mu=-2\gamma^\alpha$, (ii) $\gamma_\mu\gamma^\alpha\gamma^\beta\gamma^\mu=4g^{\alpha\beta}$, and (iii) $\gamma_\mu\gamma^\alpha\gamma^\beta\gamma^\nu\gamma^\mu=-2\gamma^\nu\gamma^\beta\gamma^\alpha$.

3.4.20 If $\mathsf{M}=\frac{1}{2}(1+\gamma_5)$, show that

$$\mathsf{M}^2=\mathsf{M}.$$

Note that γ_5 may be replaced by any other Dirac matrix (any Γ_i of Eq. (3.124)). If M is Hermitian, then this result, $\mathsf{M}^2=\mathsf{M}$, is the defining equation for a quantum mechanical projection operator.

3.4.21 Show that

$$\boldsymbol{\alpha}\times\boldsymbol{\alpha}=2i\boldsymbol{\sigma}\otimes 1_2,$$

where $\boldsymbol{\alpha}=\gamma_0\boldsymbol{\gamma}$ is a vector

$$\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\alpha_3).$$

Note that if $\boldsymbol{\alpha}$ is a polar vector (Section 2.4), then $\boldsymbol{\sigma}$ is an axial vector.

3.4.22 Prove that the 16 Dirac matrices form a linearly independent set.

3.4.23 If we assume that a given 4×4 matrix A (with constant elements) can be written as a linear combination of the 16 Dirac matrices

$$\mathsf{A}=\sum_{i=1}^{16}c_i\Gamma_i,$$

show that

$$c_i\sim\text{trace}(\mathsf{A}\Gamma_i).$$

3.4.24 If $C = i\gamma^2\gamma^0$ is the charge conjugation matrix, show that $C\gamma^\mu C^{-1} = -\tilde{\gamma}^\mu$, where $\tilde{}$ indicates transposition.

3.4.25 Let $x'_\mu = \Lambda^\nu_\mu x_\nu$ be a rotation by an angle θ about the 3-axis,

$$x'_0 = x_0, \qquad x'_1 = x_1\cos\theta + x_2\sin\theta,$$
$$x'_2 = -x_1\sin\theta + x_2\cos\theta, \qquad x'_3 = x_3.$$

Use $R = \exp(i\theta\sigma^{12}/2) = \cos\theta/2 + i\sigma^{12}\sin\theta/2$ (see Eq. (3.170b)) and show that the γ's transform just like the coordinates x^μ, that is, $\Lambda^\nu_\mu\gamma_\nu = R^{-1}\gamma_\mu R$. (Note that $\gamma_\mu = g_{\mu\nu}\gamma^\nu$ and that the γ^μ are well defined only up to a similarity transformation.) Similarly, if $x' = \Lambda x$ is a boost (pure Lorentz transformation) along the 1-axis, that is,

$$x'_0 = x_0\cosh\zeta - x_1\sinh\zeta, \qquad x'_1 = -x_0\sinh\zeta + x_1\cosh\zeta,$$
$$x'_2 = x_2, \qquad x'_3 = x_3,$$

with $\tanh\zeta = v/c$ and $B = \exp(-i\zeta\sigma^{01}/2) = \cosh\zeta/2 - i\sigma^{01}\sinh\zeta/2$ (see Eq. (3.170b)), show that $\Lambda^\nu_\mu\gamma_\nu = B\gamma_\mu B^{-1}$.

3.4.26 (a) Given $\mathbf{r}' = \mathsf{U}\mathbf{r}$, with U a unitary matrix and $\mathbf{r}$ a (column) vector with complex elements, show that the norm (magnitude) of $\mathbf{r}$ is invariant under this operation.

(b) The matrix U transforms any column vector $\mathbf{r}$ with complex elements into $\mathbf{r}'$, leaving the magnitude invariant: $\mathbf{r}^\dagger\mathbf{r} = \mathbf{r}'^\dagger\mathbf{r}'$. Show that U is unitary.

3.4.27 Write a subroutine that will test whether a complex $n \times n$ matrix is self-adjoint. In demanding equality of matrix elements $a_{ij} = a^\dagger_{ij}$, allow some small tolerance ε to compensate for truncation error of the computer.

3.4.28 Write a subroutine that will form the adjoint of a complex $M \times N$ matrix.

3.4.29 (a) Write a subroutine that will take a complex $M \times N$ matrix A and yield the product $\mathsf{A}^\dagger\mathsf{A}$.

Hint. This subroutine can call the subroutines of Exercises 3.2.41 and 3.4.28.

(b) Test your subroutine by taking A to be one or more of the Dirac matrices, Eq. (3.124).

3.5 Diagonalization of Matrices

Moment of Inertia Matrix

In many physical problems involving real symmetric or complex Hermitian matrices it is desirable to carry out a (real) orthogonal similarity transformation or a unitary transformation (corresponding to a rotation of the coordinate system) to reduce the matrix to a diagonal form, nondiagonal elements all equal to zero. One particularly direct example of this is the moment of inertia matrix I of a rigid body. From the definition of angular momentum $\mathbf{L}$ we have

$$\mathbf{L} = \mathsf{I}\boldsymbol{\omega}, \tag{3.126}$$

ω being the angular velocity.[19] The inertia matrix I is found to have diagonal components

$$I_{xx} = \sum_i m_i(r_i^2 - x_i^2), \qquad \text{and so on,} \tag{3.127}$$

the subscript i referring to mass m_i located at $\mathbf{r}_i = (x_i, y_i, z_i)$. For the nondiagonal components we have

$$I_{xy} = -\sum_i m_i x_i y_i = I_{yx}. \tag{3.128}$$

By inspection, matrix I is symmetric. Also, since I appears in a physical equation of the form (3.126), which holds for all orientations of the coordinate system, it may be considered to be a tensor (quotient rule, Section 2.3).

The key now is to orient the coordinate axes (along a body-fixed frame) so that the I_{xy} and the other nondiagonal elements will vanish. As a consequence of this orientation and an indication of it, if the angular velocity is along one such realigned **principal axis**, the angular velocity and the angular momentum will be parallel. As an illustration, the stability of rotation is used by football players when they throw the ball spinning about its long principal axis.

Eigenvectors, Eigenvalues

It is instructive to consider a geometrical picture of this problem. If the inertia matrix I is multiplied from each side by a unit vector of variable direction, $\hat{\mathbf{n}} = (\alpha, \beta, \gamma)$, then in the Dirac bracket notation of Section 3.2,

$$\langle \hat{\mathbf{n}} | \mathsf{I} | \hat{\mathbf{n}} \rangle = I, \tag{3.129}$$

where I is the moment of inertia about the direction $\hat{\mathbf{n}}$ and a positive number (scalar). Carrying out the multiplication, we obtain

$$I = I_{xx}\alpha^2 + I_{yy}\beta^2 + I_{zz}\gamma^2 + 2I_{xy}\alpha\beta + 2I_{xz}\alpha\gamma + 2I_{yz}\beta\gamma, \tag{3.130}$$

a positive definite quadratic form that must be an ellipsoid (see Fig. 3.5). From analytic geometry it is known that the coordinate axes can always be rotated to coincide with the axes of our ellipsoid. In many elementary cases, especially when symmetry is present, these new axes, called the *principal axes*, can be found by inspection. We can find the axes by locating the local extrema of the ellipsoid in terms of the variable components of $\mathbf{n}$, subject to the constraint $\hat{\mathbf{n}}^2 = 1$. To deal with the constraint, we introduce a Lagrange multiplier λ (Section 17.6). Differentiating $\langle \hat{\mathbf{n}} | \mathsf{I} | \hat{\mathbf{n}} \rangle - \lambda \langle \hat{\mathbf{n}} | \hat{\mathbf{n}} \rangle$,

$$\frac{\partial}{\partial n_j}\left(\langle \hat{\mathbf{n}} | \mathsf{I} | \hat{\mathbf{n}} \rangle - \lambda \langle \hat{\mathbf{n}} | \hat{\mathbf{n}} \rangle\right) = 2\sum_k I_{jk} n_k - 2\lambda n_j = 0, \qquad j = 1, 2, 3 \tag{3.131}$$

yields the eigenvalue equations

$$\mathsf{I} | \hat{\mathbf{n}} \rangle = \lambda | \hat{\mathbf{n}} \rangle. \tag{3.132}$$

[19] The moment of inertia matrix may also be developed from the kinetic energy of a rotating body, $T = 1/2\langle \omega | \mathsf{I} | \omega \rangle$.

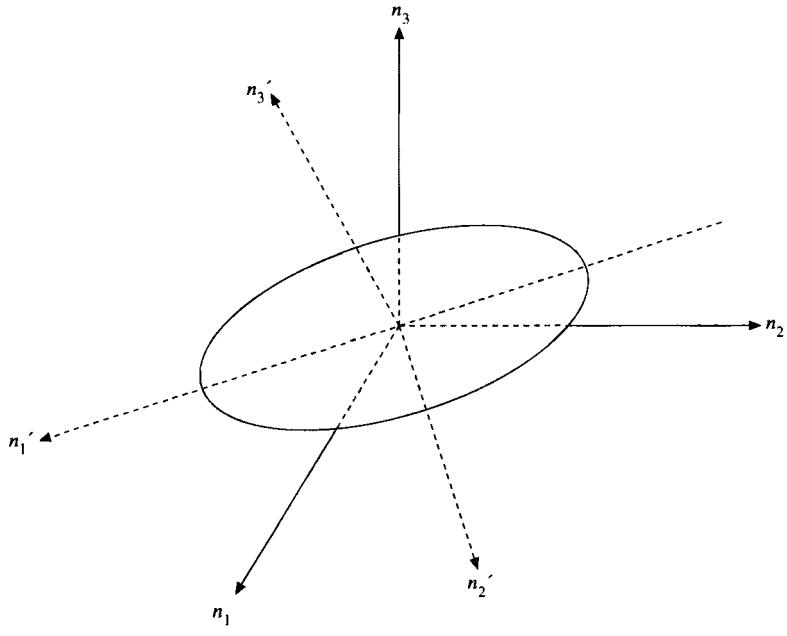

FIGURE 3.5 Moment of inertia ellipsoid.

The same result can be found by purely geometric methods. We now proceed to develop a general method of finding the diagonal elements and the principal axes.

If $\mathsf{R}^{-1} = \tilde{\mathsf{R}}$ is the real orthogonal matrix such that $\mathbf{n}' = \mathsf{R}\mathbf{n}$, or $|\mathbf{n}'\rangle = \mathsf{R}|\mathbf{n}\rangle$ in Dirac notation, are the new coordinates, then we obtain, using $\langle\mathbf{n}'|\mathsf{R} = \langle\mathbf{n}|$ in Eq. (3.132),

$$\langle\mathbf{n}|\mathsf{I}|\mathbf{n}\rangle = \langle\mathbf{n}'|\mathsf{R}\mathsf{I}\tilde{\mathsf{R}}|\mathbf{n}'\rangle = I_1' n_1'^2 + I_2' n_2'^2 + I_3' n_3'^2, \tag{3.133}$$

where the $I_i' > 0$ are the principal moments of inertia. The inertia matrix I' in Eq. (3.133) is diagonal in the new coordinates,

$$\mathsf{I}' = \mathsf{R}\mathsf{I}\tilde{\mathsf{R}} = \begin{pmatrix} I_1' & 0 & 0 \\ 0 & I_2' & 0 \\ 0 & 0 & I_3' \end{pmatrix}. \tag{3.134}$$

If we rewrite Eq. (3.134) using $\mathsf{R}^{-1} = \tilde{\mathsf{R}}$ in the form

$$\tilde{\mathsf{R}}\mathsf{I}' = \mathsf{I}\tilde{\mathsf{R}} \tag{3.135}$$

and take $\tilde{\mathsf{R}} = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ to consist of three column vectors, then Eq. (3.135) splits up into three eigenvalue equations,

$$\mathsf{I}\mathbf{v}_i = I_i'\mathbf{v}_i, \qquad i = 1, 2, 3 \tag{3.136}$$

with **eigenvalues** I_i' and **eigenvectors** $\mathbf{v}_i$. The names were introduced from the German literature on quantum mechanics. Because these equations are linear and homogeneous

(for fixed i), by Section 3.1 their determinants have to vanish:

$$\begin{vmatrix} I_{11} - I_i' & I_{12} & I_{13} \\ I_{12} & I_{22} - I_i' & I_{23} \\ I_{13} & I_{23} & I_{33} - I_i' \end{vmatrix} = 0. \tag{3.137}$$

Replacing the eigenvalue I_i' by a variable λ times the unit matrix $\mathsf{1}$, we may rewrite Eq. (3.136) as

$$\boxed{(\mathsf{I} - \lambda \mathsf{1})|\mathbf{v}\rangle = 0.} \tag{3.136'}$$

The determinant set to zero,

$$\boxed{|\mathsf{I} - \lambda \mathsf{1}| = 0,} \tag{3.137'}$$

is a cubic polynomial in λ; its three roots, of course, are the I_i'. Substituting one root at a time back into Eq. (3.136) (or (3.136')), we can find the corresponding eigenvectors. Because of its applications in astronomical theories, Eq. (3.137) (or (3.137')) is known as the **secular equation**.[20] The same treatment applies to any real symmetric matrix I, except that its eigenvalues need not all be positive. Also, the orthogonality condition in Eq. (3.87) for R say that, in geometric terms, the eigenvectors v_i are mutually orthogonal unit vectors. Indeed they form the new coordinate system. The fact that any two eigenvectors $\mathbf{v}_i, \mathbf{v}_j$ are orthogonal if $I_i' \neq I_j'$ follows from Eq. (3.136) in conjunction with the symmetry of I by multiplying with $\mathbf{v}_i$ and $\mathbf{v}_j$, respectively,

$$\langle \mathbf{v}_j|\mathsf{I}|\mathbf{v}_i\rangle = I_i' \mathbf{v}_j \cdot \mathbf{v}_i = \langle \mathbf{v}_i|\mathsf{I}|\mathbf{v}_j\rangle = I_j' \mathbf{v}_i \cdot \mathbf{v}_j. \tag{3.138a}$$

Since $I_i' \neq I_j'$ and Eq. (3.138a) implies that $(I_j' - I_i')\mathbf{v}_i \cdot \mathbf{v}_j = 0$, so $\mathbf{v}_i \cdot \mathbf{v}_j = 0$.

We can write the quadratic forms in Eq. (3.133) as a sum of squares in the original coordinates $|\mathbf{n}\rangle$,

$$\boxed{\langle \mathbf{n}|\mathsf{I}|\mathbf{n}\rangle = \langle \mathbf{n}'|\mathsf{R}\mathsf{I}\tilde{\mathsf{R}}|\mathbf{n}'\rangle = \sum_i I_i'(\mathbf{n} \cdot \mathbf{v}_i)^2,} \tag{3.138b}$$

because the rows of the rotation matrix in $\mathbf{n}' = \mathsf{R}\mathbf{n}$, or

$$\begin{pmatrix} n_1' \\ n_2' \\ n_3' \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1 \cdot \mathbf{n} \\ \mathbf{v}_2 \cdot \mathbf{n} \\ \mathbf{v}_3 \cdot \mathbf{n} \end{pmatrix}$$

componentwise, are made up of the eigenvectors $\mathbf{v}_i$. The underlying matrix identity,

$$\boxed{\mathsf{I} = \sum_i I_i' |\mathbf{v}_i\rangle\langle \mathbf{v}_i|,} \tag{3.138c}$$

[20] Equation (3.126) will take on this form when ω is along one of the principal axes. Then $\mathbf{L} = \lambda\boldsymbol{\omega}$ and $\mathsf{I}\boldsymbol{\omega} = \lambda\boldsymbol{\omega}$. In the mathematics literature λ is usually called a **characteristic value**, ω a **characteristic vector**.

may be viewed as the **spectral decomposition** of the inertia tensor (or any real symmetric matrix). Here, the word *spectral* is just another term for expansion in terms of its eigenvalues. When we multiply this eigenvalue expansion by $\langle \mathbf{n}|$ on the left and $|\mathbf{n}\rangle$ on the right we reproduce the previous relation between quadratic forms. The operator $P_i = |\mathbf{v}_i\rangle\langle\mathbf{v}_i|$ is a projection operator satisfying $P_i^2 = P_i$ that projects the ith component w_i of any vector $|\mathbf{w}\rangle = \sum_j w_j|\mathbf{v}_j\rangle$ that is expanded in terms of the eigenvector basis $|\mathbf{v}_j\rangle$. This is verified by

$$P_i|\mathbf{w}\rangle = \sum_j w_j|\mathbf{v}_i\rangle\langle\mathbf{v}_i|\mathbf{v}_j\rangle = w_i|\mathbf{v}_i\rangle = \mathbf{v}_i \cdot \mathbf{w}|\mathbf{v}_i\rangle.$$

Finally, the identity

$$\sum_i |\mathbf{v}_i\rangle\langle\mathbf{v}_i| = 1$$

expresses the completeness of the eigenvector basis according to which any vector $|\mathbf{w}\rangle = \sum_i w_i|\mathbf{v}_i\rangle$ can be expanded in terms of the eigenvectors. Multiplying the completeness relation by $|\mathbf{w}\rangle$ proves the expansion $|\mathbf{w}\rangle = \sum_i \langle\mathbf{v}_i|\mathbf{w}\rangle|\mathbf{v}_i\rangle$.

An important extension of the spectral decomposition theorem applies to commuting symmetric (or Hermitian) matrices A, B: If $[\mathsf{A}, \mathsf{B}] = 0$, then there is an orthogonal (unitary) matrix that diagonalizes both A and B; that is, both matrices have common eigenvectors if the eigenvalues are nondegenerate. The reverse of this *theorem* is also valid.

To prove this theorem we diagonalize A: $\mathsf{A}\mathbf{v}_i = a_i\mathbf{v}_i$. Multiplying each eigenvalue equation by B we obtain $\mathsf{B}\mathsf{A}\mathbf{v}_i = a_i\mathsf{B}\mathbf{v}_i = \mathsf{A}(\mathsf{B}\mathbf{v}_i)$, which says that $\mathsf{B}\mathbf{v}_i$ is an eigenvector of A with eigenvalue a_i. Hence $\mathsf{B}\mathbf{v}_i = b_i\mathbf{v}_i$ with real b_i. Conversely, if the vectors $\mathbf{v}_i$ are common eigenvectors of A and B, then $\mathsf{A}\mathsf{B}\mathbf{v}_i = \mathsf{A}b_i\mathbf{v}_i = a_ib_i\mathbf{v}_i = \mathsf{B}\mathsf{A}\mathbf{v}_i$. Since the eigenvectors $\mathbf{v}_i$ are complete, this implies $\mathsf{A}\mathsf{B} = \mathsf{B}\mathsf{A}$.

Hermitian Matrices

For complex vector spaces, Hermitian and unitary matrices play the same role as symmetric and orthogonal matrices over real vector spaces, respectively. First, let us generalize the important theorem about the diagonal elements and the principal axes for the eigenvalue equation

$$\boxed{\mathsf{A}|\mathbf{r}\rangle = \lambda|\mathbf{r}\rangle,} \tag{3.139}$$

We now show that if A is a Hermitian matrix,[21] its eigenvalues are real and its eigenvectors orthogonal.

Let λ_i and λ_j be two eigenvalues and $|\mathbf{r}_i\rangle$ and $|\mathbf{r}_j\rangle$, the corresponding eigenvectors of A, a Hermitian matrix. Then

$$\mathsf{A}|\mathbf{r}_i\rangle = \lambda_i|\mathbf{r}_i\rangle, \tag{3.140}$$

$$\mathsf{A}|\mathbf{r}_j\rangle = \lambda_j|\mathbf{r}_j\rangle. \tag{3.141}$$

[21] If A is real, the Hermitian requirement reduces to a requirement of symmetry.

Equation (3.140) is multiplied by $\langle \mathbf{r}_j |$:

$$\langle \mathbf{r}_j | \mathsf{A} | \mathbf{r}_i \rangle = \lambda_i \langle \mathbf{r}_j | \mathbf{r}_i \rangle. \tag{3.142}$$

Equation (3.141) is multiplied by $\langle \mathbf{r}_i |$ to give

$$\langle \mathbf{r}_i | \mathsf{A} | \mathbf{r}_j \rangle = \lambda_j \langle \mathbf{r}_i | \mathbf{r}_j \rangle. \tag{3.143}$$

Taking the adjoint[22] of this equation, we have

$$\langle \mathbf{r}_j | \mathsf{A}^\dagger | \mathbf{r}_i \rangle = \lambda_j^* \langle \mathbf{r}_j | \mathbf{r}_i \rangle, \tag{3.144}$$

or

$$\langle \mathbf{r}_j | \mathsf{A} | \mathbf{r}_i \rangle = \lambda_j^* \langle \mathbf{r}_j | \mathbf{r}_i \rangle \tag{3.145}$$

since A is Hermitian. Subtracting Eq. (3.145) from Eq. (3.142), we obtain

$$(\lambda_i - \lambda_j^*) \langle \mathbf{r}_j | \mathbf{r}_i \rangle = 0. \tag{3.146}$$

This is a general result for all possible combinations of i and j. First, let $j = i$. Then Eq. (3.146) becomes

$$(\lambda_i - \lambda_i^*) \langle \mathbf{r}_i | \mathbf{r}_i \rangle = 0. \tag{3.147}$$

Since $\langle \mathbf{r}_i | \mathbf{r}_i \rangle = 0$ would be a trivial solution of Eq. (3.147), we conclude that

$$\lambda_i = \lambda_i^*, \tag{3.148}$$

or λ_i is real, for all i.

Second, for $i \neq j$ and $\lambda_i \neq \lambda_j$,

$$(\lambda_i - \lambda_j) \langle \mathbf{r}_j | \mathbf{r}_i \rangle = 0, \tag{3.149}$$

or

$$\langle \mathbf{r}_j | \mathbf{r}_i \rangle = 0, \tag{3.150}$$

which means that the eigenvectors of **distinct** eigenvalues are orthogonal, Eq. (3.150) being our generalization of orthogonality in this complex space.[23]

If $\lambda_i = \lambda_j$ (degenerate case), $|\mathbf{r}_i\rangle$ is not automatically orthogonal to $|\mathbf{r}_j\rangle$, but it may be **made** orthogonal.[24] Consider the physical problem of the moment of inertia matrix again. If x_1 is an axis of rotational symmetry, then we will find that $\lambda_2 = \lambda_3$. Eigenvectors $|\mathbf{r}_2\rangle$ and $|\mathbf{r}_3\rangle$ are each perpendicular to the symmetry axis, $|\mathbf{r}_1\rangle$, but they lie anywhere in the plane perpendicular to $|\mathbf{r}_1\rangle$; that is, any linear combination of $|\mathbf{r}_2\rangle$ and $|\mathbf{r}_3\rangle$ is also an eigenvector. Consider $(a_2|\mathbf{r}_2\rangle + a_3|\mathbf{r}_3\rangle)$ with a_2 and a_3 constants. Then

$$\begin{aligned} \mathsf{A}\big(a_2|\mathbf{r}_2\rangle + a_3|\mathbf{r}_3\rangle\big) &= a_2\lambda_2|\mathbf{r}_2\rangle + a_3\lambda_3|\mathbf{r}_3\rangle \\ &= \lambda_2\big(a_2|\mathbf{r}_2\rangle + a_3|\mathbf{r}_3\rangle\big), \end{aligned} \tag{3.151}$$

[22] Note $\langle \mathbf{r}_j | = |\mathbf{r}_j\rangle^\dagger$ for complex vectors.

[23] The corresponding theory for differential operators (Sturm–Liouville theory) appears in Section 10.2. The integral equation analog (Hilbert–Schmidt theory) is given in Section 16.4.

[24] We are assuming here that the eigenvectors of the n-fold degenerate λ_i span the corresponding n-dimensional space. This may be shown by including a parameter ε in the original matrix to remove the degeneracy and then letting ε approach zero (compare Exercise 3.5.30). This is analogous to breaking a degeneracy in atomic spectroscopy by applying an external magnetic field (Zeeman effect).

as is to be expected, for x_1 is an axis of rotational symmetry. Therefore, if $|\mathbf{r}_1\rangle$ and $|\mathbf{r}_2\rangle$ are fixed, $|\mathbf{r}_3\rangle$ may simply be chosen to lie in the plane perpendicular to $|\mathbf{r}_1\rangle$ and also perpendicular to $|\mathbf{r}_2\rangle$. A general method of orthogonalizing solutions, the Gram–Schmidt process (Section 3.1), is applied to functions in Section 10.3.

The set of n orthogonal eigenvectors $|\mathbf{r}_i\rangle$ of our $n \times n$ Hermitian matrix A forms a **complete** set, spanning the n-dimensional (complex) space, $\sum_i |\mathbf{r}_i\rangle\langle\mathbf{r}_i| = 1$. This fact is useful in a variational calculation of the eigenvalues, Section 17.8.

The spectral decomposition of any Hermitian matrix A is proved by analogy with real symmetric matrices

$$\boxed{\mathsf{A} = \sum_i \lambda_i |\mathbf{r}_i\rangle\langle\mathbf{r}_i|,}$$

with real eigenvalues λ_i and orthonormal eigenvectors $|\mathbf{r}_i\rangle$.

Eigenvalues and eigenvectors are not limited to Hermitian matrices. All matrices have at least one eigenvalue and eigenvector. However, only Hermitian matrices have all eigenvectors orthogonal and all eigenvalues real.

Anti-Hermitian Matrices

Occasionally in quantum theory we encounter anti-Hermitian matrices:

$$\mathsf{A}^\dagger = -\mathsf{A}.$$

Following the analysis of the first portion of this section, we can show that

a. The eigenvalues are pure imaginary (or zero).
b. The eigenvectors corresponding to distinct eigenvalues are orthogonal.

The matrix R formed from the normalized eigenvectors is unitary. This anti-Hermitian property is preserved under unitary transformations.

Example 3.5.1 Eigenvalues and Eigenvectors of a Real Symmetric Matrix

Let

$$\mathsf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{3.152}$$

The secular equation is

$$\begin{vmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 0 \\ 0 & 0 & -\lambda \end{vmatrix} = 0, \tag{3.153}$$

or

$$-\lambda(\lambda^2 - 1) = 0, \tag{3.154}$$

expanding by minors. The roots are $\lambda = -1, 0, 1$. To find the eigenvector corresponding to $\lambda = -1$, we substitute this value back into the eigenvalue equation, Eq. (3.139),

$$\begin{pmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 0 \\ 0 & 0 & -\lambda \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{3.155}$$

With $\lambda = -1$, this yields

$$x + y = 0, \qquad z = 0. \tag{3.156}$$

Within an arbitrary scale factor and an arbitrary sign (or phase factor), $\langle \mathbf{r}_1| = (1, -1, 0)$. Note that (for real $|\mathbf{r}\rangle$ in ordinary space) the eigenvector singles out a line in space. The positive or negative sense is not determined. This indeterminancy could be expected if we noted that Eq. (3.139) is homogeneous in $|\mathbf{r}\rangle$. For convenience we will require that the eigenvectors be normalized to unity, $\langle \mathbf{r}_1|\mathbf{r}_1\rangle = 1$. With this condition,

$$\langle \mathbf{r}_1| = \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0\right) \tag{3.157}$$

is fixed except for an overall sign. For $\lambda = 0$, Eq. (3.139) yields

$$y = 0, \qquad x = 0, \tag{3.158}$$

$\langle \mathbf{r}_2| = (0, 0, 1)$ is a suitable eigenvector. Finally, for $\lambda = 1$, we get

$$-x + y = 0, \qquad z = 0, \tag{3.159}$$

or

$$\langle \mathbf{r}_3| = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right). \tag{3.160}$$

The orthogonality of $\mathbf{r}_1, \mathbf{r}_2$, and $\mathbf{r}_3$, corresponding to three distinct eigenvalues, may be easily verified.

The corresponding spectral decomposition gives

$$\mathsf{A} = (-1)\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0\right)\begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} + (+1)\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right)\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} + 0(0, 0, 1)\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$= -\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \qquad \blacksquare$$

Example 3.5.2 DEGENERATE EIGENVALUES

Consider

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{3.161}$$

The secular equation is

$$\begin{vmatrix} 1-\lambda & 0 & 0 \\ 0 & -\lambda & 1 \\ 0 & 1 & -\lambda \end{vmatrix} = 0 \tag{3.162}$$

or

$$(1-\lambda)(\lambda^2 - 1) = 0, \quad \lambda = -1, 1, 1, \tag{3.163}$$

a degenerate case. If $\lambda = -1$, the eigenvalue equation (3.139) yields

$$2x = 0, \qquad y + z = 0. \tag{3.164}$$

A suitable normalized eigenvector is

$$\langle \mathbf{r}_1 | = \left(0, \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right). \tag{3.165}$$

For $\lambda = 1$, we get

$$-y + z = 0. \tag{3.166}$$

Any eigenvector satisfying Eq. (3.166) is perpendicular to $\mathbf{r}_1$. We have an infinite number of choices. Suppose, as one possible choice, $\mathbf{r}_2$ is taken as

$$\langle \mathbf{r}_2 | = \left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \tag{3.167}$$

which clearly satisfies Eq. (3.166). Then $\mathbf{r}_3$ must be perpendicular to $\mathbf{r}_1$ and may be made perpendicular to $\mathbf{r}_2$ by[25]

$$\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2 = (1, 0, 0). \tag{3.168}$$

The corresponding spectral decomposition gives

$$A = -\left(0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)\begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} + \left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)\begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} + (1,0,0)\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$= -\begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & -\frac{1}{2} & \frac{1}{2} \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \quad \blacksquare$$

[25]The use of the cross product is limited to three-dimensional space (see Section 1.4).

Functions of Matrices

Polynomials with one or more matrix arguments are well defined and occur often. Power series of a matrix may also be defined, provided the series converge (see Chapter 5) for each matrix element. For example, if A is any $n \times n$ matrix, then the power series

$$\exp(\mathsf{A}) = \sum_{j=0}^{\infty} \frac{1}{j!}\mathsf{A}^j, \tag{3.169a}$$

$$\sin(\mathsf{A}) = \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j+1)!}\mathsf{A}^{2j+1}, \tag{3.169b}$$

$$\cos(\mathsf{A}) = \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j)!}\mathsf{A}^{2j} \tag{3.169c}$$

are well defined $n \times n$ matrices. For the Pauli matrices σ_k the **Euler identity** for real θ and $k = 1, 2$, or 3

$$\boxed{\exp(i\sigma_k\theta) = \mathsf{1}_2 \cos\theta + i\sigma_k \sin\theta,} \tag{3.170a}$$

follows from collecting all even and odd powers of θ in separate series using $\sigma_k^2 = 1$. For the 4×4 Dirac matrices $\sigma^{jk} = 1$ with $(\sigma^{jk})^2 = 1$ if $j \neq k = 1, 2$ or 3 we obtain similarly (without writing the obvious unit matrix $\mathsf{1}_4$ anymore)

$$\boxed{\exp\bigl(i\sigma^{jk}\theta\bigr) = \cos\theta + i\sigma^{jk}\sin\theta,} \tag{3.170b}$$

while

$$\boxed{\exp\bigl(i\sigma^{0k}\zeta\bigr) = \cosh\zeta + i\sigma^{0k}\sinh\zeta} \tag{3.170c}$$

holds for real ζ because $(i\sigma^{0k})^2 = 1$ for $k = 1, 2$, or 3.

For a Hermitian matrix A there is a unitary matrix U that diagonalizes it; that is, $\mathsf{UAU}^\dagger = [a_1, a_2, \ldots, a_n]$. Then the **trace formula**

$$\boxed{\det\bigl(\exp(\mathsf{A})\bigr) = \exp\bigl(\operatorname{trace}(\mathsf{A})\bigr)} \tag{3.171}$$

is obtained (see Exercises 3.5.2 and 3.5.9) from

$$\begin{aligned}\det\bigl(\exp(\mathsf{A})\bigr) &= \det\bigl(\mathsf{U}\exp(\mathsf{A})\mathsf{U}^\dagger\bigr) = \det\bigl(\exp\bigl(\mathsf{UAU}^\dagger\bigr)\bigr)\\ &= \det\exp[a_1, a_2, \ldots, a_n] = \det\bigl[e^{a_1}, e^{a_2}, \ldots, e^{a_n}\bigr]\\ &= \prod e^{a_i} = \exp\Bigl(\sum a_i\Bigr) = \exp\bigl(\operatorname{trace}(\mathsf{A})\bigr),\end{aligned}$$

using $\mathsf{UA}^i\mathsf{U}^\dagger = (\mathsf{UAU}^\dagger)^i$ in the power series Eq. (3.169a) for $\exp(\mathsf{UAU}^\dagger)$ and the product theorem for determinants in Section 3.2.

This trace formula is a special case of the **spectral decomposition law** for any (infinitely differentiable) function $f(\mathsf{A})$ for Hermitian A:

$$\boxed{f(\mathsf{A}) = \sum_i f(\lambda_i)|\mathbf{r}_i\rangle\langle\mathbf{r}_i|,}$$

where $|\mathbf{r}_i\rangle$ are the common eigenvectors of A and A^j. This eigenvalue expansion follows from $\mathsf{A}^j|\mathbf{r}_i\rangle = \lambda_i^j|\mathbf{r}_i\rangle$, multiplied by $f^{(j)}(0)/j!$ and summed over j to form the Taylor expansion of $f(\lambda_i)$ and yield $f(\mathsf{A})|\mathbf{r}_i\rangle = f(\lambda_i)|\mathbf{r}_i\rangle$. Finally, summing over i and using completeness we obtain $f(\mathsf{A})\sum_i |\mathbf{r}_i\rangle\langle\mathbf{r}_i| = \sum_i f(\lambda_i)|\mathbf{r}_i\rangle\langle\mathbf{r}_i| = f(\mathsf{A})$, q.e.d.

Example 3.5.3 EXPONENTIAL OF A DIAGONAL MATRIX

If the matrix A is diagonal like

$$\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

then its nth power is also diagonal with its diagonal, matrix elements raised to the nth power:

$$(\sigma_3)^n = \begin{pmatrix} 1 & 0 \\ 0 & (-1)^n \end{pmatrix}.$$

Then summing the exponential series, element for element, yields

$$e^{\sigma_3} = \begin{pmatrix} \sum_{n=0}^{\infty} \frac{1}{n!} & 0 \\ 0 & \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \end{pmatrix} = \begin{pmatrix} e & 0 \\ 0 & \frac{1}{e} \end{pmatrix}.$$

If we write the general diagonal matrix as $\mathsf{A} = [a_1, a_2, \ldots, a_n]$ with diagonal elements a_j, then $\mathsf{A}^m = [a_1^m, a_2^m, \ldots, a_n^m]$, and summing the exponentials elementwise again we obtain $e^{\mathsf{A}} = [e^{a_1}, e^{a_2}, \ldots, e^{a_n}]$.

Using the spectral decomposition law we obtain directly

$$e^{\sigma_3} = e^{+1}(1,0)\begin{pmatrix}1\\0\end{pmatrix} + e^{-1}(0,1)\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix} e & 0 \\ 0 & e^{-1} \end{pmatrix}. \qquad \blacksquare$$

Another important relation is the **Baker–Hausdorff formula**,

$$\exp(i\mathsf{G})\mathsf{H}\exp(-i\mathsf{G}) = \mathsf{H} + [i\mathsf{G}, \mathsf{H}] + \frac{1}{2}[i\mathsf{G}, [i\mathsf{G}, \mathsf{H}]] + \cdots, \tag{3.172}$$

which follows from multiplying the power series for $\exp(i\mathsf{G})$ and collecting the terms with the same powers of $i\mathsf{G}$. Here we define

$$[\mathsf{G}, \mathsf{H}] = \mathsf{GH} - \mathsf{HG}$$

as the **commutator** of G and H.

The preceding analysis has the advantage of exhibiting and clarifying conceptual relationships in the diagonalization of matrices. However, for matrices larger than 3×3, or perhaps 4×4, the process rapidly becomes so cumbersome that we turn to computers and

iterative techniques.[26] One such technique is the Jacobi method for determining eigenvalues and eigenvectors of real symmetric matrices. This Jacobi technique for determining eigenvalues and eigenvectors and the Gauss–Seidel method of solving systems of simultaneous linear equations are examples of relaxation methods. They are iterative techniques in which the errors may decrease or relax as the iterations continue. Relaxation methods are used extensively for the solution of partial differential equations.

Exercises

3.5.1 (a) Starting with the orbital angular momentum of the ith element of mass,

$$\mathbf{L}_i = \mathbf{r}_i \times \mathbf{p}_i = m_i \mathbf{r}_i \times (\boldsymbol{\omega} \times \mathbf{r}_i),$$

derive the inertia matrix such that $\mathbf{L} = \mathsf{I}\boldsymbol{\omega}$, $|\mathbf{L}\rangle = \mathsf{I}|\boldsymbol{\omega}\rangle$.

(b) Repeat the derivation starting with kinetic energy

$$T_i = \frac{1}{2} m_i (\boldsymbol{\omega} \times \mathbf{r}_i)^2 \qquad \left(T = \frac{1}{2}\langle\boldsymbol{\omega}|\mathsf{I}|\boldsymbol{\omega}\rangle\right).$$

3.5.2 Show that the eigenvalues of a matrix are unaltered if the matrix is transformed by a similarity transformation.
This property is not limited to symmetric or Hermitian matrices. It holds for any matrix satisfying the eigenvalue equation, Eq. (3.139). If our matrix can be brought into diagonal form by a similarity transformation, then two immediate consequences are

1. The trace (sum of eigenvalues) is invariant under a similarity transformation.
2. The determinant (product of eigenvalues) is invariant under a similarity transformation.

Note. The invariance of the trace and determinant are often demonstrated by using the Cayley–Hamilton theorem: A matrix satisfies its own characteristic (secular) equation.

3.5.3 As a converse of the theorem that Hermitian matrices have real eigenvalues and that eigenvectors corresponding to distinct eigenvalues are orthogonal, show that if

(a) the eigenvalues of a matrix are real and
(b) the eigenvectors satisfy $\mathbf{r}_i^\dagger \mathbf{r}_j = \delta_{ij} = \langle \mathbf{r}_i | \mathbf{r}_j \rangle$,

then the matrix is Hermitian.

3.5.4 Show that a real matrix that is not symmetric cannot be diagonalized by an orthogonal similarity transformation.
Hint. Assume that the nonsymmetric real matrix can be diagonalized and develop a contradiction.

[26] In higher-dimensional systems the secular equation may be strongly ill-conditioned with respect to the determination of its roots (the eigenvalues). Direct solution by computer may be very inaccurate. Iterative techniques for diagonalizing the original matrix are usually preferred. See Sections 2.7 and 2.9 of Press *et al.*, loc. cit.

3.5.5 The matrices representing the angular momentum components J_x, J_y, and J_z are all Hermitian. Show that the eigenvalues of $\mathbf{J}^2$, where $\mathbf{J}^2 = J_x^2 + J_y^2 + J_z^2$, are real and nonnegative.

3.5.6 A has eigenvalues λ_i and corresponding eigenvectors $|\mathbf{x}_i\rangle$. Show that A^{-1} has the same eigenvectors but with eigenvalues λ_i^{-1}.

3.5.7 A square matrix with zero determinant is labeled **singular**.

(a) If A is singular, show that there is at least one nonzero column vector $\mathbf{v}$ such that

$$\mathsf{A}|\mathbf{v}\rangle = 0.$$

(b) If there is a nonzero vector $|\mathbf{v}\rangle$ such that

$$\mathsf{A}|\mathbf{v}\rangle = 0,$$

show that A is a singular matrix. This means that if a matrix (or operator) has zero as an eigenvalue, the matrix (or operator) has no inverse and its determinant is zero.

3.5.8 The same similarity transformation diagonalizes each of two matrices. Show that the original matrices must commute. (This is particularly important in the matrix (Heisenberg) formulation of quantum mechanics.)

3.5.9 Two Hermitian matrices A and B have the same eigenvalues. Show that A and B are related by a unitary similarity transformation.

3.5.10 Find the eigenvalues and an orthonormal (orthogonal and normalized) set of eigenvectors for the matrices of Exercise 3.2.15.

3.5.11 Show that the inertia matrix for a single particle of mass m at (x, y, z) has a zero determinant. Explain this result in terms of the invariance of the determinant of a matrix under similarity transformations (Exercise 3.3.10) and a possible rotation of the coordinate system.

3.5.12 A certain rigid body may be represented by three point masses: $m_1 = 1$ at $(1, 1, -2)$, $m_2 = 2$ at $(-1, -1, 0)$, and $m_3 = 1$ at $(1, 1, 2)$.

(a) Find the inertia matrix.
(b) Diagonalize the inertia matrix, obtaining the eigenvalues and the principal axes (as orthonormal eigenvectors).

3.5.13 Unit masses are placed as shown in Fig. 3.6.

(a) Find the moment of inertia matrix.
(b) Find the eigenvalues and a set of orthonormal eigenvectors.
(c) Explain the degeneracy in terms of the symmetry of the system.

$$\text{ANS. } \mathsf{I} = \begin{pmatrix} 4 & -1 & -1 \\ -1 & 4 & -1 \\ -1 & -1 & 4 \end{pmatrix} \qquad \begin{aligned} \lambda_1 &= 2 \\ \mathbf{r}_1 &= (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}) \\ \lambda_2 &= \lambda_3 = 5. \end{aligned}$$

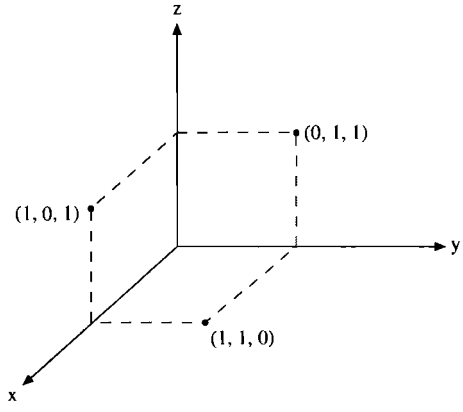

FIGURE 3.6 Mass sites for inertia tensor.

3.5.14 A mass $m_1 = 1/2$ kg is located at $(1, 1, 1)$ (meters), a mass $m_2 = 1/2$ kg is at $(-1, -1, -1)$. The two masses are held together by an ideal (weightless, rigid) rod.

(a) Find the inertia tensor of this pair of masses.
(b) Find the eigenvalues and eigenvectors of this inertia matrix.
(c) Explain the meaning, the physical significance of the $\lambda = 0$ eigenvalue. What is the significance of the corresponding eigenvector?
(d) Now that you have solved this problem by rather sophisticated matrix techniques, explain how you could obtain

(1) $\lambda = 0$ and $\lambda = ?$ — by inspection (that is, using common sense).
(2) $\mathbf{r}_{\lambda=0} = ?$ — by inspection (that is, using freshman physics).

3.5.15 Unit masses are at the eight corners of a cube $(\pm 1, \pm 1, \pm 1)$. Find the moment of inertia matrix and show that there is a triple degeneracy. This means that so far as moments of inertia are concerned, the cubic structure exhibits spherical symmetry.
Find the eigenvalues and corresponding orthonormal eigenvectors of the following matrices (as a numerical check, note that the sum of the eigenvalues equals the sum of the diagonal elements of the original matrix, Exercise 3.3.9). Note also the correspondence between $\det \mathsf{A} = 0$ and the existence of $\lambda = 0$, as required by Exercises 3.5.2 and 3.5.7.

3.5.16 $$\mathsf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

ANS. $\lambda = 0, 1, 2$.

3.5.17 $$\mathsf{A} = \begin{pmatrix} 1 & \sqrt{2} & 0 \\ \sqrt{2} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

ANS. $\lambda = -1, 0, 2$.

3.5.18 $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$.

ANS. $\lambda = -1, 1, 2$.

3.5.19 $A = \begin{pmatrix} 1 & \sqrt{8} & 0 \\ \sqrt{8} & 1 & \sqrt{8} \\ 0 & \sqrt{8} & 1 \end{pmatrix}$.

ANS. $\lambda = -3, 1, 5$.

3.5.20 $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$.

ANS. $\lambda = 0, 1, 2$.

3.5.21 $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \sqrt{2} \\ 0 & \sqrt{2} & 0 \end{pmatrix}$.

ANS. $\lambda = -1, 1, 2$.

3.5.22 $A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$.

ANS. $\lambda = -\sqrt{2}, 0, \sqrt{2}$.

3.5.23 $A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$.

ANS. $\lambda = 0, 2, 2$.

3.5.24 $A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$.

ANS. $\lambda = -1, -1, 2$.

3.5.25 $A = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$.

ANS. $\lambda = -1, 2, 2$.

3.5.26 $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$.

ANS. $\lambda = 0, 0, 3$.

3.5.27 $$\mathsf{A} = \begin{pmatrix} 5 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 0 & 2 \end{pmatrix}.$$

ANS. $\lambda = 1, 1, 6$.

3.5.28 $$\mathsf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

ANS. $\lambda = 0, 0, 2$.

3.5.29 $$\mathsf{A} = \begin{pmatrix} 5 & 0 & \sqrt{3} \\ 0 & 3 & 0 \\ \sqrt{3} & 0 & 3 \end{pmatrix}.$$

ANS. $\lambda = 2, 3, 6$.

3.5.30 (a) Determine the eigenvalues and eigenvectors of

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

Note that the eigenvalues are degenerate for $\varepsilon = 0$ but that the eigenvectors are orthogonal for all $\varepsilon \neq 0$ and $\varepsilon \to 0$.

(b) Determine the eigenvalues and eigenvectors of

$$\begin{pmatrix} 1 & 1 \\ \varepsilon^2 & 1 \end{pmatrix}.$$

Note that the eigenvalues are degenerate for $\varepsilon = 0$ and that for this (nonsymmetric) matrix the eigenvectors ($\varepsilon = 0$) do **not** span the space.

(c) Find the cosine of the angle between the two eigenvectors as a function of ε for $0 \leq \varepsilon \leq 1$.

3.5.31 (a) Take the coefficients of the simultaneous linear equations of Exercise 3.1.7 to be the matrix elements a_{ij} of matrix A (symmetric). Calculate the eigenvalues and eigenvectors.

(b) Form a matrix R whose columns are the eigenvectors of A, and calculate the triple matrix product $\tilde{\mathsf{R}}\mathsf{A}\mathsf{R}$.

ANS. $\lambda = 3.33163$.

3.5.32 Repeat Exercise 3.5.31 by using the matrix of Exercise 3.2.39.

3.5.33 Describe the geometric properties of the surface

$$x^2 + 2xy + 2y^2 + 2yz + z^2 = 1.$$

How is it oriented in three-dimensional space? Is it a conic section? If so, which kind?

Table 3.1

Matrix	Eigenvalues	Eigenvectors (for different eigenvalues)
Hermitian	Real	Orthogonal
Anti-Hermitian	Pure imaginary (or zero)	Orthogonal
Unitary	Unit magnitude	Orthogonal
Normal	If A has eigenvalue λ, $A^\dagger$ has eigenvalue λ^*	Orthogonal A and $A^\dagger$ have the same eigenvectors

3.5.34 For a Hermitian $n \times n$ matrix A with distinct eigenvalues λ_j and a function f, show that the spectral decomposition law may be expressed as

$$f(A) = \sum_{j=1}^{n} f(\lambda_j) \frac{\prod_{i \neq j}(A - \lambda_i)}{\prod_{i \neq j}(\lambda_j - \lambda_i)}.$$

This formula is due to Sylvester.

3.6 NORMAL MATRICES

In Section 3.5 we concentrated primarily on Hermitian or real symmetric matrices and on the actual process of finding the eigenvalues and eigenvectors. In this section[27] we generalize to normal matrices, with Hermitian and unitary matrices as special cases. The physically important problem of normal modes of vibration and the numerically important problem of ill-conditioned matrices are also considered.

A normal matrix is a matrix that commutes with its adjoint,

$$[A, A^\dagger] = 0.$$

Obvious and important examples are Hermitian and unitary matrices. We will show that normal matrices have orthogonal eigenvectors (see Table 3.1). We proceed in two steps.

I. Let A have an eigenvector $|\mathbf{x}\rangle$ and corresponding eigenvalue λ. Then

$$A|\mathbf{x}\rangle = \lambda|\mathbf{x}\rangle \tag{3.173}$$

or

$$(A - \lambda 1)|\mathbf{x}\rangle = 0. \tag{3.174}$$

For convenience the combination $A - \lambda 1$ will be labeled B. Taking the adjoint of Eq. (3.174), we obtain

$$\langle \mathbf{x}|(A - \lambda 1)^\dagger = 0 = \langle \mathbf{x}|B^\dagger. \tag{3.175}$$

Because

$$[(A - \lambda 1)^\dagger, (A - \lambda 1)] = [A, A^\dagger] = 0,$$

[27] Normal matrices are the largest class of matrices that can be diagonalized by unitary transformations. For an extensive discussion of normal matrices, see P. A. Macklin, Normal matrices for physicists. *Am. J. Phys.* **52**: 513 (1984).

we have

$$[\mathsf{B}, \mathsf{B}^\dagger] = 0. \tag{3.176}$$

The matrix B is also normal.

From Eqs. (3.174) and (3.175) we form

$$\langle \mathbf{x}|\mathsf{B}^\dagger \mathsf{B}|\mathbf{x}\rangle = 0. \tag{3.177}$$

This equals

$$\langle \mathbf{x}|\mathsf{B}\mathsf{B}^\dagger|\mathbf{x}\rangle = 0 \tag{3.178}$$

by Eq. (3.176). Now Eq. (3.178) may be rewritten as

$$\left(\mathsf{B}^\dagger|\mathbf{x}\rangle\right)^\dagger \left(\mathsf{B}^\dagger|\mathbf{x}\rangle\right) = 0. \tag{3.179}$$

Thus

$$\mathsf{B}^\dagger|\mathbf{x}\rangle = \left(\mathsf{A}^\dagger - \lambda^* 1\right)|\mathbf{x}\rangle = 0. \tag{3.180}$$

We see that for normal matrices, $\mathsf{A}^\dagger$ has the same eigenvectors as A but the complex conjugate eigenvalues.

II. Now, considering more than one eigenvector–eigenvalue, we have

$$\mathsf{A}|\mathbf{x}_i\rangle = \lambda_i|\mathbf{x}_i\rangle, \tag{3.181}$$

$$\mathsf{A}|\mathbf{x}_j\rangle = \lambda_j|\mathbf{x}_j\rangle. \tag{3.182}$$

Multiplying Eq. (3.182) from the left by $\langle \mathbf{x}_i|$ yields

$$\langle \mathbf{x}_i|\mathsf{A}|\mathbf{x}_j\rangle = \lambda_j \langle \mathbf{x}_i|\mathbf{x}_j\rangle. \tag{3.183}$$

Taking the transpose of Eq. (3.181), we obtain

$$\langle \mathbf{x}_i|\mathsf{A} = \left(\mathsf{A}^\dagger|\mathbf{x}_i\rangle\right)^\dagger. \tag{3.184}$$

From Eq. (3.180), with $\mathsf{A}^\dagger$ having the same eigenvectors as A but the complex conjugate eigenvalues,

$$\left(\mathsf{A}^\dagger|\mathbf{x}_i\rangle\right)^\dagger = \left(\lambda_i^*|\mathbf{x}_i\rangle\right)^\dagger = \lambda_i \langle \mathbf{x}_i|. \tag{3.185}$$

Substituting into Eq. (3.183) we have

$$\lambda_i \langle \mathbf{x}_i|\mathbf{x}_j\rangle = \lambda_j \langle \mathbf{x}_i|\mathbf{x}_j\rangle$$

or

$$(\lambda_i - \lambda_j)\langle \mathbf{x}_i|\mathbf{x}_j\rangle = 0. \tag{3.186}$$

This is the same as Eq. (3.149).

For $\lambda_i \neq \lambda_j$,

$$\langle \mathbf{x}_j|\mathbf{x}_i\rangle = 0.$$

The eigenvectors corresponding to different eigenvalues of a normal matrix are **orthogonal**. This means that a normal matrix may be diagonalized by a unitary transformation. The required unitary matrix may be constructed from the orthonormal eigenvectors as shown earlier, in Section 3.5.

The converse of this result is also true. If A can be diagonalized by a unitary transformation, then A is normal.

Normal Modes of Vibration

We consider the vibrations of a classical model of the CO_2 molecule. It is an illustration of the application of matrix techniques to a problem that does not start as a matrix problem. It also provides an example of the eigenvalues and eigenvectors of an asymmetric real matrix.

Example 3.6.1 NORMAL MODES

Consider three masses on the x-axis joined by springs as shown in Fig. 3.7. The spring forces are assumed to be linear (small displacements, Hooke's law), and the mass is constrained to stay on the x-axis.

Using a different coordinate for each mass, Newton's second law yields the set of equations

$$\begin{aligned}\ddot{x}_1 &= -\frac{k}{M}(x_1 - x_2)\\ \ddot{x}_2 &= -\frac{k}{m}(x_2 - x_1) - \frac{k}{m}(x_2 - x_3)\\ \ddot{x}_3 &= -\frac{k}{M}(x_3 - x_2).\end{aligned} \tag{3.187}$$

The system of masses is vibrating. We seek the common frequencies, ω, such that all masses vibrate at this same frequency. These are the **normal** modes. Let

$$x_i = x_{i0}e^{i\omega t}, \qquad i = 1, 2, 3.$$

Substituting this set into Eq. (3.187), we may rewrite it as

$$\begin{pmatrix} \frac{k}{M} & -\frac{k}{M} & 0 \\ -\frac{k}{m} & \frac{2k}{m} & -\frac{k}{m} \\ 0 & -\frac{k}{M} & \frac{k}{M} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = +\omega^2 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \tag{3.188}$$

with the common factor $e^{i\omega t}$ divided out. We have a matrix–eigenvalue equation with the matrix asymmetric. The secular equation is

$$\begin{vmatrix} \frac{k}{M} - \omega^2 & -\frac{k}{M} & 0 \\ -\frac{k}{m} & \frac{2k}{m} - \omega^2 & -\frac{k}{m} \\ 0 & -\frac{k}{M} & \frac{k}{M} - \omega^2 \end{vmatrix} = 0. \tag{3.189}$$

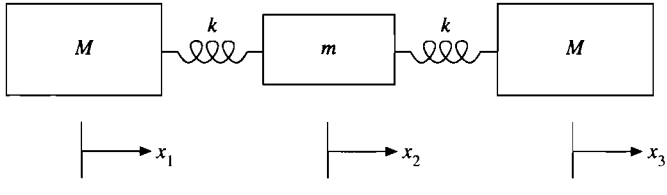

FIGURE 3.7 Double oscillator.

This leads to

$$\omega^2\left(\frac{k}{M}-\omega^2\right)\left(\omega^2-\frac{2k}{m}-\frac{k}{M}\right)=0.$$

The eigenvalues are

$$\omega^2=0, \qquad \frac{k}{M}, \qquad \frac{k}{M}+\frac{2k}{m},$$

all real.

The corresponding eigenvectors are determined by substituting the eigenvalues back into Eq. (3.188) one eigenvalue at a time. For $\omega^2=0$, Eq. (3.188), yields

$$x_1-x_2=0, \qquad -x_1+2x_2-x_3=0, \qquad -x_2+x_3=0.$$

Then we get

$$x_1=x_2=x_3.$$

This describes pure translation with no relative motion of the masses and no vibration.

For $\omega^2=k/M$, Eq. (3.188) yields

$$x_1=-x_3, \qquad x_2=0.$$

The two outer masses are moving in opposite direction. The central mass is stationary.

For $\omega^2=k/M+2k/m$, the eigenvector components are

$$x_1=x_3, \qquad x_2=-\frac{2M}{m}x_1.$$

The two outer masses are moving together. The central mass is moving opposite to the two outer ones. The net momentum is zero.

Any displacement of the three masses along the x-axis can be described as a linear combination of these three types of motion: translation plus two forms of vibration. ■

Ill-Conditioned Systems

A system of simultaneous linear equations may be written as

$$\mathsf{A}|\mathbf{x}\rangle=|\mathbf{y}\rangle \qquad \text{or} \qquad \mathsf{A}^{-1}|\mathbf{y}\rangle=|\mathbf{x}\rangle, \tag{3.190}$$

with A and $|\mathbf{y}\rangle$ known and $|\mathbf{x}\rangle$ unknown. When a small error in $|\mathbf{y}\rangle$ results in a larger error in $|\mathbf{x}\rangle$, then the matrix A is called *ill-conditioned*. With $|\delta\mathbf{x}\rangle$ an error in $|\mathbf{x}\rangle$ and $|\delta\mathbf{x}\rangle$ an error in $|\mathbf{y}\rangle$, the relative errors may be written as

$$\left[\frac{\langle\delta\mathbf{x}|\delta\mathbf{x}\rangle}{\langle\mathbf{x}|\mathbf{x}\rangle}\right]^{1/2}\le K(\mathsf{A})\left[\frac{\langle\delta\mathbf{y}|\delta\mathbf{y}\rangle}{\langle\mathbf{y}|\mathbf{y}\rangle}\right]^{1/2}. \tag{3.191}$$

Here $K(\mathsf{A})$, a property of matrix A, is labeled the **condition number**. For A Hermitian one form of the condition number is given by[28]

$$K(\mathsf{A})=\frac{|\lambda|_{\max}}{|\lambda|_{\min}}. \tag{3.192}$$

[28]G. E. Forsythe, and C. B. Moler, *Computer Solution of Linear Algebraic Systems*. Englewood Cliffs, NJ, Prentice Hall (1967).

An approximate form due to Turing[29] is

$$K(\mathsf{A}) = n[A_{ij}]_{\max}\left[A_{ij}^{-1}\right]_{\max}, \tag{3.193}$$

in which n is the order of the matrix and $[A_{ij}]_{\max}$ is the maximum element in A.

Example 3.6.1 AN ILL-CONDITIONED MATRIX

A common example of an ill-conditioned matrix is the Hilbert matrix, $H_{ij} = (i + j - 1)^{-1}$. The Hilbert matrix of order 4, H_4, is encountered in a least-squares fit of data to a third-degree polynomial. We have

$$\mathsf{H}_4 = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{pmatrix}. \tag{3.194}$$

The elements of the inverse matrix (order n) are given by

$$\left(\mathsf{H}_n^{-1}\right)_{ij} = \frac{(-1)^{i+j}}{i+j-1} \cdot \frac{(n+i-1)!(n+j-1)!}{[(i-1)!(j-1)!]^2(n-i)!(n-j)!}. \tag{3.195}$$

For $n = 4$,

$$\mathsf{H}_4^{-1} = \begin{pmatrix} 16 & -120 & 240 & -140 \\ -120 & 1200 & -2700 & 1680 \\ 240 & -2700 & 6480 & -4200 \\ -140 & 1680 & -4200 & 2800 \end{pmatrix}. \tag{3.196}$$

From Eq. (3.193) the Turing estimate of the condition number for H_4 becomes

$$\begin{aligned} K_{\text{Turing}} &= 4 \times 1 \times 6480 \\ &= 2.59 \times 10^4. \end{aligned}$$

This is a warning that an input error may be multiplied by 26,000 in the calculation of the output result. It is a statement that H_4 is ill-conditioned. If you encounter a highly ill-conditioned system, you have two alternatives (besides abandoning the problem).

(a) Try a different mathematical attack.
(b) Arrange to carry more significant figures and push through by brute force.

As previously seen, matrix eigenvector–eigenvalue techniques are not limited to the solution of strictly matrix problems. A further example of the transfer of techniques from one area to another is seen in the application of matrix techniques to the solution of Fredholm eigenvalue integral equations, Section 16.3. In turn, these matrix techniques are strengthened by a variational calculation of Section 17.8. ■

[29]Compare J. Todd, *The Condition of the Finite Segments of the Hilbert Matrix*, Applied Mathematics Series No. 313. Washington, DC: National Bureau of Standards.

Exercises

3.6.1 Show that every 2×2 matrix has two eigenvectors and corresponding eigenvalues. The eigenvectors are not necessarily orthogonal and may be degenerate. The eigenvalues are not necessarily real.

3.6.2 As an illustration of Exercise 3.6.1, find the eigenvalues and corresponding eigenvectors for

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

Note that the eigenvectors are **not** orthogonal.

ANS. $\lambda_1 = 0, \mathbf{r}_1 = (2, -1)$;
$\lambda_2 = 4, \mathbf{r}_2 = (2, 1)$.

3.6.3 If A is a 2×2 matrix, show that its eigenvalues λ satisfy the secular equation

$$\lambda^2 - \lambda \,\text{trace}(\mathsf{A}) + \det \mathsf{A} = 0.$$

3.6.4 Assuming a unitary matrix U to satisfy an eigenvalue equation $\mathsf{U}\mathbf{r} = \lambda\mathbf{r}$, show that the eigenvalues of the unitary matrix have unit magnitude. This same result holds for real orthogonal matrices.

3.6.5 Since an orthogonal matrix describing a rotation in real three-dimensional space is a special case of a unitary matrix, such an orthogonal matrix can be diagonalized by a unitary transformation.

(a) Show that the sum of the three eigenvalues is $1 + 2\cos\varphi$, where φ is the net angle of rotation about a single fixed axis.
(b) Given that one eigenvalue is 1, show that the other two eigenvalues must be $e^{i\varphi}$ and $e^{-i\varphi}$.

Our orthogonal rotation matrix (real elements) has complex eigenvalues.

3.6.6 A is an nth-order Hermitian matrix with orthonormal eigenvectors $|\mathbf{x}_i\rangle$ and real eigenvalues $\lambda_1 \le \lambda_2 \le \lambda_3 \le \cdots \le \lambda_n$. Show that for a unit magnitude vector $|\mathbf{y}\rangle$,

$$\lambda_1 \le \langle \mathbf{y}|\mathsf{A}|\mathbf{y}\rangle \le \lambda_n.$$

3.6.7 A particular matrix is both Hermitian and unitary. Show that its eigenvalues are all ± 1. *Note.* The Pauli and Dirac matrices are specific examples.

3.6.8 For his relativistic electron theory Dirac required a set of **four** anticommuting matrices. Assume that these matrices are to be Hermitian and unitary. If these are $n \times n$ matrices, show that n must be even. With 2×2 matrices inadequate (why?), this demonstrates that the smallest possible matrices forming a set of four anticommuting, Hermitian, unitary matrices are 4×4.

3.6.9 A is a normal matrix with eigenvalues λ_n and orthonormal eigenvectors $|\mathbf{x}_n\rangle$. Show that A may be written as

$$\mathsf{A} = \sum_n \lambda_n |\mathbf{x}_n\rangle\langle\mathbf{x}_n|.$$

Hint. Show that both this eigenvector form of A and the original A give the same result acting on an arbitrary vector $|\mathbf{y}\rangle$.

3.6.10 A has eigenvalues 1 and -1 and corresponding eigenvectors $\binom{1}{0}$ and $\binom{0}{1}$. Construct A.

$$\text{ANS. } \mathsf{A} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

3.6.11 A non-Hermitian matrix A has eigenvalues λ_i and corresponding eigenvectors $|\mathbf{u}_i\rangle$. The adjoint matrix $\mathsf{A}^\dagger$ has the same set of eigenvalues but **different** corresponding eigenvectors, $|\mathbf{v}_i\rangle$. Show that the eigenvectors form a **biorthogonal** set, in the sense that

$$\langle\mathbf{v}_i|\mathbf{u}_j\rangle = 0 \qquad \text{for} \quad \lambda_i^* \neq \lambda_j.$$

3.6.12 You are given a pair of equations:

$$\mathsf{A}|\mathbf{f}_n\rangle = \lambda_n|\mathbf{g}_n\rangle$$
$$\tilde{\mathsf{A}}|\mathbf{g}_n\rangle = \lambda_n|\mathbf{f}_n\rangle \qquad \text{with } \mathsf{A} \text{ real.}$$

(a) Prove that $|\mathbf{f}_n\rangle$ is an eigenvector of $(\tilde{\mathsf{A}}\mathsf{A})$ with eigenvalue λ_n^2.
(b) Prove that $|\mathbf{g}_n\rangle$ is an eigenvector of $(\mathsf{A}\tilde{\mathsf{A}})$ with eigenvalue λ_n^2.
(c) State how you know that

(1) The $|\mathbf{f}_n\rangle$ form an orthogonal set.
(2) The $|\mathbf{g}_n\rangle$ form an orthogonal set.
(3) λ_n^2 is real.

3.6.13 Prove that A of the preceding exercise may be written as

$$\mathsf{A} = \sum_n \lambda_n |\mathbf{g}_n\rangle\langle\mathbf{f}_n|,$$

with the $|\mathbf{g}_n\rangle$ and $\langle\mathbf{f}_n|$ normalized to unity.
Hint. Expand your arbitrary vector as a linear combination of $|\mathbf{f}_n\rangle$.

3.6.14 Given

$$\mathsf{A} = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 2 \\ 1 & -4 \end{pmatrix},$$

(a) Construct the transpose $\tilde{\mathsf{A}}$ and the symmetric forms $\tilde{\mathsf{A}}\mathsf{A}$ and $\mathsf{A}\tilde{\mathsf{A}}$.
(b) From $\mathsf{A}\tilde{\mathsf{A}}|\mathbf{g}_n\rangle = \lambda_n^2|\mathbf{g}_n\rangle$ find λ_n and $|\mathbf{g}_n\rangle$. Normalize the $|\mathbf{g}_n\rangle$.
(c) From $\tilde{\mathsf{A}}\mathsf{A}|\mathbf{f}_n\rangle = \lambda_n^2|\mathbf{g}_n\rangle$ find λ_n [same as (b)] and $|\mathbf{f}_n\rangle$. Normalize the $|\mathbf{f}_n\rangle$.
(d) Verify that $\mathsf{A}|\mathbf{f}_n\rangle = \lambda_n|\mathbf{g}_n\rangle$ and $\tilde{\mathsf{A}}|\mathbf{g}_n\rangle = \lambda_n|\mathbf{f}_n\rangle$.
(e) Verify that $\mathsf{A} = \sum_n \lambda_n |\mathbf{g}_n\rangle\langle\mathbf{f}_n|$.

3.6.15 Given the eigenvalues $\lambda_1 = 1$, $\lambda_2 = -1$ and the corresponding eigenvectors

$$|\mathbf{f}_1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |\mathbf{g}_1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad |\mathbf{f}_2\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad \text{and} \qquad |\mathbf{g}_2\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

(a) construct A;
(b) verify that $\mathsf{A}|\mathbf{f}_n\rangle = \lambda_n|\mathbf{g}_n\rangle$;
(c) verify that $\tilde{\mathsf{A}}|\mathbf{g}_n\rangle = \lambda_n|\mathbf{f}_n\rangle$.

$$\text{ANS. } \mathsf{A} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

3.6.16 This is a continuation of Exercise 3.4.12, where the unitary matrix U and the Hermitian matrix H are related by

$$\mathsf{U} = e^{ia\mathsf{H}}.$$

(a) If trace $\mathsf{H} = 0$, show that $\det \mathsf{U} = +1$.
(b) If $\det \mathsf{U} = +1$, show that trace $\mathsf{H} = 0$.

Hint. H may be diagonalized by a similarity transformation. Then interpreting the exponential by a Maclaurin expansion, U is also diagonal. The corresponding eigenvalues are given by $u_j = \exp(iah_j)$.
Note. These properties, and those of Exercise 3.4.12, are vital in the development of the concept of generators in group theory — Section 4.2.

3.6.17 An $n \times n$ matrix A has n eigenvalues A_i. If $\mathsf{B} = e^{\mathsf{A}}$, show that B has the same eigenvectors as A, with the corresponding eigenvalues B_i given by $B_i = \exp(A_i)$.
Note. e^{A} is defined by the Maclaurin expansion of the exponential:

$$e^{\mathsf{A}} = 1 + \mathsf{A} + \frac{\mathsf{A}^2}{2!} + \frac{\mathsf{A}^3}{3!} + \cdots.$$

3.6.18 A matrix P is a projection operator (see the discussion following Eq. (3.138c)) satisfying the condition

$$\mathsf{P}^2 = \mathsf{P}.$$

Show that the corresponding eigenvalues $(\rho^2)_\lambda$ and ρ_λ satisfy the relation

$$(\rho^2)_\lambda = (\rho_\lambda)^2 = \rho_\lambda.$$

This means that the eigenvalues of P are 0 and 1.

3.6.19 In the **matrix** eigenvector–eigenvalue equation

$$\mathsf{A}|\mathbf{r}_i\rangle = \lambda_i|\mathbf{r}_i\rangle,$$

A is an $n \times n$ Hermitian matrix. For simplicity assume that its n real eigenvalues are distinct, λ_1 being the largest. If $|\mathbf{r}\rangle$ is an approximation to $|\mathbf{r}_1\rangle$,

$$|\mathbf{r}\rangle = |\mathbf{r}_1\rangle + \sum_{i=2}^{n} \delta_i|\mathbf{r}_i\rangle,$$

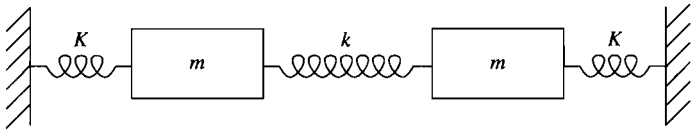

FIGURE 3.8 Triple oscillator.

show that

$$\frac{\langle \mathbf{r}|\mathsf{A}|\mathbf{r}\rangle}{\langle \mathbf{r}|\mathbf{r}\rangle} \leq \lambda_1$$

and that the error in λ_1 is of the order $|\delta_i|^2$. Take $|\delta_i| \ll 1$.
Hint. The n $|\mathbf{r}_i\rangle$ form a **complete** orthogonal set spanning the n-dimensional (complex) space.

3.6.20 Two equal masses are connected to each other and to walls by springs as shown in Fig. 3.8. The masses are constrained to stay on a horizontal line.

(a) Set up the Newtonian acceleration equation for each mass.
(b) Solve the secular equation for the eigenvectors.
(c) Determine the eigenvectors and thus the normal modes of motion.

3.6.21 Given a normal matrix A with eigenvalues λ_j, show that $\mathsf{A}^\dagger$ has eigenvalues λ_j^*, its real part $(\mathsf{A} + \mathsf{A}^\dagger)/2$ has eigenvalues $\Re(\lambda_j)$, and its imaginary part $(\mathsf{A} - \mathsf{A}^\dagger)/2i$ has eigenvalues $\Im(\lambda_j)$.

Additional Readings

Aitken, A. C., *Determinants and Matrices*. New York: Interscience (1956). Reprinted, Greenwood (1983). A readable introduction to determinants and matrices.

Barnett, S., *Matrices: Methods and Applications*. Oxford: Clarendon Press (1990).

Bickley, W. G., and R. S. H. G. Thompson, *Matrices—Their Meaning and Manipulation*. Princeton, NJ: Van Nostrand (1964). A comprehensive account of matrices in physical problems, their analytic properties, and numerical techniques.

Brown, W. C., *Matrices and Vector Spaces*. New York: Dekker (1991).

Gilbert, J. and L., *Linear Algebra and Matrix Theory*. San Diego: Academic Press (1995).

Heading, J., *Matrix Theory for Physicists*. London: Longmans, Green and Co. (1958). A readable introduction to determinants and matrices, with applications to mechanics, electromagnetism, special relativity, and quantum mechanics.

Vein, R., and P. Dale, *Determinants and Their Applications in Mathematical Physics*. Berlin: Springer (1998).

Watkins, D. S., *Fundamentals of Matrix Computations*. New York: Wiley (1991).