3

Matrix algebra

As vector methods have become standard tools for physicists, so too matrix methods are becoming very useful tools in sciences and engineering. Matrices occur in physics in at least two ways: in handling the eigenvalue problems in classical and quantum mechanics, and in the solutions of systems of linear equations. In this chapter, we introduce matrices and related concepts, and define some basic matrix algebra. In Chapter 5 we will discuss various operations with matrices in dealing with transformations of vectors in vector spaces and the operation of linear operators on vector spaces.

Definition of a matrix

A matrix consists of a rectangular block or ordered array of numbers that obeys prescribed rules of addition and multiplication. The numbers may be real or complex. The array is usually enclosed within curved brackets. Thus

$$\begin{pmatrix} 1 & 2 & 4 \\ 2 & -1 & 7 \end{pmatrix}$$

is a matrix consisting of 2 rows and 3 columns, and it is called a 2×3 (2 by 3) matrix. An $m \times n$ matrix consists of m rows and n columns, which is usually expressed in a double suffix notation:

$$\tilde{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix}. \tag{3.1}$$

Each number a_{ij} is called an element of the matrix, where the first subscript i denotes the row, while the second subscript j indicates the column. Thus, a_{23}

https://doi.org/10.1017/CBO9780511755781.004 Published online by Cambridge University Press

refers to the element in the second row and third column. The element a_{ij} should be distinguished from the element a_{ji}.

It should be pointed out that a matrix has no single numerical value; therefore it must be carefully distinguished from a determinant.

We will denote a matrix by a letter with a tilde over it, such as $\tilde{A}$ in (3.1). Sometimes we write (a_{ij}) or $(a_{ij})_{mn}$, if we wish to express explicitly the particular form of element contained in $\tilde{A}$.

Although we have defined a matrix here with reference to numbers, it is easy to extend the definition to a matrix whose elements are functions $f_i(x)$; for a 2×3 matrix, for example, we have

$$\begin{pmatrix} f_1(x) & f_2(x) & f_3(x) \\ f_4(x) & f_5(x) & f_6(x) \end{pmatrix}.$$

A matrix having only one row is called a row matrix or a row vector, while a matrix having only one column is called a column matrix or a column vector. An ordinary vector $\mathbf{A} = A_1\hat{e}_1 + A_2\hat{e}_2 + A_3\hat{e}_3$ can be represented either by a row matrix or by a column matrix.

If the numbers of rows m and columns n are equal, the matrix is called a square matrix of order n.

In a square matrix of order n, the elements $a_{11}, a_{22}, \ldots, a_{nn}$ form what is called the principal (or leading) diagonal, that is, the diagonal from the top left hand corner to the bottom right hand corner. The diagonal from the top right hand corner to the bottom left hand corner is sometimes termed the trailing diagonal. Only a square matrix possesses a principal diagonal and a trailing diagonal.

The sum of all elements down the principal diagonal is called the trace, or spur, of the matrix. We write

$$\mathrm{Tr}\,\tilde{A} = \sum_{i=1}^{n} a_{ii}.$$

If all elements of the principal diagonal of a square matrix are unity while all other elements are zero, then it is called a unit matrix (for a reason to be explained later) and is denoted by $\tilde{I}$. Thus the unit matrix of order 3 is

$$\tilde{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

A square matrix in which all elements other than those along the principal diagonal are zero is called a diagonal matrix.

A matrix with all elements zero is known as the null (or zero) matrix and is denoted by the symbol $\tilde{0}$, since it is not an ordinary number, but an array of zeros.

Four basic algebra operations for matrices

Equality of matrices

Two matrices $\tilde{A} = (a_{jk})$ and $\tilde{B} = (b_{jk})$ are equal if and only if $\tilde{A}$ and $\tilde{B}$ have the same order (equal numbers of rows and columns) and corresponding elements are equal, that is

$$a_{jk} = b_{jk} \quad \text{for all } j \text{ and } k.$$

Then we write

$$\tilde{A} = \tilde{B}.$$

Addition of matrices

Addition of matrices is defined only for matrices of the same order. If $\tilde{A} = (a_{jk})$ and $\tilde{B} = (b_{jk})$ have the same order, the sum of $\tilde{A}$ and $\tilde{B}$ is a matrix of the same order

$$\tilde{C} = \tilde{A} + \tilde{B}$$

with elements

$$c_{jk} = a_{jk} + b_{jk}. \tag{3.2}$$

We see that $\tilde{C}$ is obtained by adding corresponding elements of $\tilde{A}$ and $\tilde{B}$.

Example 3.1
If

$$\tilde{A} = \begin{pmatrix} 2 & 1 & 4 \\ 3 & 0 & 2 \end{pmatrix}, \quad \tilde{B} = \begin{pmatrix} 3 & 5 & 1 \\ 2 & 1 & -3 \end{pmatrix}$$

hen

$$\tilde{C} = \tilde{A} + \tilde{B} = \begin{pmatrix} 2 & 1 & 4 \\ 3 & 0 & 2 \end{pmatrix} + \begin{pmatrix} 3 & 5 & 1 \\ 2 & 1 & -3 \end{pmatrix} = \begin{pmatrix} 2+3 & 1+5 & 4+1 \\ 3+2 & 0+1 & 2-3 \end{pmatrix}$$
$$= \begin{pmatrix} 5 & 6 & 5 \\ 5 & 1 & -1 \end{pmatrix}.$$

From the definitions we see that matrix addition obeys the commutative and associative laws, that is, for any matrices $\tilde{A}$, $\tilde{B}$, $\tilde{C}$ of the same order

$$\tilde{A} + \tilde{B} = \tilde{B} + \tilde{A}, \quad \tilde{A} + (\tilde{B} + \tilde{C}) = (\tilde{A} + \tilde{B}) + \tilde{C}. \tag{3.3}$$

Similarly, if $\tilde{A} = (a_{jk})$ and $\tilde{B} = (b_{jk})$ have the same order, we define the difference of $\tilde{A}$ and $\tilde{B}$ as

$$\tilde{D} = \tilde{A} - \tilde{B}$$

with elements

$$d_{jk} = a_{jk} - b_{jk}. \tag{3.4}$$

Multiplication of a matrix by a number

If $\tilde{A} = (a_{jk})$ and c is a number (or scalar), then we define the product of $\tilde{A}$ and c as

$$c\tilde{A} = \tilde{A}c = (ca_{jk}); \tag{3.5}$$

we see that $c\tilde{A}$ is the matrix obtained by multiplying each element of $\tilde{A}$ by c.

We see from the definition that for any matrices and any numbers,

$$c(\tilde{A} + \tilde{B}) = c\tilde{A} + c\tilde{B}, \quad (c + k)\tilde{A} = c\tilde{A} + k\tilde{A}, \quad c(k\tilde{A}) = ck\tilde{A}. \tag{3.6}$$

Example 3.2

$$7\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} = \begin{pmatrix} 7a & 7b & 7c \\ 7d & 7e & 7f \end{pmatrix}.$$

Formulas (3.3) and (3.6) express the properties which are characteristic for a vector space. This gives vector spaces of matrices. We will discuss this further in Chapter 5.

Matrix multiplication

The matrix product $\tilde{A}\tilde{B}$ of the matrices $\tilde{A}$ and $\tilde{B}$ is defined *if and only if* the number of columns in $\tilde{A}$ is equal to the number of rows in $\tilde{B}$. Such matrices are sometimes called 'conformable'. If $\tilde{A} = (a_{jk})$ is an $n \times s$ matrix and $\tilde{B} = (b_{jk})$ is an $s \times m$ matrix, then $\tilde{A}$ and $\tilde{B}$ are conformable and their matrix product, written $\tilde{C} = \tilde{A}\tilde{B}$, is an $n \times m$ matrix formed according to the rule

$$c_{ik} = \sum_{j=1}^{s} a_{ij}b_{jk}, \quad i = 1, 2, \ldots, n \quad k = 1, 2, \ldots, m. \tag{3.7}$$

Consequently, to determine the ijth element of matrix $\tilde{C}$, the corresponding terms of the ith row of $\tilde{A}$ and jth column of $\tilde{B}$ are multiplied and the resulting products added to form c_{ij}.

Example 3.3
Let

$$\tilde{A} = \begin{pmatrix} 2 & 1 & 4 \\ -3 & 0 & 2 \end{pmatrix}, \quad \tilde{B} = \begin{pmatrix} 3 & 5 \\ 2 & -1 \\ 4 & 2 \end{pmatrix}$$

https://doi.org/10.1017/CBO9780511755781.004 Published online by Cambridge University Press

then

$$\tilde{A}\tilde{B} = \begin{pmatrix} 2\times 3 + 1\times 2 + 4\times 4 & 2\times 5 + 1\times(-1) + 4\times 2 \\ (-3)\times 3 + 0\times 2 + 2\times 4 & (-3)\times 5 + 0\times(-1) + 2\times 2 \end{pmatrix}$$

$$= \begin{pmatrix} 24 & 17 \\ -1 & -11 \end{pmatrix}.$$

The reader should master matrix multiplication, since it is used throughout the rest of the book.

In general, matrix multiplication is not commutative: $\tilde{A}\tilde{B} \neq \tilde{B}\tilde{A}$. In fact, $\tilde{B}\tilde{A}$ is often not defined for non-square matrices, as shown in the following example.

Example 3.4
If

$$\tilde{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad \tilde{B} = \begin{pmatrix} 3 \\ 7 \end{pmatrix}$$

then

$$\tilde{A}\tilde{B} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 3 \\ 7 \end{pmatrix} = \begin{pmatrix} 1\times 3 + 2\times 7 \\ 3\times 3 + 4\times 7 \end{pmatrix} = \begin{pmatrix} 17 \\ 37 \end{pmatrix}.$$

But

$$\tilde{B}\tilde{A} = \begin{pmatrix} 3 \\ 7 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

is not defined.

Matrix multiplication is associative and distributive:

$$(\tilde{A}\tilde{B})\tilde{C} = \tilde{A}(\tilde{B}\tilde{C}), \quad (\tilde{A} + \tilde{B})\tilde{C} = \tilde{A}\tilde{C} + \tilde{B}\tilde{C}.$$

To prove the associative law, we start with the matrix product $\tilde{A}\tilde{B}$, then multiply this product from the right by $\tilde{C}$:

$$\tilde{A}\tilde{B} = \sum_k a_{ik} b_{kj},$$

$$(\tilde{A}\tilde{B})\tilde{C} = \sum_j \left[\left(\sum_k a_{ik} b_{kj}\right) c_{js}\right] = \sum_k a_{ik}\left(\sum_j b_{kj} c_{js}\right) = \tilde{A}(\tilde{B}\tilde{C}).$$

Products of matrices differ from products of ordinary numbers in many remarkable ways. For example, $\tilde{A}\tilde{B} = 0$ does not imply $\tilde{A} = 0$ or $\tilde{B} = 0$. Even more bizarre is the case where $\tilde{A}^2 = 0$, $\tilde{A} \neq 0$; an example of which is

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

When you first run into Eq. (3.7), the rule for matrix multiplication, you might ask how anyone would arrive at it. It is suggested by the use of matrices in connection with linear transformations. For simplicity, we consider a very simple case: three coordinates systems in the plane denoted by the x_1x_2-system, the y_1y_2-system, and the z_1z_2-system. We assume that these systems are related by the following linear transformations

$$x_1 = a_{11}y_1 + a_{12}y_2, \quad x_2 = a_{21}y_1 + a_{22}y_2, \tag{3.8}$$

$$y_1 = b_{11}z_1 + b_{12}z_2, \quad y_2 = b_{21}z_1 + b_{22}z_2. \tag{3.9}$$

Clearly, the x_1x_2-coordinates can be obtained directly from the z_1z_2-coordinates by a single linear transformation

$$x_1 = c_{11}z_1 + c_{12}z_2, \quad x_2 = c_{21}z_1 + c_{22}z_2, \tag{3.10}$$

whose coefficients can be found by inserting (3.9) into (3.8),

$$x_1 = a_{11}(b_{11}z_1 + b_{12}z_2) + a_{12}(b_{21}z_1 + b_{22}z_2),$$
$$x_2 = a_{21}(b_{11}z_1 + b_{12}z_2) + a_{22}(b_{21}z_1 + b_{22}z_2).$$

Comparing this with (3.10), we find

$$c_{11} = a_{11}b_{11} + a_{12}b_{21}, \quad c_{12} = a_{11}b_{12} + a_{12}b_{22},$$
$$c_{21} = a_{21}b_{11} + a_{22}b_{21}, \quad c_{22} = a_{21}b_{12} + a_{22}b_{22},$$

or briefly

$$c_{jk} = \sum_{i=1}^{2} a_{ji}b_{ik}, \quad j,k = 1,2, \tag{3.11}$$

which is in the form of (3.7).

Now we rewrite the transformations (3.8), (3.9) and (3.10) in matrix form:

$$\tilde{X} = \tilde{A}\tilde{Y}, \quad \tilde{Y} = \tilde{B}\tilde{Z}, \quad \text{and} \quad \tilde{X} = \tilde{C}\tilde{Z},$$

where

$$\tilde{X} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \tilde{Y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad \tilde{Z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix},$$

$$\tilde{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad \tilde{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}, \quad \tilde{C} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}.$$

We then see that $\tilde{C} = \tilde{A}\tilde{B}$, and the elements of $\tilde{C}$ are given by (3.11).

Example 3.5

Rotations in three-dimensional space: An example of the use of matrix multiplication is provided by the representation of rotations in three-dimensional

https://doi.org/10.1017/CBO9780511755781.004 Published online by Cambridge University Press

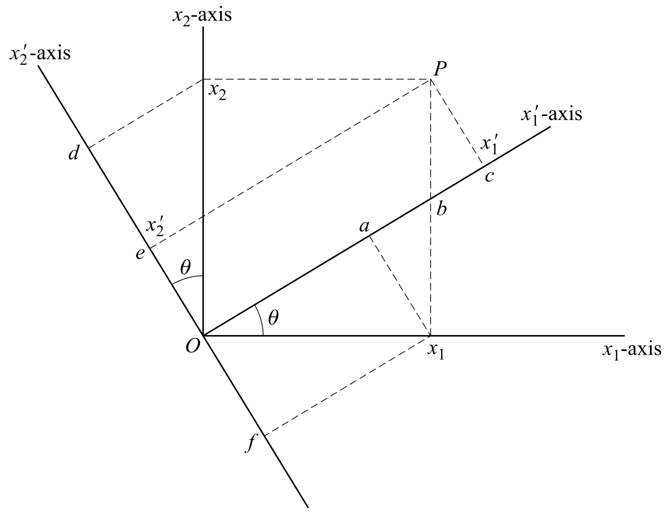

Figure 3.1. Coordinate changes by rotation.

space. In Fig. 3.1, the primed coordinates are obtained from the unprimed coordinates by a rotation through an angle θ about the x_3-axis. We see that x_1' is the sum of the projection of x_1 onto the x_1'-axis and the projection of x_2 onto the x_1'-axis:

$$x_1' = x_1 \cos\theta + x_2 \cos(\pi/2 - \theta) = x_1 \cos\theta + x_2 \sin\theta;$$

similarly

$$x_2' = x_1 \cos(\pi/2 + \theta) + x_2 \cos\theta = -x_1 \sin\theta + x_2 \cos\theta$$

and

$$x_3' = x_3.$$

We can put these in matrix form

$$X' = R_\theta X,$$

where

$$X' = \begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad R_\theta = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

https://doi.org/10.1017/CBO9780511755781.004 Published online by Cambridge University Press

The commutator

Even if matrices $\tilde{A}$ and $\tilde{B}$ are both square matrices of order n, the products $\tilde{A}\tilde{B}$ and $\tilde{B}\tilde{A}$, although both square matrices of order n, are in general quite different, since their individual elements are formed differently. For example,

$$\begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 4 & 6 \end{pmatrix} \quad \text{but} \quad \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 8 \end{pmatrix}.$$

The difference between the two products $\tilde{A}\tilde{B}$ and $\tilde{B}\tilde{A}$ is known as the commutator of $\tilde{A}$ and $\tilde{B}$ and is denoted by

$$[\tilde{A}, \tilde{B}] = \tilde{A}\tilde{B} - \tilde{B}\tilde{A}. \tag{3.12}$$

It is obvious that

$$[\tilde{B}, \tilde{A}] = -[\tilde{A}, \tilde{B}]. \tag{3.13}$$

If two square matrices $\tilde{A}$ and $\tilde{B}$ are very carefully chosen, it is possible to make the product identical. That is $\tilde{A}\tilde{B} = \tilde{B}\tilde{A}$. Two such matrices are said to commute with each other. Commuting matrices play an important role in quantum mechanics.

If $\tilde{A}$ commutes with $\tilde{B}$ and $\tilde{B}$ commutes with $\tilde{C}$, it does not necessarily follow that $\tilde{A}$ commutes with $\tilde{C}$.

Powers of a matrix

If n is a positive integer and $\tilde{A}$ is a square matrix, then $\tilde{A}^2 = \tilde{A}\tilde{A}$, $\tilde{A}^3 = \tilde{A}\tilde{A}\tilde{A}$, and in general, $\tilde{A}^n = \tilde{A}\tilde{A}\cdots\tilde{A}$ (n times). In particular, $\tilde{A}^0 = \tilde{I}$.

Functions of matrices

As we define and study various functions of a variable in algebra, it is possible to define and evaluate functions of matrices. We shall briefly discuss the following functions of matrices in this section: integral powers and exponential.

A simple example of integral powers of a matrix is polynomials such as

$$f(\tilde{A}) = \tilde{A}^2 + 3\tilde{A}^5.$$

Note that a matrix can be multiplied by itself if and only if it is a square matrix. Thus $\tilde{A}$ here is a square matrix and we denote the product $\tilde{A} \cdot \tilde{A}$ as $\tilde{A}^2$. More fancy examples can be obtained by taking series, such as

$$\tilde{S} = \sum_{k=0}^{\infty} a_k \tilde{A}^k,$$

where a_k are scalar coefficients. Of course, the sum has no meaning if it does not converge. The convergence of the matrix series means every matrix element of the

https://doi.org/10.1017/CBO9780511755781.004 Published online by Cambridge University Press

infinite sum of matrices converges to a limit. We will not discuss the general theory of convergence of matrix functions. Another very common series is defined by

$$e^{\tilde{A}} = \sum_{n=0}^{\infty} \frac{\tilde{A}^n}{n!}.$$

Transpose of a matrix

Consider an $m \times n$ matrix $\tilde{A}$, if the rows and columns are systematically changed to columns to rows, without changing the order in which they occur, the new matrix is called the transpose of matrix $\tilde{A}$. It is denoted by $\tilde{A}^{\mathrm{T}}$:

$$\tilde{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{pmatrix}, \quad \tilde{A}^{\mathrm{T}} = \begin{pmatrix} a_{11} & a_{21} & a_{31} & \ldots & a_{m1} \\ a_{12} & a_{22} & a_{32} & \ldots & a_{m2} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{2n} & a_{3n} & \ldots & a_{mn} \end{pmatrix}.$$

Thus the transpose matrix has n rows and m columns. If $\tilde{A}$ is written as (a_{jk}), then $\tilde{A}^{\mathrm{T}}$ may be written as (a_{kj}).

$$\tilde{A} = (a_{jk}), \quad \tilde{A}^{\mathrm{T}} = (a_{kj}). \tag{3.14}$$

The transpose of a row matrix is a column matrix, and vice versa.

Example 3.6

$$\tilde{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \quad \tilde{A}^{\mathrm{T}} = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}; \quad \tilde{B} = (1 \quad 2 \quad 3), \quad \tilde{B}^{\mathrm{T}} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

It is obvious that $(\tilde{A}^{\mathrm{T}})^{\mathrm{T}} = \tilde{A}$, and $(\tilde{A} + \tilde{B})^{\mathrm{T}} = \tilde{A}^{\mathrm{T}} + \tilde{B}^{\mathrm{T}}$. It is also easy to prove that the transpose of the product is the product of the transposes in reverse:

$$(\tilde{A}\tilde{B})^{\mathrm{T}} = \tilde{B}^{\mathrm{T}}\tilde{A}^{\mathrm{T}}. \tag{3.15}$$

Proof:

$$\begin{aligned} (\tilde{A}\tilde{B})^{\mathrm{T}}_{ij} &= (\tilde{A}\tilde{B})_{ji} \quad \text{by definition} \\ &= \sum_k A_{jk} B_{ki} \\ &= \sum_k B^{\mathrm{T}}_{ik} A^{\mathrm{T}}_{kj} \\ &= (\tilde{B}^{\mathrm{T}} \tilde{A}^{\mathrm{T}})_{ij} \end{aligned}$$

so that

$$(AB)^{\mathrm{T}} = B^{\mathrm{T}} A^{\mathrm{T}} \quad \text{q.e.d.}$$

Because of (3.15), even if $\tilde{A} = \tilde{A}^{\mathrm{T}}$ and $\tilde{B} = \tilde{B}^{\mathrm{T}}$, $(\tilde{A}\tilde{B})^{\mathrm{T}} \neq \tilde{A}\tilde{B}$ unless the matrices commute.

Symmetric and skew-symmetric matrices

A square matrix $\tilde{A} = (a_{jk})$ is said to be symmetric if all its elements satisfy the equations

$$a_{kj} = a_{jk}, \tag{3.16}$$

that is, $\tilde{A}$ and its transpose are equal $\tilde{A} = \tilde{A}^{\mathrm{T}}$. For example,

$$\tilde{A} = \begin{pmatrix} 1 & 5 & 7 \\ 5 & 3 & -4 \\ 7 & -4 & 0 \end{pmatrix}$$

is a third-order symmetric matrix: the elements of the ith row equal the elements of ith column, for all i.

On the other hand, if the elements of $\tilde{A}$ satisfy the equations

$$a_{kj} = -a_{jk}, \tag{3.17}$$

then $\tilde{A}$ is said to be skew-symmetric, or antisymmetric. Thus, for a skew-symmetric $\tilde{A}$, its transpose equals minus $-\tilde{A}$: $\tilde{A}^{\mathrm{T}} = -\tilde{A}$.

Since the elements a_{jj} along the principal diagonal satisfy the equations $a_{jj} = -a_{jj}$, it is evident that they must all vanish. For example,

$$\tilde{A} = \begin{pmatrix} 0 & -2 & 5 \\ 2 & 0 & 1 \\ -5 & -1 & 0 \end{pmatrix}$$

is a skew-symmetric matrix.

Any real square matrix $\tilde{A}$ may be expressed as the sum of a symmetric matrix $\tilde{R}$ and a skew-symmetric matrix $\tilde{S}$, where

$$\tilde{R} = \tfrac{1}{2}(\tilde{A} + \tilde{A}^{\mathrm{T}}) \quad \text{and} \quad \tilde{S} = \tfrac{1}{2}(\tilde{A} - \tilde{A}^{\mathrm{T}}). \tag{3.18}$$

Example 3.7

The matrix

$$\tilde{A} = \begin{pmatrix} 2 & 3 \\ 5 & -1 \end{pmatrix}$$

may be written in the form $\tilde{A} = \tilde{R} + \tilde{S}$, where

$$\tilde{R} = \frac{1}{2}(\tilde{A} + \tilde{A}^{\mathrm{T}}) = \begin{pmatrix} 2 & 4 \\ 4 & -1 \end{pmatrix} \quad \tilde{S} = \frac{1}{2}(\tilde{A} - \tilde{A}^{\mathrm{T}}) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

The product of two symmetric matrices need not be symmetric. This is because of (3.15): even if $\tilde{A} = \tilde{A}^{\mathrm{T}}$ and $\tilde{B} = \tilde{B}^{\mathrm{T}}$, $(\tilde{A}\tilde{B})^{\mathrm{T}} \neq \tilde{A}\tilde{B}$ unless the matrices commute.

A square matrix whose elements above or below the principal diagonal are all zero is called a triangular matrix. The following two matrices are triangular matrices:

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 5 & 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 6 & -1 \\ 0 & 2 & 3 \\ 0 & 0 & 4 \end{pmatrix}.$$

A square matrix $\tilde{A}$ is said to be singular if $\det \tilde{A} = 0$, and non-singular if $\det \tilde{A} \neq 0$, where $\det \tilde{A}$ is the determinant of the matrix $\tilde{A}$.

The matrix representation of a vector product

The scalar product defined in ordinary vector theory has its counterpart in matrix theory. Consider two vectors $\mathbf{A} = (A_1, A_2, A_3)$ and $\mathbf{B} = (B_1, B_2, B_3)$ the counterpart of the scalar product is given by

$$\tilde{A}\tilde{B}^{\mathrm{T}} = (A_1 \quad A_2 \quad A_3)\begin{pmatrix} B_1 \\ B_2 \\ B_3 \end{pmatrix} = A_1B_1 + A_2B_2 + A_3B_3.$$

Note that $\tilde{B}\tilde{A}^{\mathrm{T}}$ is the transpose of $\tilde{A}\tilde{B}^{\mathrm{T}}$, and, being a 1×1 matrix, the transpose equals itself. Thus a scalar product may be written in these two equivalent forms.

Similarly, the vector product used in ordinary vector theory must be replaced by something more in keeping with the definition of matrix multiplication. Note that the vector product

$$\mathbf{A} \times \mathbf{B} = (A_2B_3 - A_3B_2)\hat{e}_1 + (A_3B_1 - A_1B_3)\hat{e}_2 + (A_1B_2 - A_2B_1)\hat{e}_3$$

can be represented by the column matrix

$$\begin{pmatrix} A_2B_3 - A_3B_2 \\ A_3B_1 - A_1B_3 \\ A_1B_2 - A_2B_1 \end{pmatrix}.$$

This can be split into the product of two matrices

$$\begin{pmatrix} A_2B_3 - A_3B_2 \\ A_3B_1 - A_1B_3 \\ A_1B_2 - A_2B_1 \end{pmatrix} = \begin{pmatrix} 0 & -A_2 & A_2 \\ A_3 & 0 & -A_1 \\ -A_2 & A_1 & 0 \end{pmatrix}\begin{pmatrix} B_1 \\ B_2 \\ B_3 \end{pmatrix}$$

or

$$\begin{pmatrix} A_2B_3 - A_3B_2 \\ A_3B_1 - A_1B_3 \\ A_1B_2 - A_2B_1 \end{pmatrix} = \begin{pmatrix} 0 & -B_2 & B_2 \\ B_3 & 0 & -B_1 \\ -B_2 & B_1 & 0 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}.$$

Thus the vector product may be represented as the product of a skew-symmetric matrix and a column matrix. However, this definition only holds for 3×3 matrices.

Similarly, curl A may be represented in terms of a skew-symmetric matrix operator, given in Cartesian coordinates by

$$\nabla \times A = \begin{pmatrix} 0 & -\partial/\partial x_3 & \partial/\partial x_2 \\ \partial/\partial x_3 & 0 & -\partial/\partial x_1 \\ -\partial/\partial x_2 & \partial/\partial x_1 & 0 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}.$$

In a similar way, we can investigate the triple scalar product and the triple vector product.

The inverse of a matrix

If for a given square matrix $\tilde{A}$ there exists a matrix $\tilde{B}$ such that $\tilde{A}\tilde{B} = \tilde{B}\tilde{A} = \tilde{I}$, where $\tilde{I}$ is a unit matrix, then $\tilde{B}$ is called an inverse of matrix $\tilde{A}$.

Example 3.8

The matrix

$$\tilde{B} = \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix}$$

is an inverse of

$$\tilde{A} = \begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix},$$

since

$$\tilde{A}\tilde{B} = \begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \tilde{I}$$

and

$$\tilde{B}\tilde{A} = \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \tilde{I}.$$

An invertible matrix has a unique inverse. That is, if $\tilde{B}$ and $\tilde{C}$ are both inverses of the matrix $\tilde{A}$, then $\tilde{B} = \tilde{C}$. The proof is simple. Since $\tilde{B}$ is an inverse of $\tilde{A}$,

$\tilde{B}\tilde{A} = \tilde{I}$. Multiplying both sides on the right by $\tilde{C}$ gives $(\tilde{\mathbf{B}}\tilde{A})\tilde{C} = \tilde{I}\tilde{C} = \tilde{C}$. On the other hand, $(\tilde{B}\tilde{A})\tilde{C} = \tilde{B}(\tilde{A}\tilde{C}) = \tilde{B}\tilde{I} = \tilde{B}$, so that $\tilde{B} = \tilde{C}$. As a consequence of this result, we can now speak of *the* inverse of an invertible matrix. If $\tilde{A}$ is invertible, then its inverse will be denoted by $\tilde{A}^{-1}$. Thus

$$\tilde{A}\tilde{A}^{-1} = \tilde{A}^{-1}\tilde{A} = \tilde{I}. \tag{3.19}$$

It is obvious that the inverse of the inverse is the given matrix, that is,

$$(\tilde{A}^{-1})^{-1} = \tilde{A}. \tag{3.20}$$

It is easy to prove that the inverse of the product is the product of the inverse in reverse order, that is,

$$(\tilde{A}\tilde{B})^{-1} = \tilde{B}^{-1}\tilde{A}^{-1}. \tag{3.21}$$

To prove (3.21), we start with $\tilde{A}\tilde{A}^{-1} = \tilde{I}$, with $\tilde{A}$ replaced by $\tilde{A}\tilde{B}$, that is,

$$\tilde{A}\tilde{B}(\tilde{A}\tilde{B})^{-1} = \tilde{I}.$$

By premultiplying this by $\tilde{A}^{-1}$ we get

$$\tilde{B}(\tilde{A}\tilde{B})^{-1} = \tilde{A}^{-1}.$$

If we premultiply this by $\tilde{B}^{-1}$, the result follows.

A method for finding $\tilde{A}^{-1}$

The positive power for a square matrix $\tilde{A}$ is defined as $\tilde{A}^n = \tilde{A}\tilde{A}\cdots\tilde{A}$ (n factors) and $\tilde{A}^0 = \tilde{I}$, where n is a positive integer. If, in addition, $\tilde{A}$ is invertible, we define

$$\tilde{A}^{-n} = (\tilde{A}^{-1})^n = \tilde{A}^{-1}\tilde{A}^{-1}\cdots\tilde{A}^{-1}(n \text{ factors}).$$

We are now in position to construct the inverse of an invertible matrix $\tilde{A}$:

$$\tilde{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}.$$

The a_{jk} are known. Now let

$$\tilde{A}^{-1} = \begin{pmatrix} a'_{11} & a'_{12} & \dots & a'_{1n} \\ a'_{21} & a'_{22} & \dots & a'_{2n} \\ \vdots & \vdots & & \vdots \\ a'_{n1} & a'_{n2} & \dots & a'_{nn} \end{pmatrix}.$$

The a'_{jk} are required to construct $\tilde{A}^{-1}$. Since $\tilde{A}\tilde{A}^{-1} = \tilde{I}$, we have

$$\begin{aligned} a_{11}a'_{11} + a_{12}a'_{12} + \cdots + a_{1n}a'_{1n} &= 1, \\ a_{21}a'_{21} + a_{22}a'_{22} + \cdots + a_{2n}a'_{2n} &= 0, \\ &\vdots \\ a_{n1}a'_{n1} + a_{n2}a'_{n2} + \cdots + a_{nn}a'_{nn} &= 0. \end{aligned} \tag{3.22}$$

The solution to the above set of linear algebraic equations (3.22) may be facilitated by applying Cramer's rule. Thus

$$a'_{jk} = \frac{\text{cofactor } a_{kj}}{\det \tilde{A}}. \tag{3.23}$$

From (3.23) it is clear that $\tilde{A}^{-1}$ exists if and only if matrix $\tilde{A}$ is non-singular (that is, $\det \tilde{A} \neq 0$).

Systems of linear equations and the inverse of a matrix

As an immediate application, let us apply the concept of an inverse matrix to a system of n linear equations in n unknowns $(x_1, \ldots, x_n)$:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1; \\ a_{21}x_2 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2; \\ &\vdots \\ a_{n1}x_n + a_{n2}x_n + \cdots + a_{nn}x_n &= b_n; \end{aligned}$$

in matrix form we have

$$\tilde{A}\tilde{X} = \tilde{B}, \tag{3.24}$$

where

$$\tilde{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}, \quad \tilde{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \tilde{B} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

We can prove that the above linear system possesses a unique solution given by

$$\tilde{X} = \tilde{A}^{-1}\tilde{B}. \tag{3.25}$$

The proof is simple. If $\tilde{A}$ is non-singular it has a unique inverse $\tilde{A}^{-1}$. Now premultiplying (3.24) by $\tilde{A}^{-1}$ we obtain

$$\tilde{A}^{-1}(\tilde{A}\tilde{X}) = \tilde{A}^{-1}\tilde{B},$$

https://doi.org/10.1017/CBO9780511755781.004 Published online by Cambridge University Press

but

$$\tilde{A}^{-1}(\tilde{A}\tilde{X}) = (\tilde{A}^{-1}\tilde{A})\tilde{X} = \tilde{X}$$

so that

$$\tilde{X} = \tilde{A}^{-1}\tilde{B} \text{ is a solution to (3.24), } \tilde{A}\tilde{X} = \tilde{B}.$$

Complex conjugate of a matrix

If $\tilde{A} = (a_{jk})$ is an arbitrary matrix whose elements may be complex numbers, the complex conjugate matrix, denoted by $\tilde{A}^*$, is also a matrix of the same order, every element of which is the complex conjugate of the corresponding element of $\tilde{A}$, that is,

$$(A^*)_{jk} = a^*_{jk}. \tag{3.26}$$

Hermitian conjugation

If $\tilde{A} = (a_{jk})$ is an arbitrary matrix whose elements may be complex numbers, when the two operations of transposition and complex conjugation are carried out on $\tilde{A}$, the resulting matrix is called the hermitian conjugate (or hermitian adjoint) of the original matrix $\tilde{A}$ and will be denoted by $\tilde{A}^\dagger$. We frequently call $\tilde{A}^\dagger$ A-dagger. The order of the two operations is immaterial:

$$\tilde{A}^\dagger = (\tilde{A}^{\mathrm{T}})^* = (\tilde{A}^*)^{\mathrm{T}}. \tag{3.27}$$

In terms of the elements, we have

$$(\tilde{A}^\dagger)_{jk} = a^*_{kj}. \tag{3.27a}$$

It is clear that if $\tilde{A}$ is a matrix of order $m \times n$, then $\tilde{A}^\dagger$ is a matrix of order $n \times m$. We can prove that, as in the case of the transpose of a product, the adjoint of the product is the product of the adjoints in reverse:

$$(\tilde{A}\tilde{B})^\dagger = \tilde{B}^\dagger\tilde{A}^\dagger. \tag{3.28}$$

Hermitian/anti-hermitian matrix

A matrix $\tilde{A}$ that obeys

$$\tilde{A}^\dagger = \tilde{A} \tag{3.29}$$

is called a hermitian matrix. It is very clear the following matrices are hermitian:

$$\begin{pmatrix} 1 & -i \\ i & 2 \end{pmatrix}, \quad \begin{pmatrix} 4 & 5+2i & 6+3i \\ 5-2i & 5 & -1-2i \\ 6-3i & -1+2i & 6 \end{pmatrix}, \quad \text{where} \quad i = \sqrt{-1}.$$

https://doi.org/10.1017/CBO9780511755781.004 Published online by Cambridge University Press

Table 3.1. Operations on matrices

Operation	Matrix element		$\tilde{A}$	$\tilde{B}$	If $\tilde{B} = \tilde{A}$
Transposition	$\tilde{B} = \tilde{A}^{\mathrm{T}}$	$b_{ij} = a_{ji}$	$m \times n$	$n \times m$	Symmetric[a]
Complex conjugation	$\tilde{B} = \tilde{A}^*$	$b_{ij} = a^*_{ij}$	$m \times n$	$m \times n$	Real
Hermitian conjugation	$\tilde{B} = \tilde{A}^{\mathrm{T}*}$	$b_{ij} = a^*_{ji}$	$m \times n$	$n \times m$	Hermitian

[a] For square matrices only.

Evidently all the elements along the principal diagonal of a hermitian matrix must be real.

A hermitian matrix is also defined as a matrix whose transpose equals its complex conjugate:

$$\tilde{A}^{\mathrm{T}} = \tilde{A}^* \quad (\text{that is}, a_{kj} = a^*_{jk}). \tag{3.29a}$$

These two definitions are the same. First note that the elements in the principal diagonal of a hermitian matrix are always real. Furthermore, any real symmetric matrix is hermitian, so a real hermitian matrix is a symmetric matrix.

The product of two hermitian matrices is not generally hermitian unless they commute. This is because of property (3.28): even if $\tilde{A}^\dagger = \tilde{A}$ and $\tilde{B}^\dagger = \tilde{B}$, $(\tilde{A}\tilde{B})^\dagger \neq \tilde{A}\tilde{B}$ unless the matrices commute.

A matrix $\tilde{A}$ that obeys

$$\tilde{A}^\dagger = -\tilde{A} \tag{3.30}$$

is called an anti-hermitian (or skew-hermitian) matrix. All the elements along the principal diagonal must be pure imaginary. An example is

$$\begin{pmatrix} 6i & 5+2i & 6+3i \\ -5+2i & -8i & -1-2i \\ -6+3i & 1-2i & 0 \end{pmatrix}.$$

We summarize the three operations on matrices discussed above in Table 3.1.

Orthogonal matrix (real)

A matrix $\tilde{A} = (a_{jk})_{mn}$ satisfying the relations

$$\tilde{A}\tilde{A}^{\mathrm{T}} = \tilde{I}_n, \tag{3.31a}$$

$$\tilde{A}^{\mathrm{T}}\tilde{A} = \tilde{I}_m \tag{3.31b}$$

is called an orthogonal matrix. It can be shown that if $\tilde{A}$ is a finite matrix satisfying both relations (3.31a) and (3.31b), then $\tilde{A}$ must be square, and we have

$$\tilde{A}\tilde{A}^{\mathrm{T}} = \tilde{A}^{\mathrm{T}}\tilde{A} = \tilde{I}. \tag{3.32}$$

https://doi.org/10.1017/CBO9780511755781.004 Published online by Cambridge University Press

But if $\tilde{A}$ is an infinite matrix, then $\tilde{A}$ is orthogonal if and only if both (3.31a) and (3.31b) are simultaneously satisfied.

Now taking the determinant of both sides of Eq. (3.32), we have $(\det \tilde{A})^2 = 1$, or $\det \tilde{A} = \pm 1$. This shows that $\tilde{A}$ is non-singular, and so $\tilde{A}^{-1}$ exists. Premultiplying (3.32) by $\tilde{A}^{-1}$ we have

$$\tilde{A}^{-1} = \tilde{A}^{\mathrm{T}}. \tag{3.33}$$

This is often used as an alternative way of defining an orthogonal matrix.

The elements of an orthogonal matrix are not all independent. To find the conditions between them, let us first equate the ijth element of both sides of $\tilde{A}\tilde{A}^{\mathrm{T}} = \tilde{I}$; we find that

$$\sum_{k=1}^{n} a_{ik} a_{jk} = \delta_{ij}. \tag{3.34a}$$

Similarly, equating the ijth element of both sides of $\tilde{A}^{\mathrm{T}}\tilde{A} = \tilde{I}$, we obtain

$$\sum_{k=1}^{n} a_{ki} a_{kj} = \delta_{ij}. \tag{3.34b}$$

Note that either (3.34a) and (3.34b) gives $2n(n+1)$ relations. Thus, for a real orthogonal matrix of order n, there are only $n^2 - n(n+1)/2 = n(n-1)/2$ different elements.

Unitary matrix

A matrix $\tilde{U} = (u_{jk})_{mn}$ satisfying the relations

$$\tilde{U}\tilde{U}^{\dagger} = \tilde{I}_n, \tag{3.35a}$$

$$\tilde{U}^{\dagger}\tilde{U} = \tilde{I}_m \tag{3.35b}$$

is called a unitary matrix. If $\tilde{U}$ is a finite matrix satisfying both (3.35a) and (3.35b), then $\tilde{U}$ must be a square matrix, and we have

$$\tilde{U}\tilde{U}^{\dagger} = \tilde{U}^{\dagger}\tilde{U} = \tilde{I}. \tag{3.36}$$

This is the complex generalization of the real orthogonal matrix. The elements of a unitary matrix may be complex, for example

$$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$$

is unitary. From the definition (3.35), a real unitary matrix is orthogonal.

Taking the determinant of both sides of (3.36) and noting that $\det \tilde{U}^{\dagger} = (\det \tilde{U})^*$, we have

$$(\det \tilde{U})(\det \tilde{U})^* = 1 \quad \text{or} \quad |\det \tilde{U}| = 1. \tag{3.37}$$

https://doi.org/10.1017/CBO9780511755781.004 Published online by Cambridge University Press

This shows that the determinant of a unitary matrix can be a complex number of unit magnitude, that is, a number of the form $e^{i\alpha}$, where α is a real number. It also shows that a unitary matrix is non-singular and possesses an inverse. Premultiplying (3.35a) by $\tilde{U}^{-1}$, we get

$$\tilde{U}^{\dagger} = \tilde{U}^{-1}. \tag{3.38}$$

This is often used as an alternative way of defining a unitary matrix.

Just as in the case of an orthogonal matrix that is a special (real) case of a unitary matrix, the elements of a unitary matrix satisfy the following conditions:

$$\sum_{k=1}^{n} u_{ik} u_{jk}^{*} = \delta_{ij}, \qquad \sum_{k=1}^{n} u_{ki} u_{kj}^{*} = \delta_{ij}. \tag{3.39}$$

The product of two unitary matrices is unitary. The reason is as follows. If $\tilde{U}_1$ and $\tilde{U}_2$ are two unitary matrices, then

$$\tilde{U}_1 \tilde{U}_2 (\tilde{U}_1 \tilde{U}_2)^{\dagger} = \tilde{U}_1 \tilde{U}_2 (\tilde{U}_2^{\dagger} \tilde{U}_1^{\dagger}) = \tilde{U}_1 \tilde{U}_1^{\dagger} = \tilde{I}, \tag{3.40}$$

which shows that $U_1 U_2$ is unitary.

Rotation matrices

Let us revisit Example 3.5. Our discussion will illustrate the power and usefulness of matrix methods. We will also see that rotation matrices are orthogonal matrices. Consider a point P with Cartesian coordinates (x_1, x_2, x_3) (see Fig. 3.2). We rotate the coordinate axes about the x_3-axis through an angle θ and create a new coordinate system, the primed system. The point P now has the coordinates (x_1', x_2', x_3') in the primed system. Thus the position vector $\mathbf{r}$ of point P can be written as

$$\mathbf{r} = \sum_{i=1}^{3} x_i \hat{e}_i = \sum_{i=1}^{3} x_i' \hat{e}_i'. \tag{3.41}$$

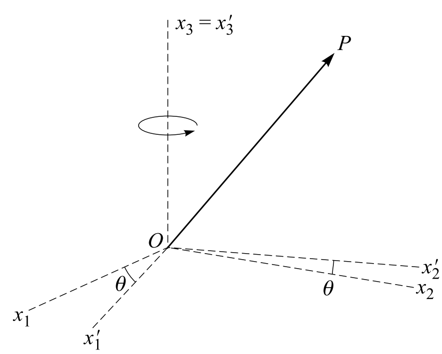

Figure 3.2. Coordinate change by rotation.

Taking the dot product of Eq. (3.41) with $\hat{e}_1'$ and using the orthonormal relation $\hat{e}_i' \cdot \hat{e}_j' = \delta_{ij}$ (where δ_{ij} is the Kronecker delta symbol), we obtain $x_1' = \mathbf{r} \cdot \hat{e}_1'$. Similarly, we have $x_2' = \mathbf{r} \cdot \hat{e}_2'$ and $x_3' = \mathbf{r} \cdot \hat{e}_3'$. Combining these results we have

$$x_i' = \sum_{j=1}^{3} \hat{e}_i' \cdot \hat{e}_j x_j = \sum_{j=1}^{3} \lambda_{ij} x_j, \qquad i = 1, 2, 3. \tag{3.42}$$

The quantities $\lambda_{ij} = \hat{e}_i' \cdot \hat{e}_j$ are called the coefficients of transformation. They are the direction cosines of the primed coordinate axes relative to the unprimed ones

$$\lambda_{ij} = \hat{e}_i' \cdot \hat{e}_j = \cos(x_i', x_j), \qquad i, j = 1, 2, 3. \tag{3.42a}$$

Eq. (3.42) can be written conveniently in the following matrix form

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \tag{3.43a}$$

or

$$\tilde{X}' = \tilde{\lambda}(\theta)\tilde{X}, \tag{3.43b}$$

where $\tilde{X}'$ and $\tilde{X}$ are the column matrices, $\tilde{\lambda}(\theta)$ is called a transformation (or rotation) matrix; it acts as a linear operator which transforms the vector $\mathbf{X}$ into the vector $\mathbf{X}'$. Strictly speaking, we should describe the matrix $\tilde{\lambda}(\theta)$ as the matrix representation of the linear operator $\hat{\lambda}$. The concept of linear operator is more general than that of matrix.

Not all of the nine quantities λ_{ij} are independent; six relations exist among the λ_{ij}, hence only three of them are independent. These six relations are found by using the fact that the magnitude of the vector must be the same in both systems:

$$\sum_{i=1}^{3} (x_i')^2 = \sum_{i=1}^{3} x_i^2. \tag{3.44}$$

With the help of Eq. (3.42), the left hand side of the last equation becomes

$$\sum_{i=1}^{3} \left(\sum_{j=1}^{3} \lambda_{ij} x_j \right) \left(\sum_{k=1}^{3} \lambda_{ik} x_k \right) = \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{k=1}^{3} \lambda_{ij} \lambda_{ik} x_j x_k,$$

which, by rearranging the summations, can be rewritten as

$$\sum_{k=1}^{3} \sum_{j=1}^{3} \left(\sum_{i=1}^{3} \lambda_{ij} \lambda_{ik} \right) x_j x_k.$$

This last expression will reduce to the right hand side of Eq. (3.43) if and only if

$$\sum_{i=1}^{3} \lambda_{ij} \lambda_{ik} = \delta_{jk}, \qquad j, k = 1, 2, 3. \tag{3.45}$$

Eq. (3.45) gives six relations among the λ_{ij}, and is known as the orthogonal condition.

If the primed coordinates system is generated by a rotation about the x_3-axis through an angle θ as shown in Fig. 3.2. Then from Example 3.5, we have

$$x_1' = x_1 \cos\theta + x_2 \sin\theta, \quad x_2' = -x_1 \sin\theta + x_2 \cos\theta, \quad x_3' = x_3. \tag{3.46}$$

Thus

$$\begin{aligned}
&\lambda_{11} = \cos\theta, \quad \lambda_{12} = \sin\theta, \quad \lambda_{13} = 0,\\
&\lambda_{21} = -\sin\theta, \quad \lambda_{22} = \cos\theta, \quad \lambda_{23} = 0,\\
&\lambda_{31} = 0, \quad \lambda_{32} = 0, \quad \lambda_{33} = 1.
\end{aligned}$$

We can also obtain these elements from Eq. (3.42a). It is obvious that only three of them are independent, and it is easy to check that they satisfy the condition given in Eq. (3.45). Now the rotation matrix takes the simple form

$$\tilde{\lambda}(\theta) = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3.47}$$

and its transpose is

$$\tilde{\lambda}^{\mathrm{T}}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Now take the product

$$\tilde{\lambda}^{\mathrm{T}}(\theta)\tilde{\lambda}(\theta) = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \tilde{I},$$

which shows that the rotation matrix is an orthogonal matrix. In fact, rotation matrices are orthogonal matrices, not limited to $\tilde{\lambda}(\theta)$ of Eq. (3.47). The proof of this is easy. Since coordinate transformations are reversible by interchanging old and new indices, we must have

$$\left(\tilde{\lambda}^{-1}\right)_{ij} = \hat{e}_i^{\mathrm{old}} \cdot \hat{e}_j^{\mathrm{new}} = \hat{e}_j^{\mathrm{new}} \cdot \hat{e}_i^{\mathrm{old}} = \lambda_{ji} = \left(\tilde{\lambda}^{\mathrm{T}}\right)_{ij}.$$

Hence rotation matrices are orthogonal matrices. It is obvious that the inverse of an orthogonal matrix is equal to its transpose.

A rotation matrix such as given in Eq. (3.47) is a continuous function of its argument θ. So its determinant is also a continuous function of θ and, in fact, it is equal to 1 for any θ. There are matrices of coordinate changes with a determinant of -1. These correspond to inversion of the coordinate axes about the origin and

https://doi.org/10.1017/CBO9780511755781.004 Published online by Cambridge University Press

change the handedness of the coordinate system. Examples of such parity transformations are

$$\tilde{P}_1 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \tilde{P}_3 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \qquad \tilde{P}_i^2 = I.$$

They change the signs of an odd number of coordinates of a fixed point **r** in space (Fig. 3.3).

What is the advantage of using matrices in describing rotation in space? One of the advantages is that successive transformations $1, 2, \ldots, m$ of the coordinate axes about the origin are described by successive matrix multiplications as far as their effects on the coordinates of a fixed point are concerned:

If $\tilde{X}^{(1)} = \tilde{\lambda}_1 \tilde{X}, \tilde{X}^{(2)} = \tilde{\lambda}_2 \tilde{X}^{(1)}, \ldots,$ then

$$\tilde{X}^{(m)} = \tilde{\lambda}_m \tilde{X}^{(m-1)} = (\tilde{\lambda}_m \tilde{\lambda}_{m-1} \cdots \tilde{\lambda}_1)\tilde{X} = \tilde{R}\tilde{X}$$

where

$$\tilde{R} = \tilde{\lambda}_m \tilde{\lambda}_{m-1} \cdots \tilde{\lambda}_1$$

is the resultant (or net) rotation matrix for the m successive transformations taken place in the specified manner.

Example 3.9

Consider a rotation of the x_1-, x_2-axes about the x_3-axis by an angle θ. If this rotation is followed by a back-rotation of the same angle in the opposite direction,

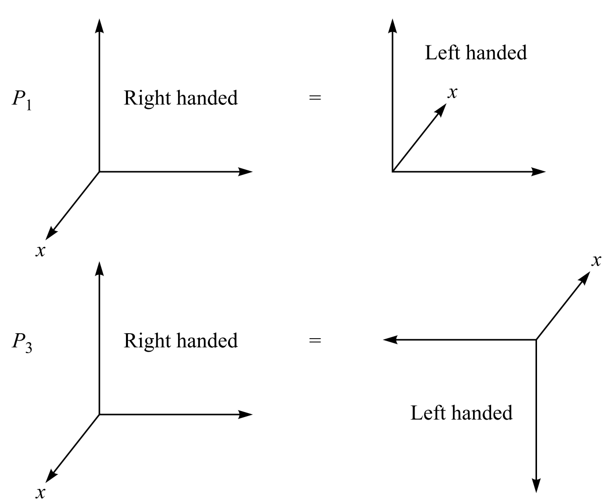

Figure 3.3. Parity transformations of the coordinate system.

https://doi.org/10.1017/CBO9780511755781.004 Published online by Cambridge University Press

that is, by $-\theta$, we recover the original coordinate system. Thus

$$\tilde{R}(-\theta)\tilde{R}(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \tilde{R}^{-1}(\theta)\tilde{R}(\theta).$$

Hence

$$\tilde{R}^{-1}(\theta) = \tilde{R}(-\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} = \tilde{R}^T(\theta),$$

which shows that a rotation matrix is an orthogonal matrix.

We would like to make one remark on rotation in space. In the above discussion, we have considered the vector to be fixed and rotated the coordinate axes. The rotation matrix can be thought of as an operator that, acting on the unprimed system, transforms it into the primed system. This view is often called the passive view of rotation. We could equally well keep the coordinate axes fixed and rotate the vector through an equal angle, but in the opposite direction. Then the rotation matrix would be thought of as an operator acting on the vector, say $\mathbf{X}$, and changing it into $\mathbf{X}'$. This procedure is called the active view of the rotation.

Trace of a matrix

Recall that the trace of a square matrix $\tilde{A}$ is defined as the sum of all the principal diagonal elements:

$$\mathrm{Tr}\,\tilde{A} = \sum_k a_{kk}.$$

It can be proved that the trace of the product of a finite number of matrices is invariant under any *cyclic* permutation of the matrices. We leave this as home work.

Orthogonal and unitary transformations

Eq. (3.42) is a linear transformation and it is called an orthogonal transformation, because the rotation matrix is an orthogonal matrix. One of the properties of an orthogonal transformation is that it preserves the length of a vector. A more useful linear transformation in physics is the unitary transformation:

$$\tilde{Y} = \tilde{U}\tilde{X} \tag{3.48}$$

in which $\tilde{X}$ and $\tilde{Y}$ are column matrices (vectors) of order $n \times 1$ and $\tilde{U}$ is a unitary matrix of order $n \times n$. One of the properties of a unitary transformation is that it

https://doi.org/10.1017/CBO9780511755781.004 Published online by Cambridge University Press

preserves the norm of a vector. To see this, premultiplying Eq. (3.48) by $\tilde{Y}^\dagger(=\tilde{X}^\dagger\tilde{U}^\dagger)$ and using the condition $\tilde{U}^\dagger\tilde{U}=\tilde{I}$, we obtain

$$\tilde{Y}^\dagger\tilde{Y}=\tilde{X}^\dagger\tilde{U}^\dagger\tilde{U}\tilde{X}=\tilde{X}^\dagger\tilde{X} \tag{3.49a}$$

or

$$\sum_{k=1}^{n} y_k^* y_k=\sum_{k=1}^{n} x_k^* x_k. \tag{3.49b}$$

This shows that the norm of a vector remains invariant under a unitary transformation. If the matrix $\tilde{U}$ of transformation happens to be real, then $\tilde{U}$ is also an orthogonal matrix and the transformation (3.48) is an orthogonal transformation, and Eqs. (3.49) reduce to

$$\tilde{Y}^{\mathrm{T}}\tilde{Y}=\tilde{X}^{\mathrm{T}}\tilde{X}, \tag{3.50a}$$

$$\sum_{k=1}^{n} y_k^2=\sum_{k=1}^{n} x_k^2, \tag{3.50b}$$

as we expected.

Similarity transformation

We now consider a different linear transformation, the similarity transformation that, we shall see later, is very useful in diagonalization of a matrix. To get the idea about similarity transformations, we consider vectors **r** and **R** in a particular basis, the coordinate system $Ox_1x_2x_3$, which are connected by a square matrix $\tilde{A}$:

$$\mathbf{R}=\tilde{A}\mathbf{r}. \tag{3.51a}$$

Now rotating the coordinate system about the origin O we obtain a new system $Ox_1'x_2'x_3'$ (a new basis). The vectors **r** and **R** have not been affected by this rotation. Their components, however, will have different values in the new system, and we now have

$$\mathbf{R}'=\tilde{A}'\mathbf{r}'. \tag{3.51b}$$

The matrix $\tilde{A}'$ in the new (primed) system is called similar to the matrix $\tilde{A}$ in the old (unprimed) system, since they perform same function. Then what is the relationship between matrices $\tilde{A}$ and $\tilde{A}'$? This information is given in the form of coordinate transformation. We learned in the previous section that the components of a vector in the primed and unprimed systems are connected by a matrix equation similar to Eq. (3.43). Thus we have

$$\mathbf{r}=\tilde{S}\mathbf{r}' \quad \text{and} \quad \mathbf{R}=\tilde{S}\mathbf{R}',$$

where $\tilde{S}$ is a non-singular matrix, the transition matrix from the new coordinate system to the old system. With these, Eq. (3.51a) becomes

$$\tilde{S}\mathbf{R}' = \tilde{A}\tilde{S}\mathbf{r}'$$

or

$$\mathbf{R}' = \tilde{S}^{-1}\tilde{A}\tilde{S}\mathbf{r}'.$$

Combining this with Eq. (3.51) gives

$$\tilde{A}' = \tilde{S}^{-1}\tilde{A}\tilde{S}, \tag{3.52}$$

where $\tilde{A}'$ and $\tilde{A}$ are similar matrices. Eq. (3.52) is called a similarity transformation.

Generalization of this idea to n-dimensional vectors is straightforward. In this case, we take $\mathbf{r}$ and $\mathbf{R}$ as two n-dimensional vectors in a particular basis, having their coordinates connected by the matrix $\tilde{A}$ (a $n \times n$ square matrix) through Eq. (3.51a). In another basis they are connected by Eq. (3.51b). The relationship between $\tilde{A}$ and $\tilde{A}'$ is given by Eq. (3.52). The transformation of $\tilde{A}$ into $\tilde{S}^{-1}\ \tilde{A}\tilde{S}$ is called a similarity transformation.

All identities involving vectors and matrices will remain invariant under a similarity transformation since this arises only in connection with a change in basis. That this is so can be seen in the following two simple examples.

Example 3.10

Given the matrix equation $\tilde{A}\tilde{B} = \tilde{C}$, and the matrices $\tilde{A}$, $\tilde{B}$, $\tilde{C}$ subjected to the same similarity transformation, show that the matrix equation is invariant.

Solution: Since the three matrices are all subjected to the same similarity transformation, we have

$$\tilde{A}' = \tilde{S}\tilde{A}\tilde{S}^{-1}, \quad \tilde{B}' = \tilde{S}\tilde{B}\tilde{S}^{-1}, \quad \tilde{C}' = \tilde{S}\tilde{C}\tilde{S}^{-1}$$

and it follows that

$$\tilde{A}'\tilde{B}' = (\tilde{S}\tilde{A}\tilde{S}^{-1})(\tilde{S}\tilde{B}\tilde{S}^{-1}) = \tilde{S}\tilde{A}\tilde{I}\tilde{B}\tilde{S}^{-1} = \tilde{S}\tilde{A}\tilde{B}\tilde{S}^{-1} = \tilde{S}\tilde{C}\tilde{S}^{-1} = \tilde{C}'.$$

Example 3.11

Show that the relation $\tilde{A}\mathbf{R} = \tilde{B}\mathbf{r}$ is invariant under a similarity transformation.

Solution: Since matrices $\tilde{A}$ and $\tilde{B}$ are subjected to the same similarity transformation, we have

$$\tilde{A}' = \tilde{S}\tilde{A}\tilde{S}^{-1}, \quad \tilde{B}' = \tilde{S}\tilde{B}\tilde{S}^{-1}$$

we also have

$$\mathbf{R}' = \tilde{S}\mathbf{R}, \qquad \mathbf{r}' = \tilde{S}\mathbf{r}.$$

Then

$$\tilde{A}'\mathbf{R}' = (\tilde{S}\tilde{A}\tilde{S}^{-1})(S\mathbf{R}) = \tilde{S}\tilde{A}\mathbf{R} \quad \text{and} \quad \tilde{B}'\mathbf{r}' = (\tilde{S}\tilde{B}\tilde{S}^{-1})(S\mathbf{r}) = \tilde{S}\tilde{B}\mathbf{r}$$

thus

$$\tilde{A}'\mathbf{R}' = \tilde{B}'\mathbf{r}'.$$

We shall see in the following section that similarity transformations are very useful in diagonalization of a matrix, and that two similar matrices have the same eigenvalues.

The matrix eigenvalue problem

As we saw in preceding sections, a linear transformation generally carries a vector $\mathbf{X} = (x_1, x_2, \ldots, x_n)$ into a vector $\mathbf{Y} = (y_1, y_2, \ldots, y_n)$. However, there may exist certain non-zero vectors for which $\tilde{A}\mathbf{X}$ is just $\mathbf{X}$ multiplied by a constant λ

$$\tilde{A}\mathbf{X} = \lambda\mathbf{X}. \tag{3.53}$$

That is, the transformation represented by the matrix (operator) $\tilde{A}$ just multiplies the vector $\mathbf{X}$ by a number λ. Such a vector is called an eigenvector of the matrix $\tilde{A}$, and λ is called an eigenvalue (German: *eigenwert*) or characteristic value of the matrix $\tilde{A}$. The eigenvector is said to 'belong' (or correspond) to the eigenvalue. And the set of the eigenvalues of a matrix (an operator) is called its eigenvalue spectrum.

The problem of finding the eigenvalues and eigenvectors of a matrix is called an eigenvalue problem. We encounter problems of this type in all branches of physics, classical or quantum. Various methods for the approximate determination of eigenvalues have been developed, but here we only discuss the fundamental ideas and concepts that are important for the topics discussed in this book.

There are two parts to every eigenvalue problem. First, we compute the eigenvalue λ, given the matrix $\tilde{A}$. Then, we compute an eigenvector X for each previously computed eigenvalue λ.

Determination of eigenvalues and eigenvectors

We shall now demonstrate that any square matrix of order n has at least 1 and at most n distinct (real or complex) eigenvalues. To this purpose, let us rewrite the system of Eq. (3.53) as

$$(\tilde{A} - \lambda\tilde{I})X = 0. \tag{3.54}$$

This matrix equation really consists of n homogeneous linear equations in the n unknown elements x_i of X:

$$\left.\begin{aligned}(a_{11}-\lambda)x_1+a_{12}x_2+\cdots+a_{1n}x_n&=0\\ a_{21}x_1+(a_{22}-\lambda)x_2+\cdots+a_{2n}x_n&=0\\ &\cdots\\ a_{n1}x_1+a_{n2}x_2+\cdots+(a_{nn}-\lambda)x_n&=0\end{aligned}\right\} \tag{3.55}$$

In order to have a non-zero solution, we recall that the determinant of the coefficients must be zero; that is,

$$\det(\tilde{A}-\lambda\tilde{I})=\begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22}-\lambda & \cdots & a_{2n}\\ \vdots & \vdots & & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda\end{vmatrix}=0. \tag{3.56}$$

The expansion of the determinant gives an nth order polynomial equation in λ, and we write this as

$$c_0\lambda^n+c_1\lambda^{n-1}+c_2\lambda^{n-2}+\cdots+c_{n-1}\lambda+c_n=0, \tag{3.57}$$

where the coefficients c_i are functions of the elements a_{jk} of $\tilde{A}$. Eq. (3.56) or (3.57) is called the characteristic equation corresponding to the matrix $\tilde{A}$. We have thus obtained a very important result: the eigenvalues of a square matrix $\tilde{A}$ are the roots of the corresponding characteristic equation (3.56) or (3.57).

Some of the coefficients c_i can be readily determined; by an inspection of Eq. (3.56) we find

$$c_0=(-1)^n,\quad c_1=(-1)^{n-1}(a_{11}+a_{22}+\cdots+a_{nn}),\quad c_n=\det\tilde{A}. \tag{3.58}$$

Now let us rewrite the characteristic polynomial in terms of its n roots $\lambda_1,\lambda_2,\ldots,\lambda_n$

$$c_0\lambda^n+c_1\lambda^{n-1}+c_2\lambda^{n-2}+\cdots+c_{n-1}\lambda+c_n=(\lambda_1-\lambda)(\lambda_2-\lambda)\cdots(\lambda_n-\lambda),$$

then we see that

$$c_1=(-1)^{n-1}(\lambda_1+\lambda_2+\cdots+\lambda_n),\quad c_n=\lambda_1\lambda_2\cdots\lambda_n. \tag{3.59}$$

Comparing this with Eq. (3.58), we obtain the following two important results on the eigenvalues of a matrix:

(1) The sum of the eigenvalues equals the trace (spur) of the matrix:

$$\lambda_1+\lambda_2+\cdots+\lambda_n=a_{11}+a_{22}+\cdots+a_{nn}\equiv\mathrm{Tr}\,\tilde{A}. \tag{3.60}$$

(2) The product of the eigenvalues equals the determinant of the matrix:

$$\lambda_1\lambda_2\cdots\lambda_n=\det\tilde{A}. \tag{3.61}$$

Once the eigenvalues have been found, corresponding eigenvectors can be found from the system (3.55). Since the system is homogeneous, if X is an eigenvector of $\tilde{A}$, then kX, where k is any constant (not zero), is also an eigenvector of $\tilde{A}$ corresponding to the same eigenvalue. It is very easy to show this. Since $\tilde{A}X = \lambda X$, multiplying by an arbitrary constant k will give $k\tilde{A}X = k\lambda X$. Now $k\tilde{A} = \tilde{A}k$ (every matrix commutes with a scalar), so we have $\tilde{A}(kX) = \lambda(kX)$, showing that kX is also an eigenvector of $\tilde{A}$ with the same eigenvalue λ. But kX is linearly dependent on X, and if we were to count all such eigenvectors separately, we would have an infinite number of them. Such eigenvectors are therefore not counted separately.

A matrix of order n does not necessarily have n linearly independent eigenvectors; some of them may be repeated. (This will happen when the characteristic polynomial has two or more identical roots.) If an eigenvalue occurs m times, m is called the multiplicity of the eigenvalue. The matrix has at most m linearly independent eigenvectors all corresponding to the same eigenvalue. Such linearly independent eigenvectors having the same eigenvalue are said to be degenerate eigenvectors; in this case, m-fold degenerate. We will deal only with those matrices that have n linearly independent eigenvectors and they are diagonalizable matrices.

Example 3.12

Find (*a*) the eigenvalues and (*b*) the eigenvectors of the matrix

$$\tilde{A} = \begin{pmatrix} 5 & 4 \\ 1 & 2 \end{pmatrix}.$$

Solution: (*a*) The eigenvalues: The characteristic equation is

$$\det(\tilde{A} - \lambda\tilde{I}) = \begin{vmatrix} 5-\lambda & 4 \\ 1 & 2-\lambda \end{vmatrix} = \lambda^2 - 7\lambda + 6 = 0$$

which has two roots

$$\lambda_1 = 6 \quad \text{and} \quad \lambda_2 = 1.$$

(*b*) The eigenvectors: For $\lambda = \lambda_1$ the system (3.55) assumes the form

$$-x_1 + 4x_2 = 0,$$
$$x_1 - 4x_2 = 0.$$

Thus $x_1 = 4x_2$, and

$$X_1 = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

is an eigenvector of $\tilde{A}$ corresponding to $\lambda_1 = 6$. In the same way we find the eigenvector corresponding to $\lambda_2 = 1$:

$$X_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Example 3.13
If $\tilde{A}$ is a non-singular matrix, show that the eigenvalues of $\tilde{A}^{-1}$ are the reciprocals of those of $\tilde{A}$ and every eigenvector of $\tilde{A}$ is also an eigenvector of $\tilde{A}^{-1}$.

Solution: Let λ be an eigenvalue of $\tilde{A}$ corresponding to the eigenvector X, so that

$$\tilde{A}X = \lambda X.$$

Since $\tilde{A}^{-1}$ exists, multiply the above equation from the left by $\tilde{A}^{-1}$

$$\tilde{A}^{-1}\tilde{A}X = \tilde{A}^{-1}\lambda X \Rightarrow X = \lambda\tilde{A}^{-1}X.$$

Since $\tilde{A}$ is non-singular, λ must be non-zero. Now dividing the above equation by λ, we have

$$\tilde{A}^{-1}X = (1/\lambda)X.$$

Since this is true for every value of $\tilde{A}$, the results follows.

Example 3.14
Show that all the eigenvalues of a unitary matrix have unit magnitude.

Solution: Let $\tilde{U}$ be a unitary matrix and X an eigenvector of $\tilde{U}$ with the eigenvalue λ, so that

$$\tilde{U}X = \lambda X.$$

Taking the hermitian conjugate of both sides, we have

$$X^\dagger\tilde{U}^\dagger = \lambda^* X^\dagger.$$

Multiplying the first equation from the left by the second equation, we obtain

$$X^\dagger\tilde{U}^\dagger\tilde{U}X = \lambda\lambda^* X^\dagger X.$$

Since $\tilde{U}$ is unitary, $\tilde{U}^\dagger\tilde{U} = \tilde{I}$, so that the last equation reduces to

$$X^\dagger X(|\lambda|^2 - 1) = 0.$$

Now $X^\dagger X$ is the square of the norm of X and hence cannot vanish unless X is a null vector and so we must have $|\lambda|^2 = 1$ or $|\lambda| = 1$, proving the desired result.

Example 3.15
Show that similar matrices have the same characteristic polynomial and hence the same eigenvalues. (Another way of stating this is to say that the eigenvalues of a matrix are invariant under similarity transformations.)

Solution: Let $\tilde{A}$ and $\tilde{B}$ be similar matrices. Thus there exists a third matrix $\tilde{S}$ such that $\tilde{B} = \tilde{S}^{-1}\tilde{A}\tilde{S}$. Substituting this into the characteristic polynomial of matrix $\tilde{B}$ which is $|\tilde{B}\lambda - \tilde{I}|$, we obtain

$$|\tilde{B} - \lambda I| = |\tilde{S}^{-1}\tilde{A}\tilde{S} - \lambda\tilde{I}| = |\tilde{S}^{-1}(\tilde{A} - \lambda\tilde{I})\tilde{S}|.$$

Using the properties of determinants, we have

$$|\tilde{S}^{-1}(\tilde{A} - \lambda\tilde{I})\tilde{S}| = |\tilde{S}^{-1}||\tilde{A} - \lambda\tilde{I}||\tilde{S}|.$$

Then it follows that

$$|\tilde{B} - \lambda\tilde{I}| = |\tilde{S}^{-1}(\tilde{A} - \lambda\tilde{I})\tilde{S}| = |\tilde{S}^{-1}||\tilde{A} - \lambda\tilde{I}||\tilde{S}| = |\tilde{A} - \lambda\tilde{I}|,$$

which shows that the characteristic polynomials of $\tilde{A}$ and $\tilde{B}$ are the same; their eigenvalues will also be identical.

Eigenvalues and eigenvectors of hermitian matrices

In quantum mechanics complex variables are unavoidable because of the form of the Schrödinger equation. And all quantum observables are represented by hermitian operators. So physicists are almost always dealing with adjoint matrices, hermitian matrices, and unitary matrices. Why are physicists interested in hermitian matrices? Because they have the following properties: (1) the eigenvalues of a hermitian matrix are real, and (2) its eigenvectors corresponding to distinct eigenvalues are orthogonal, so they can be used as basis vectors. We now proceed to prove these important properties.

(1) the eigenvalues of a hermitian matrix are real.
Let $\tilde{H}$ be a hermitian matrix and X a non-trivial eigenvector corresponding to the eigenvalue λ, so that

$$\tilde{H}X = \lambda X. \tag{3.62}$$

Taking the hermitian conjugate and note that $\tilde{H}^\dagger = \tilde{H}$, we have

$$X^\dagger\tilde{H} = \lambda^* X^\dagger. \tag{3.63}$$

Multiplying (3.62) from the left by $X^\dagger$, and (3.63) from the right by $X^\dagger$, and then subtracting, we get

$$(\lambda - \lambda^*)X^\dagger X = 0. \tag{3.64}$$

Now, since $X^\dagger X$ cannot be zero, it follows that $\lambda = \lambda^*$, or that λ is real.

(2) The eigenvectors corresponding to distinct eigenvalues are orthogonal.
Let X_1 and X_2 be eigenvectors of $\tilde{H}$ corresponding to the distinct eigenvalues λ_1 and λ_2, respectively, so that

$$\tilde{H}X_1 = \lambda_1 X_1, \tag{3.65}$$

$$\tilde{H}X_2 = \lambda_2 X_2. \tag{3.66}$$

Taking the hermitian conjugate of (3.66) and noting that $\lambda^* = \lambda$, we have

$$X_2^\dagger \tilde{H} = \lambda_2 X_2^\dagger. \tag{3.67}$$

Multiplying (3.65) from the left by $X_2^\dagger$ and (3.67) from the right by X_1, then subtracting, we obtain

$$(\lambda_1 - \lambda_2)X_2^\dagger + X_1 = 0. \tag{3.68}$$

Since $\lambda_1 = \lambda_2$, it follows that $X_2^\dagger X_1 = 0$ or that X_1 and X_2 are orthogonal.

If X is an eigenvector of $\tilde{H}$, any multiple of X, λX, is also an eigenvector of $\tilde{H}$. Thus we can normalize the eigenvector X with a properly chosen scalar λ. This means that the eigenvectors of $\tilde{H}$ corresponding to distinct eigenvalues are orthonormal. Just as the three orthogonal unit coordinate vectors $\hat{e}_1, \hat{e}_2,$ and $\hat{e}_3$ form the basis of a three-dimensional vector space, the orthonormal eigenvectors of $\tilde{H}$ may serve as a basis for a function space.

Diagonalization of a matrix

Let $\tilde{A} = (a_{ij})$ be a square matrix of order n, which has n linearly independent eigenvectors X_i with the corresponding eigenvalues λ_i: $\tilde{A}X_i = \lambda_i X_i$. If we denote the eigenvectors X_i by column vectors with elements $x_{1i}, x_{2i}, \ldots, x_{ni}$, then the eigenvalue equation can be written in matrix form:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{pmatrix} = \lambda_i \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{pmatrix}. \tag{3.69}$$

From the above matrix equation we obtain

$$\sum_{k=1}^{n} a_{jk} x_{ki} = \lambda_i x_{ji}. \tag{3.69b}$$

Now we want to diagonalize $\tilde{A}$. To this purpose, we can follow these steps. We first form a matrix $\tilde{S}$ of order $n \times n$ whose columns are the vector X_i, that is,

$$\tilde{S} = \begin{pmatrix} x_{11} & \cdots & x_{1i} & \cdots & x_{1n} \\ x_{21} & \cdots & x_{2i} & \cdots & x_{2n} \\ \vdots & & \vdots & & \vdots \\ x_{n1} & \cdots & x_{ni} & \cdots & x_{nn} \end{pmatrix}, \quad (\tilde{S})_{ij} = x_{ij}. \tag{3.70}$$

Since the vectors X_i are linear independent, $\tilde{S}$ is non-singular and $\tilde{S}^{-1}$ exists. We then form a matrix $\tilde{S}^{-1}\tilde{A}\tilde{S}$; this is a diagonal matrix whose diagonal elements are the eigenvalues of $\tilde{A}$.

To show this, we first define a diagonal matrix $\tilde{B}$ whose diagonal elements are λ_i $(i = 1, 2, \ldots, n)$:

$$\tilde{B} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}, \tag{3.71}$$

and we then demonstrate that

$$\tilde{S}^{-1}\tilde{A}\tilde{S} = \tilde{B}. \tag{3.72a}$$

Eq. (3.72a) can be rewritten by multiplying it from the left by $\tilde{S}$ as

$$\tilde{A}\tilde{S} = \tilde{S}\tilde{B}. \tag{3.72b}$$

Consider the left hand side first. Taking the jith element, we obtain

$$(\tilde{A}\tilde{S})_{ji} = \sum_{k=1}^{n} (\tilde{A})_{jk}(\tilde{S})_{ki} = \sum_{k=1}^{n} a_{jk}x_{ki}. \tag{3.73a}$$

Similarly, the jith element of the right hand side is

$$(\tilde{S}\tilde{B})_{ji} = \sum_{k=1}^{n} (\tilde{S})_{jk}(\tilde{B})_{ki} = \sum_{k=1}^{n} x_{jk}\lambda_i\delta_{ki} = \lambda_i x_{ji}. \tag{3.73b}$$

Eqs. (3.73a) and (3.73b) clearly show the validity of Eq. (3.72a).

It is important to note that the matrix $\tilde{S}$ that is able to diagonalize matrix $\tilde{A}$ is not unique. This is because we could arrange the eigenvectors $X_1, X_2, \ldots, X_n$ in any order to construct $\tilde{S}$.

We summarize the procedure for diagonalizing a diagonalizable $n \times n$ matrix $\tilde{A}$:

Step 1. Find n linearly independent eigenvectors of $\tilde{A}$, X_1, $X_2, \ldots, X_n$.
Step 2. Form the matrix $\tilde{S}$ having $X_1, X_2, \ldots, X_n$ as its column vectors.
Step 3. Find the inverse of $\tilde{S}$, $\tilde{S}^{-1}$.
Step 4. The matrix $\tilde{S}^{-1}\tilde{A}\tilde{S}$ will then be diagonal with $\lambda_1, \lambda_2, \ldots, \lambda_n$ as its successive diagonal elements, where λ_i is the eigenvalue corresponding to X_i.

Example 3.16
Find a matrix $\tilde{S}$ that diagonalizes

$$\tilde{A} = \begin{pmatrix} 3 & -2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}.$$

Solution: We have first to find the eigenvalues and the corresponding eigenvectors of matrix $\tilde{A}$. The characteristic equation of $\tilde{A}$ is

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

so that the eigenvalues of $\tilde{A}$ are $\lambda = 1$ and $\lambda = 5$.

By definition

$$\tilde{X} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

is an eigenvector of $\tilde{A}$ corresponding to λ if and only if $\tilde{X}$ is a non-trivial solution of $(\lambda\tilde{I} - \tilde{A})\tilde{X} = 0$, that is, of

$$\begin{pmatrix} \lambda-3 & 2 & 0 \\ 2 & \lambda-3 & 0 \\ 0 & 0 & \lambda-5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

If $\lambda = 5$ the above equation becomes

$$\begin{pmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 2x_1 + 2x_2 + 0x_3 \\ 2x_1 + 2x_2 + 0x_3 \\ 0x_1 + 0x_2 + 0x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Solving this system yields

$$x_1 = -s, \quad x_2 = s, \quad x_3 = t,$$

where s and t are arbitrary values. Thus the eigenvectors of $\tilde{A}$ corresponding to $\lambda = 5$ are the non-zero vectors of the form

$$\tilde{X} = \begin{pmatrix} -s \\ s \\ t \end{pmatrix} = \begin{pmatrix} -s \\ s \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ t \end{pmatrix} = s\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Since

$$\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

are linearly independent, they are the eigenvectors corresponding to $\lambda = 5$.

For $\lambda = 1$, we have

$$\begin{pmatrix} -2 & 2 & 0 \\ 2 & -2 & 0 \\ 0 & 0 & -4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} -2x_1 + 2x_2 + 0x_3 \\ 2x_1 - 2x_2 + 0x_3 \\ 0x_1 + 0x_2 - 4x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Solving this system yields

$$x_1 = t, \quad x_2 = t, \quad x_3 = 0,$$

where t is arbitrary. Thus the eigenvectors corresponding to $\lambda = 1$ are non-zero vectors of the form

$$\tilde{X} = \begin{pmatrix} t \\ t \\ 0 \end{pmatrix} = t \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

It is easy to check that the three eigenvectors

$$\tilde{X}_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \quad \tilde{X}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \tilde{X}_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix},$$

are linearly independent. We now form the matrix $\tilde{S}$ that has $\tilde{X}_1$, $\tilde{X}_2$, and $\tilde{X}_3$ as its column vectors:

$$\tilde{S} = \begin{pmatrix} -1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

The matrix $\tilde{S}^{-1}\tilde{A}\tilde{S}$ is diagonal:

$$\tilde{S}^{-1}\tilde{A}\tilde{S} = \begin{pmatrix} -1/2 & 1/2 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{pmatrix} \begin{pmatrix} 3 & -2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix} \begin{pmatrix} -1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

There is no preferred order for the columns of $\tilde{S}$. If had we written

$$\tilde{S} = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

then we would have obtained (verify)

$$\tilde{S}^{-1}\tilde{A}\tilde{S} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Example 3.17
Show that the matrix

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

is not diagonalizable.

Solution: The characteristic equation of $\tilde{A}$ is

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

Thus $\lambda = -1$ the only eigenvalue of $\tilde{A}$; the eigenvectors corresponding to $\lambda = -1$ are the solutions of

$$\begin{pmatrix} \lambda+3 & -2 \\ 2 & \lambda-1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 2 & -2 \\ 2 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

from which we have

$$2x_1 - 2x_2 = 0,$$
$$2x_1 - 2x_2 = 0.$$

The solutions to this system are $x_1 = t,\ x_2 = t$; hence the eigenvectors are of the form

$$\begin{pmatrix} t \\ t \end{pmatrix} = t\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

A does not have two linearly independent eigenvectors, and is therefore not diagonalizable.

Eigenvectors of commuting matrices

There is a theorem on eigenvectors of commuting matrices that is of great importance in matrix algebra as well as in quantum mechanics. This theorem states that:

Two commuting matrices possess a common set of eigenvectors.

We now proceed to prove it. Let $\tilde{A}$ and $\tilde{B}$ be two square matrices, each of order n, which commute with each other, that is,

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

First, let λ be an eigenvalue of $\tilde{A}$ with multiplicity 1, corresponding to the eigenvector X, so that

$$\tilde{A}X = \lambda X. \tag{3.74}$$

Multiplying both sides from the left by $\tilde{B}$

$$\tilde{B}\tilde{A}X = \lambda\tilde{B}X.$$

Because $\tilde{B}\tilde{A} = \tilde{A}\tilde{B}$, we have

$$\tilde{A}(\tilde{B}X) = (\lambda\tilde{B}X).$$

Now $\tilde{B}$ is an $n \times n$ matrix and X is an $n \times 1$ vector; hence $\tilde{B}X$ is also an $n \times 1$ vector. The above equation shows that $\tilde{B}X$ is also an eigenvector of $\tilde{A}$ with the eigenvalue λ. Now X is a non-degenerate eigenvector of $\tilde{A}$, any other vector which is an eigenvector of $\tilde{A}$ with the same eigenvalue as that of X must be multiple of X. Accordingly

$$\tilde{B}X = \mu X,$$

where μ is a scalar. Thus we have proved that:

> If two matrices commute, every non-degenerate eigenvector of one is also an eigenvector of the other, and vice versa.

Next, let λ be an eigenvalue of $\tilde{A}$ with multiplicity k. So $\tilde{A}$ has k linearly independent eigenvectors, say $X_1, X_2, \ldots, X_k$, each corresponding to λ:

$$\tilde{A}X_i = \lambda X_i, \quad 1 \le i \le k.$$

Multiplying both sides from the left by $\tilde{B}$, we obtain

$$\tilde{A}(\tilde{B}X_i) = \lambda(\tilde{B}X_i),$$

which shows again that $\tilde{B}X$ is also an eigenvector of $\tilde{A}$ with the same eigenvalue λ.

Cayley–Hamilton theorem

The Cayley–Hamilton theorem is useful in evaluating the inverse of a square matrix. We now introduce it here. As given by Eq. (3.57), the characteristic equation associated with a square matrix $\tilde{A}$ of order n may be written as a polynomial

$$f(\lambda) = \sum_{i=0}^{n} c_i \lambda^{n-i} = 0,$$

https://doi.org/10.1017/CBO9780511755781.004 Published online by Cambridge University Press

where λ are the eigenvalues given by the characteristic determinant (3.56). If we replace λ in $f(\lambda)$ by the matrix $\tilde{A}$ so that

$$f(\tilde{A}) = \sum_{i=0}^{n} c_i \tilde{A}^{n-i}.$$

The Cayley–Hamilton theorem says that

$$f(\tilde{A}) = 0 \quad \text{or} \quad \sum_{i=0}^{n} c_i \tilde{A}^{n-i} = 0, \tag{3.75}$$

that is, the matrix $\tilde{A}$ satisfies its characteristic equation.

We now formally multiply Eq. (3.75) by $\tilde{A}^{-1}$ so that we obtain

$$\tilde{A}^{-1} f(\tilde{A}) = c_0 \tilde{A}^{n-1} + c_1 \tilde{A}^{n-2} + \cdots + c_{n-1}\tilde{I} + c_n \tilde{A}^{-1} = 0.$$

Solving for $\tilde{A}^{-1}$ gives

$$\tilde{A}^{-1} = -\frac{1}{c_n}\left[\sum_{i=0}^{n-1} c_i \tilde{A}^{n-1-i}\right]; \tag{3.76}$$

we can use this to find $\tilde{A}^{-1}$ (Problem 3.28).

Moment of inertia matrix

We shall see that physically diagonalization amounts to a simplification of the problem by a better choice of variable or coordinate system. As an illustrative example, we consider the moment of inertia matrix $\tilde{I}$ of a rotating rigid body (see Fig. 3.4). A rigid body can be considered to be a many-particle system, with the

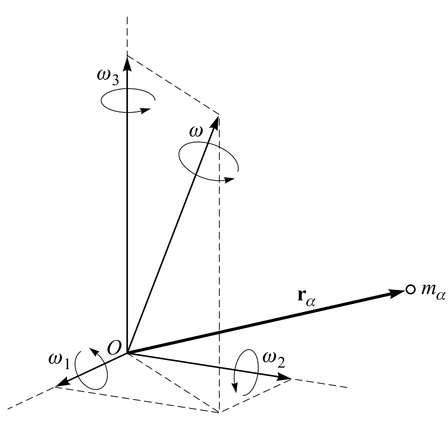

Figure 3.4. A rotating rigid body.

https://doi.org/10.1017/CBO9780511755781.004 Published online by Cambridge University Press

distance between any particle pair constant at all times. Then its angular momentum about the origin O of the coordinate system is

$$\mathbf{L} = \sum_{\alpha} m_\alpha \mathbf{r}_\alpha \times \mathbf{v}_\alpha = \sum_{\alpha} m_\alpha \mathbf{r}_\alpha \times (\boldsymbol{\omega} \times \mathbf{r}_\alpha)$$

where the subscript α refers to mass m_a located at $\mathbf{r}_\alpha = (x_{\alpha 1}, x_{\alpha 2}, x_{\alpha 3})$, and $\boldsymbol{\omega}$ the angular velocity of the rigid body.

Expanding the vector triple product by using the vector identity

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B}),$$

we obtain

$$\mathbf{L} = \sum_{\alpha} m_\alpha \lfloor r_\alpha^2 \boldsymbol{\omega} - \mathbf{r}_\alpha(\mathbf{r}_\alpha \cdot \boldsymbol{\omega}) \rfloor.$$

In terms of the components of the vectors $\mathbf{r}_\alpha$and $\boldsymbol{\omega}$, the ith component of L_i is

$$\begin{aligned} L_i &= \sum_{\alpha} m_\alpha \left[\omega_i \sum_{k=1}^{3} x_{\alpha,k}^2 - x_{\alpha,i} \sum_{j=1}^{3} x_{\alpha,j}\omega_j \right] \\ &= \sum_j \omega_j \sum_\alpha m_\alpha \left[\delta_{ij} \sum_k x_{\alpha,k}^2 - x_{\alpha,i} x_{\alpha,j} \right] = \sum_j I_{ij}\omega_j \end{aligned}$$

or

$$\tilde{L} = \tilde{I}\tilde{\omega}.$$

Both $\tilde{L}$ and $\tilde{\omega}$ are three-dimensional column vectors, while $\tilde{I}$ is a 3×3 matrix and is called the moment inertia matrix.

In general, the angular momentum vector $\mathbf{L}$ of a rigid body is not always parallel to its angular velocity $\boldsymbol{\omega}$ and $\tilde{I}$ is not a diagonal matrix. But we can orient the coordinate axes in space so that all the non-diagonal elements I_{ij} $(i \neq j)$ vanish. Such special directions are called the principal axes of inertia. If the angular velocity is along one of these principal axes, the angular momentum and the angular velocity will be parallel.

In many simple cases, especially when symmetry is present, the principal axes of inertia can be found by inspection.

Normal modes of vibrations

Another good illustrative example of the application of matrix methods in classical physics is the longitudinal vibrations of a classical model of a carbon dioxide molecule that has the chemical structure O–C–O. In particular, it provides a good example of the eigenvalues and eigenvectors of an asymmetric real matrix.

https://doi.org/10.1017/CBO9780511755781.004 Published online by Cambridge University Press

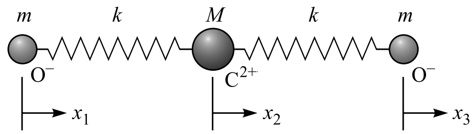

Figure 3.5. A linear symmetrical carbon dioxide molecule.

We can regard a carbon dioxide molecule as equivalent to a set of three particles jointed by elastic springs (Fig. 3.5). Clearly the system will vibrate in some manner in response to an external force. For simplicity we shall consider only longitudinal vibrations, and the interactions of the oxygen molecules with one another will be neglected, so we consider only nearest neighbor interaction. The Lagrangian function L for the system is

$$L = \tfrac{1}{2}m(\dot{x}_1^2 + \dot{x}_3^2) + \tfrac{1}{2}M\dot{x}_2^2 - \tfrac{1}{2}k(x_2 - x_1)^2 - \tfrac{1}{2}k(x_3 - x_2)^2;$$

substituting this into Lagrange's equations

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}_i}\right) - \frac{\partial L}{\partial x_i} = 0 \quad (i = 1, 2, 3),$$

we find the equations of motion to be

$$\begin{aligned}
\ddot{x}_1 &= -\frac{k}{m}(x_1 - x_2) = -\frac{k}{m}x_1 + \frac{k}{m}x_2, \\
\ddot{x}_2 &= -\frac{k}{M}(x_2 - x_1) - \frac{k}{M}(x_2 - x_3) = \frac{k}{M}x_1 - \frac{2k}{M}x_2 + \frac{k}{M}x_3, \\
\ddot{x}_3 &= \frac{k}{m}x_2 - \frac{k}{m}x_3,
\end{aligned}$$

where the dots denote time derivatives. If we define

$$\tilde{X} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \qquad \tilde{A} = \begin{pmatrix} -\dfrac{k}{m} & \dfrac{k}{m} & 0 \\ -\dfrac{k}{M} & -\dfrac{2k}{M} & \dfrac{k}{M} \\ 0 & \dfrac{k}{m} & -\dfrac{k}{m} \end{pmatrix}$$

and, furthermore, if we define the derivative of a matrix to be the matrix obtained by differentiating each matrix element, then the above system of differential equations can be written as

$$\ddot{\tilde{X}} = \tilde{A}\tilde{X}.$$

This matrix equation is reminiscent of the single differential equation $\ddot{x} = ax$, with a a constant. The latter always has an exponential solution. This suggests that we try

$$\tilde{X} = \tilde{C}e^{\omega t},$$

where ω is to be determined and

$$\tilde{C} = \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix}$$

is an as yet unknown constant matrix. Substituting this into the above matrix equation, we obtain a matrix-eigenvalue equation

$$\tilde{A}\tilde{C} = \omega^2\tilde{C}$$

or

$$\begin{pmatrix} -\frac{k}{m} & \frac{k}{m} & 0 \\ -\frac{k}{M} & -\frac{2k}{M} & \frac{k}{M} \\ 0 & \frac{k}{m} & -\frac{k}{m} \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = \omega^2 \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix}. \tag{3.77}$$

Thus the possible values of ω are the square roots of the eigenvalues of the asymmetric matrix $\tilde{A}$ with the corresponding solutions being the eigenvectors of the matrix $\tilde{A}$. The secular equation is

$$\begin{vmatrix} -\frac{k}{m} - \omega^2 & \frac{k}{m} & 0 \\ -\frac{k}{M} & -\frac{2k}{M} - \omega^2 & \frac{k}{M} \\ 0 & \frac{k}{m} & -\frac{k}{m} - \omega^2 \end{vmatrix} = 0.$$

This leads to

$$\omega^2\left(-\omega^2 + \frac{k}{m}\right)\left(-\omega^2 + \frac{k}{m} + \frac{2k}{M}\right) = 0.$$

The eigenvalues are

$$\omega^2 = 0, \quad \frac{k}{m}, \quad \text{and} \quad \frac{k}{m} + \frac{2k}{M},$$

https://doi.org/10.1017/CBO9780511755781.004 Published online by Cambridge University Press

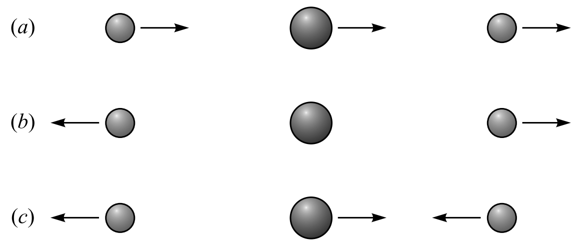

Figure 3.6. Longitudinal vibrations of a carbon dioxide molecule.

all real. The corresponding eigenvectors are determined by substituting the eigenvalues back into Eq. (3.77) one eigenvalue at a time:

(1) Setting $\omega^2 = 0$ in Eq. (3.77) we find that $C_1 = C_2 = C_3$. Thus this mode is not an oscillation at all, but is a pure translation of the system as a whole, no relative motion of the masses (Fig. 3.6(*a*)).
(2) Setting $\omega^2 = k/m$ in Eq. (3.77), we find $C_2 = 0$ and $C_3 = -C_1$. Thus the center mass M is stationary while the outer masses vibrate in opposite directions with the same amplitude (Fig. 3.6(*b*)).
(3) Setting $\omega^2 = k/m + 2k/M$ in Eq. (3.77), we find $C_1 = C_3$, and $C_2 = -2C_1(m/M)$. In this mode the two outer masses vibrate in unison and the center mass vibrates oppositely with different amplitude (Fig. 3.6(*c*)).

Direct product of matrices

Sometimes the direct product of matrices is useful. Given an $m \times m$ matrix $\tilde{A}$ and an $n \times n$ matrix $\tilde{B}$, the direct product of $\tilde{A}$ and $\tilde{B}$ is an $mn \times mn$ matrix, defined by

$$\tilde{C} = \tilde{A} \otimes \tilde{B} = \begin{pmatrix} a_{11}\tilde{B} & a_{12}\tilde{B} & \cdots & a_{1m}\tilde{B} \\ a_{21}\tilde{B} & a_{22}\tilde{B} & \cdots & a_{2m}\tilde{B} \\ \vdots & \vdots & & \vdots \\ a_{m1}\tilde{B} & a_{m2}\tilde{B} & \cdots & a_{mm}\tilde{B} \end{pmatrix}.$$

For example, if

$$\tilde{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad \tilde{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix},$$

then

https://doi.org/10.1017/CBO9780511755781.004 Published online by Cambridge University Press

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

Problems

3.1 For the pairs $\tilde{A}$ and $\tilde{B}$ given below, find $\tilde{A}+\tilde{B}$, $\tilde{A}\tilde{B}$, and $\tilde{A}^2$:

$$\tilde{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \qquad \tilde{B} = \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}.$$

3.2 Show that an n-rowed diagonal matrix $\tilde{D}$

$$\tilde{D} = \begin{pmatrix} k & 0 & \cdots & 0 \\ 0 & k & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ & & & k \end{pmatrix}$$

commutes with any n-rowed square matrix $\tilde{A}$: $\tilde{A}\tilde{D} = \tilde{D}\tilde{A} = k\tilde{A}$.

3.3 If $\tilde{A}$, $\tilde{B}$, and $\tilde{C}$ are any matrices such that the addition $\tilde{B}+\tilde{C}$ and the products $\tilde{A}\tilde{B}$ and $\tilde{A}\tilde{C}$ are defined, show that $\tilde{A}(\tilde{B}+\tilde{C}) = \tilde{A}\tilde{B}+\tilde{A}\tilde{C}$. That is, that matrix multiplication is distributive.

3.4 Given

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

show that $[\tilde{A}, \tilde{B}] = 0$, and $[\tilde{B}, \tilde{C}] = 0$, but that $\tilde{A}$ does not commute with $\tilde{C}$.

3.5 Prove that $(\tilde{A}+\tilde{B})^{\mathrm{T}} = \tilde{A}^{\mathrm{T}} + \tilde{B}^{\mathrm{T}}$.

3.6 Given

$$\tilde{A} = \begin{pmatrix} 2 & -3 \\ 0 & 4 \end{pmatrix}, \quad \tilde{B} = \begin{pmatrix} -5 & 2 \\ 2 & 1 \end{pmatrix}, \quad \text{and} \quad \tilde{C} = \begin{pmatrix} 0 & 1 & -2 \\ 3 & 0 & 4 \end{pmatrix}:$$

(*a*) Find $2\tilde{A} - 4\tilde{B}$, $2(\tilde{A} - 2\tilde{B})$
(*b*) Find $\tilde{A}^{\mathrm{T}}, \tilde{B}^{\mathrm{T}}, (\tilde{B}^{\mathrm{T}})^{\mathrm{T}}$
(*c*) Find $\tilde{C}^{\mathrm{T}}, (\tilde{C}^{\mathrm{T}})^{\mathrm{T}}$
(*d*) Is $\tilde{A}+\tilde{C}$ defined?
(*e*) Is $\tilde{C}+\tilde{C}^{\mathrm{T}}$ defined?
(*f*) Is $\tilde{A}+\tilde{A}^{\mathrm{T}}$ symmetric?

https://doi.org/10.1017/CBO9780511755781.004 Published online by Cambridge University Press

(*g*) Is $\tilde{A} - \tilde{A}^{\mathrm{T}}$ antisymmetric?

3.7 Show that the matrix

$$\tilde{A} = \begin{pmatrix} 1 & 4 & 0 \\ 2 & 5 & 0 \\ 3 & 6 & 0 \end{pmatrix}$$

is not invertible.

3.8 Show that if $\tilde{A}$ and $\tilde{B}$ are invertible matrices of the same order, then $\tilde{A}\tilde{B}$ is invertible.

3.9 Given

$$\tilde{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{pmatrix},$$

find $\tilde{A}^{-1}$ and check the answer by direct multiplication.

3.10 Prove that if $\tilde{A}$ is a non-singular matrix, then $\det(\tilde{A}^{-1}) = 1/\det(\tilde{A})$.

3.11 If $\tilde{A}$ is an invertible $n \times n$ matrix, show that $\tilde{A}X = 0$ has only the trivial solution.

3.12 Show, by computing a matrix inverse, that the solution to the following system is $x_1 = 4$, $x_2 = 1$:

$$x_1 - x_2 = 3,$$
$$x_1 + x_2 = 5.$$

3.13 Solve the system $\tilde{A}X = \tilde{B}$ if

$$\tilde{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \tilde{B} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

3.14 Given matrix $\tilde{A}$, find A*, A^{T}, and $A^{\dagger}$, where

$$\tilde{A} = \begin{pmatrix} 2+3i & 1-i & 5i & -3 \\ 1+i & 6-i & 1+3i & -1-2i \\ 5-6i & 3 & 0 & -4 \end{pmatrix}.$$

3.15 Show that:

(*a*) The matrix $\tilde{A}\tilde{A}^{\dagger}$, where $\tilde{A}$ is any matrix, is hermitian.

(*b*) $(\tilde{A}\tilde{B})^{\dagger} = \tilde{B}^{\dagger}\tilde{A}^{\dagger}$.

(*c*) If $\tilde{A}, \tilde{B}$ are hermitian, then $\tilde{A}\tilde{B} + \tilde{B}\tilde{A}$ is hermitian.

(*d*) If $\tilde{A}$ and $\tilde{B}$ are hermitian, then $i(\tilde{A}\tilde{B} - \tilde{B}\tilde{A})$ is hermitian.

3.16 Obtain the most general orthogonal matrix of order 2.
[Hint: use relations (3.34a) and (3.34b).]

3.17. Obtain the most general unitary matrix of order 2.

3.18 If $\tilde{A}\tilde{B} = 0$, show that one of these matrices must have zero determinant.

3.19 Given the Pauli spin matrices (which are very important in quantum mechanics)

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

(note that the subscripts x, y, and z are sometimes used instead of 1, 2, and 3). Show that

(a) they are hermitian,

(b) $\sigma_i^2 = \tilde{I},\ i = 1, 2, 3$

(c) as a result of (a) and (b) they are also unitary, and

(d) $[\sigma_1, \sigma_2] = 2I\sigma_3$ *et cycl.*

Find the inverses of $\sigma_1,\ \sigma_2,\ \sigma_3$.

3.20 Use a rotation matrix to show that

$$\sin(\theta_1 + \theta_2) = \sin\theta_1 \cos\theta_2 + \sin\theta_2 \cos\theta_1.$$

3.21 Show that: Tr $\tilde{A}\tilde{B} = \mathrm{Tr}\,\tilde{B}\tilde{A}$ and Tr $\tilde{A}\tilde{B}\tilde{C} = Tr\ \tilde{B}\tilde{C}\tilde{A} = \mathrm{Tr}\ \tilde{C}\tilde{A}\tilde{B}$.

3.22 Show that: (a) the trace and (b) the commutation relation between two matrices are invariant under similarity transformations.

3.23 Determine the eigenvalues and eigenvectors of the matrix

$$\tilde{A} = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Given

$$\tilde{A} = \begin{pmatrix} 5 & 7 & -5 \\ 0 & 4 & -1 \\ 2 & 8 & -3 \end{pmatrix},$$

find a matrix $\tilde{S}$ that diagonalizes $\tilde{A}$, and show that $\tilde{S}^{-1}\tilde{A}\tilde{S}$ is diagonal.

3.25 If $\tilde{A}$ and $\tilde{B}$ are square matrices of the same order, then $\det(\tilde{A}\tilde{B}) = \det(\tilde{A})\det(\tilde{B})$. Verify this theorem if

$$\tilde{A} = \begin{pmatrix} 2 & -1 \\ 3 & 2 \end{pmatrix}, \quad \tilde{B} = \begin{pmatrix} 7 & 2 \\ -3 & 4 \end{pmatrix}.$$

3.26 Find a common set of eigenvectors for the two matrices

$$\tilde{A} = \begin{pmatrix} -1 & \sqrt{6} & \sqrt{2} \\ \sqrt{6} & 0 & \sqrt{3} \\ \sqrt{2} & \sqrt{3} & -2 \end{pmatrix}, \quad \tilde{B} = \begin{pmatrix} 10 & \sqrt{6} & -\sqrt{2} \\ \sqrt{6} & 9 & \sqrt{3} \\ -\sqrt{2} & \sqrt{3} & 11 \end{pmatrix}.$$

3.27 Show that two hermitian matrices can be made diagonal if and only if they commute.

3.28 Show the validity of the Cayley–Hamilton theorem by applying it to the matrix

$$\tilde{A} = \begin{pmatrix} 5 & 4 \\ 1 & 2 \end{pmatrix};$$

then use the Cayley–Hamilton theorem to find the inverse of the matrix $\tilde{A}$.

3.29 Given

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

find the direct product of these matrices, and show that it does not commute.