CHAPTER 2

DETERMINANTS AND MATRICES

2.1 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):

$$\begin{aligned} a_1x_1 + a_2x_2 + a_3x_3 &= 0, \\ b_1x_1 + b_2x_2 + b_3x_3 &= 0, \\ c_1x_1 + c_2x_2 + c_3x_3 &= 0. \end{aligned} \tag{2.1}$$

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. (2.1), become

$$\mathbf{a}\cdot\mathbf{x} = 0, \quad \mathbf{b}\cdot\mathbf{x} = 0, \quad \mathbf{c}\cdot\mathbf{x} = 0. \tag{2.2}$$

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

Mathematical Methods for Physicists. DOI: 10.1016/B978-0-12-384654-9.00002-5

product, see Eq. (3.12) of Section 3.2)

$$D_3 = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \det(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \tag{2.3}$$

is not zero, then there is only the trivial solution $\mathbf{x} = 0$. For an introduction to the cross product of vectors, see Chapter 3: Vector Analysis, Section 3.2: Vectors in 3-D Space.

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

$$\frac{x_1}{x_3} = \frac{a_2 b_3 - a_3 b_2}{a_1 b_2 - a_2 b_1}, \quad \frac{x_2}{x_3} = -\frac{a_1 b_3 - a_3 b_1}{a_1 b_2 - a_2 b_1} \tag{2.4}$$

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,

$$a_1 x_1 + a_2 x_2 = a_3, \quad b_1 x_1 + b_2 x_2 = b_3, \tag{2.5}$$

can be reduced to the previous case by imbedding it in three-dimensional (3-D) 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. (2.5) 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 Eq. (2.4) 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

$$x_1 = \frac{a_3 b_2 - b_3 a_2}{a_1 b_2 - a_2 b_1} = \frac{\begin{vmatrix} a_3 & a_2 \\ b_3 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}}, \tag{2.6}$$

$$x_2 = \frac{a_1 b_3 - a_3 b_1}{a_1 b_2 - a_2 b_1} = \frac{\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}}{\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}}. \tag{2.7}$$

The determinant in the numerator of $x_1 (x_2)$ is obtained from the determinant of the coefficients $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$ by replacing the first (second) column vector by the vector $\begin{pmatrix} a_3 \\ b_3 \end{pmatrix}$ of the inhomogeneous side of Eq. (2.5). This is **Cramer's rule** for a set of two inhomogeneous linear equations with two unknowns.

A full understanding of the above exposition requires now that we introduce a formal definition of the determinant and show how it relates to the foregoing.

Definitions

Before defining a determinant, we need to introduce some related concepts and definitions.

- When we write two-dimensional (2-D) arrays of items, we identify the item in the nth horizontal row and the mth vertical column by the index set n, m; note that the row index is conventionally written first.
- Starting from a set of n objects in some reference order (e.g., the number sequence $1, 2, 3, \ldots, n$), we can make a **permutation** of them to some other order; the total number of distinct permutations that are possible is $n!$ (choose the first object n ways, then choose the second in $n-1$ ways, etc.).
- Every permutation of n objects can be reached from the reference order by a succession of pairwise interchanges (e.g., $1234 \rightarrow 4132$ can be reached by the successive steps $1234 \rightarrow 1432 \rightarrow 4132$). Although the number of pairwise interchanges needed for a given permutation depends on the path (compare the above example with $1234 \rightarrow 1243 \rightarrow 1423 \rightarrow 4123 \rightarrow 4132$), for a given permutation the number of interchanges will always either be **even** or **odd**. Thus a permutation can be identified as having either even or odd **parity**.
- It is convenient to introduce the **Levi-Civita symbol**, which for an n-object system is denoted by $\varepsilon_{ij\ldots}$, where ε has n subscripts, each of which identifies one of the objects. This Levi-Civita symbol is defined to be $+1$ if $ij\ldots$ represents an even permutation of the objects from a reference order; it is defined to be -1 if $ij\ldots$ represents an odd permutation of the objects, and zero if $ij\ldots$ does not represent a permutation of the objects (i.e., contains an entry duplication). Since this is an important definition, we set it out in a display format:

$$\begin{aligned} \varepsilon_{ij\ldots} &= +1, \quad ij\ldots \text{ an even permutation,} \\ &= -1, \quad ij\ldots \text{ an odd permutation,} \\ &= 0, \quad ij\ldots \text{ not a permutation.} \end{aligned} \tag{2.8}$$

We now define a determinant of **order** n to be an $n \times n$ square array of numbers (or functions), with the array conventionally written within vertical bars (not parentheses, braces, or any other type of brackets), as follows:

$$D_n = \begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ a_{31} & a_{32} & \ldots & a_{3n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix}. \tag{2.9}$$

The determinant D_n has a value that is obtained by

1. Forming all $n!$ products that can be formed by choosing one entry from each row in such a way that one entry comes from each column,
2. Assigning each product a sign that corresponds to the parity of the sequence in which the columns were used (assuming the rows were used in an ascending sequence),
3. Adding (with the assigned signs) the products.

More formally, the determinant in Eq. (2.9) is defined to have the value

$$D_n = \sum_{ij\ldots} \varepsilon_{ij\ldots} a_{1i} a_{2j} \cdots . \tag{2.10}$$

The summations in Eq. (2.10) need not be restricted to permutations, but can be assumed to range independently from 1 through n; the presence of the Levi-Civita symbol will cause only the index combinations corresponding to permutations to actually contribute to the sum.

Example 2.1.1 DETERMINANTS OF ORDERS 2 AND 3

To make the definition more concrete, we illustrate first with a determinant of order 2. The Levi-Civita symbols needed for this determinant are $\varepsilon_{12} = +1$ and $\varepsilon_{21} = -1$ (note that $\varepsilon_{11} = \varepsilon_{22} = 0$), leading to

$$D_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \varepsilon_{12} a_{11} a_{22} + \varepsilon_{21} a_{12} a_{21} = a_{11} a_{22} - a_{12} a_{21}.$$

We see that this determinant expands into $2! = 2$ terms. A specific example of a determinant of order 2 is

$$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = a_1 b_2 - b_1 a_2.$$

Determinants of order 3 expand into $3! = 6$ terms. The relevant Levi-Civita symbols are $\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = +1$, $\varepsilon_{213} = \varepsilon_{321} = \varepsilon_{132} = -1$; all other index combinations have $\varepsilon_{ijk} = 0$, so

$$D_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \sum_{ijk} \varepsilon_{ijk} a_{1i} a_{2j} a_{3k}$$

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

The expression in Eq. (2.3) is the determinant of order 3

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = a_1 b_2 c_3 - a_1 b_3 c_2 - a_2 b_1 c_3 + a_2 b_3 c_1 + a_3 b_1 c_2 - a_3 b_2 c_1.$$

Note that half of the terms in the expansion of a determinant bear negative signs. It is quite possible that a determinant of large elements will have a very small value. Here is one example:

$$\begin{vmatrix} 8 & 11 & 7 \\ 9 & 11 & 5 \\ 8 & 12 & 9 \end{vmatrix} = 1.$$

■

Properties of Determinants

The symmetry properties of the Levi-Civita symbol translate into a number of symmetries exhibited by determinants. For simplicity, we illustrate with determinants of order 3. The interchange of two columns of a determinant causes the Levi-Civita symbol multiplying each term of the expansion to change sign; the same is true if two rows are interchanged. Moreover, the roles of rows and columns may be interchanged; if a determinant with elements a_{ij} is replaced by one with elements $b_{ij} = a_{ji}$, we call the b_{ij} determinant the **transpose** of the a_{ij} determinant. Both these determinants have the same value. Summarizing:

Interchanging two rows (or two columns) changes the sign of the value of a determinant. Transposition does not alter its value.

Thus,

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = -\begin{vmatrix} a_{12} & a_{11} & a_{13} \\ a_{22} & a_{21} & a_{23} \\ a_{32} & a_{31} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{vmatrix}. \tag{2.11}$$

Further consequences of the definition in Eq. (2.10) are:

(1) Multiplication of all members of a single column (or a single row) by a constant k causes the value of the determinant to be multiplied by k,

(2) If the elements of a column (or row) are actually sums of two quantities, the determinant can be decomposed into a sum of two determinants.

Thus,

$$k\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} ka_{11} & a_{12} & a_{13} \\ ka_{21} & a_{22} & a_{23} \\ ka_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} ka_{11} & ka_{12} & ka_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}, \tag{2.12}$$

$$\begin{vmatrix} a_{11}+b_1 & a_{12} & a_{13} \\ a_{21}+b_2 & a_{22} & a_{23} \\ a_{31}+b_3 & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix}. \tag{2.13}$$

These basic properties and/or the basic definition mean that

- Any determinant with two rows equal, or two columns equal, has the value zero. To prove this, interchange the two identical rows or columns; the determinant both remains the same and changes sign, and therefore must have the value zero.
- An extension of the above is that if two rows (or columns) are proportional, the determinant is zero.
- 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. Applying Eq. (2.13), the addition does not contribute to the value of the determinant.
- If each element in a row or each element in a column is zero, the determinant has the value zero.

Laplacian Development by Minors

The fact that a determinant of order n expands into $n!$ terms means that it is important to identify efficient means for determinant evaluation. One approach is to expand in terms of **minors**. The minor corresponding to a_{ij}, denoted M_{ij}, or $M_{ij}(a)$ if we need to identify M as coming from the a_{ij}, is the determinant (of order $n-1$) produced by striking out row i and column j of the original determinant. When we expand into minors, the quantities to be used are the **cofactors** of the (ij) elements, defined as $(-1)^{i+j}M_{ij}$. The expansion can be made for any row or column of the original determinant. If, for example, we expand the determinant of Eq. (2.9) using row i, we have

$$D_n = \sum_{j=1}^{n} a_{ij}(-1)^{i+j}M_{ij}. \tag{2.14}$$

This expansion reduces the work involved in evaluation if the row or column selected for the expansion contains zeros, as the corresponding minors need not be evaluated.

Example 2.1.2 Expansion in Minors

Consider the determinant (arising in Dirac's relativistic electron theory)

$$D \equiv \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix} = \begin{vmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{vmatrix}.$$

Expanding across the top row, only one 3×3 matrix survives:

$$D = (-1)^{1+2}a_{12}M_{12}(a) = (-1)\cdot(1)\begin{vmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{vmatrix} \equiv (-1)\begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix}.$$

Expanding now across the second row, we get

$$D = (-1)(-1)^{2+3}b_{23}M_{23}(b) = \begin{vmatrix} -1 & 0 \\ 0 & -1 \end{vmatrix} = 1.$$

When we finally reached a 2×2 determinant, it was simple to evaluate it without further expansion. ■

Linear Equation Systems

We are now ready to apply our knowledge of determinants to the solution of systems of linear equations. Suppose we have the simultaneous equations

$$\begin{aligned} a_1x_1 + a_2x_2 + a_3x_3 &= h_1, \\ b_1x_1 + b_2x_2 + b_3x_3 &= h_2, \\ c_1x_1 + c_2x_2 + c_3x_3 &= h_3. \end{aligned} \tag{2.15}$$

To use determinants to help solve this equation system, we define

$$D = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}. \tag{2.16}$$

Starting from $x_1 D$, we manipulate it by (1) moving x_1 to multiply the entries of the first column of D, then (2) adding to the first column x_2 times the second column and x_3 times the third column (neither of these operations change the value). We then reach the second line of Eq. (2.17) by substituting the right-hand sides of Eqs. (2.15). These operations are illustrated here:

$$\begin{aligned} x_1 D &= \begin{vmatrix} a_1 x_1 & a_2 & a_3 \\ b_1 x_1 & b_2 & b_3 \\ c_1 x_1 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 x_1 + a_2 x_2 + a_3 x_3 & a_2 & a_3 \\ b_1 x_1 + b_2 x_2 + b_3 x_3 & b_2 & b_3 \\ c_1 x_1 + c_2 x_2 + c_3 x_3 & c_2 & c_3 \end{vmatrix} \\ &= \begin{vmatrix} h_1 & a_2 & a_3 \\ h_2 & b_2 & b_3 \\ h_3 & c_2 & c_3 \end{vmatrix}. \end{aligned} \tag{2.17}$$

If $D \neq 0$, Eq. (2.17) may now be solved for x_1:

$$x_1 = \frac{1}{D} \begin{vmatrix} h_1 & a_2 & a_3 \\ h_2 & b_2 & b_3 \\ h_3 & c_2 & c_3 \end{vmatrix}. \tag{2.18}$$

Analogous procedures starting from $x_2 D$ and $x_3 D$ give the parallel results

$$x_2 = \frac{1}{D} \begin{vmatrix} a_1 & h_1 & a_3 \\ b_1 & h_2 & b_3 \\ c_1 & h_3 & c_3 \end{vmatrix}, \quad x_3 = \frac{1}{D} \begin{vmatrix} a_1 & a_2 & h_2 \\ b_1 & b_2 & h_2 \\ c_1 & c_2 & h_3 \end{vmatrix}.$$

We see that the solution for x_i is $1/D$ times a numerator obtained by replacing the ith column of D by the right-hand-side coefficients, a result that can be generalized to an arbitrary number n of simultaneous equations. This scheme for the solution of linear equation systems is known as **Cramer's rule**.

If D is nonzero, the above construction of the x_i is definitive and unique, so that there will be exactly one solution to the equation set. If $D \neq 0$ and the equations are homogeneous (i.e., all the h_i are zero), then the unique solution is that all the x_i are zero.

Determinants and Linear Dependence

The preceding subsections go a long way toward identifying the role of the determinant with respect to linear dependence. If n linear equations in n variables, written as in Eq. (2.15), have coefficients that form a nonzero determinant, the variables are uniquely determined, meaning that the forms constituting the left-hand sides of the equations must in fact be linearly independent. However, we would still like to prove the property illustrated in the introduction to this chapter, namely that if a set of forms is linearly dependent, the determinant of their coefficients will be zero. But this result is nearly immediate. The existence of linear dependence means that there exists one equation whose coefficients are linear combinations of the coefficients of the other equations, and we may use that fact to reduce to zero the row of the determinant corresponding to that equation.

In summary, we have therefore established the following important result:

If the coefficients of n *linear forms in* n *variables form a nonzero determinant, the forms are linearly independent; if the determinant of the coefficients is zero, the forms exhibit linear dependence.*

Linearly Dependent Equations

If a set of linear forms is linearly dependent, we can distinguish three distinct situations when we consider equation systems based on these forms. First, and of most importance for physics, is the case in which all the equations are **homogeneous**, meaning that the right-hand side quantities h_i in equations of the type Eq. (2.15) are all zero. Then, one or more of the equations in the set will be equivalent to linear combinations of others, and we will have less than n equations in our n variables. We can then assign one (or in some cases, more than one) variable an arbitrary value, obtaining the others as functions of the assigned variables. We thus have a **manifold** (i.e., a parameterized set) of solutions to our equation system.

Combining the above analysis with our earlier observation that if a set of homogeneous linear equations has a nonvanishing determinant it has the unique solution that all the x_i are zero, we have the following important result:

A system of n homogeneous linear equations in n unknowns has solutions that are not identically zero only if the determinant of its coefficients vanishes. If that determinant vanishes, there will be one or more solutions that are not identically zero and are arbitrary as to scale.

A second case is where we have (or combine equations so that we have) the same linear form in two equations, but with different values of the right-hand quantities h_i. In that case the equations are mutually inconsistent, and the equation system has no solution.

A third, related case, is where we have a duplicated linear form, but with a common value of h_i. This also leads to a solution manifold.

Example 2.1.3 LINEARLY DEPENDENT HOMOGENEOUS EQUATIONS

Consider the equation set

$$x_1 + x_2 + x_3 = 0,$$
$$x_1 + 3x_2 + 5x_3 = 0,$$
$$x_1 + 2x_2 + 3x_3 = 0.$$

Here

$$D = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 3 & 5 \\ 1 & 2 & 3 \end{vmatrix} = 1(3)(3) - 1(5)(2) - 1(3)(1) - 1(1)(3) + 1(5)(1) + 1(1)(2) = 0.$$

The third equation is half the sum of the other two, so we drop it. Then,

$$\text{second equation minus first:} \quad 2x_2 + 4x_3 = 0 \longrightarrow x_2 = -2x_3,$$
$$(3\times \text{ first equation) minus second:} \quad 2x_1 - 2x_3 = 0 \longrightarrow x_1 = x_3.$$

Since x_3 can have any value, there is an infinite number of solutions, all of the form $(x_1, x_2, x_3) = \text{constant} \times (1, -2, 1)$.

Our solution illustrates an important property of homogeneous linear equations, namely that any multiple of a solution is also a solution. The solution only becomes less arbitrary if we impose a scale condition. For example, in the present case we could require the squares of the x_i to add to unity. Even then, the solution would still be arbitrary as to overall sign. ■

Numerical Evaluation

There is extensive literature on determinant evaluation. Computer codes and many references are given, for example, by Press *et al.*[1] We present here a straightforward method due to Gauss that illustrates the principles involved in all the modern evaluation methods. **Gauss elimination** is a versatile procedure that can be used for evaluating determinants, for solving linear equation systems, and (as we will see later) even for matrix inversion.

Example 2.1.4 GAUSS ELIMINATION

Our example, a 3×3 linear equation system, can easily be done in other ways, but it is used here to provide an understanding of the Gauss elimination procedure. We wish to solve

$$\begin{aligned} 3x + 2y + z &= 11 \\ 2x + 3y + z &= 13 \\ x + y + 4z &= 12. \end{aligned} \tag{2.19}$$

For convenience and for the optimum numerical accuracy, the equations are rearranged so that, to the extent possible, the largest coefficients run along the main diagonal (upper left to lower right).

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.

[1] 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.

It is convenient to start by dividing each row by its initial coefficient, converting Eq. (2.19) to

$$\begin{aligned} x + \frac{2}{3}y + \frac{1}{3}z &= \frac{11}{3} \\ x + \frac{3}{2}y + \frac{1}{2}z &= \frac{13}{2} \\ x + y + 4z &= 12. \end{aligned} \tag{2.20}$$

Now, using the first equation, we eliminate x from the second and third equations by subtracting the first equation from each of the others:

$$\begin{aligned} x + \frac{2}{3}y + \frac{1}{3}z &= \frac{11}{3} \\ \frac{5}{6}y + \frac{1}{6}z &= \frac{17}{6} \\ \frac{1}{3}y + \frac{11}{3}z &= \frac{25}{3}. \end{aligned} \tag{2.21}$$

Then we divide the second and third rows by their initial coefficients:

$$\begin{aligned} x + \frac{2}{3}y + \frac{1}{3}z &= \frac{11}{3} \\ y + \frac{1}{5}z &= \frac{17}{5} \\ y + 11z &= 25. \end{aligned} \tag{2.22}$$

Repeating the technique, we use the new second equation to eliminate y from the third equation, which can then be solved for z:

$$\begin{aligned} x + \frac{2}{3}y + \frac{1}{3}z &= \frac{11}{3} \\ y + \frac{1}{5}z &= \frac{17}{5} \\ \frac{54}{5}z &= \frac{108}{5} \quad \longrightarrow \quad z = 2. \end{aligned} \tag{2.23}$$

Now that z has been determined, we can return to the second equation, finding

$$y + \frac{1}{5} \times 2 = \frac{17}{5} \quad \longrightarrow \quad y = 3,$$

and finally, continuing to the first equation,

$$x + \frac{2}{3} \times 3 + \frac{1}{3} \times 2 = \frac{11}{3} \quad \longrightarrow \quad x = 1.$$

The technique may not seem as elegant as the use of Cramer's rule, but it is well adapted to computers and is far faster than the time spent with determinants.

If we had not kept the right-hand sides of the equation system, the Gauss elimination process would have simply brought the original determinant into triangular form (but note that our processes for making the leading coefficients unity cause corresponding changes in the value of the determinant). In the present problem, the original determinant

$$D = \begin{vmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{vmatrix}$$

was divided by 3 and by 2 going from Eq. (2.19) to (2.20), and multiplied by 6/5 and by 3 going from Eq. (2.21) to (2.22), so that D and the determinant represented by the left-hand side of Eq. (2.23) are related by

$$D = (3)(2)\left(\frac{5}{6}\right)\left(\frac{1}{3}\right)\begin{vmatrix} 1 & \frac{2}{3} & \frac{1}{3} \\ 0 & 1 & \frac{1}{5} \\ 0 & 0 & \frac{54}{5} \end{vmatrix} = \frac{5}{3}\,\frac{54}{5} = 18\,. \tag{2.24}$$

Because all the entries in the lower triangle of the determinant explicitly shown in Eq. (2.24) are zero, the only term that contributes to it is the product of the diagonal elements: To get a nonzero term, we must use the first element of the first row, then the second element of the second row, etc. It is easy to verify that the final result obtained in Eq. (2.24) agrees with the result of evaluating the original form of D. ■

Exercises

2.1.1 Evaluate the following determinants:

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

2.1.2 Test the set of linear homogeneous equations

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

to see if it possesses a nontrivial solution. In any case, find a solution to this equation set.

2.1.3 Given the pair of equations

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

(a) Show that the determinant of the coefficients vanishes.

(b) Show that the numerator determinants, see Eq. (2.18), also vanish.

(c) Find at least two solutions.

2.1.4 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) $\sum_i a_{ij} C_{ik} = \sum_i a_{ji} C_{ki} = 0,\ j \neq k$.

2.1.5 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	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}.

2.1.6 Prove that the determinant consisting of the coefficients from a set of linearly dependent forms has the value zero.

2.1.7 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, as discussed in Section 2.2.

2.1.8 Show that (in 3-D space)

(a) $\sum_i \delta_{ii} = 3$,

(b) $\sum_{ij} \delta_{ij} \varepsilon_{ijk} = 0$,

(c) $\sum_{pq} \varepsilon_{ipq} \varepsilon_{jpq} = 2\delta_{ij}$,

(d) $\sum_{ijk} \varepsilon_{ijk} \varepsilon_{ijk} = 6$.

Note. The symbol δ_{ij} is the Kronecker delta, defined in Eq. (1.164), and ε_{ijk} is the Levi-Civita symbol, Eq. (2.8).

2.1.9 Show that (in 3-D space)

$$\sum_k \varepsilon_{ijk}\varepsilon_{pqk} = \delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}.$$

Note. See Exercise 2.1.8 for definitions of δ_{ij} and ε_{ijk}.

2.2 Matrices

Matrices are 2-D arrays of numbers or functions that obey the laws that define **matrix algebra**. The subject is important for physics because it facilitates the description of linear transformations such as changes of coordinate systems, provides a useful formulation of quantum mechanics, and facilitates a variety of analyses in classical and relativistic mechanics, particle theory, and other areas. Note also that the development of a mathematics of two-dimensionally ordered arrays is a natural and logical extension of concepts involving ordered pairs of numbers (complex numbers) or ordinary vectors (one-dimensional arrays).

The most distinctive feature of matrix algebra is the rule for the multiplication of matrices. As we will see in more detail later, the algebra is defined so that a set of linear equations such as

$$a_1x_1 + a_2x_2 = h_1$$

$$b_1x_1 + b_2x_2 = h_2$$

can be written as a single matrix equation of the form

$$\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}.$$

In order for this equation to be valid, the multiplication indicated by writing the two matrices next to each other on the left-hand side has to produce the result

$$\begin{pmatrix} a_1x_1 + a_2x_2 \\ b_1x_1 + b_2x_2 \end{pmatrix}$$

and the statement of equality in the equation has to mean element-by-element agreement of its left-hand and right-hand sides. Let's move now to a more formal and precise description of matrix algebra.

Basic Definitions

A **matrix** is a set of numbers or functions in a 2-D square or rectangular array. There are no inherent limitations on the number of rows or columns. A matrix with m (horizontal) rows and n (vertical) columns is known as an $m \times n$ matrix, and the element of a matrix A in row i and column j is known as its i, j element, often labeled a_{ij}. As already observed

$$\begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix} \quad \begin{pmatrix} 4 & 2 \\ -1 & 3 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 6 & 7 & 0 \\ 1 & 4 & 3 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \begin{pmatrix} a_{11} & a_{12} \end{pmatrix}$$

FIGURE 2.1 From left to right, matrices of dimension 4×1 (column vector), 3×2, 2×3, 2×2 (square), 1×2 (row vector).

when we introduced determinants, when row and column indices or dimensions are mentioned together, it is customary to write the row indicator first. Note also that order matters, in general the i, j and j, i elements of a matrix are different, and (if $m \neq n$) $n \times m$ and $m \times n$ matrices even have different shapes. A matrix for which $n = m$ is termed **square**; one consisting of a single column (an $m \times 1$ matrix) is often called a **column vector**, while a matrix with only one row (therefore $1 \times n$) is a **row vector**. We will find that identifying these matrices as vectors is consistent with the properties identified for vectors in Section 1.7.

The arrays constituting matrices are conventionally enclosed in parentheses (not vertical lines, which indicate determinants, or square brackets). A few examples of matrices are shown in Fig. 2.1. We will usually write the symbols denoting matrices as upper-case letters in a sans-serif font (as we did when introducing A); when a matrix is known to be a column vector we often denote it by a lower-case boldface letter in a Roman font (e.g., $\mathbf{x}$).

Perhaps the most important fact to note is that the elements of a matrix are not combined with one another. A matrix is not a determinant. It is an ordered array of numbers, not a single number. To refer to the determinant whose elements are those of a square matrix A (more simply, "the determinant of A"), we can write $\det(\mathsf{A})$.

Matrices, so far just arrays of numbers, have the properties we assign to them. These properties must be specified to complete the definition of matrix algebra.

Equality

If A and B are matrices, $\mathsf{A} = \mathsf{B}$ only if $a_{ij} = b_{ij}$ for all values of i and j. A necessary but not sufficient condition for equality is that both matrices have the same dimensions.

Addition, Subtraction

Addition and subtraction are defined only for matrices A and B of the same dimensions, in which case $\mathsf{A} \pm \mathsf{B} = \mathsf{C}$, with $c_{ij} = a_{ij} \pm b_{ij}$ for all values of i and j, the elements combining according to the law of ordinary algebra (or arithmetic if they are simple numbers). This means that C will be a matrix of the same dimensions as A and B. Moreover, we see that addition is **commutative**: $\mathsf{A} + \mathsf{B} = \mathsf{B} + \mathsf{A}$. It is also **associative**, meaning that $(\mathsf{A} + \mathsf{B}) + \mathsf{C} = \mathsf{A} + (\mathsf{B} + \mathsf{C})$. A matrix with all elements zero, called a **null matrix** or **zero matrix**, can either be written as O or as a simple zero, with its matrix character and dimensions determined from the context. Thus, for all A,

$$\mathsf{A} + 0 = 0 + \mathsf{A} = \mathsf{A}. \tag{2.25}$$

Multiplication (by a Scalar)

Here what we mean by a scalar is an ordinary number or function (not another matrix). The multiplication of matrix A by the scalar quantity α produces $\mathsf{B} = \alpha\mathsf{A}$, with $b_{ij} = \alpha\, a_{ij}$ for all values of i and j. This operation is commutative, with $\alpha\,\mathsf{A} = \mathsf{A}\alpha$.

Note that the definition of multiplication by a scalar causes **each** element of matrix A to be multiplied by the scalar factor. This is in striking contrast to the behavior of determinants in which $\alpha\det(A)$ is a determinant in which the factor α multiplies only one column or one row of $\det(A)$ and not every element of the entire determinant. If A is an $n \times n$ square matrix, then

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

Matrix Multiplication (Inner Product)

Matrix multiplication is not an element-by-element operation like addition or multiplication by a scalar. Instead, it is a more complicated operation in which each element of the product is formed by combining elements of a row of the first operand with corresponding elements of a column of the second operand. This mode of combination proves to be that which is needed for many purposes, and gives matrix algebra its power for solving important problems. This **inner product** of matrices A and B is defined as

$$\mathsf{A}\,\mathsf{B} = \mathsf{C}, \quad \text{with} \quad c_{ij} = \sum_k a_{ik} b_{kj}. \tag{2.26}$$

This definition causes the ij element of C to be formed from the entire ith row of **A** and the entire jth column of **B**. Obviously this definition requires that A have the same number of columns (n) as B has rows. Note that the product will have the same number of rows as A and the same number of columns as B. Matrix multiplication is defined only if these conditions are met. The summation in Eq. (2.26) is over the range of k from 1 to n, and, more explicitly, corresponds to

$$c_{ij} = a_{i1}\, b_{1j} + a_{i2}\, b_{2j} + \cdots + a_{1n}\, b_{nj}.$$

This combination rule is of a form similar to that of the dot product of the vectors $(a_{i1}, a_{i2}, \ldots, a_{in})$ and $(b_{1j}, b_{2j}, \ldots, b_{nj})$. Because the roles of the two operands in a matrix multiplication are different (the first is processed by rows, the second by columns), the operation is in general not commutative, that is, $\mathsf{A}\,\mathsf{B} \neq \mathsf{B}\,\mathsf{A}$. In fact, $\mathsf{A}\,\mathsf{B}$ may even have a different shape than $\mathsf{B}\,\mathsf{A}$. If A and B are square, it is useful to define the **commutator** of A and B,

$$[\mathsf{A}, \mathsf{B}] = \mathsf{A}\,\mathsf{B} - \mathsf{B}\,\mathsf{A}, \tag{2.27}$$

which, as stated above, will in many cases be nonzero.

Matrix multiplication is **associative**, meaning that $(\mathsf{AB})\mathsf{C} = \mathsf{A}(\mathsf{BC})$. Proof of this statement is the topic of Exercise 2.2.26.

Example 2.2.1 MULTIPLICATION, PAULI MATRICES

These three 2×2 matrices, which occurred in early work in quantum mechanics by Pauli, are encountered frequently in physics contexts, so a familiarity with them is highly advisable. They are

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.28}$$

Let's form $\sigma_1\sigma_2$. The $1, 1$ element of the product involves the first **row** of σ_1 and the first **column** of σ_2; these are shaded and lead to the indicated computation:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \rightarrow 0(0) + 1(i) = i.$$

Continuing, we have

$$\sigma_1\sigma_2 = \begin{pmatrix} 0(0)+1(i) & 0(-i)+1(0) \\ 1(0)+0(i) & 1(-i)+0(0) \end{pmatrix} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}. \tag{2.29}$$

In a similar fashion, we can compute

$$\sigma_2\sigma_1 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}. \tag{2.30}$$

It is clear that σ_1 and σ_2 do not commute. We can construct their commutator:

$$\begin{aligned}[]
[\sigma_1, \sigma_2] = \sigma_1\sigma_2 - \sigma_2\sigma_1 &= \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} - \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} \\
&= 2i \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = 2i\sigma_3.
\end{aligned} \tag{2.31}$$

Note that not only have we verified that σ_1 and σ_2 do not commute, we have even evaluated and simplified their commutator. ■

Example 2.2.2 MULTIPLICATION, ROW AND COLUMN MATRICES

As a second example, consider

$$\mathsf{A} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \mathsf{B} = \begin{pmatrix} 4 & 5 & 6 \end{pmatrix}.$$

Let us form AB and BA:

$$\mathsf{AB} = \begin{pmatrix} 4 & 5 & 6 \\ 8 & 10 & 12 \\ 12 & 15 & 18 \end{pmatrix}, \quad \mathsf{BA} = (4 \times 1 + 5 \times 2 + 6 \times 3) = (32).$$

The results speak for themselves. Often when a matrix operation leads to a 1×1 matrix, the parentheses are dropped and the result is treated as an ordinary number or function. ■

Unit Matrix

By direct matrix multiplication, it is possible to show that a square matrix with elements of value unity on its **principal diagonal** (the elements (i, j) with $i = j$), and zeros everywhere else, will leave unchanged any matrix with which it can be multiplied. For example, the 3×3 unit matrix has the form

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

note that it is **not** a matrix all of whose elements are unity. Giving such a matrix the name **1**,

$$\mathbf{1}\mathsf{A} = \mathsf{A}\mathbf{1} = \mathsf{A}. \tag{2.32}$$

In interpreting this equation, we must keep in mind that unit matrices, which are square and therefore of dimensions $n \times n$, exist for all n; the n values for use in Eq. (2.32) must be those consistent with the applicable dimension of A. So if A is $m \times n$, the unit matrix in $\mathbf{1}\mathsf{A}$ must be $m \times m$, while that in $\mathsf{A}\mathbf{1}$ must be $n \times n$.

The previously introduced null matrices have only zero elements, so it is also obvious that for all A,

$$\mathsf{O}\,\mathsf{A} = \mathsf{A}\,\mathsf{O} = \mathsf{O}. \tag{2.33}$$

Diagonal Matrices

If a matrix D has nonzero elements d_{ij} only for $i = j$, it is said to be **diagonal**; a 3×3 example is

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

The rules of matrix multiplication cause all diagonal matrices (of the same size) to commute with each other. However, unless proportional to a unit matrix, diagonal matrices will not commute with nondiagonal matrices containing arbitrary elements.

Matrix Inverse

It will often be the case that given a square matrix A, there will be a square matrix B such that $\mathsf{A}\,\mathsf{B} = \mathsf{B}\,\mathsf{A} = \mathbf{1}$. A matrix B with this property is called the **inverse** of A and is given the name A^{-1}. If A^{-1} exists, it must be unique. The proof of this statement is simple: If B and C are both inverses of A, then

$$\mathsf{A}\,\mathsf{B} = \mathsf{B}\,\mathsf{A} = \mathsf{A}\,\mathsf{C} = \mathsf{C}\,\mathsf{A} = \mathbf{1}.$$

We now look at

$$\mathsf{C}\,\mathsf{A}\,\mathsf{B} = (\mathsf{C}\,\mathsf{A})\mathsf{B} = \mathsf{B}\,, \quad \text{but also} \quad \mathsf{C}\,\mathsf{A}\,\mathsf{B} = \mathsf{C}(\mathsf{A}\,\mathsf{B}) = \mathsf{C}\,.$$

This shows that $\mathsf{B} = \mathsf{C}$.

Every nonzero real (or complex) number α has a nonzero multiplicative inverse, often written $1/\alpha$. But the corresponding property does not hold for matrices; there exist nonzero matrices that do not have inverses. To demonstrate this, consider the following:

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

If A has an inverse, we can multiply the equation $\mathsf{AB} = \mathsf{O}$ **on the left** by A^{-1}, thereby obtaining

$$\mathsf{AB} = \mathsf{O} \quad \longrightarrow \quad \mathsf{A}^{-1}\mathsf{AB} = \mathsf{A}^{-1}\mathsf{O} \quad \longrightarrow \quad \mathsf{B} = \mathsf{O}.$$

Since we started with a matrix B that was nonzero, this is an inconsistency, and we are forced to conclude that A^{-1} does not exist. A matrix without an inverse is said to be **singular**, so our conclusion is that A is singular. Note that in our derivation, we had to be careful to multiply both members of $\mathsf{AB} = \mathsf{O}$ from the left, because multiplication is noncommutative. Alternatively, assuming B^{-1} to exist, we could multiply this equation **on the right** by B^{-1}, obtaining

$$\mathsf{AB} = \mathsf{O} \quad \longrightarrow \quad \mathsf{ABB}^{-1} = \mathsf{OB}^{-1} \quad \longrightarrow \quad \mathsf{A} = \mathsf{O}.$$

This is inconsistent with the nonzero A with which we started; we conclude that B is also singular. Summarizing, there are nonzero matrices that do not have inverses and are identified as singular.

The algebraic properties of real and complex numbers (including the existence of inverses for all nonzero numbers) define what mathematicians call a **field**. The properties we have identified for matrices are different; they form what is called a **ring**.

The numerical inversion of matrices is another topic that has been given much attention, and computer programs for matrix inversion are widely available. A closed, but cumbersome formula for the inverse of a matrix exists; it expresses the elements of A^{-1} in terms of the determinants that are the minors of $\det(\mathsf{A})$; recall that minors were defined in the paragraph immediately before Eq. (2.14). That formula, the derivation of which is in several of the Additional Readings, is

$$(\mathsf{A}^{-1})_{ij} = \frac{(-1)^{i+j} M_{ji}}{\det(\mathsf{A})}. \tag{2.34}$$

We describe here a well-known method that is computationally more efficient than Eq. (2.34), namely the Gauss-Jordan procedure.

Example 2.2.3 Gauss-Jordan Matrix Inversion

The Gauss-Jordan method is based on the fact that there exist matrices M_L such that the product $\mathsf{M}_L\mathsf{A}$ will leave an arbitrary matrix A unchanged, except 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) the interchange of two rows.

The actual matrices M_L that carry out these transformations are the subject of Exercise 2.2.21.

By using these transformations, the rows of a matrix can be altered (by matrix multiplication) in the same ways we were able to change the elements of determinants, so we can proceed in ways similar to those employed for the reduction of determinants by Gauss elimination. If A is nonsingular, the application of a succession of M_L, i.e., $\mathsf{M} = (\ldots \mathsf{M}_L'' \mathsf{M}_L' \mathsf{M}_L)$, can reduce A to a unit matrix:

$$\mathsf{M}\,\mathsf{A} = \mathbf{1}, \quad \text{or} \quad \mathsf{M} = \mathsf{A}^{-1}.$$

Thus, what we need to do is apply successive transformations to A until these transformations have reduced A to $\mathbf{1}$, keeping track of the product of these transformations. The way in which we keep track is to successively apply the transformations to a unit matrix.

Here is a concrete example. We want to invert the matrix

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

Our strategy will be to write, side by side, the matrix A and a unit matrix of the same size, and to perform the same operations on each until A has been converted to a unit matrix, which means that the unit matrix will have been changed to A^{-1}. We start with

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

We multiply the rows as necessary to set to unity all elements of the first column of the left matrix:

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

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

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

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

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

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

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

We can check our work by multiplying the original A by the calculated A^{-1} to see if we really do get the unit matrix **1**. ■

Derivatives of Determinants

The formula giving the inverse of a matrix in terms of its minors enables us to write a compact formula for the derivative of a determinant det(A) where the matrix A has elements that depend on some variable x. To carry out the differentiation with respect to the x dependence of its element a_{ij}, we write det(A) as its expansion in minors M_{ij} about the elements of row i, as in Eq. (2.14), so, appealing also to Eq. (2.34), we have

$$\frac{\partial \det(\mathsf{A})}{\partial a_{ij}} = (-1)^{i+j} M_{ij} = (\mathsf{A}^{-1})_{ji} \det(\mathsf{A}).$$

Applying now the chain rule to allow for the x dependence of all elements of A, we get

$$\frac{d \det(\mathsf{A})}{dx} = \det(\mathsf{A}) \sum_{ij} (\mathsf{A}^{-1})_{ji} \frac{da_{ij}}{dx}. \tag{2.35}$$

Systems of Linear Equations

Using the matrix inverse, we can write down formal solutions to linear equation systems. To start, we note that if A is a $n \times n$ square matrix, and $\mathbf{x}$ and $\mathbf{h}$ are $n \times 1$ column vectors, the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{h}$ is, by the rule for matrix multiplication,

$$\mathsf{A}\mathbf{x} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{pmatrix} = \mathbf{h} = \begin{pmatrix} h_1 \\ h_2 \\ \cdots \\ h_n \end{pmatrix},$$

which is entirely equivalent to a system of n linear equations with the elements of A as coefficients. If A is nonsingular, we can multiply $\mathsf{A}\mathbf{x} = \mathbf{h}$ on the left by A^{-1}, obtaining the result $\mathbf{x} = \mathsf{A}^{-1}\mathbf{h}$.

This result tells us two things: (1) that if we can evaluate A^{-1}, we can compute the solution $\mathbf{x}$; and (2) that the existence of A^{-1} means that this equation system has a unique solution. In our study of determinants we found that a linear equation system had a unique solution if and only if the determinant of its coefficients was nonzero. We therefore see that the condition that A^{-1} exists, i.e., that A is nonsingular, is the same as the condition that the determinant of A, which we write det(A), be nonzero. This result is important enough to be emphasized:

$$\text{A square matrix A is singular if and only if } \det(\mathsf{A}) = 0. \tag{2.36}$$

Determinant Product Theorem

The connection between matrices and their determinants can be made deeper by establishing a **product theorem** which states that the determinant of a product of two $n \times n$ matrices A and B is equal to the products of the determinants of the individual matrices:

$$\det(\mathsf{A}\,\mathsf{B}) = \det(\mathsf{A})\det(\mathsf{B}). \tag{2.37}$$

As an initial step toward proving this theorem, let us look at det(A B) with the elements of the matrix product written out. Showing the first two columns explicitly, we have

$$\det(\mathsf{A}\,\mathsf{B}) = \begin{vmatrix} a_{11}b_{11} + a_{12}b_{21} + \cdots + a_{1n}b_{n1} & a_{11}b_{12} + a_{12}b_{22} + \cdots + a_{1n}b_{n2} & \cdots \\ a_{21}b_{11} + a_{22}b_{21} + \cdots + a_{2n}b_{n1} & a_{21}b_{12} + a_{22}b_{22} + \cdots + a_{2n}b_{n2} & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ a_{n1}b_{11} + a_{n2}b_{21} + \cdots + a_{nn}b_{n1} & a_{n1}b_{12} + a_{n2}b_{22} + \cdots + a_{nn}b_{n2} & \cdots \end{vmatrix}.$$

Introducing the notation

$$\mathbf{a}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \cdots \\ a_{nj} \end{pmatrix}, \quad \text{this becomes} \quad \det(\mathsf{A}\,\mathsf{B}) = \begin{vmatrix} \sum_{j_1} \mathbf{a}_{j_1} b_{j_1,1} & \sum_{j_2} \mathbf{a}_{j_2} b_{j_2,2} & \cdots \end{vmatrix},$$

where the summations over $j_1, j_2, \ldots, j_n$ run independently from 1 though n. Now, calling upon Eqs. (2.12) and (2.13), we can move the summations and the factors b outside the determinant, reaching

$$\det(\mathsf{A}\,\mathsf{B}) = \sum_{j_1}\sum_{j_2}\cdots\sum_{j_n} b_{j_1,1}b_{j_2,2}\cdots b_{j_n,n}\det(\mathbf{a}_{j_1}\mathbf{a}_{j_2}\,\cdots\,\mathbf{a}_{j_n}). \tag{2.38}$$

The determinant on the right-hand side of Eq. (2.38) will vanish if any of the indices j_μ are equal; if all are unequal, that determinant will be $\pm\det(\mathsf{A})$, with the sign corresponding to the parity of the column permutation needed to put the $\mathbf{a}_j$ in numerical order. Both of these conditions are met by writing $\det(\mathbf{a}_{j_1}\mathbf{a}_{j_2} \cdots \mathbf{a}_{j_n}) = \varepsilon_{j_1\ldots j_n} \det(\mathsf{A})$, where ε is the Levi-Civita symbol defined in Eq. (2.8). The above manipulations bring us to

$$\det(\mathsf{A}\,\mathsf{B}) = \det(\mathsf{A}) \sum_{j_1\cdots j_n} \varepsilon_{j_1\ldots j_n} b_{j_1,1} b_{j_2,2} \cdots b_{j_n,n} = \det(\mathsf{A}) \det(\mathsf{B}),$$

where the final step was to invoke the definition of the determinant, Eq. (2.10). This result proves the determinant product theorem.

From the determinant product theorem, we can gain additional insight regarding singular matrices. Noting first that a special case of the theorem is that

$$\det(\mathsf{A}\,\mathsf{A}^{-1}) = \det(\mathbf{1}) = 1 = \det(\mathsf{A}) \det(\mathsf{A}^{-1}),$$

we see that

$$\det(\mathsf{A}^{-1}) = \frac{1}{\det(\mathsf{A})}. \tag{2.39}$$

It is now obvious that if $\det(\mathsf{A}) = 0$, then $\det(\mathsf{A}^{-1})$ cannot exist, meaning that A^{-1} cannot exist either. This is a direct proof that a matrix is singular if and only if it has a vanishing determinant.

Rank of a Matrix

The concept of matrix singularity can be refined by introducing the notion of the **rank** of a matrix. If the elements of a matrix are viewed as the coefficients of a set of linear forms, as in Eq. (2.1) and its generalization to n variables, a square matrix is assigned a rank equal to the number of linearly independent forms that its elements describe. Thus, a nonsingular $n \times n$ matrix will have rank n, while a $n \times n$ singular matrix will have a rank r less than n. The rank provides a measure of the extent of the singularity; if $r = n - 1$, the matrix describes one linear form that is dependent on the others; $r = n - 2$ describes a situation in which there are two forms that are linearly dependent on the others, etc. We will in Chapter 6 take up methods for systematically determining the rank of a matrix.

Transpose, Adjoint, Trace

In addition to the operations we have already discussed, there are further operations that depend on the fact that matrices are arrays. One such operation is transposition. The **transpose** of a matrix is the matrix that results from interchanging its row and column indices. This operation corresponds to subjecting the array to reflection about its principal diagonal. If a matrix is not square, its transpose will not even have the same shape as the original matrix. The transpose of A, denoted $\tilde{\mathsf{A}}$ or sometimes A^T, thus has elements

$$(\tilde{\mathsf{A}})_{ij} = a_{ji}. \tag{2.40}$$

Note that transposition will convert a column vector into a row vector, so

$$\text{if} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}, \quad \text{then} \quad \tilde{\mathbf{x}} = (x_1 \; x_2 \; \dots \; x_n).$$

A matrix that is unchanged by transposition (i.e., $\tilde{\mathsf{A}} = \mathsf{A}$) is called **symmetric**.

For matrices that may have complex elements, the **complex conjugate** of a matrix is defined as the matrix resulting if all elements of the original matrix are complex conjugated. Note that this does not change the shape or move any elements to new positions. The notation for the complex conjugate of A is A^*.

The **adjoint** of a matrix A, denoted $\mathsf{A}^\dagger$, is obtained by both complex conjugating and transposing it (the same result is obtained if these operations are performed in either order). Thus,

$$(\mathsf{A}^\dagger)_{ij} = a^*_{ji}. \tag{2.41}$$

The **trace**, a quantity defined for square matrices, is the sum of the elements on the principal diagonal. Thus, for an $n \times n$ matrix A,

$$\text{trace}(\mathsf{A}) = \sum_{i=1}^{n} a_{ii}. \tag{2.42}$$

From the rule for matrix addition, is is obvious that

$$\text{trace}(\mathsf{A} + \mathsf{B}) = \text{trace}(\mathsf{A}) + \text{trace}(\mathsf{B}). \tag{2.43}$$

Another property of the trace is that its value for a product of two matrices A and B is independent of the order of multiplication:

$$\begin{aligned} \text{trace}(\mathsf{AB}) &= \sum_i (\mathsf{AB})_{ii} = \sum_i \sum_j a_{ij} b_{ji} = \sum_j \sum_i b_{ji} a_{ij} \\ &= \sum_j (\mathsf{BA})_{jj} = \text{trace}(\mathsf{BA}). \end{aligned} \tag{2.44}$$

This holds even if $\mathsf{AB} \neq \mathsf{BA}$. Equation (2.44) means that the trace of any commutator $[\mathsf{A}, \mathsf{B}] = \mathsf{AB} - \mathsf{BA}$ is zero. Considering now the trace of the matrix product ABC, if we group the factors as $\mathsf{A}(\mathsf{BC})$, we easily see that

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

Repeating this process, we also find $\text{trace}(\mathsf{ABC}) = \text{trace}(\mathsf{CAB})$. Note, however, that we cannot equate any of these quantities to $\text{trace}(\mathsf{CBA})$ or to the trace of any other noncyclic permutation of these matrices.

Operations on Matrix Products

We have already seen that the determinant and the trace satisfy the relations

$$\det(\mathsf{AB}) = \det(\mathsf{A})\det(\mathsf{B}) = \det(\mathsf{BA}), \quad \text{trace}(\mathsf{AB}) = \text{trace}(\mathsf{BA}),$$

whether or not A and B commute. We also found that $\text{trace}(\mathsf{A}+\mathsf{B}) = \text{trace}(\mathsf{A}) + \text{trace}(\mathsf{B})$ and can easily show that $\text{trace}(\alpha\mathsf{A}) = \alpha\,\text{trace}(\mathsf{A})$, establishing that the trace is a linear operator (as defined in Chapter 5). Since similar relations do not exist for the determinant, it is **not** a linear operator.

We consider now the effect of other operations on matrix products. The transpose of a product, $(\mathsf{AB})^T$, can be shown to satisfy

$$(\mathsf{AB})^T = \tilde{\mathsf{B}}\tilde{\mathsf{A}}, \tag{2.45}$$

showing that a product is transposed by taking, in reverse order, the transposes of its factors. Note that if the respective dimensions of A and B are such as to make AB defined, it will also be true that $\tilde{\mathsf{B}}\tilde{\mathsf{A}}$ is defined.

Since complex conjugation of a product simply amounts to conjugation of its individual factors, the formula for the adjoint of a matrix product follows a rule similar to Eq. (2.45):

$$(\mathsf{AB})^\dagger = \mathsf{B}^\dagger\mathsf{A}^\dagger. \tag{2.46}$$

Finally, consider $(\mathsf{AB})^{-1}$. In order for AB to be nonsingular, neither A nor B can be singular (to see this, consider their determinants). Assuming this nonsingularity, we have

$$(\mathsf{AB})^{-1} = \mathsf{B}^{-1}\mathsf{A}^{-1}. \tag{2.47}$$

The validity of Eq. (2.47) can be demonstrated by substituting it into the obvious equation $(\mathsf{AB})(\mathsf{AB})^{-1} = \mathbf{1}$.

Matrix Representation of Vectors

The reader may have already noted that the operations of addition and multiplication by a scalar are defined in identical ways for vectors (Section 1.7) and the matrices we are calling column vectors. We can also use the matrix formalism to generate scalar products, but in order to do so we must convert one of the column vectors into a row vector. The operation of transposition provides a way to do this. Thus, letting **a** and **b** stand for vectors in $\mathbb{R}^3$,

$$\mathbf{a}\cdot\mathbf{b} \quad\longrightarrow\quad (a_1\ \ a_2\ \ a_3)\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = a_1b_1 + a_2b_2 + a_3b_3.$$

If in a matrix context we regard **a** and **b** as column vectors, the above equation assumes the form

$$\mathbf{a}\cdot\mathbf{b} \quad\longrightarrow\quad \mathbf{a}^T\mathbf{b}. \tag{2.48}$$

This notation does not really lead to significant ambiguity if we note that when dealing with matrices, we are using lower-case boldface symbols to denote **column vectors**. Note also that because $\mathbf{a}^T\mathbf{b}$ is a 1×1 matrix, it is synonymous with its transpose, which is $\mathbf{b}^T\mathbf{a}$. The matrix notation preserves the symmetry of the dot product. As in Section 1.7, the square of the magnitude of the vector corresponding to $\mathbf{a}$ will be $\mathbf{a}^T\mathbf{a}$.

If the elements of our column vectors $\mathbf{a}$ and $\mathbf{b}$ are real, then an alternate way of writing $\mathbf{a}^T\mathbf{b}$ is $\mathbf{a}^\dagger\mathbf{b}$. But these quantities are not equal if the vectors have complex elements, as will be the case in some situations in which the column vectors do not represent displacements in physical space. In that situation, the dagger notation is the more useful because then $\mathbf{a}^\dagger\mathbf{a}$ will be real and can play the role of a magnitude squared.

Orthogonal Matrices

A real matrix (one whose elements are real) is termed **orthogonal** if its transpose is equal to its inverse. Thus, if S is orthogonal, we may write

$$\mathsf{S}^{-1} = \mathsf{S}^T, \quad \text{or} \quad \mathsf{S}\mathsf{S}^T = \mathbf{1} \quad (\mathsf{S} \text{ orthogonal}). \tag{2.49}$$

Since, for S orthogonal, $\det(\mathsf{S}\mathsf{S}^T) = \det(\mathsf{S})\det(\mathsf{S}^T) = [\det(\mathsf{S})]^2 = 1$, we see that

$$\det(\mathsf{S}) = \pm 1 \quad (\mathsf{S} \text{ orthogonal}). \tag{2.50}$$

It is easy to prove that if S and S' are each orthogonal, then so also are $\mathsf{S}\mathsf{S}'$ and $\mathsf{S}'\mathsf{S}$.

Unitary Matrices

Another important class of matrices consists of matrices U with the property that $\mathsf{U}^\dagger = \mathsf{U}^{-1}$, i.e., matrices for which the adjoint is also the inverse. Such matrices are identified as **unitary**. One way of expressing this relationship is

$$\mathsf{U}\;\mathsf{U}^\dagger = \mathsf{U}^\dagger\mathsf{U} = \mathbf{1} \quad (\mathsf{U} \text{ unitary}). \tag{2.51}$$

If all the elements of a unitary matrix are real, the matrix is also orthogonal.

Since for any matrix $\det(\mathsf{A}^T) = \det(\mathsf{A})$, and therefore $\det(\mathsf{A}^\dagger) = \det(\mathsf{A})^*$, application of the determinant product theorem to a unitary matrix U leads to

$$\det(\mathsf{U})\det(\mathsf{U}^\dagger) = |\det(\mathsf{U})|^2 = 1, \tag{2.52}$$

showing that $\det(\mathsf{U})$ is a possibly complex number of magnitude unity. Since such numbers can be written in the form $\exp(i\theta)$, with θ real, the determinants of U and $\mathsf{U}^\dagger$ will, for some θ, satisfy

$$\det(\mathsf{U}) = e^{i\theta}, \quad \det(\mathsf{U}^\dagger) = e^{-i\theta}.$$

Part of the significance of the term *unitary* is associated with the fact that the determinant has unit magnitude. A special case of this relationship is our earlier observation that if U is real, and therefore also an orthogonal matrix, its determinant must be either $+1$ or -1.

Finally, we observe that if U and V are both unitary, then $\mathsf{U}\mathsf{V}$ and $\mathsf{V}\mathsf{U}$ will be unitary as well. This is a generalization of our earlier result that the matrix product of two orthogonal matrices is also orthogonal.

Hermitian Matrices

There are additional classes of matrices with useful characteristics. A matrix is identified as **Hermitian**, or, synonymously, **self-adjoint**, if it is equal to its adjoint. To be self-adjoint, a matrix H must be square, and in addition, its elements must satisfy

$$(\mathsf{H}^\dagger)_{ij} = (\mathsf{H})_{ij} \quad \longrightarrow \quad h_{ji}^* = h_{ij} \quad (\mathsf{H} \text{ is Hermitian}). \tag{2.53}$$

This condition means that the array of elements in a self-adjoint matrix exhibits a reflection symmetry about the principal diagonal: elements whose positions are connected by reflection must be complex conjugates. As a corollary to this observation, or by direct reference to Eq. (2.53), we see that the diagonal elements of a self-adjoint matrix must be real.

If all the elements of a self-adjoint matrix are real, then the condition of self-adjointness will cause the matrix also to be symmetric, so all real, symmetric matrices are self-adjoint (Hermitian).

Note that if two matrices A and B are Hermitian, it is not necessarily true that AB or BA is Hermitian; however, $\mathsf{AB} + \mathsf{BA}$, if nonzero, will be Hermitian, and $\mathsf{AB} - \mathsf{BA}$, if nonzero, will be **anti-Hermitian**, meaning that $(\mathsf{AB} - \mathsf{BA})^\dagger = -(\mathsf{AB} - \mathsf{BA})$.

Extraction of a Row or Column

It is useful to define column vectors $\hat{\mathbf{e}}_i$ which are zero except for the $(i, 1)$ element, which is unity; examples are

$$\hat{\mathbf{e}}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \cdots \\ 0 \end{pmatrix}, \quad \hat{\mathbf{e}}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \cdots \\ 0 \end{pmatrix}, \quad \text{etc.} \tag{2.54}$$

One use of these vectors is to extract a single column from a matrix. For example, if A is a 3×3 matrix, then

$$\mathsf{A}\hat{\mathbf{e}}_2 = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix}.$$

The row vector $\hat{\mathbf{e}}_i^T$ can be used in a similar fashion to extract a row from an arbitrary matrix, as in

$$\hat{\mathbf{e}}_i^T \mathsf{A} = (a_{i1} \;\; a_{i2} \;\; a_{i3}).$$

These **unit vectors** will also have many uses in other contexts.

Direct Product

A second procedure for multiplying matrices, known as the **direct** tensor or Kronecker **product**, combines a $m \times n$ matrix A and a $m' \times n'$ matrix B to make the direct product matrix $\mathsf{C} = \mathsf{A} \otimes \mathsf{B}$, which is of dimension $mm' \times nn'$ and has elements

$$C_{\alpha\beta} = A_{ij} B_{kl}, \tag{2.55}$$

with $\alpha = m'(i-1) + k$, $\beta = n'(j-1) + l$. The direct product matrix uses the indices of the first factor as major and those of the second factor as minor; it is therefore a noncommutative process. It is, however, associative.

Example 2.2.4 Direct Products

We give some specific examples. If A and B are both 2×2 matrices, we may write, first in a somewhat symbolic and then in a completely expanded form,

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

Another example is the direct product of two two-element column vectors, $\mathbf{x}$ and $\mathbf{y}$. Again writing first in symbolic, and then expanded form,

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \otimes \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1\mathbf{y} \\ x_2\mathbf{y} \end{pmatrix} = \begin{pmatrix} x_1 y_1 \\ x_1 y_2 \\ x_2 y_1 \\ x_2 y_2 \end{pmatrix}.$$

A third example is the quantity AB from Example 2.2.2. It is an instance of the special case (column vector times row vector) in which the direct and inner products coincide: $\mathsf{AB} = \mathsf{A} \otimes \mathsf{B}$. ■

If C and C' are direct products of the respective forms

$$\mathsf{C} = \mathsf{A} \otimes \mathsf{B} \quad \text{and} \quad \mathsf{C}' = \mathsf{A}' \otimes \mathsf{B}', \tag{2.56}$$

and these matrices are of dimensions such that the matrix inner products AA' and BB' are defined, then

$$\mathsf{CC}' = (\mathsf{AA}') \otimes (\mathsf{BB}'). \tag{2.57}$$

Moreover, if matrices A and B are of the same dimensions, then

$$\mathsf{C} \otimes (\mathsf{A} + \mathsf{B}) = \mathsf{C} \otimes \mathsf{A} + \mathsf{C} \otimes \mathsf{B} \quad \text{and} \quad (\mathsf{A} + \mathsf{B}) \otimes \mathsf{C} = \mathsf{A} \otimes \mathsf{C} + \mathsf{B} \otimes \mathsf{C}. \tag{2.58}$$

Example 2.2.5 Dirac Matrices

In the original, nonrelativistic formulation of quantum mechanics, agreement between theory and experiment for electronic systems required the introduction of the concept of electron spin (intrinsic angular momentum), both to provide a doubling in the number of available states and to explain phenomena involving the electron's magnetic moment. The concept was introduced in a relatively *ad hoc* fashion; the electron needed to be given spin quantum number 1/2, and that could be done by assigning it a two-component wave function, with the spin-related properties described using the Pauli matrices, which were introduced in Example 2.2.1:

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

Of relevance here is the fact that these matrices anticommute and have squares that are unit matrices:

$$\sigma_i^2 = \mathbf{1}_2, \quad \text{and} \quad \sigma_i \sigma_j + \sigma_j \sigma_i = 0, \quad i \neq j. \tag{2.59}$$

In 1927, P. A. M. Dirac developed a relativistic formulation of quantum mechanics applicable to spin-1/2 particles. To do this it was necessary to place the spatial and time variables on an equal footing, and Dirac proceeded by converting the relativistic expression for the kinetic energy to an expression that was first order in both the energy and the momentum (parallel quantities in relativistic mechanics). He started from the relativistic equation for the energy of a free particle,

$$E^2 = (p_1^2 + p_2^2 + p_3^2)c^2 + m^2c^4 = \mathbf{p}^2c^2 + m^2c^4, \tag{2.60}$$

where p_i are the components of the momentum in the coordinate directions, m is the particle mass, and c is the velocity of light. In the passage to quantum mechanics, the quantities p_i are to be replaced by the differential operators $-i\hbar\partial/\partial x_i$, and the entire equation is applied to a wave function.

It was desirable to have a formulation that would yield a two-component wave function in the nonrelativistic limit and therefore might be expected to contain the σ_i. Dirac made the observation that a key to the solution of his problem was to exploit the fact that the Pauli matrices, taken together as a vector

$$\boldsymbol{\sigma} = \sigma_1\hat{\mathbf{e}}_1 + \sigma_2\hat{\mathbf{e}}_2 + \sigma_3\hat{\mathbf{e}}_3, \tag{2.61}$$

could be combined with the vector $\mathbf{p}$ to yield the identity

$$(\boldsymbol{\sigma} \cdot \mathbf{p})^2 = \mathbf{p}^2\mathbf{1}_2, \tag{2.62}$$

where $\mathbf{1}_2$ denotes a 2×2 unit matrix. The importance of Eq. (2.62) is that, at the price of going to 2×2 matrices, we can linearize the quadratic occurrences of E and $\mathbf{p}$ in Eq. (2.60) as follows. We first write

$$E^2\mathbf{1}_2 - c^2(\boldsymbol{\sigma} \cdot \mathbf{p})^2 = m^2c^4\mathbf{1}_2. \tag{2.63}$$

We then factor the left-hand side of Eq. (2.63) and apply both sides of the resulting equation (which is a 2×2 matrix equation) to a two-component wave function that we will call ψ_1:

$$(E\mathbf{1}_2 + c\,\boldsymbol{\sigma} \cdot \mathbf{p})(E\mathbf{1}_2 - c\,\boldsymbol{\sigma} \cdot \mathbf{p})\psi_1 = m^2c^4\psi_1. \tag{2.64}$$

The meaning of this equation becomes clearer if we make the additional definition

$$(E\mathbf{1}_2 - c\,\boldsymbol{\sigma} \cdot \mathbf{p})\psi_1 = mc^2\psi_2. \tag{2.65}$$

Substituting Eq. (2.65) into Eq. (2.64), we can then write the modified Eq. (2.64) and the (unchanged) Eq. (2.65) as the equation set

$$\begin{aligned}(E\mathbf{1}_2 + c\,\boldsymbol{\sigma}\cdot\mathbf{p})\psi_2 &= mc^2\psi_1,\\ (E\mathbf{1}_2 - c\,\boldsymbol{\sigma}\cdot\mathbf{p})\psi_1 &= mc^2\psi_2;\end{aligned} \tag{2.66}$$

both these equations will need to be satisfied simultaneously.

To bring Eqs. (2.66) to the form actually used by Dirac, we now make the substitution $\psi_1 = \psi_A + \psi_B$, $\psi_2 = \psi_A - \psi_B$, and then add and subtract the two equations from each other, reaching a set of coupled equations in ψ_A and ψ_B:

$$\begin{aligned}E\psi_A - c\boldsymbol{\sigma}\cdot\mathbf{p}\psi_B &= mc^2\psi_A,\\ c\boldsymbol{\sigma}\cdot\mathbf{p}\psi_A - E\psi_B &= mc^2\psi_B.\end{aligned}$$

In anticipation of what we will do next, we write these equations in the matrix form

$$\left[\begin{pmatrix} E\mathbf{1}_2 & 0\\ 0 & -E\mathbf{1}_2\end{pmatrix} - \begin{pmatrix} 0 & c\boldsymbol{\sigma}\cdot\mathbf{p}\\ -c\boldsymbol{\sigma}\cdot\mathbf{p} & 0\end{pmatrix}\right]\begin{pmatrix}\psi_A\\ \psi_B\end{pmatrix} = mc^2\begin{pmatrix}\psi_A\\ \psi_B\end{pmatrix}. \tag{2.67}$$

We can now use the direct product notation to condense Eq. (2.67) into the simpler form

$$\left[(\sigma_3\otimes\mathbf{1}_2)E - \gamma\otimes c(\boldsymbol{\sigma}\cdot\mathbf{p})\right]\Psi = mc^2\Psi, \tag{2.68}$$

where Ψ is the **four-component** wave function built from the two-component wave functions:

$$\Psi = \begin{pmatrix}\psi_A\\ \psi_B\end{pmatrix},$$

and the terms on the left-hand side have the indicated structure because

$$\sigma_3 = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \quad \text{and we define} \quad \gamma = \begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}. \tag{2.69}$$

It has become customary to identify the matrices in Eq. (2.68) as γ^μ and to refer to them as **Dirac matrices**, with

$$\gamma^0 = \sigma_3\otimes\mathbf{1}_2 = \begin{pmatrix}\mathbf{1}_2 & 0\\ 0 & -\mathbf{1}_2\end{pmatrix} = \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & -1 & 0\\ 0 & 0 & 0 & -1\end{pmatrix}. \tag{2.70}$$

The matrices resulting from the individual components of $\boldsymbol{\sigma}$ in Eq. (2.68) are (for $i = 1, 2, 3$)

$$\gamma^i = \gamma\otimes\sigma_i = \begin{pmatrix}0 & \sigma_i\\ -\sigma_i & 0\end{pmatrix}. \tag{2.71}$$

Expanding Eq. (2.71), we have

$$\gamma^1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}, \quad \gamma^2 = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & i & 0 & 0 \\ -i & 0 & 0 & 0 \end{pmatrix},$$

$$\gamma^3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}. \tag{2.72}$$

Now that the γ^μ have been defined, we can rewrite Eq. (2.68), expanding $\boldsymbol{\sigma} \cdot \mathbf{p}$ into components:

$$\left[\gamma^0 E - c(\gamma^1 p_1 + \gamma^2 p_2 + \gamma^3 p_3)\right] \Psi = mc^2 \Psi.$$

To put this matrix equation into the specific form known as the **Dirac equation** we multiply both sides of it (on the left) by γ^0. Noting that $(\gamma^0)^2 = \mathbf{1}$ and giving $\gamma^0\gamma^i$ the new name α_i, we reach

$$\left[\gamma^0 mc^2 + c(\alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3)\right] \Psi = E\Psi. \tag{2.73}$$

Equation (2.73) is in the notation used by Dirac with the exception that he used β as the name for the matrix here called γ^0.

The Dirac gamma matrices have an algebra that is a generalization of that exhibited by the Pauli matrices, where we found that the $\sigma_i^2 = 1$ and that if $i \neq j$, then σ_i and σ_j anticommute. Either by further analysis or by direct evaluation, it is found that, for $\mu = 0, 1, 2, 3$ and $i = 1, 2, 3$,

$$(\gamma^0)^2 = 1, \quad (\gamma^i)^2 = -1, \tag{2.74}$$

$$\gamma^\mu \gamma^i + \gamma^i \gamma^\mu = 0, \quad \mu \neq i. \tag{2.75}$$

In the nonrelativistic limit, the four-component Dirac equation for an electron reduces to a two-component equation in which each component satisfies the Schrödinger equation, with the Pauli and Dirac matrices having completely disappeared. See Exercise 2.2.48. In this limit, the Pauli matrices reappear if we add to the Schrödinger equation an additional term arising from the intrinsic magnetic moment of the electron. The passage to the nonrelativistic limit provides justification for the seemingly arbitrary introduction of a two-component wavefunction and use of the Pauli matrices for discussions of spin angular momentum.

The Pauli matrices (and the unit matrix $\mathbf{1}_2$) form what is known as a **Clifford algebra**,[2] with the properties shown in Eq. (2.59). Since the algebra is based on 2×2 matrices, it can have only four members (the number of linearly independent such matrices), and is said to be of dimension 4. The Dirac matrices are members of a Clifford algebra of dimension 16. A complete basis for this Clifford algebra with convenient Lorentz transformation

[2] D. Hestenes, *Am. J. Phys.* **39**: 1013 (1971); and *J. Math. Phys.* **16**: 556 (1975).

properties consists of the 16 matrices

$$\mathbf{1}_4, \quad \boldsymbol{\gamma}^5 = i\boldsymbol{\gamma}^0\boldsymbol{\gamma}^1\boldsymbol{\gamma}^2\boldsymbol{\gamma}^3 = \begin{pmatrix} 0 & \mathbf{1}_2 \\ \mathbf{1}_2 & 0 \end{pmatrix}, \quad \boldsymbol{\gamma}^\mu \quad (\mu = 0, 1, 2, 3),$$
$$\boldsymbol{\gamma}^5\boldsymbol{\gamma}^\mu \quad (\mu = 0, 1, 2, 3), \quad \sigma^{\mu\nu} = i\boldsymbol{\gamma}^\mu\boldsymbol{\gamma}^\nu \quad (0 \le \mu < \nu \le 3). \tag{2.76}$$

■

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 converges for each matrix element. For example, if A is any $n \times n$ matrix, then the power series

$$\exp(\mathsf{A}) = \sum_{j=0}^{\infty} \frac{1}{j!}\mathsf{A}^j, \tag{2.77}$$

$$\sin(\mathsf{A}) = \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j+1)!}\mathsf{A}^{2j+1}, \tag{2.78}$$

$$\cos(\mathsf{A}) = \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j)!}\mathsf{A}^{2j} \tag{2.79}$$

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,

$$\exp(i\sigma_k\theta) = \mathbf{1}_2 \cos\theta + i\sigma_k \sin\theta, \tag{2.80}$$

follows from collecting all even and odd powers of θ in separate series using $\sigma_k^2 = 1$. For the 4×4 Dirac matrices $\sigma^{\mu\nu}$, defined in Eq. (2.76), we have for $1 \le \mu < \nu \le 3$,

$$\exp(i\sigma^{\mu\nu}\theta) = \mathbf{1}_4 \cos\theta + i\sigma^{\mu\nu} \sin\theta, \tag{2.81}$$

while

$$\exp(i\sigma^{0k}\zeta) = \mathbf{1}_4 \cosh\zeta + i\sigma^{0k} \sinh\zeta \tag{2.82}$$

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

Hermitian and unitary matrices are related in that U, given as

$$\mathsf{U} = \exp(i\mathsf{H}), \tag{2.83}$$

is unitary if H is Hermitian. To see this, just take the adjoint: $\mathsf{U}^\dagger = \exp(-i\mathsf{H}^\dagger) = \exp(-i\mathsf{H}) = [\exp(i\mathsf{H})]^{-1} = \mathsf{U}^{-1}$.

Another result which is important to identify here is that any Hermitian matrix H satisfies a relation known as the **trace formula**,

$$\det(\exp(\mathsf{H})) = \exp(\mathrm{trace}(\mathsf{H})). \tag{2.84}$$

This formula is derived at Eq. (6.27).

Finally, we note that the multiplication of two diagonal matrices produces a matrix that is also diagonal, with elements that are the products of the corresponding elements of the multiplicands. This result implies that an arbitrary function of a diagonal matrix will also be diagonal, with diagonal elements that are that function of the diagonal elements of the original matrix.

Example 2.2.6 EXPONENTIAL OF A DIAGONAL MATRIX

If a matrix A is diagonal, then its nth power is also diagonal, with the original diagonal matrix elements raised to the nth power. For example, given

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

then

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

We can now compute

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

■

A final and important result is the **Baker-Hausdorff formula**, which, among other places is used in the coupled-cluster expansions that yield highly accurate electronic structure calculations on atoms and molecules[3]:

$$\exp(-\mathsf{T})\mathsf{A}\exp(\mathsf{T}) = \mathsf{A} + [\mathsf{A},\mathsf{T}] + \frac{1}{2!}[[\mathsf{A},\mathsf{T}],\mathsf{T}] + \frac{1}{3!}[[[\mathsf{A},\mathsf{T}],\mathsf{T}],\mathsf{T}] + \cdots. \tag{2.85}$$

Exercises

2.2.1 Show that matrix multiplication is associative, $(\mathsf{AB})\mathsf{C} = \mathsf{A}(\mathsf{BC})$.

2.2.2 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] F. E. Harris, H. J. Monkhorst, and D. L. Freeman, *Algebraic and Diagrammatic Methods in Many-Fermion Theory*. New York: Oxford University Press (1992).

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

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

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

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

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

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

2.2.5 (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.

2.2.6 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, $n \neq 0$).

2.2.7 Verify the **Jacobi identity**,

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

2.2.8 Show that the matrices

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

satisfy the commutation relations

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

2.2.9 Let

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

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) $\mathsf{i}^2 = \mathsf{j}^2 = \mathsf{k}^2 = -\mathbf{1}$, where $\mathbf{1}$ 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 $\mathbf{1}$ 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 $\mathbf{1}$, where the σ_i are the Pauli spin matrices of Example 2.2.1.

2.2.10 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. Show that the product of two upper right triangular matrices is an upper right triangular matrix.

2.2.11 The three Pauli spin matrices are

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

Show that

(a) $(\sigma_i)^2 = 1_2$,

(b) $\sigma_i \sigma_j = i\sigma_k$, $(i, j, k) = (1, 2, 3)$ or a cyclic permutation thereof,

(c) $\sigma_i \sigma_j + \sigma_j \sigma_i = 2\delta_{ij}\mathbf{1}_2$; $\mathbf{1}_2$ is the 2×2 unit matrix.

2.2.12 One description of spin-1 particles uses the matrices

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

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 (cyclic permutation of indices). Using the Levi-Civita symbol, we may write

$$[\mathsf{M}_i, \mathsf{M}_j] = i\sum_k \varepsilon_{ijk}\mathsf{M}_k.$$

(b) $\mathsf{M}^2 \equiv \mathsf{M}_x^2 + \mathsf{M}_y^2 + \mathsf{M}_z^2 = 2\,\mathbf{1}_3$, where $\mathbf{1}_3$ 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 $\mathsf{L}^+ \equiv \mathsf{M}_x + i\mathsf{M}_y$ and $\mathsf{L}^- \equiv \mathsf{M}_x - i\mathsf{M}_y$.

2.2.13 Repeat Exercise 2.2.12, using the matrices for a spin of 3/2,

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

and

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

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

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

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

2.2.17 Angular momentum matrices satisfy a commutation relation

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

Show that the trace of each angular momentum matrix vanishes.

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

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

2.2.20 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 $|\mathsf{A}| \neq 0$).

2.2.21 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 ($a_{ij} \to a_{mj}$, $a_{mj} \to a_{ij}$, $j = 1, 2, 3, \ldots$).

2.2.22 Find the matrices M_R such that the product AM_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 ($a_{ji} \to a_{jm}$, $a_{jm} \to a_{ji}$, $j = 1, 2, 3, \ldots$).

2.2.23 Find the inverse of

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

2.2.24 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.

2.2.25 Given a 6×6 matrix A with elements $a_{ij} = 0.5^{|i-j|}$, $i, j = 0, 1, 2, \ldots, 5$, find A^{-1}.

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

2.2.26 Show that the product of two orthogonal matrices is orthogonal.

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

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

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

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

2.2.30 Show that

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

2.2.31 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.

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

2.2.33 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 $(\mathsf{C}) = 0$.

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

2.2.35 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,

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

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 $y = (z - z^*)/2i$.

2.2.36 A and B are two noncommuting Hermitian matrices:

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

Prove that C is Hermitian.

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

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

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

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

2.2.40 If $\boldsymbol{\sigma}$ is the vector with the σ_i as components given in Eq. (2.61), and $\mathbf{p}$ is an ordinary vector, show that

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

where $\mathbf{1}_2$ is a 2×2 unit matrix.

2.2.41 Use the equations for the properties of direct products, Eqs. (2.57) and (2.58), to show that the four matrices γ^μ, $\mu = 0, 1, 2, 3$, satisfy the conditions listed in Eqs. (2.74) and (2.75).

2.2.42 Show that $\boldsymbol{\gamma}^5$, Eq. (2.76), anticommutes with all four $\boldsymbol{\gamma}^\mu$.

2.2.43 In this problem, the summations are over $\mu = 0, 1, 2, 3$. Define $g_{\mu\nu} = g^{\mu\nu}$ by the relations

$$g_{00} = 1; \quad g_{kk} = -1, \quad k = 1, 2, 3; \quad g_{\mu\nu} = 0, \quad \mu \neq \nu;$$

and define γ_μ as $\sum g_{\nu\mu} \gamma^\mu$. Using these definitions, show that

(a) $\sum \gamma_\mu \gamma^\alpha \gamma^\mu = -2\gamma^\alpha$,

(b) $\sum \gamma_\mu \gamma^\alpha \gamma^\beta \gamma^\mu = 4 g^{\alpha\beta}$,

(c) $\sum \gamma_\mu \gamma^\alpha \gamma^\beta \gamma^\nu \gamma^\mu = -2\gamma^\nu \gamma^\beta \gamma^\alpha$.

2.2.44 If $\mathsf{M} = \frac{1}{2}(1 + \boldsymbol{\gamma}^5)$, where $\boldsymbol{\gamma}^5$ is given in Eq. (2.76), show that

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

Note that this equation is still satisfied if $\boldsymbol{\gamma}$ is replaced by any other Dirac matrix listed in Eq. (2.76).

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

2.2.46 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 (which we denote here as Γ_i)

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

show that

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

2.2.47 The matrix $\mathsf{C} = i\boldsymbol{\gamma}^2\boldsymbol{\gamma}^0$ is sometimes called the charge conjugation matrix. Show that $\mathsf{C}\boldsymbol{\gamma}^\mu\mathsf{C}^{-1} = -(\boldsymbol{\gamma}^\mu)^T$.

2.2.48 (a) Show that, by substitution of the definitions of the γ^μ matrices from Eqs. (2.70) and (2.72), that the Dirac equation, Eq. (2.73), takes the following form when written as 2×2 blocks (with ψ_L and ψ_S column vectors of dimension 2). Here

L and S stand, respectively, for "large" and "small" because of their relative size in the nonrelativistic limit):

$$\begin{pmatrix} mc^2 - E & c(\sigma_1 p_1 + \sigma_2 p_2 + \sigma_3 p_3) \\ -c(\sigma_1 p_1 + \sigma_2 p_2 + \sigma_3 p_3) & -mc^2 - E \end{pmatrix} \begin{pmatrix} \psi_L \\ \psi_S \end{pmatrix} = 0.$$

(b) To reach the nonrelativistic limit, make the substitution $E = mc^2 + \varepsilon$ and approximate $-2mc^2 - \varepsilon$ by $-2mc^2$. Then write the matrix equation as two simultaneous two-component equations and show that they can be rearranged to yield

$$\frac{1}{2m}\left(p_1^2 + p_2^2 + p_3^2\right)\psi_L = \varepsilon\psi_L,$$

which is just the Schrödinger equation for a free particle.

(c) Explain why is it reasonable to call ψ_L and ψ_S "large" and "small."

2.2.49 Show that it is consistent with the requirements that they must satisfy to take the Dirac gamma matrices to be (in 2×2 block form)

$$\gamma^0 = \begin{pmatrix} 0 & \mathbf{1}_2 \\ \mathbf{1}_2 & 0 \end{pmatrix}, \quad \gamma^i = \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix}, \quad (i = 1, 2, 3).$$

This choice for the gamma matrices is called the **Weyl representation**.

2.2.50 Show that the Dirac equation separates into independent 2×2 blocks in the Weyl representation (see Exercise 2.2.49) in the limit that the mass m approaches zero. This observation is important in the ultra relativistic regime where the rest mass is inconsequential, or for particles of negligible mass (e.g., neutrinos).

2.2.51 (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 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.

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. Gilbert, *Linear Algebra and Matrix Theory*. San Diego: Academic Press (1995).

Golub, G. H., and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore: JHU Press (1996). Detailed mathematical background and algorithms for the production of numerical software, including methods for parallel computation. A classic computer science text.

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).