CHAPTER 3

Linear Algebra

1. INTRODUCTION

In this chapter, we are going to discuss a combination of algebra and geometry which is important in many applications. As you know, problems in various fields of science and mathematics involve the solution of sets of linear equations. This sounds like algebra, but it has a useful geometric interpretation. Suppose you have solved two simultaneous linear equations and have found $x = 2$ and $y = -3$. We can think of $x = 2$, $y = -3$ as the point $(2, -3)$ in the (x, y) plane. Since two linear equations represent two straight lines, the solution is then the point of intersection of the lines. The geometry helps us to understand that sometimes there is no solution (parallel lines) and sometimes there are infinitely many solutions (both equations represent the same line).

The language of vectors is very useful in studying sets of simultaneous equations. You are familiar with quantities such as the velocity of an object, the force acting on it, or the magnetic field at a point, which have both magnitude and direction. Such quantities are called *vectors*; contrast them with such quantities as mass, time, or temperature, which have magnitude only and are called *scalars*. A vector can be represented by an arrow and labeled by a boldface letter ($\mathbf{A}$ in Figure 1.1; also see Section 4). The length of the arrow tells us the magnitude of the vector and the direction of the arrow tells us the direction of the vector. It is not necessary to use coordinate axes as in Figure 1.1; we can, for example, point a finger to tell someone which way it is to town without knowing the direction of north. This is the geometric method of discussing vectors (see Section 4). However, if we do use a coordinate system as in Figure 1.1, we can specify the vector by giving its *components* A_x and A_y which are the projections of the vector on the x axis and the y axis. Thus we have two distinct methods of defining and working with vectors. A vector may be a geometric entity (arrow), or it may be a set of numbers (components relative to a coordinate system) which we use algebraically. As we shall see, this double interpretation of everything we do makes the use of vectors a very powerful tool in applications.

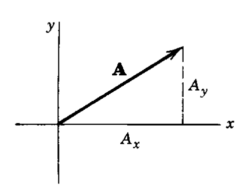


Figure 1.1

One of the great advantages of vector formulas is that they are independent of the choice of coordinate system. For example, suppose we are discussing the motion of a mass m sliding down an inclined plane. Newton's second law $\mathbf{F} = m\mathbf{a}$ is then a correct equation no matter how we choose our axes. We might, say, take the x axis horizontal and the y axis vertical, or alternatively we might take the x axis along the inclined plane and the y axis perpendicular to the plane. F_x would, of course, be different in the two cases, but for either case it would be true that $F_x = ma_x$ and $F_y = ma_y$, that is, the *vector* equation $\mathbf{F} = m\mathbf{a}$ would be true.

As we have just seen, a vector equation in two dimensions is equivalent to two component equations. In three dimensions, a vector equation is equivalent to three component equations. We will find it useful to generalize this to n dimensions and think of a set of n equations in n unknowns as the component equations for a vector equation in an n dimensional space (Section 10).

We shall also be interested in sets of linear equations which you can think of as changes of variable, say

$$\left\{\begin{aligned} x' &= ax + by, \\ y' &= cx + dy, \end{aligned}\right. \tag{1.1}$$

where a, b, c, d, are constants. Alternatively, we can think of (1.1) geometrically as telling us to move each point (x, y) to another point (x', y'), an operation we will refer to as a transformation of the plane. Or if we think of (x, y) and (x', y') as being components of vectors from the origin to the given points, then (1.1) tells us how to change each vector in the plane to another vector. Equations (1.1) could also correspond to a change of axes (say a rotation of axes around the origin) where (x, y) and (x', y') are the coordinates of the same point relative to different axes. We will learn (Sections 11 and 12) how to choose the best coordinate system or set of variables to use in solving various problems. The same methods and tools (such as matrices and determinants) which can be used to solve sets of numerical equations are what we need to work with transformations and changes of coordinate system. After we have considered 2- and 3-dimensional space, we will extend these ideas to n-dimensional space and finally to a space in which the "vectors" are functions. This generalization is of great importance in applications.

2. MATRICES; ROW REDUCTION

A matrix (plural: matrices) is just a rectangular array of quantities, usually inclosed in large parentheses, such as

$$\mathrm{A} = \begin{pmatrix} 1 & 5 & -2 \\ -3 & 0 & 6 \end{pmatrix}. \tag{2.1}$$

We will ordinarily indicate a matrix by a roman letter such as A (or B, C, M, r, etc.), but the letter does not have a numerical value; it simply stands for the array. To indicate a number in the array, we will write A_{ij} where i is the row number and j is the column number. For example, in (2.1), $A_{11} = 1$, $A_{12} = 5$, $A_{13} = -2$, $A_{21} = -3$, $A_{22} = 0$, $A_{23} = 6$. We will call a matrix with m rows and n columns an m by n matrix. Thus the matrix in (2.1) is a 2 by 3 matrix, and the matrix in (2.2) below is a 3 by 2 matrix.

Transpose of a Matrix We write

$$A^T = \begin{pmatrix} 1 & -3 \\ 5 & 0 \\ -2 & 6 \end{pmatrix}, \tag{2.2}$$

and call A^T the transpose of the matrix A in (2.1). To transpose a matrix, we simply write the rows as columns, that is, we interchange rows and columns. Note that, using index notation, we have $(A^T)_{ij} = A_{ji}$. You will find a summary of matrix notation in Section 9.

Sets of Linear Equations Historically linear algebra grew out of efforts to find efficient methods for solving sets of linear equations. As we have said, the subject has developed far beyond the solution of sets of numerical equations (which are easily solved by computer), but the ideas and methods developed for that purpose are needed in later work. A simple way to learn these techniques is to use them to solve some numerical problems by hand. In this section and the next we will develop methods of working with sets of linear equations, and introduce definitions and notation which will be useful later. Also, as you will see, we will discover how to tell whether a given set of equations has a solution or not.

▶ **Example 1.** Consider the set of equations

$$\begin{cases} 2x \phantom{{}+5y} - z = 2, \\ 6x + 5y + 3z = 7, \\ 2x - y \phantom{{}+3z} = 4. \end{cases} \tag{2.3}$$

Let's agree always to write sets of equations in this *standard form* with the x terms lined up in a column (and similarly for the other variables), and with the constants on the right hand sides of the equations. Then there are several matrices of interest connected with these equations. First is the *matrix of the coefficients* which we will call M:

$$M = \begin{pmatrix} 2 & 0 & -1 \\ 6 & 5 & 3 \\ 2 & -1 & 0 \end{pmatrix}. \tag{2.4}$$

Then there are two 3 by 1 matrices which we will call r and k:

$$r = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \qquad k = \begin{pmatrix} 2 \\ 7 \\ 4 \end{pmatrix}. \tag{2.5}$$

If we use index notation and replace x, y, z, by x_1, x_2, x_3, and call the constants k_1, k_2, k_3, then we could write the equations (2.3) in the form (Problem 1)

$$\sum_{j=1}^{3} M_{ij} x_j = k_i, \quad i = 1, 2, 3. \tag{2.6}$$

It is interesting to note that, as we will see in Section 6, this is exactly how matrices are multiplied, so we will learn to write sets of equations like (2.3) as Mr = k.

For right now we are interested in the fact that we can display all the essential numbers in equations (2.3) as a matrix known as the *augmented matrix* which we call A. Note that the first three columns of A are just the columns of M, and the fourth column is the column of constants on the right hand sides of the equations.

$$\mathrm{A} = \begin{pmatrix} 2 & 0 & -1 & 2 \\ 6 & 5 & 3 & 7 \\ 2 & -1 & 0 & 4 \end{pmatrix}. \tag{2.7}$$

Instead of working with a set of equations and writing all the variables, we can just work with the matrix (2.7). The process which we are going to show is called *row reduction* and is essentially the way your computer solves a set of linear equations. Row reduction is just a systematic way of taking linear combinations of the given equations to produce a simpler but equivalent set of equations. We will show the process, writing side-by-side the equations and the matrix corresponding to them.

(a) The first step is to use the first equation in (2.3) to eliminate the x terms in the other two equations. The corresponding matrix operation on (2.7) is to subtract 3 times the first row from the second row and subtract the first row from the third row. This gives:

$$\begin{cases} 2x \phantom{{}-y} - z = 2, \\ \phantom{2x-{}} 5y + 6z = 1, \\ - y + z = 2. \end{cases} \qquad \begin{pmatrix} 2 & 0 & -1 & 2 \\ 0 & 5 & 6 & 1 \\ 0 & -1 & 1 & 2 \end{pmatrix}$$

(b) Now it is convenient to interchange the second and third equations to get:

$$\begin{cases} 2x \phantom{{}-y} - z = 2, \\ - y + z = 2, \\ \phantom{2x-{}} 5y + 6z = 1. \end{cases} \qquad \begin{pmatrix} 2 & 0 & -1 & 2 \\ 0 & -1 & 1 & 2 \\ 0 & 5 & 6 & 1 \end{pmatrix}$$

(c) Next we use the second equation to eliminate the y terms from the other equations:

$$\begin{cases} 2x \phantom{{}-y} - z = 2, \\ - y + z = 2, \\ \phantom{2x-y+{}} 11z = 11. \end{cases} \qquad \begin{pmatrix} 2 & 0 & -1 & 2 \\ 0 & -1 & 1 & 2 \\ 0 & 0 & 11 & 11 \end{pmatrix}$$

(d) Finally, we divide the third equation by 11 and then use it to eliminate the z terms from the other equations:

$$\begin{cases} 2x = 3, \\ -y = 1, \\ z = 1. \end{cases} \qquad \begin{pmatrix} 2 & 0 & 0 & 3 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

It is customary to divide each equation by the leading coefficient so that the equations read $x = 3/2$, $y = -1$, $z = 1$. The row reduced matrix is then:

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

The important thing to understand here is that in finding a row reduced matrix we have just taken linear combinations of the original equations. This process is reversible, so the final simple equations are equivalent to the original ones. Let's summarize the allowed operations in row reducing a matrix (called *elementary row operations*).

(2.8)
i. Interchange two rows [see step (b)];
ii. Multiply (or divide) a row by a (nonzero) constant [see step (d)];
iii. Add a multiple of one row to another; this includes subtracting, that is, using a negative multiple [see steps (a) and (c)].

▶ **Example 2.** Write and row reduce the augmented matrix for the equations:

$$\left\{\begin{aligned} x - \ \ y + 4z &= 5,\\ 2x - 3y + 8z &= 4,\\ x - 2y + 4z &= 9.\end{aligned}\right. \tag{2.9}$$

This time we won't write the equations, just the augmented matrix. Remember the routine: Use the first row to clear the rest of the first column; use the new second row to clear the rest of the second column; etc. Also, since matrices are equal only if they are identical, we will not use equal signs between them. Let's use arrows.

$$\begin{pmatrix}1 & -1 & 4 & 5\\ 2 & -3 & 8 & 4\\ 1 & -2 & 4 & 9\end{pmatrix} \to \begin{pmatrix}1 & -1 & 4 & 5\\ 0 & -1 & 0 & -6\\ 0 & -1 & 0 & 4\end{pmatrix} \to \begin{pmatrix}1 & 0 & 4 & 11\\ 0 & -1 & 0 & -6\\ 0 & 0 & 0 & -20\end{pmatrix}$$

We don't need to go any farther! The last row says $0 \cdot z = -20$ which isn't true for any finite value of z. Now you see why your computer doesn't give an answer—there isn't any. We say that the equations are *inconsistent*. If this happens for a set of equations you have written for a physics problem, you know to look for a mistake.

Rank of a Matrix There is another way to discuss Example 2 using the following definition: The number of nonzero rows remaining when a matrix has been row reduced is called the *rank* of the matrix. (It is a theorem that the rank of A^T is the same as the rank of A.) Now look at the reduced augmented matrix for Example 2; it has 3 nonzero rows so its rank is 3. But the matrix M (matrix of the coefficients = first three columns of A) has only 2 nonzero rows so its rank is 2. Note that (rank of M) < (rank of A) and the equations are inconsistent.

▶ **Example 3.** Consider the equations

$$\left\{\begin{aligned} x + 2y - z &= \ \ 4,\\ 2x \qquad\ - z &= \ \ 1,\\ x - 2y \qquad &= -3.\end{aligned}\right. \tag{2.10}$$

Either by hand or by computer we row reduce the augmented matrix to get:

$$\begin{pmatrix}1 & 2 & -1 & 4\\ 2 & 0 & -1 & 1\\ 1 & -2 & 0 & -3\end{pmatrix} \to \begin{pmatrix}1 & 0 & -1/2 & 1/2\\ 0 & 1 & -1/4 & 7/4\\ 0 & 0 & 0 & 0\end{pmatrix}.$$

The last row of zeros tells us that there are infinitely many solutions. For any z we find from the first two rows that $x = (z+1)/2$ and $y = (z+7)/4$. Here we see that the rank of M and the rank of A are both 2 but the number of unknowns is 3, and we are able to find two unknowns in terms of the third.

To make this all very clear, let's look at some simple examples where the results are obvious. We write three sets of equations together with the row reduced matrices:

$$\left\{\begin{aligned} x + y &= 2,\\ x + y &= 5.\end{aligned}\right. \qquad \begin{pmatrix}1 & 1 & 2\\ 0 & 0 & 3\end{pmatrix} \tag{2.11}$$

$$\left\{\begin{aligned} x + y &= 2,\\ 2x + 2y &= 4.\end{aligned}\right. \qquad \begin{pmatrix}1 & 1 & 2\\ 0 & 0 & 0\end{pmatrix} \tag{2.12}$$

$$\left\{\begin{aligned} x + y &= 2,\\ x - y &= 4.\end{aligned}\right. \qquad \begin{pmatrix}1 & 0 & 3\\ 0 & 1 & -1\end{pmatrix} \tag{2.13}$$

In (2.11), since $x + y$ can't be equal to both 2 and 5, it is clear that there is no solution; the equations are inconsistent. Note that the last row of the reduced matrix is all zeros except for the last entry and so (rank M) < (rank A). In (2.12), the second equation is just twice the first so they are really the same equation; we say that the equations are dependent. There is an infinite set of solutions, namely all points on the line $y = 2 - x$. Note that the last line of the matrix is all zeros; this indicates linear dependence. We have (rank A) = (rank M) = 1, and we can solve for one unknown in terms of the other. Finally in (2.13) we have a set of equations with one solution, $x = 3$, $y = -1$, and we see that the row reduced matrix gives this result. Note that (rank A) = (rank M) = number of unknowns = 2.

Now let's consider the general problem of solving m equations in n unknowns. Then M has m rows (corresponding to m equations) and n columns (corresponding to n unknowns) and A has one more column (the constants). The following summary outlines the possible cases.

(2.14)

a. If (rank M) < (rank A), the equations are inconsistent and there is no solution.

b. If (rank M) = (rank A) = n (number of unknowns), there is one solution.

c. If (rank M) = (rank A) = $R < n$, then R unknowns can be found in terms of the remaining $n - R$ unknowns.

Example 4. Here is a set of equations and the row reduced matrix:

$$\left\{\begin{aligned} x + y - z &= 7,\\ 2x - y - 5z &= 2,\\ -5x + 4y + 14z &= 1,\\ 3x - y - 7z &= 5.\end{aligned}\right. \qquad \begin{pmatrix}1 & 0 & -2 & 3\\ 0 & 1 & 1 & 4\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix} \tag{2.15}$$

From the reduced matrix, the solution is $x = 3 + 2z$, $y = 4 - z$. We see that this is an example of (2.14c) with $m = 4$ (number of equations), $n = 3$ (number of unknowns), (rank M) = (rank A) = $R = 2 < n = 3$. Then by (2.14c), we solve for $R = 2$ unknowns (x and y) in terms of the $n - R = 1$ unknown (z).

PROBLEMS, SECTION 2

1. The first equation in (2.6) written out in detail is

$$M_{11}x_1 + M_{12}x_2 + M_{13}x_3 = k_1.$$

Write out the other two equations in the same way and then substitute $x_1, x_2, x_3 = x, y, z$ and the values of M_{ij} and k_i from (2.4) and (2.5) to verify that (2.6) is really (2.3).

2. As in Problem 1, write out in detail in terms of M_{ij}, x_j, and k_i, equations like (2.6) for two equations in four unknowns; for four equations in two unknowns.

For each of the following problems write and row reduce the augmented matrix to find out whether the given set of equations has exactly one solution, no solutions, or an infinite set of solutions. Check your results by computer. *Warning hint:* Be sure your equations are written in standard form. *Comment*: Remember that the point of doing these problems is not just to get an answer (which your computer will give you), but to become familiar with the terminology, ideas, and notation we are using.

3. $\begin{cases} x - 2y + 13 = 0 \\ y - 4x = 17 \end{cases}$

4. $\begin{cases} 2x + y - z = 2 \\ 4x + y - 2z = 3 \end{cases}$

5. $\begin{cases} 2x + y - z = 2 \\ 4x + 2y - 2z = 3 \end{cases}$

6. $\begin{cases} x + y - z = 1 \\ 3x + 2y - 2z = 3 \end{cases}$

7. $\begin{cases} 2x + 3y = 1 \\ x + 2y = 2 \\ x + 3y = 5 \end{cases}$

8. $\begin{cases} -x + y - z = 4 \\ x - y + 2z = 3 \\ 2x - 2y + 4z = 6 \end{cases}$

9. $\begin{cases} x - y + 2z = 5 \\ 2x + 3y - z = 4 \\ 2x - 2y + 4z = 6 \end{cases}$

10. $\begin{cases} x + 2y - z = 1 \\ 2x + 3y - 2z = -1 \\ 3x + 4y - 3z = -4 \end{cases}$

11. $\begin{cases} x - 2y = 4 \\ 5x + z = 7 \\ x + 2y - z = 3 \end{cases}$

12. $\begin{cases} 2x + 5y + z = 2 \\ x + y + 2z = 1 \\ x + 5z = 3 \end{cases}$

13. $\begin{cases} 4x + 6y - 12z = 7 \\ 5x - 2y + 4z = -15 \\ 3x + 4y - 8z = 4 \end{cases}$

14. $\begin{cases} 2x + 3y - z = -2 \\ x + 2y - z = 4 \\ 4x + 7y - 3z = 11 \end{cases}$

Find the rank of each of the following matrices.

15. $\begin{pmatrix} 1 & 1 & 2 \\ 2 & 4 & 6 \\ 3 & 2 & 5 \end{pmatrix}$

16. $\begin{pmatrix} 2 & -3 & 5 & 3 \\ 4 & -1 & 1 & 1 \\ 3 & -2 & 3 & 4 \end{pmatrix}$

17. $\begin{pmatrix} 1 & 1 & 4 & 3 \\ 3 & 1 & 10 & 7 \\ 4 & 2 & 14 & 10 \\ 2 & 0 & 6 & 4 \end{pmatrix}$

18. $\begin{pmatrix} 1 & 0 & 1 & 0 \\ -1 & -2 & -1 & 0 \\ 2 & 2 & 5 & 3 \\ 2 & 4 & 8 & 6 \end{pmatrix}$

3. DETERMINANTS; CRAMER'S RULE

We have said that a matrix is simply a display of a set of numbers; it does *not* have a numerical value. For a square matrix, however, there is a useful number called the *determinant* of the matrix. Although a computer will quickly give the value of a determinant, we need to know what this value means in order to use it in applications. [See, for example, equations (4.19), (6.24) and (8.5).] We also need to know how to work with determinants. An easy way to learn these things is to solve some numerical problems by hand. We shall outline some of the facts about determinants without proofs (for more details, see linear algebra texts).

Evaluating Determinants To indicate that we mean the determinant of a square matrix A (written det A), we replace the large parentheses inclosing A by single bars. The value of det A if A is a 1 by 1 matrix is just the value of the single element. For a 2 by 2 matrix,

$$\mathrm{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad \det \mathrm{A} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc. \tag{3.1}$$

Equation (3.1) gives the value of a second order determinant. We shall describe how to evaluate determinants of higher order.

First we need some notation and definitions. It is convenient to write an n^{th} order determinant like this:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & & & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix}. \tag{3.2}$$

Notice that a_{23} is the element in the second row and the third column; that is, the first subscript is the number of the row and the second subscript is the number of the column in which the element is. Thus the element a_{ij} is in row i and column j. As an abbreviation for the determinant in (3.2), we sometimes write simply $|a_{ij}|$, that is, the determinant whose elements are a_{ij}. In this form it looks exactly like the absolute value of the element a_{ij} and you have to tell from the context which of these meanings is intended.

If we remove one row and one column from a determinant of order n, we have a determinant of order $n-1$. Let us remove the row and column containing the element a_{ij} and call the remaining determinant M_{ij}. The determinant M_{ij} is called the *minor* of a_{ij}. For example, in the determinant

$$\begin{vmatrix} 1 & -5 & 2 \\ 7 & 3 & 4 \\ 2 & 1 & 5 \end{vmatrix}, \tag{3.3}$$

the minor of the element $a_{23} = 4$ is

$$M_{23} = \begin{vmatrix} 1 & -5 \\ 2 & 1 \end{vmatrix},$$

obtained by crossing off the row and column containing 4. The signed minor $(-1)^{i+j}M_{ij}$ is called the *cofactor* of a_{ij}. In (3.3), the element 4 is in the second row ($i = 2$) and third column ($j = 3$), so $i + j = 5$, and the cofactor of 4 is $(-1)^5 M_{23} = -11$. It is very convenient to get the proper sign (plus or minus) for the factor $(-1)^{i+j}$ by thinking of a checkerboard of plus and minus signs like this:

$$\begin{vmatrix} + & - & + & - & & & \\ - & + & - & + & & & \\ + & - & + & - & \text{etc.} & & \\ - & + & - & + & & & \\ & \text{etc.} & & & \ddots & & \\ & & & & & + & - \\ & & & & & - & + \end{vmatrix}. \tag{3.4}$$

Then the sign $(-1)^{i+j}$ to be attached to M_{ij} is just the checkerboard sign in the same position as a_{ij}. For the element a_{23}, you can see that the checkerboard sign is minus.

Now we can easily say how to find the *value of a determinant: Multiply each element of one row (or one column) by its cofactor and add the results.* It can be shown that we get the same answer whichever row or column we use.

▶ **Example 1.** Let us evaluate the determinant in (3.3) using elements of the third column. We get

$$\begin{aligned}\begin{vmatrix} 1 & -5 & 2 \\ 7 & 3 & 4 \\ 2 & 1 & 5 \end{vmatrix} &= 2\begin{vmatrix} 7 & 3 \\ 2 & 1 \end{vmatrix} - 4\begin{vmatrix} 1 & -5 \\ 2 & 1 \end{vmatrix} + 5\begin{vmatrix} 1 & -5 \\ 7 & 3 \end{vmatrix} \\ &= 2 \cdot 1 - 4 \cdot 11 + 5 \cdot 38 = 148.\end{aligned}$$

As a check, using elements of the first row, we get

$$1\begin{vmatrix} 3 & 4 \\ 1 & 5 \end{vmatrix} + 5\begin{vmatrix} 7 & 4 \\ 2 & 5 \end{vmatrix} + 2\begin{vmatrix} 7 & 3 \\ 2 & 1 \end{vmatrix} = 11 + 135 + 2 = 148.$$

The method of evaluating a determinant which we have described here is one form of Laplace's development of a determinant. If the determinant is of fourth order (or higher), using the Laplace development once gives us a set of determinants of order one less than we started with; then we use the Laplace development all over again to evaluate each of these, and so on until we get determinants of second order which we know how to evaluate. This is obviously a lot of work! We will see below how to simplify the calculation. A word of warning to anyone who has learned a special method of evaluating a third-order determinant by recopying columns to the right and multiplying along diagonals: this method *does not work* for fourth order (and higher).

Useful Facts About Determinants We state these facts without proof. (See algebra books for proofs.)

1. If each element of *one* row (or *one* column) of a determinant is multiplied by a number k, the value of the determinant is multiplied by k.
2. The value of a determinant is zero if
   (a) all elements of one row (or column) are zero; or if
   (b) two rows (or two columns) are identical; or if
   (c) two rows (or two columns) are proportional.
3. If two rows (or two columns) of a determinant are interchanged, the value of the determinant changes sign.
4. The value of a determinant is unchanged if
   (a) rows are written as columns and columns as rows; or if
   (b) we add to each element of one row, k times the corresponding element of another row, where k is any number (and a similar statement for columns).

Let us look at a few examples of the use of these facts.

▶ **Example 2.** Find the equation of a plane through the three given points $(0,0,0)$, $(1,2,5)$, and $(2,-1,0)$.

We shall verify that the answer in determinant form is

$$\begin{vmatrix} x & y & z & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 2 & 5 & 1 \\ 2 & -1 & 0 & 1 \end{vmatrix} = 0.$$

By a Laplace development using elements of the first row, we would find that this is a linear equation in x, y, z; thus it represents a plane. We need now to show that the three points are in the plane. Suppose $(x,y,z) = (0,0,0)$; then the first two rows of the determinant are identical and by Fact 2b the determinant is zero. Similarly if the point (x,y,z) is either of the other given points, two rows of the determinant are identical and the determinant is zero. Thus all three points lie in the plane.

▶ **Example 3.** Evaluate the determinant

$$D = \begin{vmatrix} 0 & a & -b \\ -a & 0 & c \\ b & -c & 0 \end{vmatrix}.$$

If we interchange rows and columns in D, then by Facts 4a and 1 we have

$$D = \begin{vmatrix} 0 & -a & b \\ a & 0 & -c \\ -b & c & 0 \end{vmatrix} = (-1)^3 \begin{vmatrix} 0 & a & -b \\ -a & 0 & c \\ b & -c & 0 \end{vmatrix},$$

where in the last step we have factored -1 out of each column by Fact 1. Thus we have $D = -D$, so $D = 0$.

We can use Facts 1 to 4 to simplify finding the value of a determinant. First we check Facts 2a, 2b, 2c, in case the determinant is trivially equal to zero. Then we try to get as many zeros as possible in some row or column in order to have fewer terms in the Laplace development. We look for rows (or columns) which can be combined (using Fact 4b) to give zeros. Although this is something like row reduction, we can operate with columns as well as rows. However, we can't just cancel a number from a row (or column); by Fact 1 we must keep it as a factor in our answer. And we must keep track of any row (or column) interchanges since by Fact 3 each interchange multiplies the determinant by (-1).

▶ **Example 4.** Evaluate the determinant

$$D = \begin{vmatrix} 4 & 3 & 0 & 1 \\ 9 & 7 & 2 & 3 \\ 4 & 0 & 2 & 1 \\ 3 & -1 & 4 & 0 \end{vmatrix}.$$

Subtract 4 times the fourth column from the first column, and subtract 2 times the fourth column from the third column to get:

$$D = \begin{vmatrix} 0 & 3 & -2 & 1 \\ -3 & 7 & -4 & 3 \\ 0 & 0 & 0 & 1 \\ 3 & -1 & 4 & 0 \end{vmatrix}.$$

Do a Laplace development using the third row:

$$D = (-1) \begin{vmatrix} 0 & 3 & -2 \\ -3 & 7 & -4 \\ 3 & -1 & 4 \end{vmatrix}. \tag{3.5}$$

Add the second row to the third row:

$$D = (-1) \begin{vmatrix} 0 & 3 & -2 \\ -3 & 7 & -4 \\ 0 & 6 & 0 \end{vmatrix}.$$

Do a Laplace development using the first column:

$$D = (-1)(-1)(-3) \begin{vmatrix} 3 & -2 \\ 6 & 0 \end{vmatrix} = (-3)[0 - 6(-2)] = -36.$$

This is the answer but you might like to look for some shorter solutions. For example, consider the determinant (3.5) above. If we immediately do another Laplace development using the first row, the minor of 3 in the first row, second column is

$$\begin{vmatrix} -3 & -4 \\ 3 & 4 \end{vmatrix}.$$

Without even evaluating it, we should recognize by Fact 2c that it is zero. Then proceeding with the Laplace development of (3.5) using the first row gives just

$$D = (-1)(-2)\begin{vmatrix} -3 & 7 \\ 3 & -1 \end{vmatrix} = 2(3-21) = -36 \quad \text{as above.}$$

Now you may be wondering why you should learn about this when your computer will do it for you. Suppose you have a determinant with elements which are algebraic expressions, and you want to write it in a different form. Then you need to know what manipulations you can do without changing its value. Also, if you know the rules, you may see that a determinant is zero without evaluating it. An easy way to learn these things is to evaluate some simple numerical determinants by hand.

Cramer's Rule This is a formula in terms of determinants for the solution of n linear equations in n unknowns when there is exactly one solution. As we said for row reduction and for evaluating determinants, your computer will quickly give you the solution of a set of linear equations when there is one. However, for theoretical purposes, we need the Cramer's rule formula, and a simple way to learn about it is to use it to solve sets of linear equations with numerical coefficients.

Let us first show the use of Cramer's rule to solve two equations in two unknowns. Then we will generalize it to n equations in n unknowns. Consider the set of equations

$$\begin{cases} a_1x + b_1y = c_1, \\ a_2x + b_2y = c_2. \end{cases} \tag{3.6}$$

If we multiply the first equation by b_2, the second by b_1, and then subtract the results and solve for x, we get (if $a_1b_2 - a_2b_1 \neq 0$)

$$x = \frac{c_1b_2 - c_2b_1}{a_1b_2 - a_2b_1}. \tag{3.7a}$$

Solving for y in a similar way, we get

$$y = \frac{a_1c_2 - a_2c_1}{a_1b_2 - a_2b_1}. \tag{3.7b}$$

Using the definition (3.1) of a second order determinant, we can write the solutions (3.7) of (3.6) in the form

$$x = \frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}, \qquad y = \frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}. \tag{3.8}$$

It is helpful in remembering (3.8) to say in words how we find the correct determinants. First, the equations must be written in standard form as for row reduction (Section 2). Then if we simply write the array of coefficients on the left-hand side of (3.6), these form the denominator determinant in (3.8). This determinant (which we shall denote by D) is called the *determinant of the coefficients*. To find the numerator determinant for x, start with D, erase the x coefficients a_1 and a_2, and replace them by the constants c_1 and c_2 from the right-hand sides of the equations. Similarly, we replace the y coefficients in D by the constant terms to find the numerator determinant in y.

▶ **Example 5.** Use (3.8) to solve the set of equations

$$\begin{cases} 2x + 3y = 3, \\ x - 2y = 5. \end{cases}$$

We find

$$D = \begin{vmatrix} 2 & 3 \\ 1 & -2 \end{vmatrix} = -4 - 3 = -7,$$

$$x = \frac{1}{D}\begin{vmatrix} 3 & 3 \\ 5 & -2 \end{vmatrix} = \frac{-6-15}{-7} = 3,$$

$$y = \frac{1}{D}\begin{vmatrix} 2 & 3 \\ 1 & 5 \end{vmatrix} = \frac{10-3}{-7} = -1.$$

> This method of solution of a set of linear equations is called Cramer's rule. It may be used to solve n equations in n unknowns if $D \neq 0$; the solution then consists of one value for each unknown. The denominator determinant D is the n by n determinant of the coefficients when the equations are arranged in standard form. The numerator determinant for each unknown is the determinant obtained by replacing the column of coefficients of that unknown in D by the constant terms from the right-hand sides of the equations. Then to find the unknowns, we must evaluate each of the determinants and divide.

Rank of a Matrix Here is another way to find the rank of a matrix (Section 2). A submatrix means a matrix remaining if we remove some rows and/or remove some columns from the original matrix. To find the rank of a matrix, we look at all the square submatrices and find their determinants. The order of the largest nonzero determinant is the rank of the matrix.

▶ **Example 6.** Find the rank of the matrix

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

We need to look at the four 3 by 3 determinants containing columns 1,2,3 or 1,2,4 or 1,3,4 or 2,3,4. We note that the first two columns are negatives of each other, so by Fact 2c the first two of these determinants are both zero. The last two determinants differ only in the sign of their first column, so we just have to look at one of them, say:

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

If we now subtract twice the first row from the third row, we have

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

and we see by Fact 2c that the determinant is zero. So the rank of the matrix is less than 3. To show that it is 2, we just have to find *one* 2 by 2 submatrix with nonzero determinant. There are several of them; find one. Thus the rank of the matrix is 2. (If we had needed to show that the rank was 1, we would have had to show that *all* the 2 by 2 submatrices had determinants equal to zero.)

PROBLEMS, SECTION 3

Evaluate the determinants in Problems 1 to 6 by the methods shown in Example 4. Remember that the reason for doing this is not just to get the answer (your computer can give you that) but to learn how to manipulate determinants correctly. Check your answers by computer.

1. $\begin{vmatrix} -2 & 3 & 4 \\ 3 & 4 & -2 \\ 5 & 6 & -3 \end{vmatrix}$

2. $\begin{vmatrix} 5 & 17 & 3 \\ 2 & 4 & -3 \\ 11 & 0 & 2 \end{vmatrix}$

3. $\begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{vmatrix}$

4. $\begin{vmatrix} -2 & 4 & 7 & 3 \\ 8 & 2 & -9 & 5 \\ -4 & 6 & 8 & 4 \\ 2 & -9 & 3 & 8 \end{vmatrix}$

5. $\begin{vmatrix} 7 & 0 & 1 & -3 & 5 \\ 2 & -1 & 0 & 1 & 4 \\ 7 & -3 & 2 & -1 & 4 \\ 8 & 6 & -2 & -7 & 4 \\ 1 & 3 & -5 & 7 & 5 \end{vmatrix}$

6. $\begin{vmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}$

7. Prove the following by appropriate manipulations using Facts 1 to 4; do not just evaluate the determinants.

$$\begin{vmatrix} 1 & a & bc \\ 1 & b & ac \\ 1 & c & ab \end{vmatrix} = \begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix} = (c-a)(b-a)(c-b)\begin{vmatrix} 1 & a & a^2 \\ 0 & 1 & b+a \\ 0 & 0 & 1 \end{vmatrix}$$
$$= (c-a)(b-a)(c-b).$$

8. Show that if, in using the Laplace development, you accidentally multiply the elements of one row by the cofactors of another row, you get zero.
Hint: Consider Fact 2b.

9. Show without computation that the following determinant is equal to zero.
Hint: Consider the effect of interchanging rows and columns.

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

10. A determinant or a square matrix is called skew-symmetric if $a_{ij} = -a_{ji}$. (The determinant in Problem 9 is an example of a skew-symmetric determinant.) Show that a skew-symmetric determinant of odd order is zero.

In Problems 11 and 12 evaluate the determinants.

11. $\begin{vmatrix} 0 & 5 & -3 & -4 & 1 \\ -5 & 0 & 2 & 6 & -2 \\ 3 & -2 & 0 & -3 & 7 \\ 4 & -6 & 3 & 0 & -3 \\ -1 & 2 & -7 & 3 & 0 \end{vmatrix}$

12. $\begin{vmatrix} 0 & 1 & 2 & -1 \\ -1 & 0 & -3 & 0 \\ -2 & 3 & 0 & 1 \\ 1 & 0 & -1 & 0 \end{vmatrix}$

13. Show that

$$\begin{vmatrix} \cos\theta & 1 & 0 \\ 1 & 2\cos\theta & 1 \\ 0 & 1 & 2\cos\theta \end{vmatrix} = \cos 3\theta.$$

14. Show that the n-rowed determinant

$$\begin{vmatrix} \cos\theta & 1 & 0 & 0 & & & 0 \\ 1 & 2\cos\theta & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 2\cos\theta & 1 & & & 0 \\ 0 & 0 & 1 & 2\cos\theta & & & 0 \\ & & \vdots & & \ddots & & \vdots \\ & & \vdots & & & 2\cos\theta & 1 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 2\cos\theta \end{vmatrix} = \cos n\theta.$$

Hint: Expand using elements of the last row or column. Use mathematical induction and the trigonometric addition formulas.

15. Use Cramer's rule to solve Problems 2.3 and 2.11.

16. In the following set of equations (from a quantum mechanics problem), A and B are the unknowns, k and K are given, and $i = \sqrt{-1}$. Use Cramer's rule to find A and show that $|A|^2 = 1$.

$$\begin{cases} A - \quad B = -1 \\ ikA - KB = ik \end{cases}$$

17. Use Cramer's rule to solve for x and t the Lorentz equations of special relativity:

$$\begin{cases} x' = \gamma(x - vt) \\ t' = \gamma(t - vx/c^2) \end{cases} \qquad \text{where} \quad \gamma^2(1 - v^2/c^2) = 1$$

Caution: Arrange the equations in standard form.

18. Find z by Cramer's rule:

$$\begin{cases} (a-b)x - (a-b)y + 3b^2z = 3ab \\ (a+2b)x - (a+2b)y - (3ab+3b^2)z = 3b^2 \\ bx + ay - (2b^2+a^2)z = 0 \end{cases}$$

▶ 4. VECTORS

Notation We shall indicate a vector by a boldface letter (for example, **A**) and a component of a vector by a subscript (for example A_x is the x component of **A**), as in Figure 4.1. Since it is not easy to handwrite boldface letters, you should write a vector with an arrow over it (for example, $\vec{A}$). It is very important to indicate clearly whether a letter represents a vector, since, as we shall see below, the same letter in italics (not boldface) is often used with a different meaning.

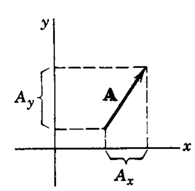


Figure 4.1

Magnitude of a Vector The length of the arrow representing a vector **A** is called the *length* or the *magnitude* of **A** (written $|\mathbf{A}|$ or A) or (see Section 10) the *norm* of **A** (written $||\mathbf{A}||$). Note the use of A to mean the magnitude of **A**; for this reason it is important to make it clear whether you mean a vector or its magnitude (which is a scalar). By the Pythagorean theorem, we find

$$
(4.1) \qquad \begin{aligned} A = |\mathbf{A}| &= \sqrt{A_x^2 + A_y^2} \qquad \text{in two dimensions, or} \\ A = |\mathbf{A}| &= \sqrt{A_x^2 + A_y^2 + A_z^2} \qquad \text{in three dimensions.} \end{aligned}
$$

▶ **Example 1.** In Figure 4.2 the force $\mathbf{F}$ has an x component of 4 lb and a y component of 3 lb. Then we write

$$
\begin{aligned} F_x &= 4 \text{ lb}, \\ F_y &= 3 \text{ lb}, \\ |\mathbf{F}| &= 5 \text{ lb}, \\ \theta &= \arctan \tfrac{3}{4}. \end{aligned}
$$

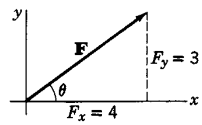


Figure 4.2

Addition of Vectors There are two ways to get the sum of two vectors. One is by the parallelogram law: To find $\mathbf{A} + \mathbf{B}$, place the tail of $\mathbf{B}$ at the head of $\mathbf{A}$ and draw the vector from the tail of $\mathbf{A}$ to the head of $\mathbf{B}$ as shown in Figures 4.3 and 4.4.

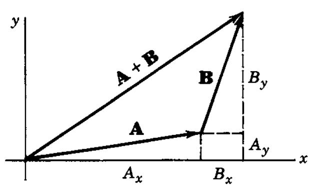


Figure 4.3

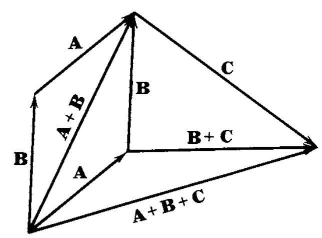


Figure 4.4

The second way of finding $\mathbf{A} + \mathbf{B}$ is to add components: $\mathbf{A} + \mathbf{B}$ has components $A_x + B_x$ and $A_y + B_y$. You should satisfy yourself from Figure 4.3 that these two methods of finding $\mathbf{A} + \mathbf{B}$ are equivalent. From Figure 4.4 and either definition of vector addition, it follows that

$$
\begin{aligned} \mathbf{A} + \mathbf{B} &= \mathbf{B} + \mathbf{A} \qquad \text{(commutative law for addition);} \\ (\mathbf{A} + \mathbf{B}) + \mathbf{C} &= \mathbf{A} + (\mathbf{B} + \mathbf{C}) \qquad \text{(associative law for addition).} \end{aligned}
$$

In other words, vectors may be added together by the usual laws of algebra.

It seems reasonable to use the symbol $3\mathbf{A}$ for the vector $\mathbf{A} + \mathbf{A} + \mathbf{A}$. By the methods of vector addition above, we can say that the vector $\mathbf{A} + \mathbf{A} + \mathbf{A}$ is a vector three times as long as $\mathbf{A}$ and in the same direction as $\mathbf{A}$ and that each component of $3\mathbf{A}$ is three times the corresponding component of $\mathbf{A}$. As a natural extension of these facts we define the vector $c\mathbf{A}$ (where c is any real positive number) to be a vector c times as long as $\mathbf{A}$ and in the same direction as $\mathbf{A}$; each component of $c\mathbf{A}$ is then c times the corresponding component of $\mathbf{A}$ (Figure 4.5).

The negative of a vector is defined as a vector of the same magnitude but in the opposite direction. Then (Figure 4.6) each component of $-\mathbf{B}$ is the negative of the corresponding component of $\mathbf{B}$. We can now define subtraction of vectors by

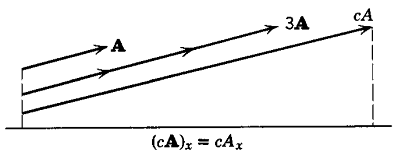


Figure 4.5

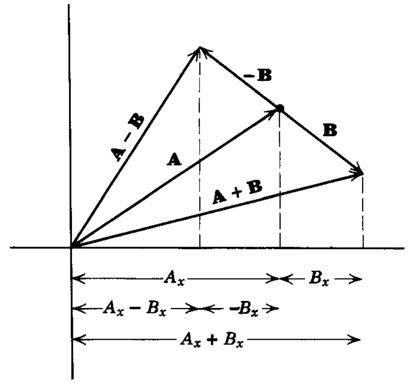


Figure 4.6

saying that $\mathbf{A}-\mathbf{B}$ means the sum of the vectors $\mathbf{A}$ and $-\mathbf{B}$. Each component of $\mathbf{A}-\mathbf{B}$ is then obtained by subtracting the corresponding components of $\mathbf{A}$ and $\mathbf{B}$, that is, $(\mathbf{A}-\mathbf{B})_x = A_x - B_x$, etc. Like addition, subtraction of vectors can be done geometrically (by the parallelogram law) or algebraically by subtracting the components (Figure 4.6).

The *zero vector* (which might arise as $\mathbf{A}=\mathbf{B}-\mathbf{B}=\mathbf{0}$, or as $\mathbf{A}=c\mathbf{B}$ with $c=0$) is a vector of zero magnitude; its components are all zero and it does not have a direction. A vector of length or magnitude 1 is called a *unit vector.* Then for any $\mathbf{A}\neq 0$, the vector $\mathbf{A}/|\mathbf{A}|$ is a unit vector. In Example 1, $\mathbf{F}/5$ is a unit vector.

We have just seen that there are two ways to combine vectors: geometric (head to tail addition), and algebraic (using components). Let us look first at an example of the geometric method; then we shall consider the algebraic method. Example 2 below illustrates the geometric method. By similar proofs, many of the facts of elementary geometry can be easily proved using vectors, with no reference to components or a coordinate system. (See Problems 3 to 8.)

▶ **Example 2.** Prove that the medians of a triangle intersect at a point two-thirds of the way from any vertex to the midpoint of the opposite side.

To prove this, we call two of the sides of the triangle $\mathbf{A}$ and $\mathbf{B}$. The third side of the triangle is then $\mathbf{A}+\mathbf{B}$ by the parallelogram law, with the directions of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{A}+\mathbf{B}$ as indicated in Figure 4.7. If we add the vector $\frac{1}{2}\mathbf{B}$ to the vector $\mathbf{A}$ (head to tail as in Figure 4.7b), we have a vector from point O to the midpoint of the opposite side of the triangle, that is, we have the median to side $\mathbf{B}$. Next, take two-thirds of this vector; we now have the vector $\frac{2}{3}(\mathbf{A}+\frac{1}{2}\mathbf{B}) = \frac{2}{3}\mathbf{A}+\frac{1}{3}\mathbf{B}$ extending from O to P in Figure 4.7b. We want to show that P is the intersection point of the three medians and also the "$\frac{2}{3}$ point" for each. We prove this by showing that P is the "$\frac{2}{3}$ point" on the median to side $\mathbf{A}$; then since $\mathbf{A}$ and $\mathbf{B}$ represent *any* two sides of the triangle, the proof holds for all three medians. The vector from R to Q (Figure 4.7c) is $\frac{1}{2}\mathbf{A}+\mathbf{B}$; this is the median to $\mathbf{A}$. The "$\frac{2}{3}$ point" on this median is the point P' (Figure 4.7d); the vector from R to P' is equal to $\frac{1}{3}(\frac{1}{2}\mathbf{A}+\mathbf{B})$. Then the vector from O to P' is $\frac{1}{2}\mathbf{A}+\frac{1}{3}(\frac{1}{2}\mathbf{A}+\mathbf{B}) = \frac{2}{3}\mathbf{A}+\frac{1}{3}\mathbf{B}$. Thus P and P' are the same point and all three medians have their "$\frac{2}{3}$ points" there. Note that we have made no reference to a coordinate system or to components in this proof.

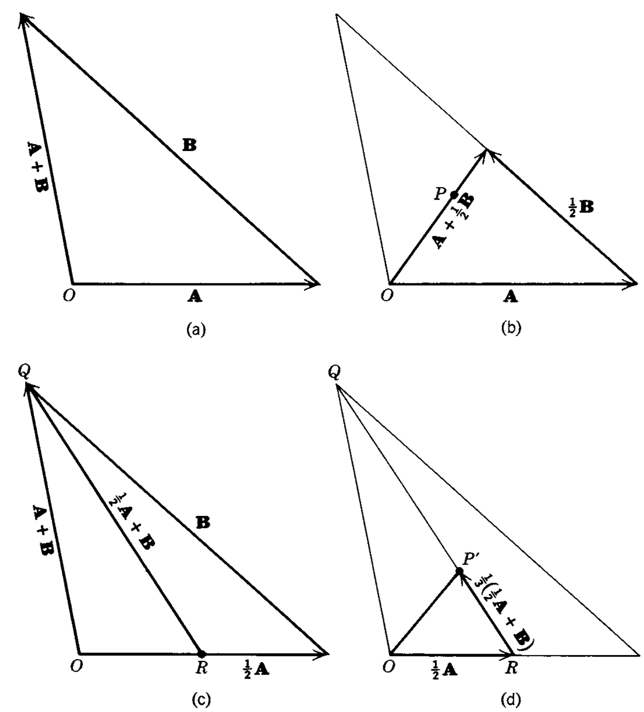


Figure 4.7

▶ PROBLEMS, SECTION 4

1. Draw diagrams and prove (4.1).

2. Given the vectors making the given angles θ with the positive x axis:

$\mathbf{A}$ of magnitude 5, $\theta = 45°$,

$\mathbf{B}$ of magnitude 3, $\theta = -30°$,

$\mathbf{C}$ of magnitude 7, $\theta = 120°$,

(a) Draw diagrams representing $2\mathbf{A}$, $\mathbf{A} - 2\mathbf{B}$, $\mathbf{C} - \mathbf{B}$, $\frac{2}{5}\mathbf{A} - \frac{1}{7}\mathbf{C}$.

(b) Draw diagrams to show that

$$\begin{array}{ll} \mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} & \mathbf{A} - (\mathbf{B} - \mathbf{C}) = (\mathbf{A} - \mathbf{B}) + \mathbf{C}, \\ (\mathbf{A} + \mathbf{B}) + \mathbf{C} = (\mathbf{A} + \mathbf{C}) + \mathbf{B}, & (\mathbf{A} + \mathbf{B})_x = \mathbf{A}_x + \mathbf{B}_x, \\ (\mathbf{B} - \mathbf{C})_x = \mathbf{B}_x - \mathbf{C}_x. & \end{array}$$

Use vectors to prove the following theorems from geometry:

3. The diagonals of a parallelogram bisect each other.

4. The line segment joining the midpoints of two sides of any triangle is parallel to the third side and half its length.

5. In a parallelogram, the two lines from one corner to the midpoints of the two opposite sides trisect the diagonal they cross.

6. In any quadrilateral (four-sided figure with sides of various lengths and—in general—four different angles), the lines joining the midpoints of opposite sides bisect each other. *Hint:* Label three sides **A**, **B**, **C**; what is the vector along the fourth side?

7. A line through the midpoint of one side of a triangle and parallel to a second side bisects the third side. *Hint:* Call parallel vectors **A** and $c\mathbf{A}$.

8. The median of a trapezoid (four-sided figure with just two parallel sides) means the line joining the midpoints of the two nonparallel sides. Prove that the median bisects both diagonals; that the median is parallel to the two parallel bases and equal to half the sum of their lengths.

We have discussed in some detail the geometric method of adding vectors (parallelogram law or head to tail addition) and its importance in stating and proving geometric and physical facts without the intrusion of a special coordinate system. There are, however, many cases in which algebraic methods (using components relative to a particular coordinate system) are better. We shall discuss this next.

Vectors in Terms of Components We consider a set of rectangular axes as in Figure 4.8. Let the vector **i** be a unit vector in the positive x direction (out of the paper toward you), and let **j** and **k** be unit vectors in the positive y and z directions. If A_x and A_y are the scalar components of a vector in the (x, y) plane, then $\mathbf{i}A_x$ and $\mathbf{j}A_y$ are its vector components, and their sum is the vector **A** (Figure 4.9).

$$\mathbf{A} = \mathbf{i}A_x + \mathbf{j}A_y.$$

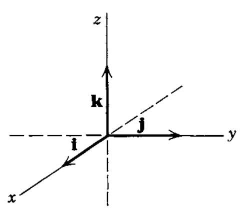


Figure 4.8

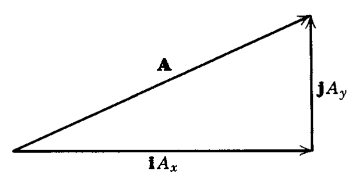


Figure 4.9

Similarly, in three dimensions

$$\mathbf{A} = \mathbf{i}A_x + \mathbf{j}A_y + \mathbf{k}A_z.$$

It is easy to add (or subtract) vectors in this form: If **A** and **B** are vectors in two dimensions, then

$$\mathbf{A} + \mathbf{B} = (\mathbf{i}A_x + \mathbf{j}A_y) + (\mathbf{i}B_x + \mathbf{j}B_y) = \mathbf{i}(A_x + B_x) + \mathbf{j}(A_y + B_y).$$

This is just the familiar result of adding components; the unit vectors **i** and **j** serve to keep track of the separate components and allow us to write **A** as a single algebraic expression. The vectors **i**, **j**, **k** are called *unit basis vectors.*

Multiplication of Vectors There are two kinds of product of two vectors. One, called the *scalar product* (or *dot product* or *inner product*), gives a result which is a scalar; the other, called the *vector product* (or *cross product*), gives a vector answer.

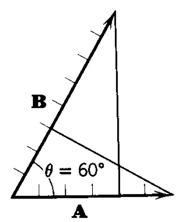
Scalar Product By definition, the scalar product of $\mathbf{A}$ and $\mathbf{B}$ (written $\mathbf{A}\cdot\mathbf{B}$) is a scalar equal to the magnitude of $\mathbf{A}$ times the magnitude of $\mathbf{B}$ times the cosine of the angle θ between $\mathbf{A}$ and $\mathbf{B}$:

$$\mathbf{A}\cdot\mathbf{B} = |\mathbf{A}|\,|\mathbf{B}|\cos\theta. \tag{4.2}$$

You should observe from (4.2) that the commutative law (4.3) holds for scalar multiplication:

$$\mathbf{A}\cdot\mathbf{B} = \mathbf{B}\cdot\mathbf{A}. \tag{4.3}$$

A useful interpretation of the dot product is shown in Figure 4.10.



$|\mathbf{B}| = 8,\ |\mathbf{A}| = 6.$
Projection of $\mathbf{B}$ on $\mathbf{A} = 4$;
$\mathbf{A}\cdot\mathbf{B} = 6\cdot 4 = 24.$
Or, projection of $\mathbf{A}$ on $\mathbf{B} = 3$;
$\mathbf{B}\cdot\mathbf{A} = 3\cdot 8 = 24.$

Figure 4.10

Since $|\mathbf{B}|\cos\theta$ is the projection of $\mathbf{B}$ on $\mathbf{A}$, we can write

$$\mathbf{A}\cdot\mathbf{B} = |\mathbf{A}| \text{ times (projection of } \mathbf{B} \text{ on } \mathbf{A}), \tag{4.4}$$

or, alternatively,

$$\mathbf{A}\cdot\mathbf{B} = |\mathbf{B}| \text{ times (projection of } \mathbf{A} \text{ on } \mathbf{B}).$$

Also we find from (4.2) that

$$\mathbf{A}\cdot\mathbf{A} = |\mathbf{A}|^2\cos 0^\circ = |\mathbf{A}|^2 = A^2. \tag{4.5}$$

Sometimes $\mathbf{A}^2$ is written instead of $|\mathbf{A}|^2$ or A^2; you should understand that the square of a vector always means the square of its magnitude or its dot product with itself.

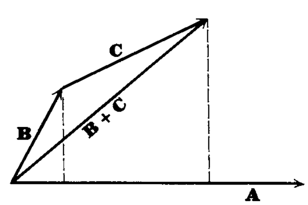


Figure 4.11

From Figure 4.11 we can see that the projection of $\mathbf{B}+\mathbf{C}$ on $\mathbf{A}$ is equal to the projection of $\mathbf{B}$ on $\mathbf{A}$ plus the projection of $\mathbf{C}$ on $\mathbf{A}$. Then by (4.4)

$$\begin{aligned} \mathbf{A}\cdot(\mathbf{B}+\mathbf{C}) &= |\mathbf{A}| \text{ times (projection of } (\mathbf{B}+\mathbf{C}) \text{ on } \mathbf{A}) \\ &= |\mathbf{A}| \text{ times (projection of } \mathbf{B} \text{ on } \mathbf{A} + \text{ projection of } \mathbf{C} \text{ on } \mathbf{A}) \\ &= \mathbf{A}\cdot\mathbf{B}+\mathbf{A}\cdot\mathbf{C}. \end{aligned} \tag{4.6}$$

This is the distributive law for scalar multiplication. By (4.3) we get also

$$(\mathbf{B}+\mathbf{C})\cdot\mathbf{A} = \mathbf{B}\cdot\mathbf{A}+\mathbf{C}\cdot\mathbf{A} = \mathbf{A}\cdot\mathbf{B}+\mathbf{A}\cdot\mathbf{C}. \tag{4.7}$$

The component form of $\mathbf{A}\cdot\mathbf{B}$ is very useful. We write

$$\mathbf{A}\cdot\mathbf{B} = (\mathbf{i}A_x+\mathbf{j}A_y+\mathbf{k}A_z)\cdot(\mathbf{i}B_x+\mathbf{j}B_y+\mathbf{k}B_z). \tag{4.8}$$

By the distributive law we can multiply this out getting nine terms such as $A_xB_x\mathbf{i}\cdot\mathbf{i}$, $A_xB_y\mathbf{i}\cdot\mathbf{j}$, and so on. Using the definition of the scalar product, we find

$$\begin{aligned} &\mathbf{i}\cdot\mathbf{i} = |\mathbf{i}|\cdot|\mathbf{i}|\cos 0^\circ = 1\cdot 1\cdot 1 = 1, \text{ and similarly, } \mathbf{j}\cdot\mathbf{j}=1,\ \mathbf{k}\cdot\mathbf{k}=1; \\ &\mathbf{i}\cdot\mathbf{j} = |\mathbf{i}|\cdot|\mathbf{j}|\cos 90^\circ = 1\cdot 1\cdot 0 = 0, \text{ and similarly, } \mathbf{i}\cdot\mathbf{k}=0,\ \mathbf{j}\cdot\mathbf{k}=0. \end{aligned} \tag{4.9}$$

Using (4.9) in (4.8), we get

$$\mathbf{A}\cdot\mathbf{B} = A_xB_x+A_yB_y+A_zB_z. \tag{4.10}$$

Equation (4.10) is an important formula which you should memorize. There are several immediate uses of this formula and of the dot product.

Angle Between Two Vectors Given the vectors, we can find the angle between them by using both (4.2) and (4.10) and solving for $\cos\theta$.

▶ **Example 3.** Find the angle between the vectors $\mathbf{A} = 3\mathbf{i}+6\mathbf{j}+9\mathbf{k}$ and $\mathbf{B} = -2\mathbf{i}+3\mathbf{j}+\mathbf{k}$. By (4.2) and (4.10) we get

$$\begin{aligned} &\mathbf{A}\cdot\mathbf{B} = |\mathbf{A}|\,|\mathbf{B}|\cos\theta = 3\cdot(-2)+6\cdot 3+9\cdot 1 = 21, \\ &|\mathbf{A}| = \sqrt{3^2+6^2+9^2} = 3\sqrt{14}, \quad |\mathbf{B}| = \sqrt{2^2+3^2+1^2} = \sqrt{14}, \\ &3\sqrt{14}\sqrt{14}\cos\theta = 21, \quad \cos\theta = \frac{1}{2}, \quad \theta = 60^\circ. \end{aligned} \tag{4.11}$$

Perpendicular and Parallel Vectors If two vectors are perpendicular, then $\cos\theta = 0$; thus

$$A_xB_x+A_yB_y+A_zB_z = 0 \qquad \text{if} \quad \mathbf{A} \text{ and } \mathbf{B} \text{ are perpendicular vectors.} \tag{4.12}$$

If two vectors are parallel, their components are proportional; thus (when no components are zero)

$$\frac{A_x}{B_x} = \frac{A_y}{B_y} = \frac{A_z}{B_z} \quad \text{if} \quad \mathbf{A} \text{ and } \mathbf{B} \text{ are parallel vectors.} \tag{4.13}$$

(Of course, if $B_x = 0$, then $A_x = 0$, etc.)

Vector Product The vector or cross product of $\mathbf{A}$ and $\mathbf{B}$ is written $\mathbf{A} \times \mathbf{B}$. By definition, $\mathbf{A} \times \mathbf{B}$ is a vector whose magnitude and direction are given as follows:

The magnitude of $\mathbf{A} \times \mathbf{B}$ is

$$|\mathbf{A} \times \mathbf{B}| = |\mathbf{A}|\,|\mathbf{B}| \sin\theta, \tag{4.14}$$

where θ is the positive angle ($\leq 180°$) between $\mathbf{A}$ and $\mathbf{B}$. The direction of $\mathbf{A} \times \mathbf{B}$ is perpendicular to the plane of $\mathbf{A}$ and $\mathbf{B}$ and in the sense $\mathbf{C}$ of advance of a right-handed screw rotated from $\mathbf{A}$ to $\mathbf{B}$ as in Figure 4.12.

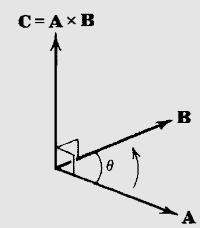


Figure 4.12

It is convenient to find the direction of $\mathbf{C} = \mathbf{A} \times \mathbf{B}$ by the following right-hand rule. Think of grasping the line $\mathbf{C}$ (or a screwdriver driving a right-handed screw in the direction $\mathbf{C}$) with the right hand. The fingers then curl in the direction of rotation of $\mathbf{A}$ into $\mathbf{B}$ (arrow in Figure 4.12) and the thumb points along $\mathbf{C} = \mathbf{A} \times \mathbf{B}$.

Perhaps the most startling result of the vector product definition is that $\mathbf{A} \times \mathbf{B}$ and $\mathbf{B} \times \mathbf{A}$ are not equal; in fact, $\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A}$. In mathematical language, vector multiplication is not commutative.

We find from (4.14) that the cross product of any two parallel (or antiparallel) vectors has magnitude $|\mathbf{A} \times \mathbf{B}| = AB \sin 0° = 0$ (or $AB \sin 180° = 0$). Thus

$$\begin{aligned} &\mathbf{A} \times \mathbf{B} = 0 \quad \text{if } \mathbf{A} \text{ and } \mathbf{B} \text{ are parallel or antiparallel,} \\ &\mathbf{A} \times \mathbf{A} = 0 \quad \text{for any } \mathbf{A}. \end{aligned} \tag{4.15}$$

Then we have the useful results

$$\mathbf{i} \times \mathbf{i} = \mathbf{j} \times \mathbf{j} = \mathbf{k} \times \mathbf{k} = 0. \tag{4.16}$$

Also from (4.14) we find

$$|\mathbf{i} \times \mathbf{j}| = |\mathbf{i}|\,|\mathbf{j}| \sin 90° = 1 \cdot 1 \cdot 1 = 1,$$

and similarly for the magnitude of the cross product of any two different unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$. From the right-hand rule and Figure 4.13, we see that the direction of $\mathbf{i} \times \mathbf{j}$ is $\mathbf{k}$, and since its magnitude is 1, we have $\mathbf{i} \times \mathbf{j} = \mathbf{k}$; however, $\mathbf{j} \times \mathbf{i} = -\mathbf{k}$. Similarly evaluating the other cross products, we find

$$
\mathbf{i}\times\mathbf{j}=\mathbf{k}\quad \mathbf{j}\times\mathbf{k}=\mathbf{i}\quad \mathbf{k}\times\mathbf{i}=\mathbf{j}.
$$
$$
\mathbf{j}\times\mathbf{i}=-\mathbf{k}\quad \mathbf{k}\times\mathbf{j}=-\mathbf{i}\quad \mathbf{i}\times\mathbf{k}=-\mathbf{j}. \tag{4.17}
$$

A good way to remember these is to write them cyclically (around a circle as indicated in Figure 4.14). Reading around the circle counterclockwise (positive θ direction), we get the positive products (for example, $\mathbf{i}\times\mathbf{j}=\mathbf{k}$); reading the other way we get the negative products (for example, $\mathbf{i}\times\mathbf{k}=-\mathbf{j}$).

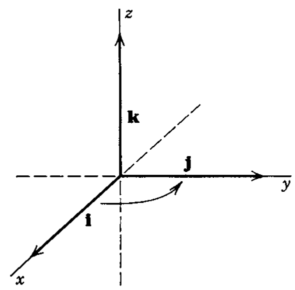


Figure 4.13

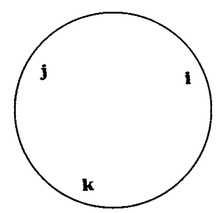


Figure 4.14

It is well to note here that the results (4.17) depend upon the way we have labeled the axes in Figure 4.13. We have arranged the (x, y, z) axes so that a rotation of the x into the y axis (through 90°) corresponds to the rotation of a right-handed screw advancing in the positive z direction. Such a coordinate system is called a *right-handed system*. If we used a left-handed system (say exchanging x and y), then all the equations in (4.17) would have their signs changed. This would be confusing; consequently, we practically always use right-handed coordinate systems, and we must be careful about this in drawing diagrams. (See Chapter 10, Section 6.)

To write $\mathbf{A}\times\mathbf{B}$ in component form we need the distributive law, namely

$$
\mathbf{A}\times(\mathbf{B}+\mathbf{C})=\mathbf{A}\times\mathbf{B}+\mathbf{A}\times\mathbf{C}. \tag{4.18}
$$

(see Problem 7.18).

Then we find

$$
\begin{aligned}
\mathbf{A}\times\mathbf{B}&=(\mathbf{i}A_x+\mathbf{j}A_y+\mathbf{k}A_z)\times(\mathbf{i}B_x+\mathbf{j}B_y+\mathbf{k}B_z)\\
&=\mathbf{i}(A_yB_z-A_zB_y)+\mathbf{j}(A_zB_x-A_xB_z)+\mathbf{k}(A_xB_y-A_yB_x)\\
&=\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\A_x&A_y&A_z\\B_x&B_y&B_z\end{vmatrix}.
\end{aligned} \tag{4.19}
$$

The second line in (4.19) is obtained by multiplying out the first line (getting nine products) and using (4.16) and (4.17). The determinant in (4.19) is the most convenient way to remember the component form of the vector product. You should verify that multiplying out the determinant using the elements of the first row gives the result in the line above it.

Since $\mathbf{A} \times \mathbf{B}$ is a vector perpendicular to $\mathbf{A}$ and to $\mathbf{B}$, we can use (4.19) to find a vector perpendicular to two given vectors.

▶ **Example 4.** Find a vector perpendicular to both $\mathbf{A} = 2\mathbf{i} + \mathbf{j} - \mathbf{k}$ and $\mathbf{B} = \mathbf{i} + 3\mathbf{j} - 2\mathbf{k}$.

$$\mathbf{A} \times \mathbf{B} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 2 & 1 & -1 \\ 1 & 3 & -2 \end{vmatrix} = \mathbf{i}(-2+3) - \mathbf{j}(-4+1) + \mathbf{k}(6-1)$$
$$= \mathbf{i} + 3\mathbf{j} + 5\mathbf{k}.$$

▶ PROBLEMS, SECTION 4

9. Let $\mathbf{A} = 2\mathbf{i} + 3\mathbf{j}$ and $\mathbf{B} = 4\mathbf{i} - 4\mathbf{j}$. Show graphically, and find algebraically, the vectors $-\mathbf{A}$, $3\mathbf{B}$, $\mathbf{A} - \mathbf{B}$, $\mathbf{B} + 2\mathbf{A}$, $\frac{1}{2}(\mathbf{A} + \mathbf{B})$.

10. If $\mathbf{A} + \mathbf{B} = 4\mathbf{j} - \mathbf{i}$ and $\mathbf{A} - \mathbf{B} = \mathbf{i} + 3\mathbf{j}$, find $\mathbf{A}$ and $\mathbf{B}$ algebraically. Show by a diagram how to find $\mathbf{A}$ and $\mathbf{B}$ geometrically.

11. Let $3\mathbf{i} - \mathbf{j} + 4\mathbf{k}$, $7\mathbf{j} - 2\mathbf{k}$, $\mathbf{i} - 3\mathbf{j} + \mathbf{k}$ be three vectors with tails at the origin. Then their heads determine three points A, B, C in space which form a triangle. Find vectors representing the sides AB, BC, CA in that order and direction (for example, A to B, not B to A) and show that the sum of these vectors is zero.

12. Find the angle between the vectors $\mathbf{A} = -2\mathbf{i} + \mathbf{j} - 2\mathbf{k}$ and $\mathbf{B} = 2\mathbf{i} - 2\mathbf{j}$.

13. If $\mathbf{A} = 4\mathbf{i} - 3\mathbf{k}$ and $\mathbf{B} = -2\mathbf{i} + 2\mathbf{j} - \mathbf{k}$, find the scalar projection of $\mathbf{A}$ on $\mathbf{B}$, the scalar projection of $\mathbf{B}$ on $\mathbf{A}$, and the cosine of the angle between $\mathbf{A}$ and $\mathbf{B}$.

14. Find the angles between (a) the space diagonals of a cube; (b) a space diagonal and an edge; (c) a space diagonal and a diagonal of a face.

15. Let $\mathbf{A} = 2\mathbf{i} - \mathbf{j} + 2\mathbf{k}$. (a) Find a *unit* vector in the same direction as $\mathbf{A}$. *Hint:* Divide $\mathbf{A}$ by $|\mathbf{A}|$. (b) Find a vector in the same direction as $\mathbf{A}$ but of magnitude 12. (c) Find a vector perpendicular to $\mathbf{A}$. *Hint:* There are *many* such vectors; you are to find one of them. (d) Find a unit vector perpendicular to $\mathbf{A}$. See hint in (a).

16. Find a unit vector in the same direction as the vector $\mathbf{A} = 4\mathbf{i} - 2\mathbf{j} + 4\mathbf{k}$, and another unit vector in the same direction as $\mathbf{B} = -4\mathbf{i} + 3\mathbf{k}$. Show that the vector sum of these unit vectors bisects the angle between $\mathbf{A}$ and $\mathbf{B}$. *Hint:* Sketch the rhombus having the two unit vectors as adjacent sides.

17. Find three vectors (none of them parallel to a coordinate axis) which have lengths and directions such that they could be made into a right triangle.

18. Show that $2\mathbf{i} - \mathbf{j} + 4\mathbf{k}$ and $5\mathbf{i} + 2\mathbf{j} - 2\mathbf{k}$ are orthogonal (perpendicular). Find a third vector perpendicular to both.

19. Find a vector perpendicular to both $\mathbf{i} - 3\mathbf{j} + 2\mathbf{k}$ and $5\mathbf{i} - \mathbf{j} - 4\mathbf{k}$.

20. Find a vector perpendicular to both $\mathbf{i} + \mathbf{j}$ and $\mathbf{i} - 2\mathbf{k}$.

21. Show that $\mathbf{B}|\mathbf{A}| + \mathbf{A}|\mathbf{B}|$ and $\mathbf{A}|\mathbf{B}| - \mathbf{B}|\mathbf{A}|$ are orthogonal.

22. Square $(\mathbf{A} + \mathbf{B})$; interpret your result geometrically. *Hint:* Your answer is a law which you learned in trigonometry.

23. If $\mathbf{A} = 2\mathbf{i} - 3\mathbf{j} + \mathbf{k}$ and $\mathbf{A} \cdot \mathbf{B} = 0$, does it follow that $\mathbf{B} = 0$? (Either prove that it does or give a specific example to show that it doesn't.) Answer the same question if $\mathbf{A} \times \mathbf{B} = \mathbf{0}$. And again answer the same question if $\mathbf{A} \cdot \mathbf{B} = 0$ *and* $\mathbf{A} \times \mathbf{B} = \mathbf{0}$.

24. What is the value of $(\mathbf{A} \times \mathbf{B})^2 + (\mathbf{A} \cdot \mathbf{B})^2$? *Comment*: This is a special case of Lagrange's identity. (*See* Chapter 6, Problem 3.12b, page 284.)

Use vectors as in Problems 3 to 8, and also the dot and cross product, to prove the following theorems from geometry.

25. The sum of the squares of the diagonals of a parallelogram is equal to twice the sum of the squares of two adjacent sides of the parallelogram.

26. The median to the base of an isosceles triangle is perpendicular to the base.

27. In a kite (four-sided figure made up of two pairs of equal adjacent sides), the diagonals are perpendicular.

28. The diagonals of a rhombus (four-sided figure with all sides of equal length) are perpendicular and bisect each other.

5. LINES AND PLANES

A great deal of analytic geometry can be simplified by the use of vector notation. Such things as equations of lines and planes, and distances between points or between lines and planes often occur in physics and it is very useful to be able to find them quickly. We shall talk about three-dimensional space most of the time although the ideas apply also to two dimensions. In analytic geometry a point is a set of three coordinates (x, y, z); we shall think of this point as the *head* of a vector $\mathbf{r} = \mathbf{i}x + \mathbf{j}y + \mathbf{k}z$ *with tail at the origin.* Most of the time the *vector* will be in the background of our minds and we shall not draw it; we shall just plot the point (x, y, z) which is the head of the vector. In other words, the point (x, y, z) and the vector $\mathbf{r}$ will be synonymous. We shall also use vectors joining two points. In Figure 5.1 the vector $\mathbf{A}$ from $(1, 2, 3)$ to (x, y, z) is

$$\mathbf{A} = \mathbf{r} - \mathbf{C} = (x, y, z) - (1, 2, 3) = (x - 1, y - 2, z - 3) \quad \text{or}$$
$$\mathbf{A} = \mathbf{i}x + \mathbf{j}y + \mathbf{k}z - (\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}) = \mathbf{i}(x - 1) + \mathbf{j}(y - 2) + \mathbf{k}(z - 3).$$

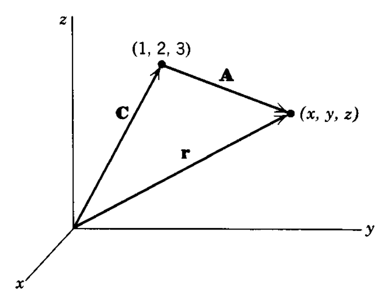


Figure 5.1

Thus we have two ways of writing vector equations; we may choose the one we prefer. Note the possible advantage of writing $(1, 0, -2)$ for $\mathbf{i} - 2\mathbf{k}$; since the zero is explicitly written, there is less chance of accidentally confusing $\mathbf{i} - 2\mathbf{k}$ with $\mathbf{i} - 2\mathbf{j} = (1, -2, 0)$. On the other hand, $5\mathbf{j}$ is simpler than $(0, 5, 0)$.

In two dimensions, we write the equation of a straight line through (x_0, y_0) with slope m as

$$\frac{y - y_0}{x - x_0} = m. \tag{5.1}$$

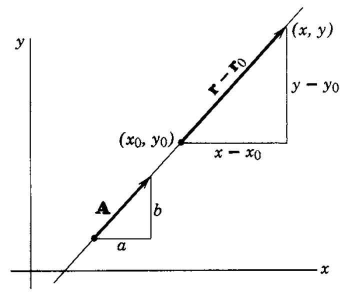


Figure 5.2

Suppose, instead of the slope, we are given a vector in the direction of the line, say $\mathbf{A} = \mathbf{i}a+\mathbf{j}b$ (Figure 5.2). Then the line through (x_0, y_0) and in the direction $\mathbf{A}$ is determined and we should be able to write its equation. The directed line segment from (x_0, y_0) to any point (x, y) on the line is the vector $\mathbf{r} - \mathbf{r}_0$ with components $x - x_0$ and $y - y_0$:

$$\mathbf{r} - \mathbf{r}_0 = \mathbf{i}(x - x_0) + \mathbf{j}(y - y_0). \tag{5.2}$$

This vector is parallel to $\mathbf{A} = \mathbf{i}a + \mathbf{j}b$. Now if two vectors are parallel, their components are proportional. Thus we can write (for $a, b \neq 0$)

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} \quad \text{or} \quad \frac{y - y_0}{x - x_0} = \frac{b}{a}. \tag{5.3}$$

This is the equation of the given straight line. As a check we see that the slope of the line is $m = b/a$, so (5.3) is the same as (5.1).

Another way to write this equation is to say that if $\mathbf{r} - \mathbf{r}_0$ and $\mathbf{A}$ are parallel vectors, one is some scalar multiple of the other, that is,

$$\mathbf{r} - \mathbf{r}_0 = \mathbf{A}t, \quad \text{or} \quad \mathbf{r} = \mathbf{r}_0 + \mathbf{A}t, \tag{5.4}$$

where t is the scalar multiple. We can think of t as a parameter; the component form of (5.4) is a set of parametric equations of the line, namely

$$\begin{aligned} x - x_0 &= at, \\ y - y_0 &= bt, \end{aligned} \quad \text{or} \quad \begin{aligned} x &= x_0 + at, \\ y &= y_0 + bt. \end{aligned} \tag{5.5}$$

Eliminating t yields the equation of the line in (5.3).

In three dimensions, the same ideas can be used. We want the equations of a straight line through a given point (x_0, y_0, z_0) and parallel to a given vector $\mathbf{A} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$. If (x, y, z) is any point on the line, the vector joining (x_0, y_0, z_0) and (x, y, z) is parallel to $\mathbf{A}$. Then its components $x - x_0$, $y - y_0$, $z - z_0$ are proportional to the components a, b, c of $\mathbf{A}$ and we have

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c} \qquad \text{(symmetric equations of a straight line, } a,\, b,\, c \neq 0\text{).} \tag{5.6}$$

If c, for instance, happens to be zero, we would have to write (5.6) in the form

$$\text{(5.7)} \qquad \frac{x-x_0}{a} = \frac{y-y_0}{b}, \quad z = z_0 \qquad \text{(symmetric equations of a straight line when } c = 0\text{).}$$

As in the two-dimensional case, equations (5.6) and (5.7) could be written

$$\text{(5.8)} \qquad \mathbf{r} = \mathbf{r}_0 + \mathbf{A}t, \quad \text{or} \quad \begin{cases} x = x_0 + at, \\ y = y_0 + bt, \\ z = z_0 + ct, \end{cases} \qquad \text{(parametric equations of a straight line).}$$

The parametric equations (5.8) have a particularly useful interpretation when the parameter t means time. Consider a particle m (electron, billiard ball, or star) moving along the straight line L in Figure 5.3. Position yourself at the origin and watch m move from P_0 to P along L. Your line of sight is the vector $\mathbf{r}$; it swings from $\mathbf{r}_0$ at $t = 0$ to $\mathbf{r} = \mathbf{r}_0 + \mathbf{A}t$ at time t. Note that the velocity of m is $d\mathbf{r}/dt = \mathbf{A}$; $\mathbf{A}$ is a vector along the line of motion.

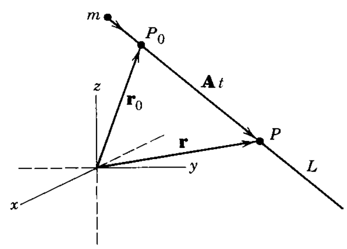


Figure 5.3

Going back to two dimensions, suppose we want the equation of a straight line L through the point (x_0, y_0) and perpendicular to a given vector $\mathbf{N} = a\mathbf{i}+b\mathbf{j}$. As above, the vector

$$\mathbf{r} - \mathbf{r}_0 = (x - x_0)\mathbf{i} + (y - y_0)\mathbf{j}$$

lies along the line. This time we want this vector perpendicular to $\mathbf{N}$; recall that two vectors are perpendicular if their dot product is zero. Setting the dot product of $\mathbf{N}$ and $\mathbf{r} - \mathbf{r}_0$ equal to zero gives

$$\text{(5.9)} \qquad a(x - x_0) + b(y - y_0) = 0 \quad \text{or} \quad \frac{y - y_0}{x - x_0} = -\frac{a}{b}.$$

This is the desired equation of the straight line L perpendicular to $\mathbf{N}$. As a check, note from Figure 5.4 that the slope of the line L is

$$\tan\theta = -\cot\phi = -a/b.$$

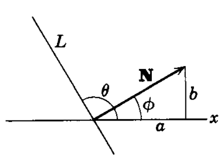


Figure 5.4

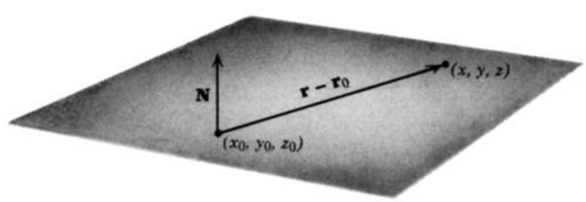


Figure 5.5

In three dimensions, we use this method to write the equation of a plane. If (x_0, y_0, z_0) is a given point in the plane and (x, y, z) is any other point in the plane, the vector (Figure 5.5)

$$\mathbf{r} - \mathbf{r}_0 = (x - x_0)\mathbf{i} + (y - y_0)\mathbf{j} + (z - z_0)\mathbf{k}$$

is in the plane. If $\mathbf{N} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$ is normal (perpendicular) to the plane, then $\mathbf{N}$ and $\mathbf{r} - \mathbf{r}_0$ are perpendicular, so the equation of the plane is $\mathbf{N} \cdot (\mathbf{r} - \mathbf{r}_0) = 0$, or

(5.10)
$$\begin{aligned} a(x - x_0) + b(y - y_0) + c(z - z_0) &= 0, \\ \text{or} \quad ax + by + cz &= d, \end{aligned} \qquad \text{(equation of a plane)}$$

where $d = ax_0 + by_0 + cz_0$.

If we are given equations like the ones above, we can read backwards to find $\mathbf{A}$ or $\mathbf{N}$. Thus we can say that the equations (5.6), (5.7), and (5.8) are the equations of a straight line which is parallel to the vector $\mathbf{A} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$, and either equation in (5.10) is the equation of a plane perpendicular to the vector $\mathbf{N} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$.

► **Example 1.** Find the equation of the plane through the three points $A(-1, 1, 1)$, $B(2, 3, 0)$, $C(0, 1, -2)$.

A vector joining any pair of the given points lies in the plane. Two such vectors are $\overrightarrow{AB} = (2, 3, 0) - (-1, 1, 1) = (3, 2, -1)$ and $\overrightarrow{AC} = (1, 0, -3)$. The cross product of these two vectors is perpendicular to the plane. This is

$$\mathbf{N} = (\overrightarrow{AB}) \times (\overrightarrow{AC}) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 3 & 2 & -1 \\ 1 & 0 & -3 \end{vmatrix} = -6\mathbf{i} + 8\mathbf{j} - 2\mathbf{k}.$$

Now we write the equation of the plane with normal direction $\mathbf{N}$ through one of the given points, say B, using (5.10):

$$-6(x - 2) + 8(y - 3) - 2z = 0 \quad \text{or} \quad 3x - 4y + z + 6 = 0.$$

(Note that we could have divided $\mathbf{N}$ by -2 to save arithmetic.)

► **Example 2.** Find the equations of a line through $(1, 0, -2)$ and perpendicular to the plane of Example 1.

The vector $3\mathbf{i} - 4\mathbf{j} + \mathbf{k}$ is perpendicular to the plane of Example 1 and so parallel to the desired line. Thus by (5.6) the symmetric equations of the line are

$$\frac{(x - 1)}{3} = \frac{y}{-4} = \frac{(z + 2)}{1}.$$

By (5.8) the parametric equations of the line are $\mathbf{r} = \mathbf{i} - 2\mathbf{k} + (3\mathbf{i} - 4\mathbf{j} + \mathbf{k})t$ or, if you like, $\mathbf{r} = (1, 0, -2) + (3, -4, 1)t$.

Vectors give us a very convenient way of finding distances between points and lines or planes. Suppose we want to find the (perpendicular) distance from a point P

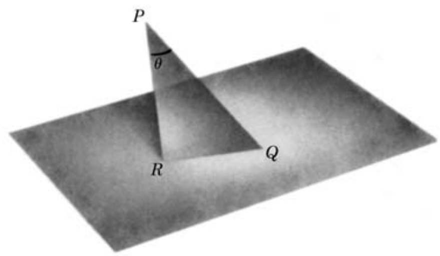


Figure 5.6

to the plane (5.10). (See Figure 5.6.) We pick *any* point Q we like in the plane (just by looking at the equation of the plane and thinking of some simple numbers x, y, z that satisfy it). The distance PR is what we want. Since PR and RQ are perpendicular (because PR is perpendicular to the plane), we have from Figure 5.6

$$PR = PQ\cos\theta. \tag{5.11}$$

From the equation of the plane, we can find a vector $\mathbf{N}$ normal to the plane. If we divide $\mathbf{N}$ by its magnitude, we have a unit vector normal to the plane; we denote this unit vector by $\mathbf{n}$. Then $|\overrightarrow{PQ}\cdot\mathbf{n}| = (PQ)\cos\theta$, which is what we need in (5.11) to find PR. (We have put in absolute value signs because $\overrightarrow{PQ}\cdot\mathbf{n}$ might be negative, whereas $(PQ)\cos\theta$, with θ acute as in Figure 5.6, is positive.)

▶ **Example 3.** Find the distance from the point $P(1, -2, 3)$ to the plane $3x - 2y + z + 1 = 0$.

One point in the plane is $(1, 2, 0)$; call this point Q. Then the vector from P to Q is

$$\overrightarrow{PQ} = (1, 2, 0) - (1, -2, 3) = (0, 4, -3) = 4\mathbf{j} - 3\mathbf{k}.$$

From the equation of the plane we get the normal vector

$$\mathbf{N} = 3\mathbf{i} - 2\mathbf{j} + \mathbf{k}.$$

We get $\mathbf{n}$ by dividing $\mathbf{N}$ by $|\mathbf{N}| = \sqrt{14}$. Then we have

$$\begin{aligned}|PR| &= \left|\overrightarrow{PQ}\cdot\mathbf{n}\right| = \left|(4\mathbf{j} - 3\mathbf{k})\cdot(3\mathbf{i} - 2\mathbf{j} + \mathbf{k})/\sqrt{14}\right| \\ &= \left|(-8-3)/\sqrt{14}\right| = 11/\sqrt{14}.\end{aligned}$$

We can find the distance from a point P to a line in a similar way. In Figure 5.7 we want the perpendicular distance PR. We select any point on the line [that is, we pick any (x, y, z) satisfying the equations of the line]; call this point Q. Then (see Figure 5.7) $PR = PQ\sin\theta$. Let $\mathbf{A}$ be a vector along the line and $\mathbf{u}$ a unit vector along the line (obtained by dividing $\mathbf{A}$ by its magnitude). Then

$$\left|\overrightarrow{PQ}\times\mathbf{u}\right| = |PQ|\sin\theta,$$

so we get

$$|PR| = \left|\overrightarrow{PQ}\times\mathbf{u}\right|.$$

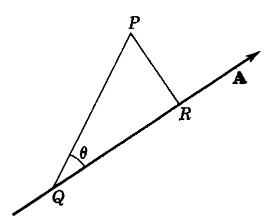


Figure 5.7

▶ **Example 4.** Find the distance from $P(1,2,-1)$ to the line joining $P_1(0,0,0)$ and $P_2(-1,0,2)$.

Let $\mathbf{A} = \overrightarrow{P_1P_2} = -\mathbf{i} + 2\mathbf{k}$; this is a vector along the line. Then a unit vector along the line is $\mathbf{u} = (1/\sqrt{5})(-\mathbf{i} + 2\mathbf{k})$. Let us take Q to be $P_1(0,0,0)$. Then $\overrightarrow{PQ} = -\mathbf{i} - 2\mathbf{j} + \mathbf{k}$, so we get for the distance $|PR|$:

$$|PR| = \frac{1}{\sqrt{5}}|(-\mathbf{i} - 2\mathbf{j} + \mathbf{k}) \times (-\mathbf{i} + 2\mathbf{k})| = \frac{1}{\sqrt{5}}|-4\mathbf{i} + \mathbf{j} - 2\mathbf{k}| = \sqrt{21/5}.$$

It is also straightforward to find the distance between two skew lines (and if you really want to appreciate vectors, just look up this calculation in an analytic geometry book that doesn't use vectors!). Pick two points P and Q, one on each line (Figure 5.8). Then $|\overrightarrow{PQ} \cdot \mathbf{n}|$, where $\mathbf{n}$ is a unit vector perpendicular to both lines, is the distance we want. Now if $\mathbf{A}$ and $\mathbf{B}$ are vectors along the two lines, then $\mathbf{A} \times \mathbf{B}$ is perpendicular to both, and $\mathbf{n}$ is just $\mathbf{A} \times \mathbf{B}$ divided by $|\mathbf{A} \times \mathbf{B}|$.

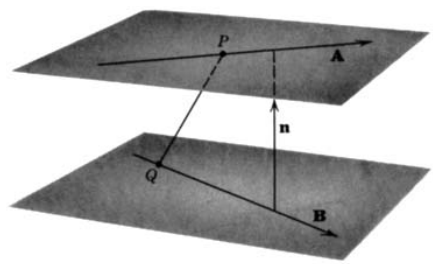


Figure 5.8

▶ **Example 5.** Find the distance between the lines $\mathbf{r} = \mathbf{i} - 2\mathbf{j} + (\mathbf{i} - \mathbf{k})t$ and $\mathbf{r} = 2\mathbf{j} - \mathbf{k} + (\mathbf{j} - \mathbf{i})t$.

If we write the first line as $\mathbf{r} = \mathbf{r}_0 + \mathbf{A}t$, then (the head of) $\mathbf{r}_0$ is a simple choice for P, so we have

$$P = (1, -2, 0) \quad \text{and} \quad \mathbf{A} = \mathbf{i} - \mathbf{k}.$$

Similarly, from the second line we find

$$Q = (0, 2, -1) \quad \text{and} \quad \mathbf{B} = \mathbf{j} - \mathbf{i}.$$

Then $\mathbf{A} \times \mathbf{B} = \mathbf{i} + \mathbf{j} + \mathbf{k}$ and $\mathbf{n} = \left(1/\sqrt{3}\right)(\mathbf{i} + \mathbf{j} + \mathbf{k})$. Also

$$\overrightarrow{PQ} = (0,2,-1) - (1,-2,0) = (-1,4,-1) = -\mathbf{i} + 4\mathbf{j} - \mathbf{k}.$$

Thus we get for the distance between the lines

$$\left|\overrightarrow{PQ} \cdot \mathbf{n}\right| = \left|(-\mathbf{i} + 4\mathbf{j} - \mathbf{k}) \cdot (\mathbf{i} + \mathbf{j} + \mathbf{k})/\sqrt{3}\right| = |-1 + 4 - 1|/\sqrt{3} = 2/\sqrt{3}.$$

▶ **Example 6.** Find the direction of the line of intersection of the planes $x - 2y + 3z = 4$ and $2x + y - z = 5$.

The desired line lies in both planes, and so is perpendicular to the two normal vectors to the planes, namely $\mathbf{i} - 2\mathbf{j} + 3\mathbf{k}$ and $2\mathbf{i} + \mathbf{j} - \mathbf{k}$. Then the direction of the line is that of the cross product of these normal vectors; this is $-\mathbf{i} + 7\mathbf{j} + 5\mathbf{k}$.

▶ **Example 7.** Find the cosine of the angle between the planes of Example 6.

The angle between the planes is the same as the angle between the normals to the planes. Thus our problem is to find the angle between the vectors $\mathbf{A} = \mathbf{i}-2\mathbf{j}+3\mathbf{k}$ and $\mathbf{B} = 2\mathbf{i}+\mathbf{j}-\mathbf{k}$. Since $\mathbf{A}\cdot\mathbf{B} = |\mathbf{A}|\,|\mathbf{B}|\cos\theta$, we have $-3 = \sqrt{14}\sqrt{6}\cos\theta$, and so $\cos\theta = -\sqrt{3/28}$. This gives the obtuse angle between the planes; the corresponding acute angle is $\pi - \theta$, or $\arccos\sqrt{3/28}$.

▶ PROBLEMS, SECTION 5

In Problems 1 to 5, all lines are in the (x, y) plane.

1. Write the equation of the straight line through $(2, -3)$ with slope $3/4$, in the parametric form $\mathbf{r} = \mathbf{r}_0 + \mathbf{A}t$.

2. Find the slope of the line whose parametric equation is $\mathbf{r} = (\mathbf{i}-\mathbf{j}) + (2\mathbf{i}+3\mathbf{j})t$.

3. Write, in parametric form [as in Problem 1], the equation of the straight line that joins $(1, -2)$ and $(3, 0)$.

4. Write, in parametric form, the equation of the straight line that is perpendicular to $\mathbf{r} = (2\mathbf{i}+4\mathbf{j}) + (\mathbf{i}-2\mathbf{j})t$ and goes through $(1, 0)$.

5. Write, in parametric form, the equation of the y axis.

Find the symmetric equations (5.6) or (5.7) and the parametric equations (5.8) of a line, and/or the equation (5.10) of the plane satisfying the following given conditions.

6. Line through $(1, -1, -5)$ and $(2, -3, -3)$.

7. Line through $(2, 3, 4)$ and $(5, 1, -2)$.

8. Line through $(0, -2, 4)$ and $(3, -2, -1)$.

9. Line through $(-1, 3, 7)$ and $(-1, -2, 7)$.

10. Line through $(3, 4, -1)$ and parallel to $2\mathbf{i}-3\mathbf{j}+6\mathbf{k}$.

11. Line through $(4, -1, 3)$ and parallel to $\mathbf{i}-2\mathbf{k}$.

12. Line through $(5, -4, 2)$ and parallel to the line $\mathbf{r} = \mathbf{i}-\mathbf{j}+(5\mathbf{i}-2\mathbf{j}+\mathbf{k})t$.

13. Line through $(3, 0, -5)$ and parallel to the line $\mathbf{r} = (2, 1, -5) + (0, -3, 1)t$.

14. Plane containing the triangle ABC of Problem 4.11.

15. Plane through the origin and the points in Problem 8.

16. Plane through the point and perpendicular to the line in Problem 12.

17. Plane through the point and perpendicular to the line in Problem 13.

18. Plane containing the two parallel lines in Problem 12.

19. Plane containing the two parallel lines in Problem 13.

20. Plane containing the three points $(0, 1, 1)$, $(2, 1, 3)$, and $(4, 2, 1)$.

In Problems 21 to 23, find the angle between the given planes.

21. $2x + 6y - 3z = 10$ and $5x + 2y - z = 12$.

22. $2x - y - z = 4$ and $3x - 2y - 6z = 7$.

23. $2x + y - 2z = 3$ and $3x - 6y - 2z = 4$.

24. Find a point on *both* the planes (that is, on their line of intersection) in Problem 21. Find a vector parallel to the line of intersection. Write the equations of the line of intersection of the planes. Find the distance from the origin to the line.

25. As in Problem 24, find the equations of the line of intersection of the planes in Problem 22. Find the distance from the point $(2, 1, -1)$ to the line.

26. As in Problem 24, find the equations of the line of intersection of the planes in Problem 23. Find the distance from the point $(1, 0, 0)$ to the line.

27. Find the equation of the plane through $(2, 3, -2)$ and perpendicular to both planes in Problem 21.

28. Find the equation of the plane through $(-4, -1, 2)$ and perpendicular to both planes in Problem 22.

29. Find a point on the plane $2x - y - z = 13$. Find the distance from $(7, 1, -2)$ to the plane.

30. Find the distance from the origin to the plane $3x - 2y - 6z = 7$.

31. Find the distance from $(-2, 4, 5)$ to the plane $2x + 6y - 3z = 10$.

32. Find the distance from $(3, -1, 2)$ to the plane $5x - y - z = 4$.

33. Find the perpendicular distance between the two parallel lines in Problem 12.

34. Find the distance (perpendicular is understood) between the two parallel lines in Problem 13.

35. Find the distance from $(2, 5, 1)$ to the line in Problem 10.

36. Find the distance from $(3, 2, 5)$ to the line in Problem 11.

37. Determine whether the lines

$$\frac{x-1}{2} = \frac{y+3}{1} = \frac{z-4}{-3} \quad \text{and} \quad \frac{x+3}{4} = \frac{y+4}{1} = \frac{8-z}{4}$$

intersect. *Two suggestions:* (1) Can you find the intersection point, if any? (2) Consider the distance between the lines.

38. Find the angle between the lines in Problem 37.

In Problems 39 and 40, show that the given lines intersect and find the acute angle between them.

39. $\mathbf{r} = 2\mathbf{j} + \mathbf{k} + (3\mathbf{i} - \mathbf{k})t_1$ and $\mathbf{r} = 7\mathbf{i} + 2\mathbf{k} + (2\mathbf{i} - \mathbf{j} + \mathbf{k})t_2$.

40. $\mathbf{r} = (5, -2, 0) + (1, -1, -1)t_1$ and $\mathbf{r} = (4, -4, -1) + (0, 3, 2)t_2$.

In Problems 41 to 44, find the distance between the two given lines.

41. $\mathbf{r} = (4, 3, -1) + (1, 1, 1)t$ and $\mathbf{r} = (4, -1, 1) + (1, -2, -1)t$.

42. The line that joins $(0, 0, 0)$ to $(1, 2, -1)$, and the line that joins $(1, 1, 1)$ to $(2, 3, 4)$.

43. $\dfrac{x-1}{2} = \dfrac{y+2}{3} = \dfrac{2z-1}{4}$ and $\dfrac{x+2}{-1} = \dfrac{2-y}{2}, \quad z = \dfrac{1}{2}$.

44. The x axis and $\mathbf{r} = \mathbf{j} - \mathbf{k} + (2\mathbf{i} - 3\mathbf{j} + \mathbf{k})t$.

45. A particle is traveling along the line $(x - 3)/2 = (y + 1)/(-2) = z - 1$. Write the equation of its path in the form $\mathbf{r} = \mathbf{r}_0 + \mathbf{A}t$. Find the distance of closest approach of the particle to the origin (that is, the distance from the origin to the line). If t represents time, show that the time of closest approach is $t = -(\mathbf{r}_0 \cdot \mathbf{A})/|\mathbf{A}|^2$. Use this value to check your answer for the distance of closest approach. *Hint:* See Figure 5.3. If P is the point of closest approach, what is $\mathbf{A} \cdot \mathbf{r}$?

6. MATRIX OPERATIONS

In Section 2 we used matrices simply as arrays of numbers. Now we want to go farther into the subject and discuss the meaning and use of multiplying a matrix by a number and of combining matrices by addition, subtraction, multiplication, and even (in a sense) division. We will see that we may be able to find functions of matrices such as e^{M}. These are, of course, all questions of definition, but we shall show some applications which might suggest reasonable definitions; or alternatively, given the definitions, we shall see what applications we can make of the matrix operations.

Matrix Equations Let us first emphasize again that two matrices are equal *only* if they are identical. Thus the matrix equation

$$\begin{pmatrix} x & r & u \\ y & s & v \end{pmatrix} = \begin{pmatrix} 2 & 1 & -5 \\ 3 & -7i & 1-i \end{pmatrix}$$

is really the set of six equations

$$x = 2, \qquad y = 3, \qquad r = 1, \qquad s = -7i, \qquad u = -5, \qquad v = 1 - i.$$

(Recall similar situations we have met before: The equation $z = x + iy = 2 - 3i$ is equivalent to the two real equations $x = 2$, $y = -3$; a vector equation in three dimensions is equivalent to three component equations.) In complicated problems involving many numbers or variables, it is often possible to save a great deal of writing by using a single matrix equation to replace a whole set of ordinary equations. Any time it is possible to so abbreviate the writing of a mathematical equation (like using a single letter for a complicated parenthesis) it not only saves time but often enables us to think more clearly.

Multiplication of a Matrix by a Number A convenient way to display the components of the vector $\mathbf{A} = 2\mathbf{i} + 3\mathbf{j}$ is to write them as elements of a matrix, either

$$\mathrm{A} = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \quad \text{called a column matrix or column vector,}$$

or

$$\mathrm{A}^{\mathrm{T}} = \begin{pmatrix} 2 & 3 \end{pmatrix} \quad \text{called a row matrix or row vector.}$$

The row matrix A^{T} is the transpose of the column matrix A. Observe the notation we are using here: We will often use the same letter for a vector and its column matrix, but we will usually write the letter representing the matrix as A (roman, not boldface), the vector as boldface $\mathbf{A}$, and the length of the vector as italic A.

Now suppose we want a vector of twice the length of $\mathbf{A}$ and in the same direction; we would write this as $2\mathbf{A} = 4\mathbf{i} + 6\mathbf{j}$. Then we would like to write its matrix representation as

$$2\mathrm{A} = 2\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \end{pmatrix}, \qquad 2\mathrm{A}^{\mathrm{T}} = 2(2 \quad 3) = (4 \quad 6).$$

This is, in fact, exactly how a matrix is multiplied by a number: *every* element of the matrix is multiplied by the number. Thus

$$k\begin{pmatrix} a & c & e \\ b & d & f \end{pmatrix} = \begin{pmatrix} ka & kc & ke \\ kb & kd & kf \end{pmatrix}$$

and

$$\begin{pmatrix} -\frac{1}{2} & \frac{3}{4} \\ -1 & -\frac{5}{8} \end{pmatrix} = -\frac{1}{8}\begin{pmatrix} 4 & -6 \\ 8 & 5 \end{pmatrix}.$$

Note carefully a difference between determinants and matrices: multiplying a matrix by a number k means multiplying every element by k, but multiplying just *one* row of a determinant by k multiplies the determinant by k. Thus $\det(k\mathrm{A}) = k^2 \det \mathrm{A}$ for a 2 by 2 matrix, $\det(k\mathrm{A}) = k^3 \det \mathrm{A}$ for a 3 by 3 matrix, and so on.

Addition of Matrices When we add vectors algebraically, we add them by components. Matrices are added in the same way, by adding corresponding elements. For example,

$$(6.1) \qquad \begin{pmatrix} 1 & 3 & -2 \\ 4 & 7 & 1 \end{pmatrix} + \begin{pmatrix} 2 & -1 & 4 \\ 3 & -7 & -2 \end{pmatrix} = \begin{pmatrix} 1+2 & 3-1 & -2+4 \\ 4+3 & 7-7 & 1-2 \end{pmatrix} = \begin{pmatrix} 3 & 2 & 2 \\ 7 & 0 & -1 \end{pmatrix}.$$

Note that if we add $\mathrm{A} + \mathrm{A}$ we would get $2\mathrm{A}$ in accord with our definition of twice a matrix above. Suppose we have

$$\mathrm{A} = \begin{pmatrix} 1 & 3 & -2 \\ 4 & 7 & 1 \end{pmatrix} \qquad \text{and} \qquad \mathrm{B} = \begin{pmatrix} 2 & -1 \\ 3 & 5 \end{pmatrix}.$$

In this case we cannot add A and B; we say that the sum is undefined or meaningless.

In applications, then, matrices are useful in representing things which are added by components. Suppose, for example, that, in (6.1), the columns represent displacements of three particles. The first particle is displaced by $\mathbf{i}+4\mathbf{j}$ (first column of the first matrix) and later by $2\mathbf{i}+3\mathbf{j}$ (first column of the second matrix). The total displacement is then $3\mathbf{i}+7\mathbf{j}$ (first column of the sum of the matrices). Similarly the second and third columns represent displacements of the second and third particles.

Multiplication of Matrices Let us start by defining the product of two matrices and then see what use we can make of the process. Here is a simple example to show what is meant by the product $\mathrm{AB} = \mathrm{C}$ of two matrices A and B:

$$(6.2a) \qquad \mathrm{AB} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{pmatrix} = \mathrm{C}.$$

Observe that in the product matrix C, the element in the first row and first column is obtained by multiplying each element of the first row in A times the corresponding element in the first column of B and adding the results. This is referred to as "row times column" multiplication; when we compute $ae+bg$, we say that we have "multiplied the first row of A times the first column of B." Next examine the element $af+bh$ in the first row and second column of C; it is the "first row of A times the second column of B." Similarly, $ce+dg$ in the second row and first column of C is the "second row of A times the first column of B," and $cf+dh$ in the second row and second column of C is the "second row of A times the second column of B." Thus all the elements of C may be obtained by using the following simple rule:

(6.2b) The element in row i and column j of the product matrix AB is equal to row i of A times column j of B. In index notation

$$(\mathrm{AB})_{ij} = \sum_k A_{ik} B_{kj}.$$

Here is another useful way of saying this: Think of the elements in a row (or a column) of a matrix as the components of a vector. Then row times column multiplication for the matrix product AB corresponds to finding the dot product of a row vector of A and a column vector of B.

It is not necessary for matrices to be square in order for us to multiply them. Consider the following example.

▶ **Example 1.** Find the product of A and B if

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

Following the rule we have stated, we get

$$\begin{aligned} \mathrm{AB} &= \begin{pmatrix} 4 & 2 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 5 & 3 \\ 2 & 7 & -4 \end{pmatrix} \\ &= \begin{pmatrix} 4\cdot 1 + 2\cdot 2 & 4\cdot 5 + 2\cdot 7 & 4\cdot 3 + 2(-4) \\ -3\cdot 1 + 1\cdot 2 & -3\cdot 5 + 1\cdot 7 & -3\cdot 3 + 1(-4) \end{pmatrix} \\ &= \begin{pmatrix} 8 & 34 & 4 \\ -1 & -8 & -13 \end{pmatrix}. \end{aligned}$$

Notice that the third column in B caused us no difficulty in following our rule; we simply multiplied each row of A times the third column of B to obtain the elements in the third column of AB. But suppose we tried to find the product BA. In B a row contains 3 elements, while in A a column contains only two; thus we are not able to apply the "row times column" method. Whenever this happens, we say that B is *not conformable* with respect to A, and the product BA is not defined (that is, it is meaningless and we do not use it).

The product AB (in that order) can be found if and only if the number of elements in a row of A equals the number of elements in a column of B; the matrices A, B in that order are then called *conformable*. (Observe that the number of rows in A and of columns in B have nothing to do with the question of whether we can find AB or not.)

▶ **Example 2.** Find AB and BA, given

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

Note that here the matrices are conformable in both orders, so we can find both AB and BA.

$$\begin{aligned}
\mathrm{AB} &= \begin{pmatrix} 3 & -1 \\ -4 & 2 \end{pmatrix}\begin{pmatrix} 5 & 2 \\ -7 & 3 \end{pmatrix} \\
&= \begin{pmatrix} 3\cdot 5 - 1(-7) & 3\cdot 2 - 1\cdot 3 \\ -4\cdot 5 + 2(-7) & -4\cdot 2 + 2\cdot 3 \end{pmatrix} = \begin{pmatrix} 22 & 3 \\ -34 & -2 \end{pmatrix}. \\
\mathrm{BA} &= \begin{pmatrix} 5 & 2 \\ -7 & 3 \end{pmatrix}\begin{pmatrix} 3 & -1 \\ -4 & 2 \end{pmatrix} \\
&= \begin{pmatrix} 5\cdot 3 + 2(-4) & 5(-1) + 2\cdot 2 \\ -7\cdot 3 + 3(-4) & -7(-1) + 3\cdot 2 \end{pmatrix} = \begin{pmatrix} 7 & -1 \\ -33 & 13 \end{pmatrix}.
\end{aligned}$$

Observe that AB is *not* the same as BA. We say that matrix multiplication is *not commutative*, or that, in general, matrices do not commute under multiplication. (Of course, two particular matrices may happen to commute.) We define the *commutator* of the matrices A and B by

(6.3) $$[\mathrm{A}, \mathrm{B}] = \mathrm{AB} - \mathrm{BA} = \text{ \textit{commutator} of A and B.}$$

(Commutators are of interest in classical and quantum mechanics.) Since matrices do not in general commute, be careful not to change the order of factors in a product of matrices unless you know they commute. For example

$$(\mathrm{A} - \mathrm{B})(\mathrm{A} + \mathrm{B}) = \mathrm{A}^2 + \mathrm{AB} - \mathrm{BA} - \mathrm{B}^2 = \mathrm{A}^2 - \mathrm{B}^2 + [\mathrm{A}, \mathrm{B}].$$

This is not equal to $\mathrm{A}^2 - \mathrm{B}^2$ when A and B don't commute. Also see the discussion just after (6.17). On the other hand, the associative law is valid, that is, $\mathrm{A}(\mathrm{BC}) = (\mathrm{AB})\mathrm{C}$, so we can write either as simply ABC. Also the distributive law holds: $\mathrm{A}(\mathrm{B} + \mathrm{C}) = \mathrm{AB} + \mathrm{AC}$ and $(\mathrm{A} + \mathrm{B})\mathrm{C} = \mathrm{AC} + \mathrm{BC}$ as we have been assuming above. (See Section 9.)

Zero Matrix The *zero* or *null* matrix means one with all its elements equal to zero. It is often abbreviated by 0, but we must be careful about this. For example:

(6.4) $$\text{If} \quad \mathrm{M} = \begin{pmatrix} 2 & -4 \\ 1 & -2 \end{pmatrix}, \quad \text{then} \quad \mathrm{M}^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

so we have $\mathrm{M}^2 = 0$, but $\mathrm{M} \neq 0$. Also see Problems 9 and 10.

Identity Matrix or Unit Matrix This is a square matrix with every element of the main diagonal (upper left to lower right) equal to 1 and all other elements equal to zero. For example

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{6.5}$$

is a unit or identity matrix of order 3 (that is, three rows and three columns). An identity or unit matrix is called 1 or I or U or E in various references. You should satisfy yourself that in multiplication, a unit matrix acts like the number 1, that is, if A is any matrix and I is the unit matrix conformable with A in the order in which we multiply, then $\mathrm{IA} = \mathrm{AI} = \mathrm{A}$ (Problem 11).

Operations with Determinants We do not define addition for determinants. However, multiplication is useful; we multiply determinants the same way we multiply matrices. It can be shown that if A and B are square matrices of the same order, then

$$\det \mathrm{AB} = \det \mathrm{BA} = (\det \mathrm{A}) \cdot (\det \mathrm{B}). \tag{6.6}$$

Look at Example 2 above to see that (6.6) is true even when matrices AB and BA are not equal, that is, when A and B do not commute.

Applications of Matrix Multiplication We can now write sets of simultaneous linear equations in a very simple form using matrices. Consider the matrix equation

$$\begin{pmatrix} 1 & 0 & -1 \\ -2 & 3 & 0 \\ 1 & -3 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 5 \\ 1 \\ -10 \end{pmatrix}. \tag{6.7}$$

If we multiply the first two matrices, we have

$$\begin{pmatrix} x - z \\ -2x + 3y \\ x - 3y + 2z \end{pmatrix} = \begin{pmatrix} 5 \\ 1 \\ -10 \end{pmatrix}. \tag{6.8}$$

Now recall that two matrices are equal only if they are identical. Thus (6.8) is the set of three equations

$$\begin{cases} x \phantom{{}+3y} - z = 5 \\ -2x + 3y \phantom{{}+2z} = 1 \\ x - 3y + 2z = -10 \end{cases}. \tag{6.9}$$

Consequently (6.7) is the matrix form for the set of equations (6.9). In this way we can write any set of linear equations in matrix form. If we use letters to represent the matrices in (6.7),

$$\mathrm{M} = \begin{pmatrix} 1 & 0 & -1 \\ -2 & 3 & 0 \\ 1 & -3 & 2 \end{pmatrix}, \quad \mathrm{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \mathrm{k} = \begin{pmatrix} 5 \\ 1 \\ -10 \end{pmatrix}, \tag{6.10}$$

then we can write (6.7) or (6.9) as

$$\mathrm{Mr} = \mathrm{k}. \tag{6.11}$$

Or, in index notation, we can write $\sum_j M_{ij}x_j = k_i$. [Review Section 2, equations (2.3) to (2.6).] Note that (6.11) could represent any number of equations or unknowns (say 100 equations in 100 unknowns!). Thus we have a great simplification in notation which may help us to think more clearly about a problem. For example, if (6.11) were an ordinary algebraic equation, we would solve it for r to get

$$\mathrm{r} = \mathrm{M}^{-1}\mathrm{k}. \tag{6.12}$$

Since M is a matrix, (6.12) only makes sense if we can give a meaning to M^{-1} such that (6.12) gives the solution of (6.7) or (6.9). Let's try to do this.

Inverse of a Matrix The reciprocal or inverse of a number x is x^{-1} such that the product $xx^{-1} = 1$. We define the inverse of a matrix M (if it has one) as the matrix M^{-1} such that MM^{-1} and $\mathrm{M}^{-1}\mathrm{M}$ are both equal to a unit matrix I. Note that only square matrices can have inverses (otherwise we could not multiply both MM^{-1} and $\mathrm{M}^{-1}\mathrm{M}$). Actually, some square matrices do not have inverses either. You can see from (6.6) that if $\mathrm{M}^{-1}\mathrm{M} = \mathrm{I}$, then $(\det \mathrm{M}^{-1})(\det \mathrm{M}) = \det \mathrm{I} = 1$. If two numbers have product $= 1$, then neither of them is zero; thus $\det \mathrm{M} \neq 0$ is a requirement for M to have an inverse.

If a matrix has an inverse we say that it is *invertible*; if it doesn't have an inverse, it is called *singular*. For simple numerical matrices your computer will easily produce the inverse of an invertible matrix. However, for theoretical purposes, we need a formula for the inverse; let's discuss this. The cofactor of an element in a square matrix M means exactly the same thing as the cofactor of that element in $\det \mathrm{M}$ [see (3.3) and (3.4)]. Thus, the cofactor C_{ij} of the element m_{ij} in row i and column j is a number equal to $(-1)^{i+j}$ times the value of the determinant remaining when we cross off row i and column j. Then to find M^{-1}: Find the cofactors C_{ij} of all elements, write the matrix C whose elements are C_{ij}, transpose it (interchange rows and columns), and divide by $\det \mathrm{M}$. (See Problem 23.)

$$\mathrm{M}^{-1} = \frac{1}{\det \mathrm{M}}\mathrm{C}^{\mathrm{T}} \quad \text{where } C_{ij} = \text{ cofactor of } m_{ij} \tag{6.13}$$

Although (6.13) is particularly useful in theoretical work, you should practice using it (as we said for Cramer's rule) on simple numerical problems in order to learn what the formula means.

▶ **Example 3.** For the matrix M of the coefficients in equations (6.7) or (6.9), find M^{-1}.

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

We find $\det \mathrm{M} = 3$. The cofactors of the elements are:

$$\begin{array}{llll}
1^{\text{st}} \text{ row}: & \begin{vmatrix} 3 & 0 \\ -3 & 2 \end{vmatrix} = 6, & -\begin{vmatrix} -2 & 0 \\ 1 & 2 \end{vmatrix} = 4, & \begin{vmatrix} -2 & 3 \\ 1 & -3 \end{vmatrix} = 3. \\
2^{\text{nd}} \text{ row}: & -\begin{vmatrix} 0 & -1 \\ -3 & 2 \end{vmatrix} = 3, & \begin{vmatrix} 1 & -1 \\ 1 & 2 \end{vmatrix} = 3, & -\begin{vmatrix} 1 & 0 \\ 1 & -3 \end{vmatrix} = 3. \\
3^{\text{rd}} \text{ row}: & \begin{vmatrix} 0 & -1 \\ 3 & 0 \end{vmatrix} = 3, & -\begin{vmatrix} 1 & -1 \\ -2 & 0 \end{vmatrix} = 2, & \begin{vmatrix} 1 & 0 \\ -2 & 3 \end{vmatrix} = 3.
\end{array}$$

Then

$$\mathrm{C} = \begin{pmatrix} 6 & 4 & 3 \\ 3 & 3 & 3 \\ 3 & 2 & 3 \end{pmatrix} \quad \text{so} \quad \mathrm{M}^{-1} = \frac{1}{\det \mathrm{M}} \mathrm{C}^{\mathrm{T}} = \frac{1}{3} \begin{pmatrix} 6 & 3 & 3 \\ 4 & 3 & 2 \\ 3 & 3 & 3 \end{pmatrix}.$$

Now we can use M^{-1} to solve equations (6.9). By (6.12), the solution is given by the column matrix $\mathrm{r} = \mathrm{M}^{-1}\mathrm{k}$, so we have

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 6 & 3 & 3 \\ 4 & 3 & 2 \\ 3 & 3 & 3 \end{pmatrix} \begin{pmatrix} 5 \\ 1 \\ -10 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ -4 \end{pmatrix},$$

or $x = 1$, $y = 1$, $z = -4$. (See Problem 12.)

Rotation Matrices As another example of matrix multiplication, let's consider a case where we know the answer, just to see that our definition of matrix multiplication works the way we want it to. You probably know the rotation equations [for reference, see the next section, equation (7.12) and Figure 7.4]. Equation (7.12) gives the matrix which rotates the vector $\mathbf{r} = \mathbf{i}x + \mathbf{j}y$ through angle θ to become the vector $\mathbf{R} = \mathbf{i}X + \mathbf{j}Y$. Suppose we further rotate $\mathbf{R}$ through angle ϕ to become $\mathbf{R}' = \mathbf{i}X' + \mathbf{j}Y'$. We could write the matrix equations for the rotations in the form $\mathrm{R} = \mathrm{Mr}$ and $\mathrm{R}' = \mathrm{M}'\mathrm{R}$ where M and M′ are the rotation matrices (7.12) for rotation through angles θ and ϕ. Then, solving for R′ in terms of r, we get $\mathrm{R}' = \mathrm{M}'\mathrm{Mr}$. We expect the matrix product M′M to give us the matrix for a rotation through the angle $\theta + \phi$, that is we expect to find

$$\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{pmatrix}. \tag{6.14}$$

It is straightforward to multiply the two matrices (Problem 25) and verify (by using trigonometric identities) that (6.14) is correct. Also note that these two rotation matrices commute (that is, rotation through angle θ and then through angle ϕ gives the same result as rotation through ϕ followed by rotation through θ). This is true in this problem in two dimensions. As we will see in Section 7, rotation matrices in three dimensions do not in general commute if the two rotation axes are different. (See Problems 7.30 and 7.31.) But all rotations in the $(x,\ y)$ plane are rotations about the z axis and so they commute.

Functions of Matrices Since we now know how to multiply matrices and how to add them, we can evaluate any power of a matrix A and so evaluate a polynomial in A. The constant term c or $c\mathrm{A}^0$ in a polynomial is defined to mean c times the unit matrix I [see (6.16) below].

► **Example 4.**

$$\text{(6.15)} \qquad \text{If} \quad \mathrm{A} = \begin{pmatrix} 1 & \sqrt{2} \\ -\sqrt{2} & -1 \end{pmatrix}, \quad \text{then} \quad \mathrm{A}^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -\mathrm{I},$$
$$\mathrm{A}^3 = -\mathrm{A},\ \mathrm{A}^4 = \mathrm{I}, \quad \text{and so on.}$$

(Verify these powers and the fact that higher powers simply repeat these four results: A, $-$I, $-$A, I, over and over.) Then we can find (Problem 28)

$$\text{(6.16)} \qquad \begin{aligned} f(\mathrm{A}) &= 3 - 2\mathrm{A}^2 - \mathrm{A}^3 - 5\mathrm{A}^4 + \mathrm{A}^6 \\ &= 3\mathrm{I} + 2\mathrm{I} + \mathrm{A} - 5\mathrm{I} - \mathrm{I} = \mathrm{A} - \mathrm{I} = \begin{pmatrix} 0 & \sqrt{2} \\ -\sqrt{2} & -2 \end{pmatrix}. \end{aligned}$$

We can extend this to other functions by expanding a given $f(x)$ in a power series if all the series we need to use happen to converge. For example, the series for e^z converges for all z, so we can find $e^{k\mathrm{A}}$ when A is a given matrix and k is any number, real or complex. Let A be the matrix in (6.15). Then (Problem 28), we find

$$\text{(6.17)} \qquad \begin{aligned} e^{k\mathrm{A}} &= 1 + k\mathrm{A} + \frac{k^2\mathrm{A}^2}{2!} + \frac{k^3\mathrm{A}^3}{3!} + \frac{k^4\mathrm{A}^4}{4!} + \frac{k^5\mathrm{A}^5}{5!} + \cdots \\ &= (1 - \frac{k^2}{2!} + \frac{k^4}{4!} + \cdots)\mathrm{I} + (k - \frac{k^3}{3!} + \frac{k^5}{5!})\mathrm{A} \\ &= (\cos k)\mathrm{I} + (\sin k)\mathrm{A} = \begin{pmatrix} \cos k + \sin k & \sqrt{2}\sin k \\ -\sqrt{2}\sin k & \cos k - \sin k \end{pmatrix}. \end{aligned}$$

A word of warning about functions of two matrices when A and B don't commute: Familiar formulas may mislead you; see (6.3) and the discussion following it. Be sure to write $(\mathrm{A}+\mathrm{B})^2 = \mathrm{A}^2 + \mathrm{AB} + \mathrm{BA} + \mathrm{B}^2$; don't write 2AB. Similarly, you can show that $e^{\mathrm{A}+\mathrm{B}}$ is not the same as $e^{\mathrm{A}}e^{\mathrm{B}}$ when A and B don't commute (see Problem 29 and Problem 15.34).

► PROBLEMS, SECTION 6

In Problems 1 to 3, find AB, BA, A + B, A − B, A^2, B^2, 5A, 3B. Observe that $\mathrm{AB} \neq \mathrm{BA}$. Show that $(\mathrm{A}-\mathrm{B})(\mathrm{A}+\mathrm{B}) \neq (\mathrm{A}+\mathrm{B})(\mathrm{A}-\mathrm{B}) \neq \mathrm{A}^2 - \mathrm{B}^2$. Show that $\det \mathrm{AB} = \det \mathrm{BA} = (\det \mathrm{A})(\det \mathrm{B})$, but that $\det(\mathrm{A}+\mathrm{B}) \neq \det \mathrm{A} + \det \mathrm{B}$. Show that $\det(5\mathrm{A}) \neq 5 \det \mathrm{A}$, and find n so that $\det(5\mathrm{A}) = 5^n \det \mathrm{A}$. Find similar results for det(3B). Remember that the point of doing these simple problems by hand is to learn how to manipulate determinants and matrices correctly. Check your answers by computer.

1. $\mathrm{A} = \begin{pmatrix} 3 & 1 \\ 2 & 5 \end{pmatrix}, \qquad \mathrm{B} = \begin{pmatrix} -2 & 2 \\ 1 & 4 \end{pmatrix}.$

2. $\mathrm{A} = \begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix}, \qquad \mathrm{B} = \begin{pmatrix} -1 & 4 \\ 0 & 2 \end{pmatrix}.$

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

4. Given the matrices

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

compute or mark as meaningless all products of two of these matrices (AB, BA, A^2, etc.); of three of them (ABC, A^2C, A^3, etc.).

5. Compute the product of each of the matrices in Problem 4 with its transpose [see (2.2) or (9.1)] in both orders, that is AA^T and A^TA, etc.

6. The Pauli spin matrices in quantum mechanics are

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

(You will probably find these called σ_x, σ_y, σ_z in your quantum mechanics texts.) Show that $A^2 = B^2 = C^2 =$ a unit matrix. Also show that any two of these matrices anticommute, that is, AB = −BA, etc. Show that the commutator of A and B, that is, AB − BA, is $2i$C, and similarly for other pairs in cyclic order.

7. Find the matrix product

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

By evaluating this in two ways, verify the associative law for matrix multiplication, that is, A(BC) = (AB)C, which justifies our writing just ABC.

8. Show, by multiplying the matrices, that the following equation represents an ellipse.

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 5 & -7 \\ 7 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 30.$$

9. Find AB and BA given

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}, \quad B = \begin{pmatrix} 10 & 4 \\ -5 & -2 \end{pmatrix}.$$

Observe that AB is the null matrix; if we call it 0, then AB = 0, but neither A nor B is 0. Show that A is singular.

10. Given

$$C = \begin{pmatrix} 7 & 6 \\ 2 & 3 \end{pmatrix}, \quad D = \begin{pmatrix} -3 & 2 \\ 7 & 5 \end{pmatrix}$$

and A as in Problem 9, show that AC = AD, but C ≠ D and A ≠ 0.

11. Show that the unit matrix I has the property that we associate with the number 1, that is, IA = A and AI = A, assuming that the matrices are conformable.

12. For the matrices in Example 3, verify that MM^{-1} and $M^{-1}M$ both equal a unit matrix. Multiply $M^{-1}k$ to verify the solution of equations (6.9).

In Problems 13 to 16, use (6.13) to find the inverse of the given matrix.

13. $\begin{pmatrix} 6 & 9 \\ 3 & 5 \end{pmatrix}$

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

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

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

17. Given the matrices

$$\mathrm{A} = \begin{pmatrix} 1 & -1 & 1 \\ 4 & 0 & -1 \\ 4 & -2 & 0 \end{pmatrix}, \quad \mathrm{B} = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 2 & 1 & 2 \end{pmatrix}.$$

(a) Find A^{-1}, B^{-1}, $\mathrm{B}^{-1}\mathrm{AB}$, and $\mathrm{B}^{-1}\mathrm{A}^{-1}\mathrm{B}$.

(b) Show that the last two matrices are inverses, that is, that their product is the unit matrix.

18. Problem 17(b) is a special case of the general theorem that the inverse of a product of matrices is the product of the inverses in reverse order. Prove this. *Hint:* Multiply ABCD times $\mathrm{D}^{-1}\mathrm{C}^{-1}\mathrm{B}^{-1}\mathrm{A}^{-1}$ to show that you get a unit matrix.

In Problems 19 to 22, solve each set of equations by the method of finding the inverse of the coefficient matrix. *Hint:* See Example 3.

19. $\begin{cases} x - 2y = 5 \\ 3x + y = 15 \end{cases}$

20. $\begin{cases} 2x + 3y = -1 \\ 5x + 4y = 8 \end{cases}$

21. $\begin{cases} x + 2z = 8 \\ 2x - y = -5 \\ x + y + z = 4 \end{cases}$

22. $\begin{cases} x - y + z = 4 \\ 2x + y - z = -1 \\ 3x + 2y + 2z = 5 \end{cases}$

23. Verify formula (6.13). *Hint:* Consider the product of the matrices MC^{T}. Use Problem 3.8.

24. Use the method of solving simultaneous equations by finding the inverse of the matrix of coefficients, together with the formula (6.13) for the inverse of a matrix, to obtain Cramer's rule.

25. Verify (6.14) by multiplying the matrices and using trigonometric addition formulas.

26. In (6.14), let $\theta = \phi = \pi/2$ and verify the result numerically.

27. Do Problem 26 if $\theta = \pi/2$, $\phi = \pi/4$.

28. Verify the calculations in (6.15), (6.16), and (6.17).

29. Show that if A and B are matrices which don't commute, then $e^{\mathrm{A}+\mathrm{B}} \neq e^{\mathrm{A}}e^{\mathrm{B}}$, but if they do commute then the relation holds. *Hint:* Write out several terms of the infinite series for e^{A}, e^{B}, and $e^{\mathrm{A}+\mathrm{B}}$ and do the multiplications carefully assuming that A and B don't commute. Then see what happens if they do commute.

30. For the Pauli spin matrix A in Problem 6, find the matrices $\sin k\mathrm{A}$, $\cos k\mathrm{A}$, $e^{k\mathrm{A}}$, and $e^{ik\mathrm{A}}$ where $i = \sqrt{-1}$.

31. Repeat Problem 30 for the Pauli spin matrix C in Problem 6. *Hint*: Show that if a matrix is diagonal, say $\mathrm{D} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$, then $f(\mathrm{D}) = \begin{pmatrix} f(a) & 0 \\ 0 & f(b) \end{pmatrix}$.

32. For the Pauli spin matrix B in Problem 6, find $e^{i\theta\mathrm{B}}$ and show that your result is a rotation matrix. Repeat the calculation for $e^{-i\theta\mathrm{B}}$.

7. LINEAR COMBINATIONS, LINEAR FUNCTIONS, LINEAR OPERATORS

Given two vectors $\mathbf{A}$ and $\mathbf{B}$, the vector $3\mathbf{A} - 2\mathbf{B}$ is called a "linear combination" of $\mathbf{A}$ and $\mathbf{B}$. In general, a linear combination of $\mathbf{A}$ and $\mathbf{B}$ means $a\mathbf{A} + b\mathbf{B}$ where a and b are scalars. Geometrically, if $\mathbf{A}$ and $\mathbf{B}$ have the same tail and do not lie along a line, then they determine a plane. You should satisfy yourself that all linear combinations of $\mathbf{A}$ and $\mathbf{B}$ then lie in the plane. It is also true that every vector in the plane can be written as a linear combination of $\mathbf{A}$ and $\mathbf{B}$; we shall consider this in Section 8. The vector $\mathbf{r} = \mathbf{i}x + \mathbf{j}y + \mathbf{k}z$ with tail at the origin (which we used in writing equations of lines and planes) is a linear combination of the unit basis vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$.

A function of a vector, say $f(\mathbf{r})$, is called linear if

(7.1) $$f(\mathbf{r}_1 + \mathbf{r}_2) = f(\mathbf{r}_1) + f(\mathbf{r}_2), \quad \text{and} \quad f(a\mathbf{r}) = af(\mathbf{r}),$$

where a is a scalar.

For example, if $\mathbf{A} = 2\mathbf{i} + 3\mathbf{j} - \mathbf{k}$ is a given vector, then $f(\mathbf{r}) = \mathbf{A} \cdot \mathbf{r} = 2x + 3y - z$ is a linear function because

$$f(\mathbf{r}_1 + \mathbf{r}_2) = \mathbf{A} \cdot (\mathbf{r}_1 + \mathbf{r}_2) = \mathbf{A} \cdot \mathbf{r}_1 + \mathbf{A} \cdot \mathbf{r}_2 = f(\mathbf{r}_1) + f(\mathbf{r}_2), \quad \text{and}$$

$$f(a\mathbf{r}) = \mathbf{A} \cdot (a\mathbf{r}) = a\mathbf{A} \cdot \mathbf{r} = af(\mathbf{r}).$$

On the other hand, $f(\mathbf{r}) = |\mathbf{r}|$ is *not* a linear function, because the length of the sum of two vectors is not in general the sum of their lengths. That is,

$$f(\mathbf{r}_1 + \mathbf{r}_2) = |\mathbf{r}_1 + \mathbf{r}_2| \neq |\mathbf{r}_1| + |\mathbf{r}_2| = f(\mathbf{r}_1) + f(\mathbf{r}_2),$$

as you can see from Figure 7.1. Also note that although we call $y = mx + b$ a linear equation (it is the equation of a straight line), the function $f(x) = mx + b$ is not linear (unless $b = 0$) because

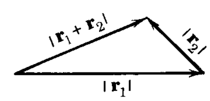


Figure 7.1

$$f(x_1 + x_2) = m(x_1 + x_2) + b \neq (mx_1 + b) + (mx_2 + b) = f(x_1) + f(x_2).$$

We can also consider vector functions of a vector $\mathbf{r}$. The magnetic field at each point (x, y, z), that is, at the head of the vector $\mathbf{r}$, is a vector $\mathbf{B} = \mathbf{i}B_x + \mathbf{j}B_y + \mathbf{k}B_z$. The components B_x, B_y, B_z may vary from point to pint, that is, they are functions of (x, y, z) or $\mathbf{r}$. Then

$\mathbf{F}(\mathbf{r})$ is a linear vector function if

(7.2) $$\mathbf{F}(\mathbf{r}_1 + \mathbf{r}_2) = \mathbf{F}(\mathbf{r}_1) + \mathbf{F}(\mathbf{r}_2) \quad \text{and} \quad \mathbf{F}(a\mathbf{r}) = a\mathbf{F}(\mathbf{r}),$$

where a is a scalar.

For example, $\mathbf{F}(\mathbf{r}) = b\mathbf{r}$ (where b is a scalar) is a linear vector function of $\mathbf{r}$.

You know from calculus that

$$\frac{d}{dx}[f(x)+g(x)] = \frac{d}{dx}f(x) + \frac{d}{dx}g(x) \quad \text{and} \qquad (7.3)$$
$$\frac{d}{dx}[kf(x)] = k\frac{d}{dx}f(x),$$

where k is a constant. We say that d/dx is a "linear operator" [compare (7.3) with (7.1) and (7.2)]. An "operator" or "operation" simply means a rule or some kind of instruction telling us what to do with whatever follows it. In other words, a linear operator is a linear function. Then

O is a linear operator if

$$O(A+B) = O(A) + O(B) \quad \text{and} \quad O(kA) = kO(A), \qquad (7.4)$$

where k is a number, and A and B are numbers, functions, vectors, and so on. Many of the errors people make happen because they assume that operators are linear when they are not (see problems).

▶ **Example 1.** Is square root a linear operator? We are asking, is $\sqrt{A+B}$ the same as $\sqrt{A}+\sqrt{B}$? The answer is no; taking the square root is not a linear operation.

▶ **Example 2.** Is taking the complex conjugate a linear operation? We want to know whether $\overline{A+B} = \bar{A}+\bar{B}$ and $\overline{kA} = k\bar{A}$. The first equation is true; the second equation is true if we restrict k to real numbers.

Matrix Operators, Linear Transformations Consider the set of equations

$$\begin{cases} X = ax + by, \\ Y = cx + dy, \end{cases} \quad \text{or} \quad \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}, \quad \text{or} \quad \mathrm{R} = \mathrm{Mr}, \qquad (7.5)$$

where a, b, c, d, are constants. For every point (x, y), these equations give us a point (X, Y). If we think of each point of the (x, y) plane being moved to some other point (with some points like the origin not being moved), we can call this process a *mapping* or *transformation* of the plane into itself. All the information about this transformation is contained in the matrix M. We say that this matrix is an operator which maps the plane into itself. Any matrix can be thought of as an operator on (conformable) column matrices r. Since

$$\mathrm{M}(\mathrm{r}_1+\mathrm{r}_2) = \mathrm{Mr}_1 + \mathrm{Mr}_2 \quad \text{and} \quad \mathrm{M}(k\mathrm{r}) = k(\mathrm{Mr}), \qquad (7.6)$$

the matrix M is a linear operator.

Equations (7.5) can be interpreted geometrically in two ways. In Figure 7.2, we have one set of coordinate axes and the vector **r** has been changed to the vector **R** by the transformation (7.5). In Figure 7.3, we have *two* sets of coordinate axes,

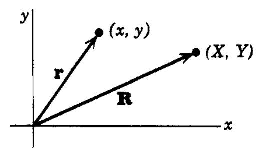


Figure 7.2

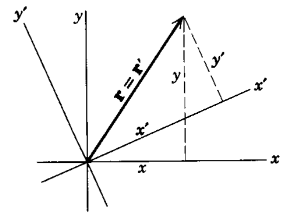


Figure 7.3

(x, y) and (x', y'), and *one* vector $\mathbf{r} = \mathbf{r}'$ with coordinates relative to each set of axes. This time the transformation

$$(7.7) \qquad \begin{cases} x' = ax + by, \\ y' = cx + dy, \end{cases} \quad \text{or} \quad \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad \text{or} \quad \mathrm{r}' = \mathrm{Mr},$$

tells us how to get the components of the vector $\mathbf{r} = \mathbf{r}'$ relative to axes (x', y') when we know its components relative to axes (x, y).

Orthogonal Transformations We shall be particularly interested in the special case of a linear transformation which preserves the length of a vector. We call (7.7) an *orthogonal transformation* if

$$(7.8) \qquad x'^2 + y'^2 = x^2 + y^2,$$

and similarly for (7.5). You can see from the figures that this requirement says that the length of a vector is not changed by an orthogonal transformation. In Figure 7.2, the vector would be rotated (or perhaps reflected) with its length held fixed (that is $R = r$ for an orthogonal transformation). In Figure 7.3, the axes are rotated (or reflected), while the vector stays fixed. The matrix M of an orthogonal transformation is called an *orthogonal matrix*. Let's show that the inverse of an orthogonal matrix equals its transpose; in symbols

$$(7.9) \qquad \mathrm{M}^{-1} = \mathrm{M}^{\mathrm{T}}, \quad \mathrm{M} \text{ orthogonal}.$$

From (7.8) and (7.7) we have

$$\begin{aligned} x'^2 + y'^2 &= (ax + by)^2 + (cx + dy)^2 \\ &= (a^2 + c^2)x^2 + 2(ab + cd)xy + (b^2 + d^2)y^2 \equiv x^2 + y^2. \end{aligned}$$

Thus we must have $a^2 + c^2 = 1$, $b^2 + d^2 = 1$, $ab + cd = 0$. Then

$$(7.10) \qquad \begin{aligned} \mathrm{M}^{\mathrm{T}}\mathrm{M} &= \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= \begin{pmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

Since $\mathrm{M^TM}$ is the unit matrix, M and $\mathrm{M^T}$ are inverse matrices as we claimed in (7.9). We have defined an orthogonal transformation in two dimensions and we have proved (7.9) for the 2-dimensional case. However, a square matrix of any order is called orthogonal if it satisfies (7.9), and you can easily show that the corresponding transformation preserves the lengths of vectors (Problem 9.24).

Now if we write (7.9) as $\mathrm{M^TM = I}$ and use the facts from Section 3 that $\det(\mathrm{M^TM}) = (\det \mathrm{M^T})(\det \mathrm{M})$ and $\det \mathrm{M^T} = \det \mathrm{M}$, we have $(\det \mathrm{M})^2 = \det(\mathrm{M^TM}) = \det \mathrm{I} = 1$, so

$$\det \mathrm{M} = \pm 1, \quad \text{M orthogonal.} \tag{7.11}$$

This is true for M of any order since we have used only the definition (7.9) of an orthogonal matrix and some properties of determinants. As we shall see, $\det \mathrm{M} = 1$ corresponds geometrically to a rotation, and $\det \mathrm{M} = -1$ means that a reflection is involved.

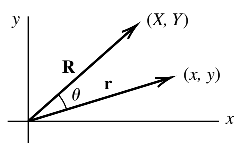


Figure 7.4

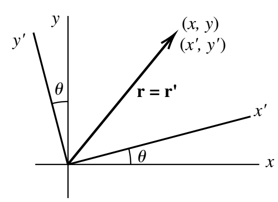


Figure 7.5

Rotations in 2 Dimensions In Figure 7.4, we have sketched the vector $\mathbf{r} = (x, y)$, and the vector $\mathbf{R} = (X, Y)$ which is the vector $\mathbf{r}$ rotated by angle θ. We write in matrix form the equations relating the components of $\mathbf{r}$ and $\mathbf{R}$ (Problem 19).

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad \text{vector rotated.} \tag{7.12}$$

In Figure 7.5, we have sketched two sets of axes with the primed axes rotated by angle θ with respect to the unprimed axes. The vector $\mathbf{r} = (x, y)$, and the vector $\mathbf{r}' = (x', y')$ are the same vector, but with components relative to different axes. These components are related by the equations (Problem 20).

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad \text{axes rotated.} \tag{7.13}$$

Both equations (7.12) and equations (7.13) are referred to as "rotation equations" and the θ matrices are called "rotation matrices". To distinguish them, we refer to the rotation (7.12) as an "active" transformation (vectors rotated), and to (7.13) as a "passive" transformation (vectors not moved but their components changed because the axes are rotated). Equations (7.7) or (7.13) are also referred to as a "change of basis". (Remember that we called $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ unit basis vectors; here we have changed from the $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ basis to the $\mathbf{i}'$, $\mathbf{j}'$, $\mathbf{k}'$ basis. Also see Section 10.) Observe that the matrices in (7.12) and (7.13) are inverses of each other. You can see from the figures why this must be so. The rotation of a vector in, say, the counterclockwise direction produces the same result as the rotation of the axes in the opposite (clockwise) direction.

We note that $\det M = \cos^2\theta + \sin^2\theta = 1$ for a rotation matrix. Any 2 by 2 orthogonal matrix with determinant 1 corresponds to a rotation, and any 2 by 2 orthogonal matrix with determinant $= -1$ corresponds to a reflection through a line.

▶ **Example 3.** Find what transformation corresponds to each of the following matrices.

$$\mathrm{A} = \frac{1}{2}\begin{pmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{pmatrix}, \quad \mathrm{B} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathrm{C} = \mathrm{AB}, \quad \mathrm{D} = \mathrm{BA}. \tag{7.14}$$

First we can show that all these matrices are orthogonal, and that $\det \mathrm{A} = 1$, but the determinants of the other three are -1 (Problem 21). Thus A is a rotation and B, C and D are reflections. Let's view these as active transformations (fixed axes, vectors rotated or reflected). Then by comparing A with (7.12), we have $\cos\theta = -1/2$, $\sin\theta = -\frac{1}{2}\sqrt{3}$, so this is a rotation of 240° (or −120°). Alternatively, we could ask what happens to the vector $\mathbf{i}$. We multiply matrix A times the column matrix $\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ and get

$$\frac{1}{2}\begin{pmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} -1 \\ -\sqrt{3} \end{pmatrix} \quad \text{or} \quad -\frac{1}{2}(\mathbf{i} + \mathbf{j}\sqrt{3}),$$

which is $\mathbf{i}$ rotated by 240° as we had before.

Now B operating on $\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$ leaves x fixed and changes the sign of y (check this); that is, B corresponds to a reflection through the x axis.

We find $\mathrm{C} = \mathrm{AB}$ and $\mathrm{D} = \mathrm{BA}$ by multiplying the matrices (Problem 21).

$$\mathrm{C} = \mathrm{AB} = \frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ -\sqrt{3} & 1 \end{pmatrix}, \quad \mathrm{D} = \mathrm{BA} = \frac{1}{2}\begin{pmatrix} -1 & \sqrt{3} \\ \sqrt{3} & 1 \end{pmatrix}. \tag{7.15}$$

We know that these are reflections since they have determinant $= -1$. To find the line through which the plane is reflected, we realize that the vectors along that line are unchanged by the reflection, so we want to find x and y, that is vector $\mathbf{r}$, which is mapped to itself by the transformation. For matrix C we write $\mathrm{Cr} = \mathrm{r}$.

$$\frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ -\sqrt{3} & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}. \tag{7.16}$$

You can verify (Problem 21) that the two equations in (7.16) are really the same equation, namely $y = -x\sqrt{3}$. Vectors along this line, say $\mathbf{i} - \mathbf{j}\sqrt{3}$, are not changed by the reflection [see (7.17)] so this is the reflection line. As further verification we can show [see (7.17)] that a vector perpendicular to this line, say $\mathbf{i}\sqrt{3} + \mathbf{j}$, is changed into its negative, that is, it is reflected through the line.

$$\begin{aligned} \frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ -\sqrt{3} & 1 \end{pmatrix}\begin{pmatrix} 1 \\ -\sqrt{3} \end{pmatrix} &= \begin{pmatrix} 1 \\ -\sqrt{3} \end{pmatrix}, \\ \frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ -\sqrt{3} & 1 \end{pmatrix}\begin{pmatrix} \sqrt{3} \\ 1 \end{pmatrix} &= \begin{pmatrix} -\sqrt{3} \\ -1 \end{pmatrix}. \end{aligned} \tag{7.17}$$

Comment: The solution of the equation $\mathrm{Cr} = \mathrm{r}$ is an example of an eigenvalue, eigenvector problem. We shall discuss such problems in detail in Section 11.

We can analyze the transformation D in the same way we did C to find (Problem 21) that the reflection line is $y = x\sqrt{3}$. Note that matrices A and B do not commute and the transformations C and D are different.

Rotations and Reflections in 3 Dimensions Let's consider 3 by 3 orthogonal matrices as active transformations rotating or reflecting vectors $\mathbf{r} = (x, y, z)$. A simple form for a rotation matrix is

$$A = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{7.18}$$

You should satisfy yourself that this transformation produces a rotation of vectors about the z axis through angle θ. We can then find the rotation angle from (7.12) as we did in 2 dimensions. Similarly the matrix

$$B = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{7.19}$$

produces a rotation about the z axis of angle θ together with a reflection through the (x, y) plane, and again we can find the rotation angle as in 2 dimensions.

We will show in Section 11 that any 3 by 3 orthogonal matrix with determinant $= 1$ can be written in the form (7.18) by choosing the z axis as the rotation axis, and any 3 by 3 orthogonal matrix with determinant $= -1$ can be written in the form (7.19). For now, let's look at a few simple problems we can do just by considering how the matrix maps certain vectors.

Example 4. The matrix for a rotation about the y axis is

$$F = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}. \tag{7.20}$$

You should satisfy yourself that the entry $-\sin\theta$ is in the right place for an active transformation. Let $\theta = 90°$; then the matrix F in (7.20) maps the vector $\mathbf{i} = (1, 0, 0)$ to the vector $-\mathbf{k} = (0, 0, -1)$; this is correct for a $90°$ rotation around the y axis. Check that $(0, 0, 1)$ is mapped to $(1, 0, 0)$.

Example 5. Find the mappings produced by the matrices

$$G = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad K = \begin{pmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}. \tag{7.21}$$

First we find that the determinants are 1 so these are rotations. For G, either by inspection or by solving Gr = r as in (7.16), we find that the vector $(1, 0, 1)$ is unchanged and so $\mathbf{i} + \mathbf{k}$ is the rotation axis. Now G^2 is the identity matrix (corresponding to a $360°$ rotation); thus the rotation angle for G is $180°$.

Similarly for K, we find that the vector $(1, -1, 1)$ is unchanged by the transformation so $\mathbf{i} - \mathbf{j} + \mathbf{k}$ is the rotation axis. Now verify that K maps $\mathbf{i}$ to $-\mathbf{j}$, and $-\mathbf{j}$ to $\mathbf{k}$, and $\mathbf{k}$ to $\mathbf{i}$ (or, alternatively that K^3 is the identity matrix) so the rotation angle for K^3 is $\pm 360°$. From the geometry we see that the rotation $\mathbf{i} \to -\mathbf{j} \to \mathbf{k} \to \mathbf{i}$ is a rotation of $-120°$ about $\mathbf{i} - \mathbf{j} + \mathbf{k}$. (Also see Section 11.)

▶ **Example 6.** Find the mapping produced by the matrix

$$\mathrm{L} = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Since $\det \mathrm{L} = -1$, this is a reflection through some plane. The vector perpendicular to the reflection plane is reversed by the reflection, so we ask for a vector satisfying $\mathrm{Lr} = -\mathrm{r}$. Either by solving these equations or by inspection we find $\mathbf{r} = (1, 1, 0) = \mathbf{i} + \mathbf{j}$. The reflecting plane is the plane through the origin perpendicular to this vector, that is, the plane $x + y = 0$ (see Section 5).

▶ PROBLEMS, SECTION 7

Are the following linear functions? Prove your conclusions by showing that $f(\mathbf{r})$ satisfies both of the equations (7.1) or that it does not satisfy at least one of them.

1. $f(\mathbf{r}) = \mathbf{A} \cdot \mathbf{r} + 3$, where $\mathbf{A}$ is a given vector.

2. $f(\mathbf{r}) = \mathbf{A} \cdot (\mathbf{r} - \mathbf{k}z)$.

3. $\mathbf{r} \cdot \mathbf{r}$.

Are the following linear vector functions? Prove your conclusions using (7.2).

4. $\mathbf{F}(\mathbf{r}) = \mathbf{r} - \mathbf{i}x = \mathbf{j}y + \mathbf{k}z$.

5. $\mathbf{F}(\mathbf{r}) = \mathbf{A} \times \mathbf{r}$, where $\mathbf{A}$ is a given vector.

6. $\mathbf{F}(\mathbf{r}) = \mathbf{r} + \mathbf{A}$, where $\mathbf{A}$ is a given vector.

Are the following operators linear?

7. Definite integral with respect to x from 0 to 1; the objects being operated on are functions of x.

8. Find the logarithm; operate on positive real numbers.

9. Find the square; operate on numbers or on functions.

10. Find the reciprocal; operate on numbers or on functions.

11. Find the absolute value; operate on complex numbers.

12. Let D stand for $\dfrac{d}{dx}$, D^2 for $\dfrac{d^2}{dx^2}$, $D^3 = \dfrac{d^3}{dx^3}$, and so on. Are D, D^2, D^3 linear? Operate on functions of x which can be differentiated as many times as needed.

13. (a) As in Problem 12, is $D^2 + 2D + 1$ linear?

(b) Is $x^2D^2 - 2xD + 7$ a linear operator?

14. Find the maximum; operate on functions of x.

15. Find the transpose; operate on matrices.

16. Find the inverse; operate on square matrices.

17. Find the determinant; operate on square matrices.

18. With the cross product of two vectors defined by (4.14), show that finding the cross product is a linear operation, that is, show that (4.18) is valid. *Warning hint:* Don't try to prove it by writing out components: Writing, for example, $\mathbf{i}A_x \times (\mathbf{j}B_y + \mathbf{k}B_z) = \mathbf{i}A_x \times \mathbf{j}B_y + \mathbf{i}A_x \times \mathbf{k}B_z$ would be assuming what you're trying to prove. *Further hints:* First show that (4.18) is valid if $\mathbf{B}$ and $\mathbf{C}$ are both perpendicular to $\mathbf{A}$ by sketching (in the plane perpendicular to $\mathbf{A}$) the vectors $\mathbf{B}$, $\mathbf{C}$, $\mathbf{B}+\mathbf{C}$, and their vector products with $\mathbf{A}$. Then do the general case by first showing that $\mathbf{A}\times\mathbf{B}$ and $\mathbf{A}\times\mathbf{B}_\perp$ (where $\mathbf{B}_\perp$ is the vector component of $\mathbf{B}$ perpendicular to $\mathbf{A}$) have the same magnitude and the same direction.

19. If we multiply a complex number $z = re^{i\phi}$ by $e^{i\theta}$, we get $e^{i\theta}z = re^{i(\phi+\theta)}$, that is, a complex number with the same r but with its angle increased by θ. We can say that the vector $\mathbf{r}$ from the origin to the point $z = x + iy$ has been rotated by angle θ as in Figure 7.4 to become the vector $\mathbf{R}$ from the origin to the point $Z = X + iY$. Then we can write $X + iY = e^{i\theta}z = e^{i\theta}(x + iy)$. Take real and imaginary parts of this equation to obtain equations (7.12).

20. Verify equations (7.13) using Figure 7.5. *Hints*: Write $\mathbf{r}' = \mathbf{r}$ as $\mathbf{i}'x' + \mathbf{j}'y' = \mathbf{i}x + \mathbf{j}y$ and take the dot product of this equation with $\mathbf{i}'$ and with $\mathbf{j}'$ to get x' and y'. Evaluate the dot products of the unit vectors in terms of θ using Figure 7.5. For example, $\mathbf{i}'\cdot\mathbf{j}$ is the cosine of the angle between the x' axis and the y axis.

21. Do the details of Example 3 as follows:

(a) Verify that the four matrices in (7.14) are all orthogonal and verify the stated values of their determinants.

(b) Verify the products C = AB and D = BA in (7.15).

(c) Solve (7.16) to find the reflection line.

(d) Analyze the transformation D as we did C.

Let each of the following matrices represent an active transformation of vectors in the (x, y) plane (axes fixed, vectors rotated or reflected). As in Example 3, show that each matrix is orthogonal, find its determinant, and find the rotation angle, or find the line of reflection.

22. $\dfrac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$

23. $\dfrac{1}{2}\begin{pmatrix} -\sqrt{3} & 1 \\ -1 & -\sqrt{3} \end{pmatrix}$

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

25. $\dfrac{1}{3}\begin{pmatrix} -1 & 2\sqrt{2} \\ 2\sqrt{2} & 1 \end{pmatrix}$

26. $\dfrac{1}{5}\begin{pmatrix} 3 & 4 \\ 4 & -3 \end{pmatrix}$

27. $\dfrac{1}{\sqrt{2}}\begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix}$

28. Write the matrices which produce a rotation θ about the x axis, or that rotation combined with a reflection through the (y, z) plane. [Compare (7.18) and (7.19) for rotation about the z axis.]

29. Construct the matrix corresponding to a rotation of 90° about the y axis together with a reflection through the (x, z) plane.

30. For the matrices G and K in (7.21), find the matrices R = GK and S = KG. Note that R $\neq S$. (In 3 dimensions, rotations about two different axes do not in general commute.) Find what geometric transformations are produced by R and S.

31. To see a physical example of non-commuting rotations, do the following experiment. Put a book on your desk and imagine a set of rectangular axes with the x and y axes in the plane of the desk with the z axis vertical. Place the book in the first quadrant with the x and y axes along the edges of the book. Rotate the book 90° about the x axis and then 90° about the z axis; note its position. Now repeat the experiment, this time rotating 90° about the z axis first, and then 90° about the x axis; note the different result. Write the matrices representing the 90° rotations and multiply them in both orders. In each case, find the axis and angle of rotation.

For each of the following matrices, find its determinant to see whether it produces a rotation or a reflection. If a rotation, find the axis and angle of rotation. If a reflection, find the reflecting plane and the rotation (if any) about the normal to this plane.

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

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

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

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

8. LINEAR DEPENDENCE AND INDEPENDENCE

We say that the three vectors $\mathbf{A} = \mathbf{i} + \mathbf{j}$, $\mathbf{B} = \mathbf{i} + \mathbf{k}$, and $\mathbf{C} = 2\mathbf{i} + \mathbf{j} + \mathbf{k}$ are *linearly dependent* because $\mathbf{A} + \mathbf{B} - \mathbf{C} = \mathbf{0}$. The two vectors $\mathbf{i}$ and $\mathbf{j}$ are *linearly independent* because there are no numbers a and b (not *both* zero) such that the linear combination $a\mathbf{i} + b\mathbf{j}$ is zero. In general, a set of vectors is linearly dependent if some linear combination of them is zero (with not *all* the coefficients equal to zero). In the simple examples above, it was easy to see by inspection whether the vectors were linearly independent or not. In more complicated cases, we need a method of determining linear dependence. Consider the set of vectors

$$(1, 4, -5), (5, 2, 1), (2, -1, 3), \text{ and } (3, -6, 11); \tag{8.1}$$

We want to know whether they are linearly dependent, and if so, we want to find a smaller linearly independent set. Let us row reduce the matrix whose rows are the given vectors (see Section 2):

$$\begin{pmatrix} 1 & 4 & -5 \\ 5 & 2 & 1 \\ 2 & -1 & 3 \\ 3 & -6 & 11 \end{pmatrix} \rightarrow \begin{pmatrix} 9 & 0 & 7 \\ 0 & -9 & 13 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{8.2}$$

In row reduction, we are forming linear combinations of the rows by elementary row operations [see (2.8)]. All these operations are reversible, so we could, if we liked, reverse our calculations and combine the two vectors $(9, 0, 7)$ and $(0, -9, 13)$ to obtain each of the four original vectors (Problem 1). Thus there are only two independent vectors in (8.1); we refer to these independent vectors as basis vectors since all the original vectors can be written in terms of them (see Section 10). Note that the rank (see Section 2) of the matrix in (8.2) is equal to the number of independent or basis vectors.

Linear Independence of Functions By a definition similar to that for vectors, we say that the functions $f_1(x)$, $f_2(x)$, $\cdots$, $f_n(x)$ are linearly dependent if some linear combination of them is identically zero, that is, if there are constants k_1, k_2, $\cdots$, k_n, not all zero, such that

$$k_1 f_1(x) + k_2 f_2(x) + \cdots + k_n f_n(x) \equiv 0. \tag{8.3}$$

For example, $\sin^2 x$ and $(1 - \cos^2 x)$ are linearly dependent since

$$\sin^2 x - (1 - \cos^2 x) \equiv 0.$$

But $\sin x$ and $\cos x$ are linearly independent since there are no numbers k_1 and k_2, not both zero, such that

$$k_1 \sin x + k_2 \cos x \tag{8.4}$$

is zero for *all* x (Problem 8).

We shall be particularly interested in knowing that a given set of functions is linearly independent. For this purpose the following theorem is useful (Problems 8 to 16, and Chapter 8, Section 5).

If $f_1(x)$, $f_2(x)$, $\cdots$, $f_n(x)$ have derivatives of order $n-1$, and if the determinant

$$W = \begin{vmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ f_1''(x) & f_2''(x) & \cdots & f_n''(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{vmatrix} \not\equiv 0, \tag{8.5}$$

then the functions are linearly independent. (See Problem 16.) The determinant W is called the *Wronskian* of the functions.

▶ **Example 1.** Using (8.5), show that the functions 1, x, $\sin x$ are linearly independent.

We write and evaluate the Wronskian,

$$W = \begin{vmatrix} 1 & x & \sin x \\ 0 & 1 & \cos x \\ 0 & 0 & -\sin x \end{vmatrix} = -\sin x.$$

Since $-\sin x$ is not identically equal to zero, the functions are linearly independent.

▶ **Example 2.** Now let's compute the Wronskian for a case when the functions are linearly dependent.

$$W = \begin{vmatrix} x & \sin x & 2x - 3\sin x \\ 1 & \cos x & 2 - 3\cos x \\ 0 & -\sin x & 3\sin x \end{vmatrix} = \begin{vmatrix} x & \sin x & 2x \\ 1 & \cos x & 2 \\ 0 & -\sin x & 0 \end{vmatrix} = (\sin x)(2x - 2x) \equiv 0,$$

as we expected. However, note that "functions dependent" implies $W \equiv 0$, but $W \equiv 0$ does not necessarily imply "functions dependent". (See Problem 16.)

Homogeneous Equations In Section 2 we considered sets of linear equations. Here we want to consider the special case of such equations when the constants on the right hand sides are all zero; these are called homogeneous equations. We write the homogeneous equations corresponding to (2.12) and (2.13) together with the row reduced matrices:

$$\text{(8.6)} \qquad \begin{cases} x + y = 0 \\ x - y = 0 \end{cases} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\text{(8.7)} \qquad \begin{cases} \;\;x + \;\;y = 0 \\ 2x + 2y = 0 \end{cases} \qquad \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We can draw several conclusions from these examples. Note that in (8.6) the only solution is $x = y = 0$; the rank of the matrix is 2, the same as the number of unknowns. In (8.7), the rank of the matrix is 1; this is less than the number of unknowns. This reflects what we could see in (8.7), that we really have just one equation in two unknowns; all the points on a line satisfy $x + y = 0$. In (8.8) we summarize the facts for homogeneous equations:

(8.8) Homogeneous equations are never inconsistent; they always have the solution "all unknowns = 0" (often called the "trivial solution"). If the number of independent equations (that is, the rank of the matrix) is the same as the number of unknowns, this is the only solution. If the rank of the matrix is less than the number of unknowns, there are infinitely many solutions.

A very important special case is a set of n homogeneous equations in n unknowns. By (8.8), these equations have only the trivial solution unless the rank of the matrix is less than n. This means that at least one row of the row reduced n by n matrix of the coefficients is a zero row. But then the determinant D of the coefficients is zero. Thus we have an important result (see Problems 21 to 25; also see Section 11):

(8.9) A system of n homogeneous equations in n unknowns has solutions other than the trivial solution if and only if the determinant of the coefficients is zero.

Solutions in Vector Form Geometrically, solutions of sets of linear equations may be points or lines or planes.

▶ **Example 3.** In Section 2, Example 4, we solved equations (2.15):

$$\text{(8.10)} \qquad x = 3 + 2z, \qquad y = 4 - z.$$

This solution set consists of all points on the line which is the intersection of these two planes. An interesting way to write the solution is the vector form

$$\text{(8.11)} \qquad \mathbf{r} = (x, y, z) = (3 + 2z, 4 - z, z) = (3, 4, 0) + (2, -1, 1)z.$$

If we put $z = t$, this is the parametric form of the equations of a straight line, $\mathbf{r} = \mathbf{r}_0 + \mathbf{A}t$ [see (5.8)].

Now let's consider the homogeneous equations (zero right hand sides) corresponding to equations (2.15). The equations and the row reduced matrix are:

$$\begin{pmatrix} 1 & 1 & -1 \\ 2 & -1 & -5 \\ -5 & 4 & 14 \\ 3 & -1 & -7 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{8.12}$$

so the solutions are

$$x = 2z, \quad y = -z, \quad \text{or} \quad \mathbf{r} = (2, -1, 1)z. \tag{8.13}$$

Comparing (8.11) and (8.13), we see that the solution of the homogeneous equations $\mathrm{Mr} = 0$ is a straight line through the origin; the solution of the equations $\mathrm{Mr} = \mathrm{k}$ is a parallel straight line through the point $(3, 4, 0)$. We could say that the solution of $\mathrm{Mr} = \mathrm{k}$ is the solution of the corresponding homogeneous equations plus the particular solution $\mathbf{r} = (3, 4, 0)$.

Here is an example of an important use of (8.9).

Example 4. For what values of λ does the following set of equations have nontrivial solutions for x and y? For each value of λ find the corresponding relation between x and y. This is an example of an *eigenvalue* problem; we shall discuss such problems in detail in Sections 11 and 12. The values of λ are called eigenvalues and the corresponding vectors (x, y) are called eigenvectors.

$$\begin{cases} (1-\lambda)x + 2y = 0, \\ 2x + (4-\lambda)y = 0. \end{cases} \tag{8.14}$$

By (8.9), we set the determinant M of the coefficients equal to zero. Then we solve for λ, and for each value of λ we solve for x and y.

$$\begin{vmatrix} 1-\lambda & 2 \\ 2 & 4-\lambda \end{vmatrix} = \lambda^2 - 5\lambda + 4 - 4 = \lambda(\lambda - 5) = 0, \quad \lambda = 0, 5.$$

For $\lambda = 0$, we find $x + 2y = 0$. For $\lambda = 5$, we find $2x - y = 0$. In vector notation the eigenvectors are: For $\lambda = 0$, $\mathbf{r} = (2, -1)s$, and for $\lambda = 5$, $\mathbf{r} = (1, 2)t$, where s and t are parameters in these vector equations of straight lines through the origin.

PROBLEMS, SECTION 8

1. Write each of the vectors (8.1) as a linear combination of the vectors $(9, 0, 7)$ and $(0, -9, 13)$. *Hint:* To get the right x component in $(1, 4, -5)$, you have to use $(1/9)(9, 0, 7)$. How do you get the right y component? Is the z component now correct?

In Problems 2 to 4, find out whether the given vectors are dependent or independent; if they are dependent, find a linearly independent subset. Write each of the given vectors as a linear combination of the independent vectors.

2. $(1, -2, 3)$, $(1, 1, 1)$, $(-2, 1, -4)$, $(3, 0, 5)$

3. $(0,1,1),\ (-1,5,3),\ (1,0,2),\ (2,-15,1)$

4. $(3,5,-1),\ (1,4,2),\ (-1,0,5),\ (6,14,5)$

5. Show that any vector $\mathbf{V}$ in a plane can be written as a linear combination of two non-parallel vectors $\mathbf{A}$ and $\mathbf{B}$ in the plane; that is, find a and b so that $\mathbf{V} = a\mathbf{A}+b\mathbf{B}$. *Hint:* Find the cross products $\mathbf{A}\times\mathbf{V}$ and $\mathbf{B}\times\mathbf{V}$; what are $\mathbf{A}\times\mathbf{A}$ and $\mathbf{B}\times\mathbf{B}$? Take components perpendicular to the plane to show that

$$a = \frac{(\mathbf{B}\times\mathbf{V})\cdot\mathbf{n}}{(\mathbf{B}\times\mathbf{A})\cdot\mathbf{n}}$$

where $\mathbf{n}$ is normal to the plane, and a similar formula for b.

6. Use Problem 5 to write $\mathbf{V} = 3\mathbf{i}+5\mathbf{j}$ as a linear combination of $\mathbf{A} = 2\mathbf{i}+\mathbf{j}$ and $\mathbf{B} = 3\mathbf{i}-2\mathbf{j}$. Show that the formulas in Problem 5, written as a quotient of 2 by 2 determinants, are just the Cramer's rule solution of simultaneous equations for a and b.

7. As in Problem 6, write $\mathbf{V} = 4\mathbf{i}-5\mathbf{j}$ in terms of the basis vectors $\mathbf{i}-4\mathbf{j}$ and $5\mathbf{i}+2\mathbf{j}$.

In Problems 8 to 15, use (8.5) to show that the given functions are linearly independent.

8. $\sin x,\ \cos x$

9. $e^{ix},\ \sin x$

10. $x,\ e^x,\ xe^x$

11. $\sin x,\ \cos x,\ x\sin x,\ x\cos x$

12. $1,\ x^2,\ x^4,\ x^6$

13. $\sin x,\ \sin 2x$

14. $e^{ix},\ e^{-ix}$

15. $e^x,\ e^{ix},\ \cosh x$

16. (a) Prove that if the Wronskian (8.5) is not identically zero, then the functions f_1, f_2, $\ldots$, f_n are linearly independent. Note that this is equivalent to proving that if the functions are linearly dependent, then W is identically zero. *Hints:* Suppose (8.3) were true; you want to find the k's. Differentiate (8.3) repeatedly until you have a set of n equations for the n unknown k's. Then use (8.9).

(b) In part (a) you proved that if $W \not\equiv 0$, then the functions are linearly independent. You might think that if $W \equiv 0$, the functions would be linearly dependent. This is not necessarily true; if $W \equiv 0$, the functions might be either dependent or independent. For example, consider the functions x^3 and $|x^3|$ on the interval $(-1,1)$. Show that $W \equiv 0$, but the functions are not linearly dependent on $(-1,1)$. (Sketch them.) On the other hand, they are linearly dependent (in fact identical) on $(0,1)$.

In Problems 17 to 20, solve the sets of homogeneous equations by row reducing the matrix.

17. $$\begin{cases} x - 2y + 3z = 0 \\ x + 4y - 6z = 0 \\ 2x + 2y - 3z = 0 \end{cases}$$

18. $$\begin{cases} 2x + 3z = 0 \\ 4x + 2y + 5z = 0 \\ x - y + 2z = 0 \end{cases}$$

19. $$\begin{cases} 3x + y + 3z + 6w = 0 \\ 4x - 7y - 3z + 5w = 0 \\ x + 3y + 4z - 3w = 0 \\ 3x + 2z + 7w = 0 \end{cases}$$

20. $$\begin{cases} 2x - 3y + 5z = 0 \\ x + 2y - z = 0 \\ x - 5y + 6z = 0 \\ 4x + y + 3z = 0 \end{cases}$$

21. Find a condition for four points in space to lie in a plane. Your answer should be in the form a determinant which must be equal to zero. *Hint:* The equation of a plane is of the form $ax+by+cz = d$, where a, b, c, d are constants. The four points (x_1, y_1, z_1), (x_2, y_2, z_2), etc., are all to satisfy this equation. When can you find a, b, c, d not all zero?

22. Find a condition for three lines in a plane to intersect in one point. *Hint:* See Problem 21. Write the equation of a line as $ax + by = c$. Assume that no two of the lines are parallel.

Using (8.9), find the values of λ such that the following equations have nontrivial solutions, and for each λ, solve the equations. (See Example 4.)

23. $\begin{cases} (4-\lambda)x - 2y = 0 \\ -2x + (7-\lambda)y = 0 \end{cases}$

24. $\begin{cases} (6-\lambda)x + 3y = 0 \\ 3x - (2+\lambda)y = 0 \end{cases}$

25. $\begin{cases} -(1+\lambda)x + y + 3z = 0, \\ x + (2-\lambda)y = 0, \\ 3x + (2-\lambda)z = 0. \end{cases}$

For each of the following, write the solution in vector form [see (8.11) and (8.13)].

26. $\begin{cases} 2x - 3y + 5z = 3 \\ x + 2y - z = 5 \\ x - 5y + 6z = -2 \\ 4x + y + 3z = 13 \end{cases}$

27. $\begin{cases} x - y + 2z = 3 \\ -2x + 2y - z = 0 \\ 4x - 4y + 5z = 6 \end{cases}$

28. $\begin{cases} 2x + y - 5z = 7 \\ x - 2y = 1 \\ 3x - 5y - z = 4 \end{cases}$

9. SPECIAL MATRICES AND FORMULAS

In this section we want to discuss various terms used in work with matrices, and prove some important formulas. First we list for reference needed definitions and facts about matrices.

There are several special matrices which are related to a given matrix A. We outline in (9.1) what these matrices are called, what notations are used for them, and how we get them from A.

(9.1)	*Name of Matrix*	*Notations for it*	*How to get it from* A
	Transpose of A, or A transpose	A^{T} or $\widetilde{\mathrm{A}}$ or A' or A^t	Interchange rows and columns in A.
	Complex conjugate of A	$\bar{\mathrm{A}}$ or A^*	Take the complex conjugate of each element.
	Transpose conjugate, Hermitian conjugate, adjoint (Problem 9), Hermitian adjoint.	$\mathrm{A}^\dagger$ (A dagger)	Take the complex conjugate of each element and transpose.
	Inverse of A	A^{-1}	See Formula (6.13).

There is another set of names for special types of matrices. In (9.2), we list these and their definitions for reference.

(9.2)	A matrix is called	if it satisfies the condition(s)
	real	$A = \bar{A}$
	symmetric	$A = A^T$, A real (matrix = its transpose)
	skew-symmetric or antisymmetric	$A = -A^T$, A real
	orthogonal	$A^{-1} = A^T$, A real (inverse = transpose)
	pure imaginary	$A = -\bar{A}$
	Hermitian	$A = A^\dagger$ (matrix = its transpose conjugate)
	anti-Hermitian	$A = -A^\dagger$
	unitary	$A^{-1} = A^\dagger$ (inverse = transpose conjugate)
	normal	$AA^\dagger = A^\dagger A$ (A and $A^\dagger$ commute)

Now let's consider some examples and proofs using these terms.

Index Notation We are going to need index notation in some of our work below, so for reference we restate the rule in (6.2b) for matrix multiplication.

$$(AB)_{ij} = \sum_k A_{ik}B_{kj}. \tag{9.3}$$

Study carefully the index notation for "row times column" multiplication. To find the element in row i and column j of the product matrix AB, we multiply row i of A times column j of B. Note that the k's (the sum is over k) are next to each other in (9.3). If we should happen to have $\sum_k B_{kj}A_{ik}$, we should rewrite it as $\sum_k A_{ik}B_{kj}$ (with the k's next to each other) to recognize it as an element of the matrix AB (not BA). We will see an example of this in (9.10) below.

Kronecker δ The *Kronecker* δ is defined by

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \tag{9.4}$$

For example, $\delta_{11} = 1$, $\delta_{12} = 0$, $\delta_{22} = 1$, $\delta_{31} = 0$, and so on. In this notation a unit matrix is one whose elements are δ_{ij} and we can write

$$I = (\delta_{ij}). \tag{9.5}$$

(*Also see* Chapter 10, Section 5.) The Kronecker δ notation is useful for other purposes. For example, since (for positive integers m and n)

$$\int_{-\pi}^{\pi} \cos nx \cos mx \, dx = \begin{cases} \pi, & \text{if } m = n, \\ 0, & \text{if } m \neq n, \end{cases} \tag{9.6a}$$

we can write

$$\int_{-\pi}^{\pi} \cos nx \cos mx \, dx = \pi \cdot \delta_{nm}. \tag{9.6b}$$

This is the same as (9.6a) because $\delta_{nm} = 0$ if $m \neq n$, and $\delta_{nm} = 1$ if $m = n$.

Using the Kronecker δ, we can give a formal proof that for any matrix M and a conformable unit matrix I, the product of I and M is just M. Using index notation and equations (9.3) and (9.4), we have

$$(\mathrm{IM})_{ij} = \sum_k \delta_{ik} M_{kj} = M_{ij} \quad \text{or} \quad \mathrm{IM} = \mathrm{M} \tag{9.7}$$

since $\delta_{ik} = 0$ unless $k = i$.

More Useful Theorems Let's use index notation to prove the associative law for matrix multiplication, that is

$$\mathrm{A(BC)} = \mathrm{(AB)C} = \mathrm{ABC}. \tag{9.8}$$

First we write $(\mathrm{BC})_{kj} = \sum_l B_{kl} C_{lj}$. Then we have

$$\begin{aligned} [\mathrm{A(BC)}]_{ij} &= \sum_k A_{ik} (\mathrm{BC})_{kj} = \sum_k A_{ik} \sum_l B_{kl} C_{lj} \\ &= \sum_k \sum_l A_{ik} B_{kl} C_{lj} = (\mathrm{ABC})_{ij} \end{aligned} \tag{9.9}$$

which is the index notation for A(BC) = ABC as in (9.8). We can prove (AB)C = ABC in a similar way (Problem 1).

In formulas we may want the transpose of the product of two matrices. First note that $A^{\mathrm{T}}_{ik} = A_{ki}$ [see (2.1) or (9.1)]. Then

$$\begin{aligned} (\mathrm{AB})^{\mathrm{T}}_{ik} &= (\mathrm{AB})_{ki} = \sum_j A_{kj} B_{ji} = \sum_j A^{\mathrm{T}}_{jk} B^{\mathrm{T}}_{ij} \\ &= \sum_j B^{\mathrm{T}}_{ij} A^{\mathrm{T}}_{jk} = (\mathrm{B}^{\mathrm{T}} \mathrm{A}^{\mathrm{T}})_{ik}, \quad \text{or,} \\ (\mathrm{AB})^{\mathrm{T}} &= \mathrm{B}^{\mathrm{T}} \mathrm{A}^{\mathrm{T}}. \end{aligned} \tag{9.10}$$

The theorem applies to a product of any number of matrices (see Problem 8b). For example

$$(\mathrm{ABCD})^{\mathrm{T}} = \mathrm{D}^{\mathrm{T}} \mathrm{C}^{\mathrm{T}} \mathrm{B}^{\mathrm{T}} \mathrm{A}^{\mathrm{T}}. \tag{9.11}$$

The transpose of a product of matrices is equal to the product of the transposes in reverse order.

A similar theorem is true for the inverse of a product (see Section 6, Problem 18).

(9.12) $$(\mathrm{ABCD})^{-1} = \mathrm{D}^{-1}\mathrm{C}^{-1}\mathrm{B}^{-1}\mathrm{A}^{-1}.$$

The inverse of a product of matrices is equal to the product of the inverses in reverse order.

Trace of a Matrix The *trace* (or *spur*) or a square matrix A (written Tr A) is the sum of the elements on the main diagonal. Thus the trace of a unit n by n matrix is n, and the trace of the matrix M in (6.10) is 6. It is a theorem that the trace of a product of matrices is not changed by permuting them in cyclic order. For example

(9.13) $$\mathrm{Tr}(\mathrm{ABC}) = \mathrm{Tr}(\mathrm{BCA}) = \mathrm{Tr}(\mathrm{CAB}).$$

We can prove this as follows:

$$\begin{aligned}\mathrm{Tr}(\mathrm{ABC}) &= \sum_i (\mathrm{ABC})_{ii} = \sum_i \sum_j \sum_k A_{ij}B_{jk}C_{ki} \\ &= \sum_i \sum_j \sum_k B_{jk}C_{ki}A_{ij} = \mathrm{Tr}(\mathrm{BCA}) \\ &= \sum_i \sum_j \sum_k C_{ki}A_{ij}B_{jk} = \mathrm{Tr}(\mathrm{CAB}).\end{aligned}$$

Warning: Tr(ABC) is *not* equal to Tr(ACB) in general.

Theorem: If H is a Hermitian matrix, then $\mathrm{U} = e^{i\mathrm{H}}$ is a unitary matrix. (This is an important relation in quantum mechanics.) By (9.2) we need to prove that $\mathrm{U}^\dagger = \mathrm{U}^{-1}$ if $\mathrm{H}^\dagger = \mathrm{H}$. First, $e^{i\mathrm{H}}e^{-i\mathrm{H}} = e^{i\mathrm{H}-i\mathrm{H}}$ since H commutes with itself—see Problem 6.29. But this is e^0 which is the unit matrix [see Section 6] so $\mathrm{U}^{-1} = e^{-i\mathrm{H}}$. To find $\mathrm{U}^\dagger = (e^{i\mathrm{H}})^\dagger$, we expand $\mathrm{U} = e^{i\mathrm{H}}$ in a power series to get $\mathrm{U} = \sum_k (i\mathrm{H})^k/k!$ and then take the transpose conjugate. To do this we just need to realize that the transpose of a sum of matrices is the sum of the transposes, and that the transpose of a power of a matrix, say $(\mathrm{M}^n)^\mathrm{T}$ is equal to $(\mathrm{M}^\mathrm{T})^n$ (Problem 9.21). Also recall from Chapter 2 that you find the complex conjugate of an expression by changing the signs of all the i's. This means that $(i\mathrm{H})^\dagger = -i\mathrm{H}^\dagger = -i\mathrm{H}$ since H is Hermitian. Then summing the series we get $\mathrm{U}^\dagger = e^{-i\mathrm{H}}$, which is just what we found for U^{-1} above. Thus $\mathrm{U}^\dagger = \mathrm{U}^{-1}$, so U is a unitary matrix. (Also see Problem 11.61.)

▶ PROBLEMS, SECTION 9

1. Use index notation as in (9.9) to prove the second part of the associative law for matrix multiplication: (AB)C = ABC.

2. Use index notation to prove the distributive law for matrix multiplication, namely: A(B + C) = AB + AC.

3. Given the following matrix, find the transpose, the inverse, the complex conjugate, and the transpose conjugate of A. Verify that $AA^{-1} = A^{-1}A =$ the unit matrix.

$$A = \begin{pmatrix} 1 & 0 & 5i \\ -2i & 2 & 0 \\ 1 & 1+i & 0 \end{pmatrix},$$

4. Repeat Problem 3 given

$$A = \begin{pmatrix} 0 & 2i & -1 \\ -i & 2 & 0 \\ 3 & 0 & 0 \end{pmatrix}.$$

5. Show that the product AA^T is a symmetric matrix.

6. Give numerical examples of: a symmetric matrix; a skew-symmetric matrix; a real matrix; a pure imaginary matrix.

7. Write each of the items in the second column of (9.2) in index notation.

8. (a) Prove that $(AB)^\dagger = B^\dagger A^\dagger$. *Hint:* See (9.10).

(b) Verify (9.11), that is, show that (9.10) applies to a product of any number of matrices. *Hint:* Use (9.10) and (9.8).

9. In (9.1) we have defined the adjoint of a matrix as the transpose conjugate. This is the usual definition except in algebra where the adjoint is defined as the transposed matrix of cofactors [see (6.13)]. Show that the two definitions are the same for a unitary matrix with determinant $= +1$.

10. Show that if a matrix is orthogonal and its determinant is $+1$, then each element of the matrix is equal to its own cofactor. *Hint:* Use (6.13) and the definition of an orthogonal matrix.

11. Show that a real Hermitian matrix is symmetric. Show that a real unitary matrix is orthogonal. *Note:* Thus we see that Hermitian is the complex analogue of symmetric, and unitary is the complex analogue of orthogonal. (See Section 11.)

12. Show that the definition of a Hermitian matrix ($A = A^\dagger$) can be written $a_{ij} = \bar{a}_{ji}$ (that is, the diagonal elements are real and the other elements have the property that $a_{12} = \bar{a}_{21}$, etc.). Construct an example of a Hermitian matrix.

13. Show that the following matrix is a unitary matrix.

$$\begin{pmatrix} (1+i\sqrt{3})/4 & \dfrac{\sqrt{3}}{2\sqrt{2}}(1+i) \\ \dfrac{-\sqrt{3}}{2\sqrt{2}}(1+i) & (\sqrt{3}+i)/4 \end{pmatrix}$$

14. Use (9.11) and (9.12) to simplify $(AB^TC)^T$, $(C^{-1}MC)^{-1}$, $(AH)^{-1}(AHA^{-1})^3(HA^{-1})^{-1}$.

15. (a) Show that the Pauli spin matrices (Problem 6.6) are Hermitian.

(b) Show that the Pauli spin matrices satisfy the Jacobi identity $\big[A,[B,C]\big] + \big[B,[C,A]\big] + \big[C,[A,B]\big] = 0$ where $[A,B]$ is the commutator of A, B [*see* (6.3)].

(c) Generalize (b) to prove the Jacobi identity for any (conformable) matrices A, B, C. *Also see* Chapter 6, Problem 3.14.

16. Let $C_{ij} = (-1)^{i+j}M_{ij}$ be the cofactor of element a_{ij} in the determinant A. Show that the statement of Laplace's development and the statement of Problem 3.8 can be combined in the equations

$$\sum_j a_{ij}C_{kj} = \delta_{ik} \cdot \det A, \quad \text{or} \quad \sum_i a_{ij}C_{ik} = \delta_{jk} \cdot \det A.$$

17. (a) Show that if A and B are symmetric, then AB is not symmetric unless A and B commute.

(b) Show that a product of orthogonal matrices is orthogonal.

(c) Show that if A and B are Hermitian, then AB is not Hermitian unless A and B commute.

(d) Show that a product of unitary matrices is unitary.

18. If A and B are symmetric matrices, show that their commutator is antisymmetric [see equation (6.3)].

19. (a) Prove that $\mathrm{Tr(AB)} = \mathrm{Tr(BA)}$. *Hint:* See proof of (9.13).

(b) Construct matrices A, B, C for which $\mathrm{Tr(ABC)} \neq \mathrm{Tr(CBA)}$, but verify that $\mathrm{Tr(ABC)} = \mathrm{Tr(CAB)}$.

(c) If S is a symmetric matrix and A is an antisymmetric matrix, show that $\mathrm{Tr(SA)} = 0$. *Hint:* Consider $\mathrm{Tr(SA)^T}$ and prove that $\mathrm{Tr(SA)} = -\mathrm{Tr(SA)}$.

20. Show that the determinant of a unitary matrix is a complex number with absolute value $= 1$. *Hint:* See proof of equation (7.11).

21. Show that the transpose of a sum of matrices is equal to the sum of the transposes. Also show that $(\mathrm{M}^n)^\mathrm{T} = (\mathrm{M}^\mathrm{T})^n$. *Hint:* Use (9.11) and (9.8).

22. Show that a unitary matrix is a normal matrix, that is, that it commutes with its transpose conjugate [see (9.2)]. Also show that orthogonal, symmetric, antisymmetric, Hermitian, and anti-Hermitian matrices are normal.

23. Show that the following matrices are Hermitian whether A is Hermitian or not: $\mathrm{AA}^\dagger$, $\mathrm{A} + \mathrm{A}^\dagger$, $i(\mathrm{A} - \mathrm{A}^\dagger)$.

24. Show that an orthogonal transformation preserves the length of vectors. *Hint:* If r is the column matrix of vector $\mathbf{r}$ [see (6.10)], write out $\mathrm{r^Tr}$ to show that it is the square of the length of $\mathbf{r}$. Similarly $\mathrm{R^TR} = |\mathbf{R}|^2$ and you want to show that $|\mathbf{R}|^2 = |\mathbf{r}|^2$, that is, $\mathrm{R^TR} = \mathrm{r^Tr}$ if $\mathrm{R} = \mathrm{Mr}$ and M is orthogonal. Use (9.11).

25. (a) Show that the inverse of an orthogonal matrix is orthogonal. *Hint:* Let $\mathrm{A} = \mathrm{O}^{-1}$; from (9.2), write the condition for O to be orthogonal and show that A satisfies it.

(b) Show that the inverse of a unitary matrix is unitary. See hint in (a).

(c) If H is Hermitian and U is unitary, show that $\mathrm{U^{-1}HU}$ is Hermitian.

10. LINEAR VECTOR SPACES

We have used extensively the vector $\mathbf{r} = \mathbf{i}x+\mathbf{j}y+\mathbf{k}z$ to mean a vector from the origin to the point (x, y, z). There is a one-to-one correspondence between the vectors $\mathbf{r}$ and the points (x, y, z); the collection of all such points or all such vectors makes up the 3-dimensional space often called R_3 (R for real) or V_3 (V for vector) or E_3 (E for Euclidean). Similarly, we can consider a 2-dimensional space V_2 of vectors $\mathbf{r} = \mathbf{i}x + \mathbf{j}y$ or points (x, y) making up the (x, y) plane. V_2 might also mean *any* plane through the origin. And V_1 means all the vectors from the origin to points on some line through the origin.

We also use x, y, z to mean the variables or unknowns in a problem. Now applied problems often involve more than three variables. By extension of the idea of V_3, it is convenient to call an ordered set of n numbers a point or vector in the n-dimensional space V_n. For example, the 4-vectors of special relativity are ordered sets of four numbers; we say that space-time is 4-dimensional. A point of the *phase space* used in classical and quantum mechanics is an ordered set of six numbers, the three components of the position of a particle and the three components of its momentum; thus the phase space of a particle is the 6-dimensional space V_6.

In such cases, we can't represent the variables as coordinates of a point in *physical* space since physical space has only three dimensions. But it is convenient and customary to extend our geometrical *terminology* anyway. Thus we use the terms *variables* and *coordinates* interchangeably and speak, for example, of a "point in 5-dimensional space," meaning an ordered set of values of five variables, and similarly for any number of variables. In three dimensions, we think of the coordinates of a point as the components of a vector from the origin to the point. By analogy, we call an ordered set of five numbers a "vector in 5-dimensional space" or an ordered set of n numbers a "vector in n-dimensional space."

Much of the geometrical terminology which is familiar in two and three dimensions can be extended to problems in n dimensions (that is, n variables) by using the algebra which parallels the geometry. For example, the distance from the origin to the point (x, y, z) is $\sqrt{x^2 + y^2 + z^2}$. By analogy in a problem in the five variables x, y, z, u, v, we define the distance from the origin $(0, 0, 0, 0, 0)$ to the point (x, y, z, u, v) as $\sqrt{x^2 + y^2 + z^2 + u^2 + v^2}$. By using the algebra which goes with the geometry, we can easily extend such ideas as the length of a vector, the dot product of two vectors, and therefore the angle between the vectors and the idea of orthogonality, etc. We saw in Section 7, that an orthogonal transformation in two or three dimensions corresponds to a rotation. Thus we might say, in a problem in n variables, that a linear transformation (that is a linear change of variables) satisfying "sum of squares of new variables = sum of squares of old variables" [compare (7.8)] corresponds to a "rotation in *n-dimensional space*."

▶ **Example 1.** Find the distance between the points $(3, 0, 5, -2, 1)$ and $(0, 1, -2, 3, 0)$.

Generalizing what we would do in three dimensions, we find $d^2 = (3-0)^2 + (0-1)^2 + (5+2)^2 + (-2-3)^2 + (1-0)^2 = 9 + 1 + 49 + 25 + 1 = 85$, $d = \sqrt{85}$.

If we start with several vectors, and find linear combinations of them in the algebraic way (by components), then we say that the original set of vectors and all their linear combinations form a *linear vector space* (or just *vector space* or *linear space* or *space*). Note that if $\mathbf{r}$ is one of our original vectors, then $\mathbf{r} - \mathbf{r} = \mathbf{0}$ is one of the linear combinations; thus the zero vector (that is, the origin) must be a point in every vector space. A line or plane not passing through the origin is not a vector space.

Subspace, Span, Basis, Dimension Suppose we start with the four vectors in (8.1). We showed in (8.2) that they are all linear combinations of the two vectors $(9, 0, 7)$ and $(0, -9, 13)$. Now two linearly independent vectors (remember their tails are at the origin) determine a plane; all linear combinations of the two vectors lie in the plane. [The plane we are talking about in this example is the plane through the three points $(9, 0, 7)$, $(0, -9, 13)$, and the origin.] Since all the vectors making up this plane V_2 are also part of 3-dimensional space V_3, we call V_2 a *subspace* of V_3. Similarly any line lying in this plane and passing through the origin is a subspace of V_2 and of V_3. We say that either the original four vectors or the two independent ones *span* the space V_2; a set of vectors spans a space if all the vectors in the space can be written as linear combinations of the spanning set. A set of *linearly independent* vectors which span a vector space is called a *basis*. Here the vectors $(9, 0, 7)$ and $(0, -9, 13)$ are one possible choice as a basis for the space V_2; another choice would be any two of the original vectors since in (8.2) no two of the vectors are dependent.

The *dimension* of a vector space is equal to the number of basis vectors. Note that this statement implies (correctly—see Problem 8) that no matter how you pick the basis vectors for a given vector space, there will always be the same number of them. This number is the dimension of the space. In 3 dimensions, we have frequently used the unit basis vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ which can also be written as $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$. Then in, say 5 dimensions, a corresponding set of unit basis vectors would be $(1, 0, 0, 0, 0)$, $(0, 1, 0, 0, 0)$, $(0, 0, 1, 0, 0)$, $(0, 0, 0, 1, 0)$, $(0, 0, 0, 0, 1)$. You should satisfy yourself that these five vectors are linearly independent and span a 5 dimensional space.

▶ **Example 2.** Find the dimension of the space spanned by the following vectors, and a basis for the space: $(1, 0, 1, 5, -2)$, $(0, 1, 0, 6, -3)$, $(2, -1, 2, 4, 1)$, $(3, 0, 3, 15, -6)$.

We write the matrix whose rows are the components of the vectors and row reduce it to find that there are three linearly independent vectors: $(1, 0, 1, 5, 0)$, $(0, 1, 0, 6, 0)$, $(0, 0, 0, 0, 1)$. These three vectors are a basis for the space which is therefore 3-dimensional.

Inner Product, Norm, Orthogonality Recall from (4.10) that the scalar (or dot or inner) product of two vectors $\mathbf{A} = (A_1, A_2, A_3)$ and $\mathbf{B} = (B_1, B_2, B_3)$ is $A_1B_1 + A_2B_2 + A_3B_3 = \sum_{i=1}^{3} A_iB_i$. This is very easy to generalize to n dimensions. By definition, the inner product of two vectors in n dimensions is given by

$$\mathbf{A} \cdot \mathbf{B} = (\text{Inner product of } \mathbf{A} \text{ and } \mathbf{B}) = \sum_{i=1}^{n} A_iB_i. \tag{10.1}$$

Similarly, generalizing (4.1), we can define the length or *norm* of a vector in n dimensions by the formula:

$$A = \text{Norm of } \mathbf{A} = ||\mathbf{A}|| = \sqrt{\mathbf{A} \cdot \mathbf{A}} = \sqrt{\sum_{i=1}^{n} A_i^2}. \tag{10.2}$$

In 3 dimensions, we also write the scalar product as $AB \cos\theta$ [see (4.2)] so if two vectors are orthogonal (perpendicular) their scalar product is $AB \cos \pi/2 = 0$. We generalize this to n dimensions by saying that two vectors in n dimensions are orthogonal if their inner product is zero.

$$\mathbf{A} \text{ and } \mathbf{B} \text{ are orthogonal if } \sum_{i=1}^{n} A_iB_i = 0. \tag{10.3}$$

Schwarz Inequality In 2 or 3 dimensions we can find the angle between two vectors [see (4.11)] from the formula $\mathbf{A}\cdot\mathbf{B} = AB\cos\theta$. It is tempting to use the same formula in n dimensions, but before we do we should be sure that the resulting value of $\cos\theta$ will satisfy $|\cos\theta| \le 1$, that is

$$|\mathbf{A}\cdot\mathbf{B}| \le AB, \quad \text{or} \quad \left|\sum_{i=1}^{n} A_i B_i\right| \le \sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}. \tag{10.4}$$

This is called the Schwarz inequality (for n-dimensional Euclidean space). We can prove it as follows. First note that if $\mathbf{B} = \mathbf{0}$, (10.4) just says $0 \le 0$ which is certainly true. For $\mathbf{B} \ne \mathbf{0}$, we consider the vector $\mathbf{C} = B\mathbf{A} - (\mathbf{A}\cdot\mathbf{B})\mathbf{B}/B$, and find $\mathbf{C}\cdot\mathbf{C}$. Now $\mathbf{C}\cdot\mathbf{C} = \sum C_i^2 \ge 0$, so we have

$$\begin{aligned}\mathbf{C}\cdot\mathbf{C} &= B^2(\mathbf{A}\cdot\mathbf{A}) - 2B(\mathbf{A}\cdot\mathbf{B})(\mathbf{A}\cdot\mathbf{B})/B + (\mathbf{A}\cdot\mathbf{B})^2(\mathbf{B}\cdot\mathbf{B})/B^2 \\ &= A^2B^2 - 2(\mathbf{A}\cdot\mathbf{B})^2 + (\mathbf{A}\cdot\mathbf{B})^2 = A^2B^2 - (\mathbf{A}\cdot\mathbf{B})^2 = C^2 \ge 0,\end{aligned} \tag{10.5}$$

which gives (10.4). Thus, if we like, we can define the cosine of the angle between two vectors in n dimensions by $\cos\theta = \mathbf{A}\cdot\mathbf{B}/(AB)$. Note that equality holds in Schwarz's inequality if and only if $\cos\theta = \pm 1$, that is, when $\mathbf{A}$ and $\mathbf{B}$ are parallel or antiparallel, say $\mathbf{B} = k\mathbf{A}$.

▶ **Example 3.** Find the cosine of the angle between each pair of the 3 basis vectors we found in Example 2.

By (10.2) we find that the norms of the first two basis vectors are $\sqrt{1+1+25} = \sqrt{27}$ and $\sqrt{1+36} = \sqrt{37}$. By (10.1), the inner product of these two vectors is $1\cdot 0 + 0\cdot 1 + 1\cdot 0 + 5\cdot 6 + 0\cdot 0 = 30$. Thus $\cos\theta = 30/(\sqrt{27\cdot 37}) \simeq 0.949$, which, we note, is < 1 as Schwarz's inequality says. The third basis vector in Example 2 is orthogonal to the other two since the inner products are zero, that is, $\cos\theta = 0$.

Orthonormal Basis; Gram-Schmidt Method We call a set of vectors *orthonormal* if they are all mutually *ortho*gonal (perpendicular), and each vector is *normal*ized (that is, its norm is one—it has unit length). For example, the vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, form an orthonormal set. If we have a set of basis vectors for a space, it is often convenient to take combinations of them to form an orthonormal basis. The Gram-Schmidt method is a systematic process for doing this. It is very simple in idea although the details of carrying it out can get messy. Suppose we have basis vectors $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$. Normalize $\mathbf{A}$ to get the first vector of a set of orthonormal basis vectors. To get a second basis vector, subtract from $\mathbf{B}$ its component along $\mathbf{A}$; what remains is orthogonal to $\mathbf{A}$. [See equation (4.4) and Figure 4.10.] Normalize this remainder to find the second vector of an orthonormal basis. Similarly, subtract from $\mathbf{C}$ its components along $\mathbf{A}$ and $\mathbf{B}$ to find a third vector orthogonal to both $\mathbf{A}$ and $\mathbf{B}$ and normalize this third vector. We now have 3 mutually orthogonal unit vectors; this is the desired set of orthonormal basis vectors. In a space of higher dimension, this process can be continued. (We will see a use for this method in Section 11; see degeneracy, pages 152–153.)

▶ **Example 4.** Given the basis vectors **A**, **B**, **C**, below, use the Gram-Schmidt method to find an orthonormal set of basis vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$. Following the outline above, we find

$$\begin{aligned}
\mathbf{A} &= (0,0,5,0); & \mathbf{e}_1 &= \mathbf{A}/A = (0,0,1,0);\\
\mathbf{B} &= (2,0,3,0); & \mathbf{B}-(\mathbf{e}_1\cdot\mathbf{B})\mathbf{e}_1 &= \mathbf{B}-3\mathbf{e}_1 = (2,0,0,0);\\
& & \mathbf{e}_2 &= (1,0,0,0);\\
\mathbf{C} &= (7,1,-5,3); & \mathbf{C}-(\mathbf{e}_1\cdot\mathbf{C})\mathbf{e}_1-(\mathbf{e}_2\cdot\mathbf{C})\mathbf{e}_2 &= \mathbf{C}-(-5)\mathbf{e}_1-7\mathbf{e}_2\\
& & &= (0,1,0,3);\\
& & \mathbf{e}_3 &= (0,1,0,3)/\sqrt{10}.
\end{aligned}$$

Complex Euclidean Space In applications it is useful to allow vector components to be complex. For example, in three dimensions we might consider vectors like $(5+2i, 3-i, 1+i)$. Let's go back and see what modifications are needed in this case. In (10.2), we want the quantity under the square root sign to be positive. To assure this, we replace the square of A_i by the absolute square of A_i, that is by $|A_i|^2 = A_i^* A_i$ where A_i^* is the complex conjugate of A_i (see Chapter 2). Similarly, in (10.1) and (10.3), we replace $A_i B_i$ by $A_i^* B_i$. Thus we define

$$\text{(10.6)} \qquad (\text{Inner product of } \mathbf{A} \text{ and } \mathbf{B}) = \sum_{i=1}^{n} A_i^* B_i$$

$$\text{(10.7)} \qquad (\text{Norm of } \mathbf{A}) = ||\mathbf{A}|| = \sqrt{\sum_{i=1}^{n} A_i^* A_i}$$

$$\text{(10.8)} \qquad \mathbf{A} \text{ and } \mathbf{B} \text{ are orthogonal if } \sum_{i=1}^{n} A_i^* B_i = 0.$$

The Schwarz inequality becomes (see Problem 6)

$$\text{(10.9)} \qquad \left|\sum_{i=1}^{n} A_i^* B_i\right| \le \sqrt{\sum_{i=1}^{n} A_i^* A_i}\sqrt{\sum_{i=1}^{n} B_i^* B_i}.$$

Note that we can write the inner product in matrix form. If A is a column matrix with elements A_i, then the transpose conjugate matrix $\mathrm{A}^\dagger$ is a row matrix with elements A_i^*. Using this notation we can write $\sum A_i^* B_i = \mathrm{A}^\dagger \mathrm{B}$ (Problem 9).

▶ **Example 5.** Given $\mathbf{A} = (3i, 1-i, 2+3i, 1+2i)$, $\mathbf{B} = (-1, 1+2i, 3-i, i)$, $\mathbf{C} = (4-2i, 2-i, 1, i-2)$, we find by (10.6) to (10.8):

$$\begin{aligned}
(\text{Inner product of } \mathbf{A} \text{ and } \mathbf{B}) &= (-3i)(-1) + (1+i)(1+2i)\\
&\quad + (2-3i)(3-i) + (1-2i)i = 4-4i.
\end{aligned}$$

$$\begin{aligned}(\text{Norm of } \mathbf{A})^2 &= (-3i)(3i) + (1+i)(1-i) + (2-3i)(2+3i) + (1-2i)(1+2i)\\ &= 9+2+13+5 = 29, \quad ||\mathbf{A}|| = \sqrt{29}.\\ (\text{Norm of } \mathbf{B})^2 &= 1+5+10+1 = 17, \quad ||\mathbf{B}|| = \sqrt{17}.\end{aligned}$$

Note that $|4-4i| = 4\sqrt{2} < \sqrt{29}\sqrt{17}$ in accord with the Schwarz inequality (10.9).

$$\begin{aligned}(\text{Inner product of } \mathbf{B} \text{ and } \mathbf{C}) &= (-1)(4-2i) + (1-2i)(2-i) + (3+i)(1)\\ &\quad + (-i)(i-2) = -4+2i-5i+3+i+1+2i = 0.\end{aligned}$$

Thus by (10.8), **B** and **C** are orthogonal.

PROBLEMS, SECTION 10

1. Find the distance between the points

(a) $(4,-1,2,7)$ and $(2,3,1,9)$;

(b) $(-1,5,-3,2,4)$ and $(2,6,2,7,6)$;

(c) $(5,-2,3,3,1,0)$ and $(0,1,5,7,2,1)$.

2. For the given sets of vectors, find the dimension of the space spanned by them and a basis for this space.

(a) $(1,-1,0,0)$, $(0,-2,5,1)$, $(1,-3,5,1)$, $(2,-4,5,1)$;

(b) $(0,1,2,0,0,4)$, $(1,1,3,5,-3,5)$, $(1,0,0,5,0,1)$, $(-1,1,3,-5,-3,3)$, $(0,0,1,0,-3,0)$;

(c) $(0,10,-1,1,10)$, $(2,-2,-4,0,-3)$, $(4,2,0,4,5)$, $(3,2,0,3,4)$, $(5,-4,5,6,2)$.

3. (a) Find the cosines of the angles between pairs of vectors in Problem 2(a).

(b) Find two orthogonal vectors in Problem 2(b).

4. For each given set of basis vectors, use the Gram-Schmidt method to find an orthonormal set.

(a) $\mathbf{A} = (0,2,0,0)$, $\mathbf{B} = (3,-4,0,0)$, $\mathbf{C} = (1,2,3,4)$.

(b) $\mathbf{A} = (0,0,0,7)$, $\mathbf{B} = (2,0,0,5)$, $\mathbf{C} = (3,1,1,4)$.

(c) $\mathbf{A} = (6,0,0,0)$, $\mathbf{B} = (1,0,2,0)$, $\mathbf{C} = (4,1,9,2)$.

5. By (10.6) and (10.7), find the norms of **A** and **B** and the inner product of **A** and **B**, and note that the Schwarz inequality (10.9) is satisfied:

(a) $\mathbf{A} = (3+i, 1, 2-i, -5i, i+1)$, $\mathbf{B} = (2i, 4-3i, 1+i, 3i, 1)$;

(b) $\mathbf{A} = (2, 2i-3, 1+i, 5i, i-2)$, $\mathbf{B} = (5i-2, 1, 3+i, 2i, 4)$.

6. Write out the proof of the Schwarz inequality (10.9) for a complex Euclidean space. *Hint:* Follow the proof of (10.4) in (10.5), replacing the definitions of norm and inner product in (10.1) and (10.2) by the definitions in (10.6) and (10.7). Remember that norms are real and ≥ 0.

7. Show that, in n-dimensional space, any $n+1$ vectors are linearly dependent. *Hint:* See Section 8.

8. Show that two different sets of basis vectors for the same vector space must contain the same number of vectors. *Hint:* Suppose a basis for a given vector space contains n vectors. Use Problem 7 to show that there cannot be more than n vectors in a basis for this space. Conversely, if there were a correct basis with less than n vectors, what can you say about the claimed n-vector basis?

9. Write equations (10.6) to (10.9) in matrix form as discussed just after (10.9).

10. Prove that $||\mathbf{A}+\mathbf{B}|| \le ||\mathbf{A}|| + ||\mathbf{B}||$. This is called the triangle inequality; in two or three dimensions, it simply says that the length of one side of a triangle $\le$ sum of the lengths of the other 2 sides. *Hint:* To prove it in n-dimensional space, write the square of the desired inequality using (10.2) and also use the Schwarz inequality (10.4). Generalize the theorem to complex Euclidean space by using (10.7) and (10.9).

11. EIGENVALUES AND EIGENVECTORS; DIAGONALIZING MATRICES

We can give the following physical interpretation to Figure 7.2 and equations (7.5). Suppose the (x, y) plane is covered by an elastic membrane which can be stretched, shrunk, or rotated (with the origin fixed). Then any point (x, y) of the membrane becomes some point (X, Y) after the deformation, and we can say that the matrix M describes the deformation. Let us now ask whether there are any vectors such that $\mathbf{R} = \lambda\mathbf{r}$ where $\lambda =$ const. Such vectors are called *eigenvectors* (or *characteristic* vectors) of the transformation, and the values of λ are called the *eigenvalues* (or *characteristic* values) of the matrix M of the transformation.

Eigenvalues To illustrate finding eigenvalues, let's consider the transformation

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{11.1}$$

The eigenvector condition $\mathbf{R} = \lambda\mathbf{r}$ is, in matrix notation,

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix},$$

or written out in equation form:

$$\begin{aligned} 5x - 2y &= \lambda x, \\ -2x + 2y &= \lambda y, \end{aligned} \qquad \text{or} \qquad \begin{aligned} (5-\lambda)x - 2y &= 0, \\ -2x + (2-\lambda)y &= 0. \end{aligned} \tag{11.2}$$

These equations are homogeneous. Recall from (8.9) that a set of homogeneous equations has solutions other than $x = y = 0$ only if the determinant of the coefficients is zero. Thus we want

$$\begin{vmatrix} 5-\lambda & -2 \\ -2 & 2-\lambda \end{vmatrix} = 0. \tag{11.3}$$

This is called the *characteristic equation* of the matrix M, and the determinant in (11.3) is called the *secular determinant*.

> To obtain the characteristic equation of a matrix M, we subtract λ from the elements on the main diagonal of M, and then set the determinant of the resulting matrix equal to zero.

We solve (11.3) for λ to find the characteristic values of M:

$$\begin{aligned} &(5-\lambda)(2-\lambda) - 4 = \lambda^2 - 7\lambda + 6 = 0, \\ &\lambda = 1 \quad \text{or} \quad \lambda = 6. \end{aligned} \tag{11.4}$$

Eigenvectors Substituting the λ values from (11.4) into (11.2), we get:

$$\begin{array}{lll} & 2x - y = 0 & \text{from either of the equations (11.2) when } \lambda = 1; \\ (11.5) & & \\ & x + 2y = 0 & \text{from either of the equations (11.2) when } \lambda = 6. \end{array}$$

We were looking for vectors $\mathbf{r} = \mathbf{i}x + \mathbf{j}y$ such that the transformation (11.1) would give an $\mathbf{R}$ parallel to $\mathbf{r}$. What we have found is that *any* vector $\mathbf{r}$ with x and y components satisfying either of the equations (11.5) has this property. Since equations (11.5) are equations of straight lines through the origin, such vectors lie along these lines (Figure 11.1). Then equations (11.5) show that any vector $\mathbf{r}$ from the origin to a point on $x + 2y = 0$ is changed by the transformation (11.1) to a vector in the same direction but six times as long, and any vector from the origin to a point on $2x - y = 0$ is unchanged by the transformation (11.1). These vectors (along $x + 2y = 0$ and $2x - y = 0$) are the eigenvectors of the transformation. Along these two directions (and only these), the deformation of the elastic membrane was a pure stretch with no shear (rotation).

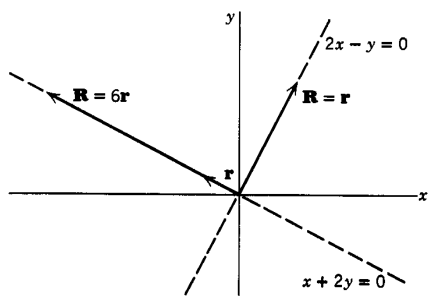


Figure 11.1

Diagonalizing a Matrix We next write (11.2) once with $\lambda = 1$, and again with $\lambda = 6$, using subscripts 1 and 2 to identify the corresponding eigenvectors:

$$(11.6) \qquad \begin{aligned} 5x_1 - 2y_1 &= x_1, & 5x_2 - 2y_2 &= 6x_2, \\ -2x_1 + 2y_1 &= y_1, & -2x_2 + 2y_2 &= 6y_2. \end{aligned}$$

These four equations can be written as one matrix equation, as you can easily verify by multiplying out both sides (Problem 1):

$$(11.7) \qquad \begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix}.$$

All we really can say about (x_1, y_1) is that $2x_1 - y_1 = 0$; however, it is convenient to pick numerical values of x_1 and y_1 to make $\mathbf{r}_1 = (x_1, y_1)$ a unit vector, and similarly for $\mathbf{r}_2 = (x_2, y_2)$. Then we have

$$(11.8) \qquad x_1 = \frac{1}{\sqrt{5}}, \quad y_1 = \frac{2}{\sqrt{5}}, \quad x_2 = \frac{-2}{\sqrt{5}}, \quad y_2 = \frac{1}{\sqrt{5}},$$

and (11.7) becomes

$$(11.9)\qquad \begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix}.$$

Representing these matrices by letters we can write

$$\mathrm{MC} = \mathrm{CD}, \qquad \text{where}$$

$$(11.10)\qquad \mathrm{M} = \begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix}, \quad \mathrm{C} = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}, \quad \mathrm{D} = \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix}.$$

If, as here, the determinant of C is not zero, then C has an inverse C^{-1}; let us multiply (11.10) by C^{-1} and remember that $\mathrm{C}^{-1}\mathrm{C}$ is the unit matrix; then $\mathrm{C}^{-1}\mathrm{MC} = \mathrm{C}^{-1}\mathrm{CD} = \mathrm{D}$.

$$(11.11)\qquad \mathrm{C}^{-1}\mathrm{MC} = \mathrm{D}.$$

The matrix D has elements different from zero only down the main diagonal; it is called a *diagonal matrix.* The matrix D is called *similar* to M, and when we obtain D given M, we say that we have *diagonalized* M *by a similarity transformation.*

We shall see shortly that this amounts physically to a simplification of the problem by a better choice of variables. For example, in the problem of the membrane, it is simpler to describe the deformation if we use axes along the eigenvectors. Later we shall see more examples of the use of the diagonalization process.

Observe that it is easy to find D; we need only solve the characteristic equation of M. Then D is a matrix with these characteristic values down the main diagonal and zeros elsewhere. We can also find C (with more work), but for many purposes only D is needed.

Note that the order of the eigenvalues down the main diagonal of D is arbitrary; for example we could write (11.6) as

$$(11.12)\qquad \begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} x_2 & x_1 \\ y_2 & y_1 \end{pmatrix} = \begin{pmatrix} x_2 & x_1 \\ y_2 & y_1 \end{pmatrix} \begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}$$

instead of (11.7). Then (11.11) still holds, with a different C, of course, and with

$$\mathrm{D} = \begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}$$

instead of as in (11.10) (Problem 1).

Meaning of C **and** D To see more clearly the meaning of (11.11) let us find what the matrices C and D mean physically. We consider two sets of axes (x, y) and (x', y') with (x', y') rotated through θ from (x, y) (Figure 11.2). The (x, y) and (x', y') coordinates of *one* point (or components of one vector $\mathbf{r} = \mathbf{r}'$) relative to the two systems are related by (7.13). Solving (7.13) for x and y, we have

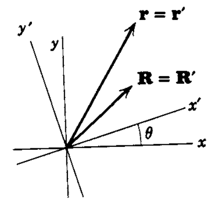


Figure 11.2

$$x = x' \cos\theta - y' \sin\theta,$$
$$y = x' \sin\theta + y' \cos\theta, \tag{11.13}$$

or in matrix notation

$$\mathrm{r} = \mathrm{Cr}' \qquad \text{where} \qquad \mathrm{C} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{11.14}$$

This equation is true for *any* single vector with components given in the two systems. Suppose we have another vector $\mathbf{R} = \mathbf{R}'$ (Figure 11.2) with components X, Y and X', Y'; these components are related by

$$\mathrm{R} = \mathrm{CR}'. \tag{11.15}$$

Now let M be a matrix which describes a deformation of the plane in the (x, y) system. Then the equation

$$\mathrm{R} = \mathrm{Mr} \tag{11.16}$$

says that the vector $\mathbf{r}$ becomes the vector $\mathbf{R}$ after the deformation, both vectors given relative to the (x, y) axes. Let us ask how we can describe the deformation in the (x', y') system, that is, what matrix carries $\mathbf{r}'$ into $\mathbf{R}'$? We substitute (11.14) and (11.15) into (11.16) and find $\mathrm{CR}' = \mathrm{MCr}'$ or

$$\mathrm{R}' = \mathrm{C}^{-1}\mathrm{MCr}'. \tag{11.17}$$

Thus the answer to our question is that

> $\mathrm{D} = \mathrm{C}^{-1}\mathrm{MC}$ is the matrix which describes in the (x', y') system the same deformation that M describes in the (x, y) system.

Next we want to show that if the matrix C is chosen to make $\mathrm{D} = \mathrm{C}^{-1}\mathrm{MC}$ a diagonal matrix, then the new axes (x', y') are along the directions of the eigenvectors of M. Recall from (11.10) that the columns of C are the components of the unit eigenvectors. If the eigenvectors are perpendicular, as they are in our example (see Problem 2) then the new axes (x', y') along the eigenvector directions are a set of perpendicular axes rotated

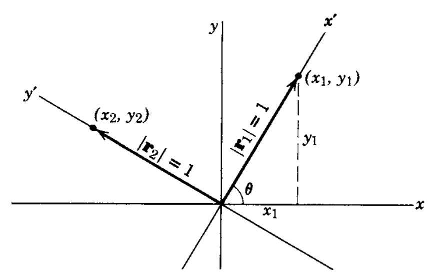


Figure 11.3

from axes (x, y) by some angle θ (Figure 11.3). The unit eigenvectors $\mathbf{r}_1$ and $\mathbf{r}_2$ are shown in Figure 11.3; from the figure we find

$$
\begin{aligned}
&x_1 = |\mathbf{r}_1|\cos\theta = \cos\theta, \quad x_2 = -|\mathbf{r}_2|\sin\theta = -\sin\theta\\
&y_1 = |\mathbf{r}_1|\sin\theta = \sin\theta, \quad y_2 = |\mathbf{r}_2|\cos\theta = \cos\theta;\\
&\mathrm{C} = \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.
\end{aligned} \tag{11.18}
$$

Thus, the matrix C which diagonalizes M is the rotation matrix C in (11.14) when the (x', y') axes are along the directions of the eigenvectors of M.

Relative to these new axes, the diagonal matrix D describes the deformation. For our example we have

$$
\begin{aligned}
&\mathrm{R}' = \mathrm{D}\mathrm{r}' \qquad \text{or} \qquad \begin{pmatrix} X' \\ Y' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix}\begin{pmatrix} x' \\ y' \end{pmatrix} \qquad \text{or}\\
&X' = x', \qquad Y' = 6y'.
\end{aligned} \tag{11.19}
$$

In words, (11.19) says that [in the (x', y') system] each point (x', y') has its x' coordinate unchanged by the deformation and its y' coordinate multiplied by 6, that is, the deformation is simply a stretch in the y' direction. This is a simpler description of the deformation and clearer physically than the description given by (11.1).

You can see now why the order of eigenvalues down the main diagonal in D is arbitrary and why (11.12) is just as satisfactory as (11.7). The new axes (x', y') are along the eigenvectors, but it is unimportant which eigenvector we call x' and which we call y'. In doing a problem we simply select a D with the eigenvalues of M in some (arbitrary) order down the main diagonal. Our choice of D then determines which eigenvector direction is called the x' axis and which is called y'.

It was unnecessary in the above discussion to have the x' and y' axes perpendicular, although this is the most useful case. If $\mathrm{r} = \mathrm{C}\mathrm{r}'$ but C is just any (nonsingular) matrix [not necessarily the orthogonal rotation matrix as in (11.14)], then (11.17) still follows. That is, $\mathrm{C}^{-1}\mathrm{MC}$ describes the deformation using (x', y') axes. But if C is not an orthogonal matrix, then the (x', y') axes are not perpendicular (Figure 11.4) and $x^2 + y^2 \neq x'^2 + y'^2$, that is, the transformation is not a rotation of axes. Recall that C is the matrix of unit eigenvectors; if these are perpendicular, then C is an orthogonal matrix (Problem 6). It can be shown that this will be the case if and only if the matrix M is symmetric. [See equation (11.27) and the discussion just before it. Also see Problems 33 to 35, and Problem 15.25.]

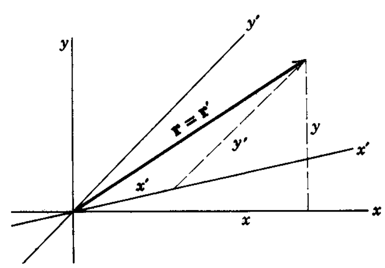


Figure 11.4

Degeneracy For a symmetric matrix, we have seen that the eigenvectors corresponding to different eigenvalues are orthogonal. If two (or more) eigenvalues are the same, then that eigenvalue is called *degenerate*. Degeneracy means that two (or more) independent eigenvectors correspond to the same eigenvalue.

Example 1. Consider the following matrix:

$$M = \begin{pmatrix} 1 & -4 & 2 \\ -4 & 1 & -2 \\ 2 & -2 & -2 \end{pmatrix}. \tag{11.20}$$

The eigenvalues of M are $\lambda = 6, -3, -3$, and the eigenvector corresponding to $\lambda = 6$ is $(2, -2, 1)$ (Problem 36). For $\lambda = -3$, the eigenvector condition is $2x - 2y + z = 0$. This is a plane orthogonal to the $\lambda = 6$ eigenvector, and any vector in this plane is an eigenvector corresponding to $\lambda = -3$. That is, the $\lambda = -3$ eigenspace is a plane. It is convenient to choose two orthogonal eigenvectors as basis vectors in this $\lambda = -3$ eigenplane, for example $(1, 1, 0)$ and $(-1, 1, 4)$. (See Problem 36.)

You may ask how you find these orthogonal eigenvectors except by inspection. Recall that the cross product of two vectors is perpendicular to both of them. Thus in the present case we could pick one vector in the $\lambda = -3$ eigenplane and then take its cross product with the $\lambda = 6$ eigenvector. This gives a second vector in the $\lambda = -3$ eigenplane, perpendicular to the first one we picked. However, this only works in three dimensions; if we are dealing with spaces of higher dimension (see Section 10), then we need another method. Suppose we first write down just any two (different) vectors in the eigenplane not trying to make them orthogonal. Then we can use the Gram-Schmidt method (see Section 10) to find an orthogonal set. For example, in the problem above, suppose you had thought of (or your computer had given you) the vectors $\mathbf{A} = (1, 1, 0)$ and $\mathbf{B} = (-1, 0, 2)$ which are vectors in the $\lambda = -3$ eigenplane but not orthogonal to each other. Following the Gram-Schmidt method, we find

$$\mathbf{A} = (1, 1, 0), \qquad \mathbf{e} = \mathbf{A}/A = (1, 1, 0)/\sqrt{2},$$
$$\mathbf{B} - (\mathbf{e} \cdot \mathbf{B})\mathbf{e} = (-1, 0, 2) - \frac{-1}{2}(1, 1, 0) = \left(\frac{-1}{2}, \frac{1}{2}, 2\right),$$

or $(-1, 1, 4)$ as we had above. For a degenerate subspace of dimension $m > 2$, we just need to write down m linearly independent eigenvectors, and then find an orthogonal set by the Gram-Schmidt method.

Diagonalizing Hermitian Matrices We have seen how to diagonalize symmetric matrices by orthogonal similarity transformations. The complex analogue of a symmetric matrix ($S^T = S$) is a Hermitian matrix ($H^\dagger = H$) and the complex analogue of an orthogonal matrix ($O^T = O^{-1}$) is a unitary matrix ($U^\dagger = U^{-1}$). So let's discuss diagonalizing Hermitian matrices by unitary similarity transformations. This is of great importance in quantum mechanics.

Although Hermitian matrices may have complex off-diagonal elements, the eigenvalues of a Hermitian matrix are always real. Let's prove this. (Refer to Section 9 for definitions and theorems as needed.) Let H be a Hermitian matrix, and let r be the column matrix of a non-zero eigenvector of H corresponding to the eigenvalue λ. Then the eigenvector condition is $Hr = \lambda r$. We want to take the transpose conjugate (dagger) of this equation. Using the complex conjugate of equation (9.10), we get $(Hr)^\dagger = r^\dagger H^\dagger = r^\dagger H$ since $H^\dagger = H$ for a Hermitian matrix. The transpose conjugate of λr is $\lambda^* r^\dagger$ (since λ is a number, we just need to take its complex conjugate). Now we have the two equations

$$Hr = \lambda r \qquad \text{and} \qquad r^\dagger H = \lambda^* r^\dagger. \tag{11.21}$$

Multiply the first equation in (11.21) on the left [see discussion following (10.9)] by the row matrix $\mathrm{r}^\dagger$ and the second equation on the right by the column matrix r to get

$$\mathrm{r}^\dagger \mathrm{H}\mathrm{r} = \lambda \mathrm{r}^\dagger \mathrm{r} \qquad \text{and} \qquad \mathrm{r}^\dagger \mathrm{H}\mathrm{r} = \lambda^* \mathrm{r}^\dagger \mathrm{r}. \tag{11.22}$$

Subtracting the two equations we find $(\lambda - \lambda^*)\mathrm{r}^\dagger\mathrm{r} = 0$. Since we assumed $\mathrm{r} \neq 0$, we have $\lambda^* = \lambda$, that is, λ is real.

We can also show that for a Hermitian matrix the eigenvectors corresponding to two different eigenvalues are orthogonal. Start with the two eigenvector conditions,

$$\mathrm{H}\mathrm{r}_1 = \lambda_1 \mathrm{r}_1 \qquad \text{and} \qquad \mathrm{H}\mathrm{r}_2 = \lambda_2 \mathrm{r}_2. \tag{11.23}$$

From these we can show (Problem 37)

$$\mathrm{r}_1^\dagger \mathrm{H}\mathrm{r}_2 = \lambda_1 \mathrm{r}_1^\dagger \mathrm{r}_2 = \lambda_2 \mathrm{r}_1^\dagger \mathrm{r}_2, \quad \text{or} \quad (\lambda_1 - \lambda_2)\mathrm{r}_1^\dagger \mathrm{r}_2 = 0. \tag{11.24}$$

Thus if $\lambda_1 \neq \lambda_2$, then the inner product of r_1 and r_2 is zero, that is, they are orthogonal [see (10.8)].

We can also prove that if a matrix M has real eigenvalues and can be diagonalized by a unitary similarity transformation, then it is Hermitian. In symbols, we write $\mathrm{U}^{-1}\mathrm{M}\mathrm{U} = \mathrm{D}$, and find the transpose conjugate of this equation to get (Problem 38)

$$(\mathrm{U}^{-1}\mathrm{M}\mathrm{U})^\dagger = \mathrm{U}^{-1}\mathrm{M}^\dagger\mathrm{U} = \mathrm{D}^\dagger = \mathrm{D}. \tag{11.25}$$

Thus $\mathrm{U}^{-1}\mathrm{M}\mathrm{U} = \mathrm{D} = \mathrm{U}^{-1}\mathrm{M}^\dagger\mathrm{U}$, so $\mathrm{M} = \mathrm{M}^\dagger$, which says that M is Hermitian. So we have proved that

(11.26) A matrix has real eigenvalues and can be diagonalized by a unitary similarity transformation if and only if it is Hermitian.

Since a real Hermitian matrix is a symmetric matrix and a real unitary matrix is an orthogonal matrix, the corresponding statement for symmetric matrices is (Problem 39).

(11.27) A matrix has real eigenvalues and can be diagonalized by an orthogonal similarity transformation if and only if it is symmetric.

Recall from (9.2) and Problem 9.22 that normal matrices include symmetric, Hermitian, orthogonal, and unitary matrices (as well as some others). It may be useful to know the following general theorem which we state without proof [see, for example, Am. J. Phys. **52**, 513–515 (1984)].

(11.28) A matrix can be diagonalized by a unitary similarity transformation if and only if it is normal.

Example 2. To illustrate diagonalizing a Hermitian matrix by a unitary similarity transformation, we consider the matrix

$$\mathrm{H} = \begin{pmatrix} 2 & 3-i \\ 3+i & -1 \end{pmatrix}. \tag{11.29}$$

(Verify that H is Hermitian.) We follow the same routine we used to find the eigenvalues and eigenvectors of a symmetric matrix. The eigenvalues are given by

$$\begin{aligned} &(2-\lambda)(-1-\lambda) - (3+i)(3-i) = 0, \\ &\lambda^2 - \lambda - 12 = 0, \quad \lambda = -3, 4. \end{aligned}$$

For $\lambda = -3$, an eigenvector satisfies the equations

$$\begin{aligned} &\begin{pmatrix} 5 & 3-i \\ 3+i & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0, \quad \text{or} \\ &5x + (3-i)y = 0, \quad (3+i)x + 2y = 0. \end{aligned}$$

These equations are satisfied by $x = 2$, $y = (-3-i)$. A choice for the unit eigenvector is $(2, -3-i)/\sqrt{14}$. For $\lambda = 4$, we find similarly the equations

$$-2x + (3-i)y = 0, \quad (3+i)x - 5y = 0,$$

which are satisfied by $y = 2$, $x = 3-i$, so a unit eigenvector is $(3-i, 2)/\sqrt{14}$. We can verify that the two eigenvectors are orthogonal (as we proved above that they must be) by finding that their inner product [see (10.8)] is $(2, -3-i)^* \cdot (3-i, 2) = 2(3-i) + 2(-3+i) = 0$. As in (11.10) we write the unit eigenvectors as the columns of a matrix U which diagonalizes H by a similarity transformation.

$$\mathrm{U} = \frac{1}{\sqrt{14}} \begin{pmatrix} 2 & 3-i \\ -3-i & 2 \end{pmatrix}, \quad \mathrm{U}^\dagger = \frac{1}{\sqrt{14}} \begin{pmatrix} 2 & -3+i \\ 3+i & 2 \end{pmatrix}$$

You can easily verify that $\mathrm{U}^\dagger\mathrm{U}$ = the unit matrix, so $\mathrm{U}^{-1} = \mathrm{U}^\dagger$. Then (Problem 40)

$$\mathrm{U}^{-1}\mathrm{HU} = \mathrm{U}^\dagger\mathrm{HU} = \begin{pmatrix} -3 & 0 \\ 0 & 4 \end{pmatrix}, \tag{11.30}$$

that is, H is diagonalized by a unitary similarity transformation.

Orthogonal Transformations in 3 Dimensions In Section 7, we considered the active rotation and/or reflection of vectors $\mathbf{r}$ which was produced by a given 3 by 3 orthogonal matrix. Study Equations (7.18) and (7.19) carefully to see that, acting on a column vector $\mathbf{r}$, they rotate the vector by angle θ around the z axis and/or reflect it through the (x, y) plane. We would now like to see how to find the effect of more complicated orthogonal matrices. We can do this by using an orthogonal similarity transformation to write a given orthogonal matrix relative to a new coordinate system in which the rotation axis is the z axis, and/or the (x, y) plane is the reflecting plane (in vector space language, this is a change of basis). Then a comparison with (7.18) or (7.19) gives the rotation angle. Recall how we construct a C matrix so that $\mathrm{C}^{-1}\mathrm{MC}$ describes the same transformation relative to a new set of axes that M described relative to the original axes: The columns of the C matrix are the components of unit vectors along the new axes [see (11.18) and Figure 11.3].

▶ **Example 3.** Consider the following matrices.

$$\text{(11.31)} \qquad \mathrm{A} = \frac{1}{2}\begin{pmatrix} 1 & \sqrt{2} & 1 \\ -\sqrt{2} & 0 & \sqrt{2} \\ 1 & -\sqrt{2} & 1 \end{pmatrix}, \qquad \mathrm{B} = \frac{1}{3}\begin{pmatrix} -2 & -1 & -2 \\ 2 & -2 & -1 \\ 1 & 2 & -2 \end{pmatrix}$$

You can verify that A and B are both orthogonal, and that $\det \mathrm{A} = 1$, $\det \mathrm{B} = -1$ (Problem 45). Thus A is a rotation matrix while B involves a reflection (and perhaps also a rotation). For A, a vector along the rotation axis is not affected by the transformation so we find the rotation axis by solving the equation $\mathrm{Ar} = \mathrm{r}$. We did this in Section 7, but now you should recognize this as an eigenvector equation. We want the eigenvector corresponding to the eigenvalue 1. By hand or by computer (Problem 45) we find that the eigenvector of A corresponding to $\lambda = 1$ is $(1, 0, 1)$ or $\mathbf{i} + \mathbf{k}$; this is the rotation axis. We want the new z axis to lie along this direction, so we take the elements of the third column of matrix C to be the components of the unit vector $\mathbf{u} = (1, 0, 1)/\sqrt{2}$. For the first column (new x axis) we choose a unit vector perpendicular to the rotation axis, say $\mathbf{v} = (1, 0, -1)/\sqrt{2}$, and for the second column (new y axis), we use the cross product $\mathbf{u} \times \mathbf{v} = (0, 1, 0)$ (so that the new axes form a right-handed orthogonal triad). This gives (Problem 45)

$$\text{(11.32)} \qquad \mathrm{C} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 & 1 \\ 0 & \sqrt{2} & 0 \\ -1 & 0 & 1 \end{pmatrix}, \quad \mathrm{C}^{-1}\mathrm{AC} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Comparing this result with (7.18), we see that $\cos\theta = 0$ and $\sin\theta = -1$, so the rotation is $-90°$ around the axis $\mathbf{i} + \mathbf{k}$ (or, if you prefer, $+90°$ around $-\mathbf{i} - \mathbf{k}$).

▶ **Example 4.** For the matrix B, a vector perpendicular to the reflection plane is reversed in direction by the reflection. Thus we want to solve the equation $\mathrm{Br} = -\mathrm{r}$, that is, to find the eigenvector corresponding to $\lambda = -1$. You can verify (Problem 45) that this is the vector $(1, -1, 1)$ or $\mathbf{i} - \mathbf{j} + \mathbf{k}$. The reflection is through the plane $x - y + z = 0$, and the rotation (if any) is about the vector $\mathbf{i} - \mathbf{j} + \mathbf{k}$. As we did for matrix A, we construct a matrix C from this vector and two perpendicular vectors, to get (Problem 45)

$$\text{(11.33)} \qquad \mathrm{C} = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{3}} \\ \frac{-2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{pmatrix}, \qquad \mathrm{C}^{-1}\mathrm{BC} = \begin{pmatrix} \frac{-1}{2} & \frac{-\sqrt{3}}{2} & 0 \\ \frac{\sqrt{3}}{2} & \frac{-1}{2} & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Compare this with (7.19) to get $\cos\theta = -\frac{1}{2}$, $\sin\theta = \frac{\sqrt{3}}{2}$, so matrix B produces a rotation of $120°$ around $\mathbf{i} - \mathbf{j} + \mathbf{k}$ and a reflection through the plane $x - y + z = 0$.

You may have discovered that matrices A and B have two complex eigenvalues (see Problem 46). The corresponding eigenvectors are also complex, and we didn't use them because this would take us into complex vector space (see Section 10, and Problem 47) and our rotation and reflection problems are in ordinary real 3-dimensional space. (Note also that we did not diagonalize A and B, but just used similarity transformations to display them relative to rotated axes.) However, when all the eigenvalues of an orthogonal matrix are real (see Problem 48), then this process produces a diagonalized matrix with the eigenvalues down the main diagonal.

Example 5. Consider the matrix

$$\mathrm{F} = \frac{1}{7}\begin{pmatrix} 2 & 6 & 3 \\ 6 & -3 & 2 \\ 3 & 2 & -6 \end{pmatrix}. \tag{11.34}$$

You can verify (Problem 49) that $\det \mathrm{F} = 1$, that the rotation axis (eigenvector corresponding to the eigenvalue $\lambda = 1$) is $3\mathbf{i} + 2\mathbf{j} + \mathbf{k}$, and that the other two eigenvalues are $-1, -1$. Then the diagonalized F (relative to axes with the new z axis along the rotation axis) is

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{11.35}$$

Comparing this with equation (7.18), we see that $\cos\theta = -1$, $\sin\theta = 0$, so F produces a rotation of $180°$ about $3\mathbf{i} + 2\mathbf{j} + \mathbf{k}$.

An even easier way to find the rotation angle in this problem is to use the trace of F (Problem 50). From (7.18) and (11.34) we have $2\cos\theta + 1 = -1$. Thus $\cos\theta = -1$, $\theta = 180°$ as before. This method gives $\cos\theta$ for any rotation or reflection matrix, but unless $\cos\theta = \pm 1$, we also need more information (say the value of $\sin\theta$) to determine whether θ is positive or negative.

Powers and Functions of Matrices In Section 6 we found functions of some matrices A for which it was easy to find the powers because they repeated periodically [see equations (6.15) to (6.17)]. When this doesn't happen, it isn't so easy to find powers directly (Problem 58). But it is easy to find powers of a diagonal matrix, and you can also show that (Problem 57)

$$\mathrm{M}^n = \mathrm{C}\mathrm{D}^n\mathrm{C}^{-1}, \quad \text{where} \quad \mathrm{C}^{-1}\mathrm{M}\mathrm{C} = \mathrm{D}, \quad \mathrm{D}\ \text{diagonal}. \tag{11.36}$$

This result is useful not just for evaluating powers and functions of numerical matrices but also for proving theorems (Problem 60).

Example 6. We can show that if, as above, $\mathrm{C}^{-1}\mathrm{M}\mathrm{C} = \mathrm{D}$, then

$$\det e^{\mathrm{M}} = e^{\mathrm{Tr}(\mathrm{M})}. \tag{11.37}$$

As in (6.17) we define e^{M} by its power series. For each term of the series $\mathrm{M}^n = \mathrm{C}\mathrm{D}^n\mathrm{C}^{-1}$ by (11.36), so $e^{\mathrm{M}} = \mathrm{C}e^{\mathrm{D}}\mathrm{C}^{-1}$. By (6.6), the determinant of a product = the product of the determinants, and $\det \mathrm{C}\mathrm{C}^{-1} = 1$, so we have $\det e^{\mathrm{M}} = \det e^{\mathrm{D}}$. Now the matrix e^{D} is diagonal and the diagonal elements are e^{λ_i} where λ_i are the eigenvalues of M. Thus $\det e^{\mathrm{D}} = e^{\lambda_1}e^{\lambda_2}e^{\lambda_3}\cdots = e^{\mathrm{Tr}\,\mathrm{D}}$. But by (9.13), $\mathrm{Tr}\,\mathrm{D} = \mathrm{Tr}(\mathrm{C}\mathrm{C}^{-1}\mathrm{M}) = \mathrm{Tr}\,\mathrm{M}$, so we have (11.37).

Simultaneous Diagonalization Can we diagonalize two (or more) matrices using the same similarity transformation? Sometimes we can, namely if, and only if, they commute. Let's see why this is true. Recall that the diagonalizing C matrix has columns which are mutually orthogonal unit eigenvectors of the matrix being diagonalized. Suppose we can find the same set of eigenvectors for two matrices F and G; then the same C will diagonalize both. So the problem amounts to showing how to find a common set of eigenvectors for F and G if they commute.

▶ **Example 7.** Let's start by diagonalizing F. Suppose r (a column matrix) is the eigenvector corresponding to the eigenvalue λ, that is, $\mathrm{Fr} = \lambda\mathrm{r}$. Multiply this on the left by G and use $\mathrm{GF} = \mathrm{FG}$ (matrices commute) to get

$$\mathrm{GFr} = \lambda\mathrm{Gr}, \qquad \text{or} \qquad \mathrm{F(Gr)} = \lambda(\mathrm{Gr}). \tag{11.38}$$

This says that Gr is an eigenvector of F corresponding to the eigenvalue λ. If λ is not degenerate (that is if there is just one eigenvector corresponding to λ) then Gr must be the same vector as r (except maybe for length), that is, Gr is a multiple of r, or $\mathrm{Gr} = \lambda'\mathrm{r}$. This is the eigenvector equation for G; it says that r is an eigenvector of G. If all eigenvalues of F are non-degenerate, then F and G have the same set of eigenvectors, and so can be diagonalized by the same C matrix.

▶ **Example 8.** Now suppose that there are two (or more) linearly independent eigenvectors corresponding to the eigenvalue λ of F. Then every vector in the degenerate eigenspace corresponding to λ is an eigenvector of matrix F (see discussion of degeneracy above). Next consider matrix G. Corresponding to all non-degenerate F eigenvalues we already have the same set of eigenvectors for G as for F. So we just have to find the eigenvectors of G in the degenerate eigenspace of F. Since all vectors in this subspace are eigenvectors of F, we are free to choose ones which are eigenvectors of G. Thus we now have the same set of eigenvectors for both matrices, and so we can construct a C matrix which will diagonalize both F and G. For the converse, see Problem 62.

▶ PROBLEMS, SECTION 11

1. Verify (11.7). Also verify (11.12) and find the corresponding different C in (11.11). *Hint:* To find C, start with (11.12) instead of (11.7) and follow through the method of getting (11.10) from (11.7).

2. Verify that the two eigenvectors in (11.8) are perpendicular, and that C in (11.10) satisfies the condition (7.9) for an orthogonal matrix.

3. **(a)** If C is orthogonal and M is symmetric, show that $\mathrm{C}^{-1}\mathrm{MC}$ is symmetric.

(b) If C is orthogonal and M antisymmetric, show that $\mathrm{C}^{-1}\mathrm{MC}$ is antisymmetric.

4. Find the inverse of the rotation matrix in (7.13); you should get C in (11.14). Replace θ by $-\theta$ in (7.13) to see that the matrix C corresponds to a rotation through $-\theta$.

5. Show that the C matrix in (11.10) does represent a rotation by finding the rotation angle. Write equations (7.13) and (11.13) for this rotation.

6. Show that if C is a matrix whose columns are the components (x_1, y_1) and (x_2, y_2) of two perpendicular vectors each of unit length, then C is an orthogonal matrix. *Hint:* Find $\mathrm{C^TC}$.

7. Generalize Problem 6 to three dimensions; to n dimensions.

8. Show that under the transformation (11.1), all points (x, y) on a given straight line through the origin go into points (X, Y) on another straight line through the origin. *Hint:* Solve (11.1) for x and y in terms of X and Y and substitute into the equation $y = mx$ to get an equation $Y = kX$, where k is a constant. *Further hint:* If $\mathrm{R} = \mathrm{Mr}$, then $\mathrm{r} = \mathrm{M}^{-1}\mathrm{R}$.

9. Show that $\det(\mathrm{C}^{-1}\mathrm{MC}) = \det \mathrm{M}$. *Hints:* See (6.6). What is the product of $\det(\mathrm{C}^{-1})$ and $\det \mathrm{C}$? Thus show that the product of the eigenvalues of M is equal to $\det \mathrm{M}$.

10. Show that $\mathrm{Tr}(\mathrm{C}^{-1}\mathrm{MC}) = \mathrm{Tr}\,\mathrm{M}$. *Hint:* See (9.13). Thus show that the sum of the eigenvalues of M is equal to $\mathrm{Tr}\,\mathrm{M}$.

11. Find the inverse of the transformation $x' = 2x - 3y$, $y' = x + y$, that is, find x, y in terms of x', y'. (*Hint:* Use matrices.) Is the transformation orthogonal?

Find the eigenvalues and eigenvectors of the following matrices. Do some problems by hand to be sure you understand what the process means. Then check your results by computer.

12. $\begin{pmatrix} 1 & 3 \\ 2 & 2 \end{pmatrix}$ **13.** $\begin{pmatrix} 2 & 2 \\ 2 & -1 \end{pmatrix}$ **14.** $\begin{pmatrix} 3 & -2 \\ -2 & 0 \end{pmatrix}$

15. $\begin{pmatrix} 2 & 3 & 0 \\ 3 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ **16.** $\begin{pmatrix} 2 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & -1 \end{pmatrix}$ **17.** $\begin{pmatrix} 5 & 0 & 2 \\ 0 & 3 & 0 \\ 2 & 0 & 5 \end{pmatrix}$

18. $\begin{pmatrix} -1 & 1 & 3 \\ 1 & 2 & 0 \\ 3 & 0 & 2 \end{pmatrix}$ **19.** $\begin{pmatrix} 1 & 2 & 2 \\ 2 & 3 & 0 \\ 2 & 0 & 3 \end{pmatrix}$ **20.** $\begin{pmatrix} -1 & 2 & 1 \\ 2 & 3 & 0 \\ 1 & 0 & 3 \end{pmatrix}$

21. $\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}$ **22.** $\begin{pmatrix} -3 & 2 & 2 \\ 2 & 1 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ **23.** $\begin{pmatrix} 13 & 4 & -2 \\ 4 & 13 & -2 \\ -2 & -2 & 10 \end{pmatrix}$

24. $\begin{pmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{pmatrix}$ **25.** $\begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \\ -1 & 1 & -1 \end{pmatrix}$ **26.** $\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$

Let each of the following matrices M describe a deformation of the (x, y) plane. For each given M find: the eigenvalues and eigenvectors of the transformation, the matrix C which diagonalizes M and specifies the rotation to new axes (x', y') along the eigenvectors, and the matrix D which gives the deformation relative to the new axes. Describe the deformation relative to the new axes.

27. $\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$ **28.** $\begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}$ **29.** $\begin{pmatrix} 3 & 4 \\ 4 & 9 \end{pmatrix}$

30. $\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$ **31.** $\begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}$ **32.** $\begin{pmatrix} 6 & -2 \\ -2 & 3 \end{pmatrix}$

33. Find the eigenvalues and eigenvectors of the real symmetric matrix

$$\mathrm{M} = \begin{pmatrix} A & H \\ H & B \end{pmatrix}.$$

Show that the eigenvalues are real and the eigenvectors are perpendicular.

34. By multiplying out $M = CDC^{-1}$ where C is the rotation matrix (11.14) and D is the diagonal matrix

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix},$$

show that if M can be diagonalized by a rotation, then M is symmetric.

35. The characteristic equation for a second-order matrix M is a quadratic equation. We have considered in detail the case in which M is a real symmetric matrix and the roots of the characteristic equation (eigenvalues) are real, positive, and unequal. Discuss some other possibilities as follows:

(a) M real and symmetric, eigenvalues real, one positive and one negative. Show that the plane is reflected in one of the eigenvector lines (as well as stretched or shrunk). Consider as a simple special case

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

(b) M real and symmetric, eigenvalues equal (and therefore real). Show that M must be a multiple of the unit matrix. Thus show that the deformation consists of dilation or shrinkage in the radial direction (the same in all directions) with no rotation (and reflection in the origin if the root is negative).

(c) M real, *not* symmetric, eigenvalues real and not equal. Show that in this case the eigenvectors are not orthogonal. *Hint:* Find their dot product.

(d) M real, *not* symmetric, eigenvalues complex. Show that all vectors are rotated, that is, there are no (real) eigenvectors which are unchanged in direction by the transformation. Consider the characteristic equation of a rotation matrix as a special case.

36. Verify the eigenvalues and eigenvectors of matrix M in (11.20). Find some other pairs of orthogonal eigenvectors in the $\lambda = -3$ eigenplane.

37. Starting with (11.23), obtain (11.24). *Hints:* Take the transpose conjugate (dagger) of the first equation in (11.23), (remember that H is Hermitian and the λ's are real) and multiply on the right by r_2. Multiply the second equation in (11.23) on the left by $r_1^\dagger$.

38. Verify equation (11.25). *Hint:* Remember from Section 9 that the transpose conjugate (dagger) of a product of matrices is the product of the transpose conjugates in reverse order and that $U^\dagger = U^{-1}$. Also remember that we have assumed real eigenvalues, so D is a real diagonal matrix.

39. Write out the detailed proof of (11.27). *Hint:* Follow the proof of (11.26) in equations (11.21) to (11.25), replacing the Hermitian matrix H by a symmetric matrix M which is real. However, don't assume that the eigenvalues λ are real until you prove it.

40. Verify the details as indicated in diagonalizing H in (11.29).

Verify that each of the following matrices is Hermitian. Find its eigenvalues and eigenvectors, write a unitary matrix U which diagonalizes H by a similarity transformation, and show that $U^{-1}HU$ is the diagonal matrix of eigenvalues.

41. $\begin{pmatrix} 2 & i \\ -i & 2 \end{pmatrix}$

42. $\begin{pmatrix} 3 & 1-i \\ 1+i & 2 \end{pmatrix}$

43. $\begin{pmatrix} 1 & 2i \\ -2i & -2 \end{pmatrix}$

44. $\begin{pmatrix} -2 & 3+4i \\ 3-4i & -2 \end{pmatrix}$

45. Verify the details in the discussion of the matrices in (11.31).

46. We have seen that an orthogonal matrix with determinant 1 has at least one eigenvalue $= 1$, and an orthogonal matrix with determinant $= -1$ has at least one eigenvalue $= -1$. Show that the other two eigenvalues in both cases are $e^{i\theta}$, $e^{-i\theta}$, which, of course, includes the real values 1 (when $\theta = 0$), and -1 (when $\theta = \pi$). *Hint:* See Problem 9, and remember that rotations and reflections do not change the length of vectors so eigenvalues must have absolute value $= 1$.

47. Find a unitary matrix U which diagonalizes A in (11.31) and verify that $\mathrm{U}^{-1}\mathrm{AU}$ is diagonal with the eigenvalues down the main diagonal.

48. Show that an orthogonal matrix M with all real eigenvalues is symmetric. *Hints*: Method 1. When the eigenvalues are real, so are the eigenvectors, and the unitary matrix which diagonalizes M is orthogonal. Use (11.27). Method 2. From Problem 46, note that the only real eigenvalues of an orthogonal M are ± 1. Thus show that $\mathrm{M} = \mathrm{M}^{-1}$. Remember that M is orthogonal to show that $\mathrm{M} = \mathrm{M}^{\mathrm{T}}$.

49. Verify the results for F in the discussion of (11.34).

50. Show that the trace of a rotation matrix equals $2\cos\theta + 1$ where θ is the rotation angle, and the trace of a reflection matrix equals $2\cos\theta - 1$. *Hint*: See equations (7.18) and (7.19), and Problem 10.

Show that each of the following matrices is orthogonal and find the rotation and/or reflection it produces as an operator acting on vectors. If a rotation, find the axis and angle; if a reflection, find the reflecting plane and the rotation, if any, about the normal to that plane.

51. $\dfrac{1}{11}\begin{pmatrix} 2 & 6 & 9 \\ 6 & 7 & -6 \\ 9 & -6 & 2 \end{pmatrix}$

52. $\dfrac{1}{2}\begin{pmatrix} -1 & -1 & \sqrt{2} \\ 1 & 1 & \sqrt{2} \\ \sqrt{2} & -\sqrt{2} & 0 \end{pmatrix}$

53. $\dfrac{1}{3}\begin{pmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{pmatrix}$

54. $\dfrac{1}{2}\begin{pmatrix} 1 & \sqrt{2} & -1 \\ \sqrt{2} & 0 & \sqrt{2} \\ 1 & -\sqrt{2} & -1 \end{pmatrix}$

55. $\dfrac{1}{9}\begin{pmatrix} -1 & 8 & 4 \\ -4 & -4 & 7 \\ -8 & 1 & -4 \end{pmatrix}$

56. $\dfrac{1}{2\sqrt{2}}\begin{pmatrix} 2 & \sqrt{2} & \sqrt{2} \\ -\sqrt{2} & 1+\sqrt{2} & 1-\sqrt{2} \\ -\sqrt{2} & 1-\sqrt{2} & 1+\sqrt{2} \end{pmatrix}$

57. Show that if D is a diagonal matrix, then D^n is the diagonal matrix with elements equal to the n^{th} power of the elements of D. Also show that if $\mathrm{D} = \mathrm{C}^{-1}\mathrm{MC}$, then $\mathrm{D}^n = \mathrm{C}^{-1}\mathrm{M}^n\mathrm{C}$, so $\mathrm{M}^n = \mathrm{CD}^n\mathrm{C}^{-1}$. *Hint*: For $n = 2$, $(\mathrm{C}^{-1}\mathrm{MC})^2 = \mathrm{C}^{-1}\mathrm{MCC}^{-1}\mathrm{MC}$; what is CC^{-1}?

58. Note in Section 6 [see (6.15)] that, for the given matrix A, we found $\mathrm{A}^2 = -\mathrm{I}$, so it was easy to find all the powers of A. It is not usually this easy to find high powers of a matrix directly. Try it for the square matrix M in equation (11.1). Then use the method outlined in Problem 57 to find M^4, M^{10}, e^{M}.

59. Repeat the last part of Problem 58 for the matrix $\mathrm{M} = \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}$.

60. The Caley-Hamilton theorem states that "A matrix satisfies its own characteristic equation." Verify this theorem for the matrix M in equation (11.1). *Hint*: Substitute the matrix M for λ in the characteristic equation (11.4) and verify that you have a correct matrix equation. *Further hint*: Don't do all the arithmetic. Use (11.36) to write the left side of your equation as $\mathrm{C}(\mathrm{D}^2 - 7\mathrm{D} + 6)\mathrm{C}^{-1}$ and show that the parenthesis $= 0$. Remember that, by definition, the eigenvalues satisfy the characteristic equation.

61. At the end of Section 9 we proved that if H is a Hermitian matrix, then the matrix $e^{i\mathrm{H}}$ is unitary. Give another proof by writing $\mathrm{H} = \mathrm{CDC}^{-1}$, remembering that now C is unitary and the eigenvalues in D are real. Show that $e^{i\mathrm{D}}$ is unitary and that $e^{i\mathrm{H}}$ is a product of three unitary matrices. See Problem 9.17d.

62. Show that if matrices F and G can be diagonalized by the same C matrix, then they commute. *Hint*: Do diagonal matrices commute?

12. APPLICATIONS OF DIAGONALIZATION

We next consider some examples of the use of the diagonalization process. A central conic section (ellipse or hyperbola) with center at the origin has the equation

$$Ax^2 + 2Hxy + By^2 = K, \tag{12.1}$$

where A, B, H and K are constants. In matrix form this can be written

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} A & H \\ H & B \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = K \quad \text{or} \quad \begin{pmatrix} x & y \end{pmatrix} \mathrm{M} \begin{pmatrix} x \\ y \end{pmatrix} = K \tag{12.2}$$

if we call

$$\begin{pmatrix} A & H \\ H & B \end{pmatrix} = \mathrm{M}$$

(as you can verify by multiplying out the matrices). We want to choose the principal axes of the conic as our reference axes in order to write the equation in simpler form. Consider Figure 11.2; let the axes (x', y') be rotated by some angle θ from (x, y). Then the (x, y) and (x', y') coordinates of a point are related by (11.13) or (11.14):

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} x \\ y \end{pmatrix} = \mathrm{C} \begin{pmatrix} x' \\ y' \end{pmatrix}. \tag{12.3}$$

By (9.11) the transpose of (12.3) is

$$\begin{aligned} &\begin{pmatrix} x & y \end{pmatrix} = \begin{pmatrix} x' & y' \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \quad \text{or} \\ &\begin{pmatrix} x & y \end{pmatrix} = \begin{pmatrix} x' & y' \end{pmatrix} \mathrm{C}^{\mathrm{T}} = \begin{pmatrix} x' & y' \end{pmatrix} \mathrm{C}^{-1} \end{aligned} \tag{12.4}$$

since C is an orthogonal matrix. Substituting (12.3) and (12.4) into (12.2), we get

$$\begin{pmatrix} x' & y' \end{pmatrix} \mathrm{C}^{-1}\mathrm{MC} \begin{pmatrix} x' \\ y' \end{pmatrix} = K. \tag{12.5}$$

If C is the matrix which diagonalizes M, then (12.5) is the equation of the conic relative to its principal axes.

▶ **Example 1.** Consider the conic

$$5x^2 - 4xy + 2y^2 = 30. \tag{12.6}$$

In matrix form this can be written

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 30. \tag{12.7}$$

We have here the same matrix,

$$\mathrm{M} = \begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix},$$

whose eigenvalues we found in Section 11. In that section we found a C such that

$$\mathrm{C}^{-1}\mathrm{MC} = \mathrm{D} = \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix}.$$

Then the equation (12.5) of the conic relative to principal axes is

$$\begin{pmatrix} x' & y' \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} = x'^2 + 6y'^2 = 30. \tag{12.8}$$

Observe that changing the order of 1 and 6 in D would give $6x'^2 + y'^2 = 30$ as the new equation of the ellipse instead of (12.8). This amounts simply to interchanging the x' and y' axes.

By comparing the matrix C of the unit eigenvectors in (11.10) with the rotation matrix in (11.14), we see that the rotation angle θ (Figure 11.3) from the original axes (x, y) to the principal axes (x', y') is

$$\theta = \arccos \frac{1}{\sqrt{5}}. \tag{12.9}$$

Notice that in writing the conic section equation in matrix form (12.2) and (12.7), we split the xy term evenly between the two nondiagonal elements of the matrix; this made M symmetric. Recall (end of Section 11) that M can be diagonalized by a similarity transformation $\mathrm{C}^{-1}\mathrm{MC}$ with C an orthogonal matrix (that is, by a rotation of axes) if and only if M is symmetric. We choose M symmetric (by splitting the xy term in half) to make our process work.

Although for simplicity we have been working in two dimensions, the same ideas apply to three (or more) dimensions (that is, three or more variables). As we have said (Section 10), although we can represent only three coordinates in physical space, it is very convenient to use the same geometrical terminology even though the number of variables is greater than three. Thus if we diagonalize a matrix of any order, we still use the terms eigenvalues, eigenvectors, principal axes, rotation to principal axes, etc.

▶ **Example 2.** Rotate to principal axes the quadric surface

$$x^2 + 6xy - 2y^2 - 2yz + z^2 = 24.$$

In matrix form this equation is

$$\begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} 1 & 3 & 0 \\ 3 & -2 & -1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 24.$$

The characteristic equation of this matrix is

$$\begin{vmatrix} 1-\lambda & 3 & 0 \\ 3 & -2-\lambda & -1 \\ 0 & -1 & 1-\lambda \end{vmatrix} = 0 = -\lambda^3 + 13\lambda - 12$$

$$= -(\lambda - 1)(\lambda + 4)(\lambda - 3).$$

The characteristic values are

$$\lambda = 1, \qquad \lambda = -4, \qquad \lambda = 3.$$

Relative to the principal axes (x', y', z') the quadric surface equation becomes

$$\begin{pmatrix} x' & y' & z' \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = 24$$

or

$${x'}^2 - 4{y'}^2 + 3{z'}^2 = 24.$$

From this equation we can identify the quadric surface (hyperboloid of one sheet) and sketch its size and shape using (x', y', z') axes without finding their relation to the original (x, y, z) axes. However, if we do want to know the relation between the two sets of axes, we find the C matrix in the following way. Recall from Section 11 that C is the matrix whose columns are the components of the unit eigenvectors. One of the eigenvectors can be found by substituting the eigenvalue $\lambda = 1$ into the equations

$$\begin{pmatrix} 1 & 3 & 0 \\ 3 & -2 & -1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \lambda x \\ \lambda y \\ \lambda z \end{pmatrix}$$

and solving for x, y, z. Then $\mathbf{i}x + \mathbf{j}y + \mathbf{k}z$ is an eigenvector corresponding to $\lambda = 1$, and by dividing it by its magnitude we get a *unit* eigenvector (Problem 8). Repeating this process for each of the other values of λ, we get the following three unit eigenvectors:

$$\begin{aligned} &\left(\frac{1}{\sqrt{10}}, 0, \frac{3}{\sqrt{10}}\right) && \text{when} \quad \lambda = 1; \\ &\left(\frac{-3}{\sqrt{35}}, \frac{5}{\sqrt{35}}, \frac{1}{\sqrt{35}}\right) && \text{when} \quad \lambda = -4; \\ &\left(\frac{-3}{\sqrt{14}}, \frac{-2}{\sqrt{14}}, \frac{1}{\sqrt{14}}\right) && \text{when} \quad \lambda = 3. \end{aligned}$$

Then the rotation matrix C is

$$\mathrm{C} = \begin{pmatrix} \dfrac{1}{\sqrt{10}} & \dfrac{-3}{\sqrt{35}} & \dfrac{-3}{\sqrt{14}} \\ 0 & \dfrac{5}{\sqrt{35}} & \dfrac{-2}{\sqrt{14}} \\ \dfrac{3}{\sqrt{10}} & \dfrac{1}{\sqrt{35}} & \dfrac{1}{\sqrt{14}} \end{pmatrix}$$

The numbers in C are the cosines of the nine angles between the (x, y, z) and (x', y', z') axes. (Compare Figure 11.3 and the discussion of it.)

A useful physical application of this method occurs in discussing vibrations. We illustrate this with a simple problem.

▶ **Example 3.** Find the characteristic vibration frequencies for the system of masses and springs shown in Figure 12.1.

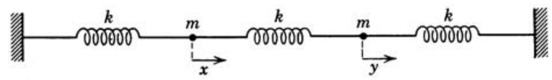


Figure 12.1

Let x and y be the coordinates of the two masses at time t relative to their equilibrium positions, as shown in Figure 12.1. We want to write the equations of motion (mass times acceleration = force) for the two masses (see Chapter 2, end of Section 16). We *can* just write the forces by inspection as we did in Chapter 2, but for more complicated problems it is useful to have a systematic method. First write the potential energy; for a spring this is $V = \frac{1}{2}ky^2$ where y is the compression or extension of the spring from its equilibrium length. Then the force exerted on a mass attached to the spring is $-ky = -dV/dy$. If V is a function of two (or more) variables, say x and y as in Figure 12.1, then the forces on the two masses are $-\partial V/\partial x$ and $-\partial V/\partial y$ (and so on for more variables). For Figure 12.1, the extension or compression of the middle spring is $x - y$ so its potential energy is $\frac{1}{2}k(x-y)^2$. For the other two springs, the potential energies are $\frac{1}{2}kx^2$ and $\frac{1}{2}ky^2$ so the total potential energy is

$$V = \frac{1}{2}kx^2 + \frac{1}{2}k(x-y)^2 + \frac{1}{2}ky^2 = k(x^2 - xy + y^2). \tag{12.10}$$

In writing the equations of motion it is convenient to use a dot to indicate a time derivative (as we often use a prime to mean an x derivative). Thus $\dot{x} = dx/dt$, $\ddot{x} = d^2x/dt^2$, etc. Then the equations of motion are

$$\begin{cases} m\ddot{x} = -\partial V/\partial x = -2kx + ky, \\ m\ddot{y} = -\partial V/\partial y = \quad kx - 2ky. \end{cases} \tag{12.11}$$

In a *normal* or *characteristic* mode of vibration, the x and y vibrations have the same frequency. As in Chapter 2, equations (16.22), we assume solutions $x = x_0 e^{i\omega t}$, $y = y_0 e^{i\omega t}$, with the same frequency ω for both x and y. [Or, if you prefer, we could replace $e^{i\omega t}$ by $\sin\omega t$ or $\cos\omega t$ or $\sin(\omega t + \alpha)$, etc.] Note that (for any of these solutions),

$$\ddot{x} = -\omega^2 x, \quad \text{and} \quad \ddot{y} = -\omega^2 y. \tag{12.12}$$

Substituting (12.12) into (12.11) we get (Problem 10)

$$\begin{cases} -m\omega^2 x = -2kx + ky, \\ -m\omega^2 y = \quad kx - 2ky. \end{cases} \tag{12.13}$$

In matrix form these equations are

$$\lambda \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{with} \quad \lambda = \frac{m\omega^2}{k}. \tag{12.14}$$

Note that this is an eigenvalue problem (see Section 11). To find the eigenvalues λ, we write

$$\begin{vmatrix} 2-\lambda & -1 \\ -1 & 2-\lambda \end{vmatrix} = 0 \tag{12.15}$$

and solve for λ to find $\lambda = 1$ or $\lambda = 3$. Thus [by the definition of λ in (12.14)] the characteristic frequencies are

$$\omega_1 = \sqrt{\frac{k}{m}} \quad \text{and} \quad \omega_2 = \sqrt{\frac{3k}{m}}. \tag{12.16}$$

The eigenvectors (not normalized) corresponding to these eigenvalues are:

(12.17) For $\lambda = 1$: $y = x$ or $\mathbf{r} = (1, 1)$; for $\lambda = 3$: $y = -x$ or $\mathbf{r} = (1, -1)$.

Thus at frequency ω_1 (with $y = x$), the two masses oscillate back and forth together like this $\rightarrow\rightarrow$ and then like this $\leftarrow\leftarrow$. At frequency ω_2 (with $y = -x$), they oscillate in opposite directions like this $\leftarrow\rightarrow$ and then like this $\rightarrow\leftarrow$. These two especially simple ways in which the system can vibrate, each involving just one vibration frequency, are called the characteristic (or normal) modes of vibration; the corresponding frequencies are called the characteristic (or normal) frequencies of the system.

The problem we have just done shows an important method which can be used in many different applications. There are numerous examples of vibration problems in physics—in acoustics: the vibrations of strings of musical instruments, of drumheads, of the air in organ pipes or in a room; in mechanics and its engineering applications: vibrations of mechanical systems all the way from the simple pendulum to complicated structures like bridges and airplanes; in electricity: the vibrations of radio waves, of electric currents and voltages as in a tuned radio; and so on. In such problems, it is often useful to find the characteristic vibration frequencies of the system under consideration and the characteristic modes of vibration. More complicated vibrations can then be discussed as combinations of these simpler normal modes of vibration.

▶ **Example 4.** In Example 3 and Figure 12.1, the two masses were equal and all the spring constants were the same. Changing the spring constants to different values doesn't cause any problems but when the masses are different, there is a possible difficulty which we want to discuss. Consider an array of masses and springs as in Figure 12.1 but with the following masses and spring constants: $2k$, $2m$, $6k$, $3m$, $3k$. We want to find the characteristic frequencies and modes of vibration. Following our work in Example 3, we write the potential energy V, find the forces, write the equations of motion, and substitute $\ddot{x} = -\omega^2 x$, and $\ddot{y} = -\omega^2 y$, in order to find the characteristic frequencies. (*Do the details*: Problem 11.)

$$V = \frac{1}{2}2kx^2 + \frac{1}{2}6k(x-y)^2 + \frac{1}{2}3ky^2 = \frac{1}{2}k(8x^2 - 12xy + 9y^2) \tag{12.18}$$

$$\begin{cases} 2m\ddot{x} = -\partial V/\partial x, \\ 3m\ddot{y} = -\partial V/\partial y, \end{cases} \quad \text{or} \quad \begin{cases} -2m\omega^2 x = -k(8x - 6y), \\ -3m\omega^2 y = -k(-6x + 9y). \end{cases} \tag{12.19}$$

Next divide each equation by its mass and write the equations in matrix form.

$$\omega^2 \begin{pmatrix} x \\ y \end{pmatrix} = \frac{k}{m} \begin{pmatrix} 4 & -3 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{12.20}$$

With $\lambda = m\omega^2/k$, the eigenvalues of the square matrix are $\lambda = 1$ and $\lambda = 6$. Thus the characteristic frequencies of vibration are

$$\omega_1 = \sqrt{\frac{k}{m}} \quad \text{and} \quad \omega_2 = \sqrt{\frac{6k}{m}}. \tag{12.21}$$

The corresponding eigenvectors are:

(12.22) For $\lambda = 1$: $y = x$ or $\mathbf{r} = (1, 1)$; for $\lambda = 6$: $3y = -2x$ or $\mathbf{r} = (3, -2)$.

Thus at frequency ω_1 the two masses oscillate back and forth together with equal amplitudes like this $\leftarrow\leftarrow$ and then like this $\rightarrow\rightarrow$. At frequency ω_2 the two masses oscillate in opposite directions with amplitudes in the ratio 3 to 2 like this $\leftarrow\rightarrow$ and then like this $\rightarrow\leftarrow$.

Now we seem to have solved the problem; where is the difficulty? Note that the square matrix in (12.20) is not symmetric [and compare (12.14) where the square matrix was symmetric]. In Section 11 we discussed the fact that (for real matrices) only symmetric matrices have orthogonal eigenvectors and can be diagonalized by an orthogonal transformation. Here note that the eigenvectors in Example 3 were orthogonal [dot product of $(1, 1)$ and $(1, -1)$ is zero] but the eigenvectors for (12.20) are not orthogonal [dot product of $(1, 1)$ and $(3, -2)$ is not zero]. If we want orthogonal eigenvectors, we can make the change of variables (also see Example 6)

$$X = x\sqrt{2}, \quad Y = y\sqrt{3}, \tag{12.23}$$

where the constants are the square roots of the numerical factors in the masses $2m$ and $3m$. (Note that geometrically this just amounts to different changes in scale along the two axes, not to a rotation.) Then (12.20) becomes

$$\omega^2 \begin{pmatrix} X \\ Y \end{pmatrix} = \frac{k}{m} \begin{pmatrix} 4 & -\sqrt{6} \\ -\sqrt{6} & 3 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}. \tag{12.24}$$

By inspection we see that the characteristic equation for the square matrix in (12.24) is the same as the characteristic equation for (12.20) so the eigenvalues and the characteristic frequencies are the same as before (as they must be by physical reasoning). However the (12.24) matrix is symmetric and so we know that its eigenvectors are orthogonal. By direct substitution of (12.23) into (12.22), [or by solving for the eigenvectors in the (12.24) matrix] we find the eigenvectors in the X, Y coordinates:

(12.25) For $\lambda = 1$: $\mathbf{R} = (X, Y) = (\sqrt{2}, -\sqrt{3})$; for $\lambda = 6$: $\mathbf{R} = (3\sqrt{2}, 2\sqrt{3})$.

As expected, these eigenvectors are orthogonal.

Example 5. Let's consider a model of a linear triatomic molecule in which we approximate the forces between the atoms by forces due to springs (Figure 12.2).

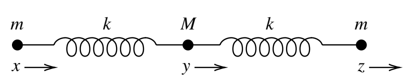


Figure 12.2

As in Example 3, let x, y, z be the coordinates of the three masses relative to their equilibrium positions. We want to find the characteristic vibration frequencies of the molecule. Following our work in Examples 3 and 4, we find (Problem 12)

$$V = \tfrac{1}{2}k(x-y)^2 + \tfrac{1}{2}k(y-z)^2 = \tfrac{1}{2}k(x^2+2y^2+z^2-2xy-2yz), \tag{12.26}$$

$$\begin{cases} m\ddot{x} = -\partial V/\partial x = -k(x-y), \\ M\ddot{y} = -\partial V/\partial y = -k(2y-x-z), \\ m\ddot{z} = -\partial V/\partial z = -k(z-y), \end{cases}$$
$$\text{or} \tag{12.27}$$
$$\begin{cases} -m\omega^2 x = -k(x-y), \\ -M\omega^2 y = -k(2y-x-z), \\ -m\omega^2 z = -k(z-y). \end{cases}$$

We are going to consider several different ways of solving this problem in order to learn some useful techniques. First of all, if we add the three equations we get

$$m\ddot{x} + M\ddot{y} + m\ddot{z} = 0. \tag{12.28}$$

Physically (12.28) says that the center of mass is at rest or moving at constant speed (that is, has zero acceleration). Since we are just interested in vibrational motion, let's assume that the center of mass is at rest at the origin. Then we have $mx + My + mz = 0$. Solving this equation for y gives

$$y = -\frac{m}{M}(x+z). \tag{12.29}$$

Substitute (12.29) into the second set of equations in (12.27) to get the x and z equations

$$\begin{aligned} -m\omega^2 x &= -k(1+\frac{m}{M})x - k\frac{m}{M}z, \\ -m\omega^2 z &= -k\frac{m}{M}x - k(1+\frac{m}{M})z. \end{aligned} \tag{12.30}$$

In matrix form equations (12.30) become [compare (12.14)]

$$\lambda\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1+\frac{m}{M} & \frac{m}{M} \\ \frac{m}{M} & 1+\frac{m}{M} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \quad \text{with} \quad \lambda = \frac{m\omega^2}{k}. \tag{12.31}$$

We solve this eigenvalue problem to find

$$\omega_1 = \sqrt{\frac{k}{m}}, \qquad \omega_2 = \sqrt{\frac{k}{m}\left(1+\frac{2m}{M}\right)}. \tag{12.32}$$

For ω_1 we find $z = -x$, and consequently by (12.29), $y = 0$. For ω_2, we find $z = x$ and so $y = -\frac{2m}{M}x$. Thus at frequency ω_1, the central mass M is at rest and the two masses m vibrate in opposite directions like this $\leftarrow m\ M\ m\rightarrow$ and then like this $m\rightarrow\ M\ \leftarrow m$. At the higher frequency ω_2, the central mass M moves in one direction while the two masses m move in the opposite direction, first like this $m\rightarrow \leftarrow M\ m\rightarrow$ and then like this $\leftarrow m\ M\rightarrow \leftarrow m$.

Now suppose that we had not thought about eliminating the translational motion and had set this problem up as a 3 variable problem. Let's go back to the second set of equations in (12.27), and divide the x and z equations by m and the y equation by M. Then in matrix form these equations can be written as

$$\omega^2 \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \frac{k}{m} \begin{pmatrix} 1 & -1 & 0 \\ \frac{-m}{M} & \frac{2m}{M} & \frac{-m}{M} \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{12.33}$$

With $\lambda = m\omega^2/k$, the eigenvalues of the square matrix are $\lambda = 0, 1, 1 + \frac{2m}{M}$, and the corresponding eigenvectors are (check these)

$$\begin{aligned} &\text{For } \lambda = 0,\ \mathbf{r} = (1,1,1); \\ &\text{for } \lambda = 1,\ \mathbf{r} = (1,0,-1); \\ &\text{for } \lambda = 1 + \frac{2m}{M},\ \mathbf{r} = (1, -\frac{2m}{M}, 1). \end{aligned} \tag{12.34}$$

We recognize the $\lambda = 0$ solution as corresponding to translation both because $\omega = 0$ (so there is no vibration), and because $\mathbf{r} = (1,1,1)$ says that any motion is the same for all three masses. The other two modes of vibration are the same ones we had above. We note that the square matrix in (12.33) is not symmetric and so, as expected, the eigenvectors in (12.34) are not an orthogonal set. However, the last two (which correspond to vibrations) are orthogonal so if we are just interested in modes of vibration we can ignore the translation eigenvector. If we want to consider all motion of the molecule along its axis (both translation and vibration), and want an orthogonal set of eigenvectors, we can make the change of variables discussed in Example 4, namely

$$X = x, \qquad Y = y\sqrt{\frac{M}{m}}, \qquad Z = z. \tag{12.35}$$

Then the eigenvectors become

$$(1, \sqrt{M/m}\,, 1), \quad (1,0,-1), \quad (1, -2\sqrt{m/M}\,, 1) \tag{12.36}$$

which are an orthogonal set. The first eigenvector (corresponding to translation) may seem confusing, looking as if the central mass M doesn't move with the others (as it must for pure translation). But remember from Example 4 that changes of variable like (12.23) and (12.35) correspond to changes of scale, so in the XYZ system we are not using the same measuring stick to find the position of the central mass as for the other two masses. Their physical displacements are actually all the same.

Example 6. Let's consider Example 4 again in order to illustrate a very compact form for the eigenvalue equation. Satisfy yourself (Problem 13) that we can write the potential energy V in (12.18) as

$$V = \frac{1}{2}k\mathrm{r}^{\mathrm{T}}\mathrm{V}\mathrm{r} \quad \text{where} \quad V = \begin{pmatrix} 8 & -6 \\ -6 & 9 \end{pmatrix}, \quad \mathrm{r} = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \mathrm{r}^{\mathrm{T}} = \begin{pmatrix} x & y \end{pmatrix}. \tag{12.37}$$

Similarly the kinetic energy $T = \frac{1}{2}(2m\dot{x}^2 + 3m\dot{y}^2)$ can be written as

$$T = \frac{1}{2}m\dot{\mathrm{r}}^{\mathrm{T}}\mathrm{T}\dot{\mathrm{r}} \quad \text{where} \quad \mathrm{T} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}, \quad \dot{\mathrm{r}} = \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix}, \quad \dot{\mathrm{r}}^{\mathrm{T}} = \begin{pmatrix} \dot{x} & \dot{y} \end{pmatrix}. \tag{12.38}$$

(Notice that the T matrix is diagonal and is a unit matrix when the masses are equal; otherwise T has the mass factors along the main diagonal and zeros elsewhere.) Now using the matrices T and V, we can write the equations of motion (12.19) as

$$m\omega^2 \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = k \begin{pmatrix} 8 & -6 \\ -6 & 9 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{or}$$

(12.39) $$\lambda \mathrm{Tr} = \mathrm{Vr} \quad \text{where} \quad \lambda = \frac{m\omega^2}{k}.$$

We can think of (12.39) as the basic eigenvalue equation. If T is a unit matrix, then we just have $\lambda \mathrm{r} = \mathrm{Vr}$ as in (12.14). If not, then we can multiply (12.39) by T^{-1} to get

(12.40) $$\lambda \mathrm{r} = \mathrm{T}^{-1}\mathrm{Vr} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/3 \end{pmatrix} \begin{pmatrix} 8 & -6 \\ -6 & 9 \end{pmatrix} \mathrm{r} = \begin{pmatrix} 4 & -3 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

as in (12.20). However, we see that this matrix is not symmetric and so the eigenvectors will not be orthogonal. If we want the eigenvectors to be orthogonal as in (12.23), we choose new variables so that the T matrix is the unit matrix, that is variables X and Y so that

(12.41) $$T = \tfrac{1}{2}(2m\dot{x}^2 + 3m\dot{y}^2) = \tfrac{1}{2}m(\dot{X}^2 + \dot{Y}^2).$$

But this means that we want $X^2 = 2x^2$ and $Y^2 = 3y^2$ as in (12.23), or in matrix form,

$$\mathrm{R} = \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} x\sqrt{2} \\ y\sqrt{3} \end{pmatrix} = \begin{pmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{3} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathrm{T}^{1/2}\mathrm{r} \quad \text{or}$$

(12.42) $$\mathrm{r} = \mathrm{T}^{-1/2}\mathrm{R} = \begin{pmatrix} 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{3} \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

Substituting (12.42) into (12.39), we get $\lambda \mathrm{T}\mathrm{T}^{-1/2}\mathrm{R} = \mathrm{V}\mathrm{T}^{-1/2}\mathrm{R}$. Then multiplying on the left by $\mathrm{T}^{-1/2}$ and noting that $\mathrm{T}^{-1/2}\mathrm{T}\mathrm{T}^{-1/2} = \mathrm{I}$, we have

(12.43) $$\lambda \mathrm{R} = \mathrm{T}^{-1/2}\mathrm{V}\mathrm{T}^{-1/2}\mathrm{R}$$

as the eigenvalue equation in terms of the new variables X and Y. Substituting the numerical $\mathrm{T}^{-1/2}$ from (12.42) into (12.43) gives the result we had in (12.24).

We have simply demonstrated that (12.39) and (12.43) give compact forms of the eigenvalue equations for Example 4. However, it is straightforward to show that these equations are just a compact summary of the equations of motion for any similar vibrations problem, in any number of variables, just by writing the potential and kinetic energy matrices and comparing the equations of motion in matrix form.

▶ **Example 7.** Find the characteristic frequencies and the characteristic modes of vibration for the system of masses and springs shown in Figure 12.3, where the motion is along a vertical line.

Let's use the simplified method of Example 6 for this problem. We first write the expressions for the kinetic energy and the potential energy as in previous examples.

(Note carefully that we measure x and y from the equilibrium positions of the masses when they are hanging at rest; then the gravitational forces are already balanced and gravitational potential energy does not come into the expression for V.)

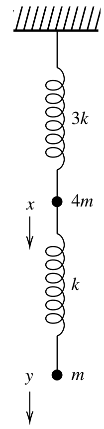


Figure 12.3

$$T = \tfrac{1}{2}m(4\dot{x}^2 + \dot{y}^2),$$
$$V = \tfrac{1}{2}k[3x^2 + (x-y)^2] = \tfrac{1}{2}k(4x^2 - 2xy + y^2). \tag{12.44}$$

The corresponding matrices are [see equations (12.37) and (12.38)]:

$$\mathrm{T} = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathrm{V} = \begin{pmatrix} 4 & -1 \\ -1 & 1 \end{pmatrix}. \tag{12.45}$$

As in equation (12.40), we find $\mathrm{T}^{-1}\mathrm{V}$ and its eigenvalues and eigenvectors.

$$\mathrm{T}^{-1}\mathrm{V} = \begin{pmatrix} 1/4 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 4 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1/4 \\ -1 & 1 \end{pmatrix}, \quad \lambda = \frac{m\omega^2}{k} = \frac{1}{2}, \ \frac{3}{2}.$$

$$\text{For } \omega = \sqrt{\frac{k}{2m}}, \ \mathbf{r} = (1,2); \quad \text{for } \omega = \sqrt{\frac{3k}{2m}}, \ \mathbf{r} = (1,-2). \tag{12.46}$$

As expected (since $\mathrm{T}^{-1}\mathrm{V}$ is not symmetric), the eigenvectors are not orthogonal. If we want orthogonal eigenvectors, we make the change of variables $X = 2x$, $Y = y$, to find the eigenvectors $\mathbf{R} = (1,1)$ and $\mathbf{R} = (1,-1)$ which are orthogonal. Alternatively, we can find the matrix $\mathrm{T}^{-1/2}\mathrm{V}\mathrm{T}^{-1/2}$

$$\begin{pmatrix} 1/2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 4 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1/2 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}, \tag{12.47}$$

and find its eigenvalues and eigenvectors.

PROBLEMS, SECTION 12

1. Verify that (12.2) multiplied out is (12.1).

Find the equations of the following conics and quadric surfaces relative to principal axes.

2. $2x^2 + 4xy - y^2 = 24$

3. $8x^2 + 8xy + 2y^2 = 35$

4. $3x^2 + 8xy - 3y^2 = 8$

5. $5x^2 + 3y^2 + 2z^2 + 4xz = 14$

6. $x^2 + y^2 + z^2 + 4xy + 2xz - 2yz = 12$

7. $x^2 + 3y^2 + 3z^2 + 4xy + 4xz = 60$

8. Carry through the details of Example 2 to find the unit eigenvectors. Show that the resulting rotation matrix C is orthogonal. *Hint:* Find $\mathrm{CC^T}$.

9. For Problems 2 to 7, find the rotation matrix C which relates the principal axes and the original axes. See Example 2.

10. Verify equations (12.13) and (12.14). Solve (12.15) to find the eigenvalues and verify (12.16). Find the corresponding eigenvectors as stated in (12.17).

11. Verify the details of Example 4, equations (12.18) to (12.25).

12. Verify the details of Example 5, equations (12.26) to (12.36).

13. Verify the details of Example 6, equations (12.37) to (12.43).

Find the characteristic frequencies and the characteristic modes of vibration for systems of masses and springs as in Figure 12.1 and Examples 3, 4, and 6 for the following arrays.

14. $k, m, 2k, m, k$

15. $5k, m, 2k, m, 2k$

16. $4k, m, 2k, m, k$

17. $3k, 3m, 2k, 4m, 2k$

18. $2k, m, k, 5m, 10k$

19. $4k, 2m, k, m, k$

20. Carry through the details of Example 7.

Find the characteristic frequencies and the characteristic modes of vibration as in Example 7 for the following arrays of masses and springs, reading from top to bottom in a diagram like Figure 12.3.

21. $3k, m, 2k, m$

22. $4k, 3m, k, m$

23. $2k, 4m, k, 2m$

▶ 13. A BRIEF INTRODUCTION TO GROUPS

We will not go very far into group theory—there are whole books on the subject as well as on its applications in physics. But since so many of the ideas we are discussing in this chapter are involved, it is interesting to have a quick look at groups.

▶ **Example 1.** Think about the four numbers ± 1, $\pm i$. Notice that no matter what products and powers of them we compute, we never get any numbers besides these four. This property of a set of elements with a law of combination is called *closure*. Now think about these numbers written in polar form: $e^{i\pi/2}$, $e^{i\pi}$, $e^{3i\pi/2}$, $e^{2i\pi} = 1$, or the corresponding rotations of a vector (in the xy plane with tail at the origin), or the set of rotation matrices corresponding to these successive 90° rotations of a vector (Problem 1). Note also that these numbers are the four fourth roots of 1, so we could write them as A, A^2, A^3, $A^4 = 1$. All these sets are examples of groups, or more precisely, they are all *representations* of the same group known as the *cyclic group of order* 4. We will be particularly interested in groups of matrices, that is, in matrix representations of groups, since this is very important in applications. Now just what is a group?

Definition of a Group A group is a set $\{A, B, C, \cdots\}$ of elements—which may be numbers, matrices, operations (such as the rotations above)—together with a law of combination of two elements (often called the "product" and written as AB—see discussion below) subject to the following four conditions.

1. Closure: The combination of any two elements is an element of the group.
2. Associative law: The law of combination satisfies the associative law: $(AB)C = A(BC)$.
3. Unit element: There is a unit element I with the property that $IA = AI = A$ for every element of the group.

4. Inverses: Every element of the group has an inverse in the group; that is, for any element A there is an element B such that $AB = BA = I$.

We can easily verify that these four conditions are satisfied for the set ± 1, $\pm i$ under multiplication.

1. We have already discussed closure.
2. Multiplication of numbers is associative.
3. The unit element is 1.
4. The numbers i and $-i$ are inverses since their product is 1; -1 is its own inverse, and 1 is its own inverse.

Thus the set ± 1, $\pm i$, under the operation of multiplication, is a group. The *order of a finite group* is the number of elements in the group. When the elements of a group of order n are of the form A, A^2, A^3, $\cdots$, $A^n = 1$, it is called a *cyclic group*. Thus the group ± 1, $\pm i$, under multiplication, is a cyclic group of order 4 as we claimed above.

A *subgroup* is a subset which is itself a group. The whole group, or the unit element, are called *trivial subgroups*; any other subgroup is called a *proper subgroup*. The group ± 1, $\pm i$ has the proper subgroup ± 1.

Product, Multiplication Table In the definition of a group and in the discussion so far, we have used the term "product" and have written AB for the combination of two elements. However, terms like "product" or "multiplication" are used here in a generalized sense to refer to whatever the operation is for combining group elements. In applications, group elements are often matrices and the operation is matrix multiplication. In general mathematical group theory, the operation might be, for example, addition of two elements, and that sounds confusing to say "product" when we mean sum! Look at one of the first examples we discussed, namely the rotation of a vector by angles $\pi/2$, π, $3\pi/2$, 2π or 0. If the group elements are rotation matrices, then we multiply them, but if the group elements are the angles, then we add them. But the physical problem is exactly the same in both cases. So remember that group multiplication refers to the law of combination for the group rather than just to ordinary multiplication in arithmetic.

Multiplication tables for groups are very useful; equations (13.1), (13.2), and (13.4) show some examples. Look at (13.1) for the group ± 1, $\pm i$. The first column and the top row (set off by lines) list the group elements. The sixteen possible products of these elements are in the body of the table. Note that each element of the group appears exactly once in each row and in each column (Problem 3). At the intersection of the row starting with i and the column headed by $-i$, you find the product $(i)(-i) = 1$, and similarly for the other products.

$$
\begin{array}{r|rrrr}
 & 1 & i & -1 & -i \\
\hline
1 & 1 & i & -1 & -i \\
i & i & -1 & -i & 1 \\
-1 & -1 & -i & 1 & i \\
-i & -i & 1 & i & -1
\end{array}
\tag{13.1}
$$

In (13.2) below, note that you add the angles as we discussed above. However, it's not quite just adding—it's really the familiar process of adding angles until you get to 2π and then starting over again at zero. In mathematical language this is called adding (mod 2π) and we write $\pi/2 + 3\pi/2 \equiv 0 \pmod{2\pi}$. Hours on an ordinary clock add in a similar way. If it's 10 o'clock and then 4 hours elapse, the clock says it's 2 o'clock. We write $10 + 4 \equiv 2 \pmod{12}$. (See Problems 6 and 7 for more examples.)

$$\text{(13.2)}\qquad \begin{array}{c|cccc} & 0 & \pi/2 & \pi & 3\pi/2 \\ \hline 0 & 0 & \pi/2 & \pi & 3\pi/2 \\ \pi/2 & \pi/2 & \pi & 3\pi/2 & 0 \\ \pi & \pi & 3\pi/2 & 0 & \pi/2 \\ 3\pi/2 & 3\pi/2 & 0 & \pi/2 & \pi \end{array}$$

Two groups are called *isomorphic* if their multiplication tables are identical except for the names we attach to the elements [compare (13.1) and (13.2)]. Thus all the 4-element groups we have discussed so far are isomorphic to each other, that is, they are really all the same group. However, there are two different groups of order 4, the cyclic group we have discussed, and another group called the 4's group (see Problem 4).

Symmetry Group of the Equilateral Triangle Consider three identical atoms at the corners of an equilateral triangle in the xy plane, with the center of the triangle at the origin as shown in Figure 13.1. What rotations and reflections of vectors in the xy plane (as in Section 7) will produce an identical array of atoms? By considering Figure 13.1, we see that there are three possible rotations: 0°, 120°, 240°, and three possible reflections, through the three lines F, G, H (lines along the altitudes of the triangle). Think of moving just the triangle (that is, the atoms), leaving the axes and the lines F, G, H fixed in the background. As in Section 7, we can write a 2 by 2 rotation or reflection matrix for each of these six transformations and set up a multiplication table to show that they do form a group of order 6. This group is called the symmetry group of the equilateral triangle. We find (Problem 8)

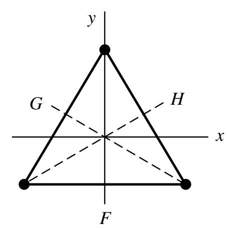


Figure 13.1

$$\text{(13.3)}\qquad \begin{array}{ll}
\text{Identity, } 0^\circ \text{ rotation} & \mathrm{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\[2ex]
120^\circ \text{ rotation} & \mathrm{A} = \frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix} \\[2ex]
240^\circ \text{ rotation} & \mathrm{B} = \frac{1}{2}\begin{pmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{pmatrix} \\[2ex]
\text{Reflection through line } F \ (y \text{ axis}) & \mathrm{F} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \\[2ex]
\text{Reflection through line } G & \mathrm{G} = \frac{1}{2}\begin{pmatrix} 1 & -\sqrt{3} \\ -\sqrt{3} & -1 \end{pmatrix} \\[2ex]
\text{Reflection through line } H & \mathrm{H} = \frac{1}{2}\begin{pmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix}
\end{array}$$

The group multiplication table is:

(13.4)

	I	A	B	F	G	H
I	I	A	B	F	G	H
A	A	B	I	G	H	F
B	B	I	A	H	F	G
F	F	H	G	I	B	A
G	G	F	H	A	I	B
H	H	G	F	B	A	I

Note here that GF = A, but FG = B, not surprising since we know that matrices don't always commute. In group theory, if every two group elements commute, the group is called *Abelian*. Our previous group examples have all been Abelian, but the group in (13.4) is not Abelian.

This is just one example of a symmetry group. Group theory is so important in applications because it offers a systematic way of using the symmetry of a physical problem to simplify the solution. As we have seen, groups can be represented by sets of matrices, and this is widely used in applications.

Conjugate Elements, Class, Character Two group elements A and B are called *conjugate* elements if there is a group element C such that $C^{-1}AC = B$. By letting C be successively one group element after another, we can find all the group elements conjugate to A. This set of conjugate elements is called a *class*. Recall from Section 11 that if A is a matrix describing a transformation (such as a rotation or some sort of mapping of a space onto itself), then $B = C^{-1}AC$ describes the same mapping but relative to a different set of axes (different basis). Thus all the elements of a class really describe the same mapping, just relative to different bases.

► **Example 2.** Find the classes for the group in (13.3) and (13.4). We find the elements conjugate to F as follows [use (13.4) to find inverses and products]:

$$\begin{aligned} &I^{-1}FI = F;\\ &A^{-1}FA = BFA = BH = G;\\ &B^{-1}FB = AFB = AG = H;\\ &F^{-1}FF = F;\\ &G^{-1}FG = GFG = GB = H;\\ &H^{-1}FH = HFH = HA = G. \end{aligned} \tag{13.5}$$

Thus the elements F, G, and H are conjugate to each other and form one class. You can easily show (Problem 12) that elements A and B are another class, and the unit element I is a class by itself. Now notice what we observed above. The elements F, G, and H all just interchange two atoms, that is, all of them do the same thing, just seen from a different viewpoint. The elements A and B rotate the atoms, A by 120° and B by 240° which is the same as 120° looked at upside down. And finally the unit element I leaves things unchanged so it is a class by itself. Notice that a class is not a group (except for the class consisting of I) since a group must contain the unit element. So a class is a subset of a group, but not a subgroup.

Recall from (9.13) and Problem 11.10 that the trace of a matrix (sum of diagonal elements) is not changed by a similarity transformation. Thus all the matrices of a class have the same trace. Observe that this is true for the group (13.3): Matrix I has trace = 2, A and B have trace $= -\frac{1}{2} - \frac{1}{2} = -1$, and F, G, and H have trace = 0. In this connection, the trace of a matrix is called its *character*, so we see that all matrices of a class have the same character. Also note that we could write the matrices (13.3) in (infinitely) many other ways by rotating the reference axes, that is, by performing similarity transformations. But since similarity transformations do not change the trace, that is, the character, we now have a number attached to each class which is independent of the particular choice of coordinate system (basis). Classes and their associated character are very important in applications of group theory.

One more number is important here, and that is the dimension of a representation. In (13.3), we used 2 by 2 matrices (2 dimensions), but it would be possible to work in 3 dimensions. Then, for example, the A matrix would describe a 120° rotation around the z axis and would be

$$\text{A} = \begin{pmatrix} -1/2 & -\sqrt{3}/2 & 0 \\ \sqrt{3}/2 & -1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{13.6}$$

and the other matrices in (13.3) would have a similar form, called *block diagonalized*. But now the traces of all the matrices are increased by 1. To avoid having any ambiguity about character, we use what are called "irreducible representations" in finding character; let's discuss this.

Irreducible Representations A 2-dimensional representation is called *reducible* if all the group matrices can be diagonalized by the same unitary similarity transformation (that is, the same change of basis). For example, the matrices in Problem 1 and the matrices in Problem 4 both give 2-dimensional reducible representations of their groups (see Problems 13, 15, and 16). On the other hand, the matrices in (13.3) cannot be simultaneously diagonalized (see Problem 13), so (13.3) is called a 2-dimensional *irreducible representation* of the equilateral triangle symmetry group. If a group of 3 by 3 matrices can all be either diagonalized or put in the form of (13.6) (block diagonalized) by the same unitary similarity transformation, then the representation is called reducible; if not, it is a 3-dimensional irreducible representation. For still larger matrices, imagine the matrices block diagonalized with blocks along the main diagonal which are the matrices of irreducible representations.

Thus we see that any representation is made up of irreducible representations. For each irreducible representation, we find the character of each class. Such lists are known as character tables, but their construction is beyond our scope.

Infinite Groups Here we survey some examples of infinite groups as well as some sets which are not groups.

(13.7)

(a) The set of all integers, positive, negative, and zero, under ordinary addition, is a group. *Proof*: The sum of two integers is an integer. Ordinary addition obeys the associative law. The unit element is 0. The inverse of the integer N is −N since N + (−N) = 0.

(b) The same set under ordinary multiplication is not a group because 0 has no inverse. But even if we omit 0, the inverses of the other integers are fractions which are not in the set.

(c) Under ordinary multiplication, the set of all rational numbers except zero, is a group. *Proof*: The product of two rational numbers is a rational number. Ordinary multiplication is associative. The unit element is 1, and the inverse of a rational number is just its reciprocal.

Similarly, you can show that the following sets are groups under ordinary multiplication (Problem 17): All real numbers except zero, all complex numbers except zero, all complex numbers $re^{i\theta}$ with $r = 1$.

(d) Ordinary subtraction or division cannot be group operations because they don't satisfy the associative law; for example, $x-(y-z) \neq (x-y)-z$. (Problem 18.)

(e) The set of all orthogonal 2 by 2 matrices under matrix multiplication is a group called O(2). If the matrices are required to be rotation matrices, that is, have determinant $+1$, the set is a group called SO(2) (the S stands for special). Similarly, the following sets of matrices are groups under matrix multiplication: The set of all orthogonal 3 by 3 matrices, called O(3); its subgroup SO(3) with determinant $= 1$; or the corresponding sets of orthogonal matrices of any dimension n, called O(n) and SO(n). (Problem 19.)

(f) The set of all unitary n by n matrices, $n = 1, 2, 3, \cdots$, called U(n), is a group under matrix multiplication, and its subgroup SU(n) of unitary matrices with determinant $= 1$ is also a group. *Proof*: We have repeatedly noted that matrix multiplication is associative and that the unit matrix is the unit element of a group of matrices. So we just need to check closure and inverses. The product of two unitary matrices is unitary (see Section 9). If two matrices have determinant $= 1$, their product has determinant $= 1$ [see equation (6.6)]. The inverse of a unitary matrix is unitary (see Problem 9.25).

PROBLEMS, SECTION 13

1. Write the four rotation matrices for rotations of vectors in the xy plane through angles 90°, 180°, 270°, 360° (or 0°) [see equation (7.12)]. Verify that these 4 matrices under matrix multiplication satisfy the four group requirements and are a matrix representation of the cyclic group of order 4. Write their multiplication table and compare with Equations (13.1) and (13.2).

2. Following the text discussion of the cyclic group of order 4, and Problem 1, discuss

(a) the cyclic group of order 3 (see Chapter 2, Problem 10.32);

(b) the cyclic group of order 6.

3. Show that, in a group multiplication table, each element appears exactly once in each row and in each column. *Hint:* Suppose that an element appears twice, and show that this leads to a contradiction, namely that two elements assumed different are the same element.

4. Show that the matrices

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

under matrix multiplication, form a group. Write the group multiplication table to see that this group (called the 4's group) is not isomorphic to the cyclic group of order 4 in Problem 1. Show that the 4's group is Abelian but not cyclic.

5. Consider the group of order 4 with unit element I and other elements A, B, C, where $AB = BA = C$, and $A^2 = B^2 = I$. Write the group multiplication table and verify that it is a group. There are two groups of order 4 (discussed in Problems 1 and 4). To which is this one isomorphic? *Hint:* Compare the multiplication tables.

6. Consider the integers 0, 1, 2, 3 under addition (mod 4). Write the group "multiplication" table and show that you have a group of order 4. Is this group isomorphic to the cyclic group of order 4 or to the 4's group?

7. Consider the set of numbers 1, 3, 5, 7 with multiplication (mod 8) as the law of combination. Write the multiplication table to show that this is a group. [To multiply two numbers (mod 8), you multiply them and then take the remainder after dividing by 8. For example, $5 \times 7 = 35 \equiv 3 (\text{mod } 8)$.] Is this group isomorphic to the cyclic group of order 4 or to the 4's group?

8. Verify (13.3) and (13.4). *Hints:* For the rotation and reflection matrices, see Section 7. In checking the multiplication table, be sure you are multiplying the matrices in the right order. Remember that matrices are operators on the vectors in the plane (Section 7), and matrices may not commute. GFA means apply A, then F, then G.

9. Show that any cyclic group is Abelian. *Hint:* Does a matrix commute with itself?

10. As we did for the equilateral triangle, find the symmetry group of the square. *Hints:* Draw the square with its center at the origin and its sides parallel to the x and y axes. Find a set of eight 2 by 2 matrices (4 rotation and 4 reflection) which map the square onto itself, and write the multiplication table to show that you have a group.

11. Do Problem 10 for a rectangle. Note that now only two rotations and 2 reflections leave the rectangle unchanged. So you have a group of order 4. To which is it isomorphic, the cyclic group or the 4's group?

12. Verify (13.5) and then also show that A, B are the elements of a class, and that I is a class by itself. Show that it will always be true in any group that I is a class by itself. *Hint:* What is $\mathrm{C}^{-1}\mathrm{IC}$ for any element C of a group?

13. Using the discussion of simultaneous diagonalization at the end of Section 11, show that the 2-dimensional matrices in Problems 1 and 4 are reducible representations of their groups, and the matrices in (13.5) give an irreducible representation of the equilateral triangle symmetry group. *Hint:* Look at the multiplication tables to see which matrices commute.

14. Use the multiplication table you found in Problem 10 to find the classes in the symmetry group of a square. Show that the 2 by 2 matrices you found are an irreducible representation of the group (see Problem 13), and find the character of each class for that representation. Note that it is possible for the character to be the same for two classes, but it is not possible for the character of two elements of the same class to be different.

15. By Problem 13, you know that the matrices in Problem 4 are a reducible representation of the 4's group, that is they can all be diagonalized by the same unitary similarity transformation (in this case orthogonal since the matrices are symmetric). Demonstrate this directly by finding the matrix C and diagonalizing all 4 matrices.

16. Do Problem 15 for the group of matrices you found in Problem 1. Be careful here—you are working in a complex vector space and your C matrix will be unitary but not orthogonal (see Sections 10 and 11). *Comment:* Not surprisingly, the numbers 1, i, -1, $-i$ give a 1-dimensional representation—note that a single number can be thought of as a 1-dimensional matrix.

17. Verify that the sets listed in (13.7c) are groups.

18. Show that division cannot be a group operation. *Hint:* See (13.7d).

19. Verify that the sets listed in (13.7e) are groups. *Hint:* See the proofs in (13.7f).

20. Is the set of all orthogonal 3-by-3 matrices with determinant $= -1$ a group? If so, what is the unit element?

21. Is the group SO(2) Abelian? What about SO(3)? *Hint:* See the discussion following equation (6.14).

14. GENERAL VECTOR SPACES

In this section we are going to introduce a generalization of our picture of vector spaces which is of great importance in applications. This will be merely an introduction because the ideas here will be used in many of the following chapters as you will discover. The basic idea will be to set up an outline of the requirements for 3-dimensional vector spaces (as we listed the requirements for a group), and then show that these familiar 3-dimensional vector space requirements are satisfied by sets of things like functions or matrices which we would not ordinarily think of as vectors.

Definition of a Vector Space A vector space is a set of elements $\{\mathbf{U}, \mathbf{V}, \mathbf{W}, \cdots\}$ called vectors, together with two operations: addition of vectors, and multiplication of a vector by a scalar (which for our purposes will be a real or a complex number), and subject to the following requirements:

1. Closure: The sum of any two vectors is a vector in the space.
2. Vector addition is:
   (a) commutative: $\mathbf{U} + \mathbf{V} = \mathbf{V} + \mathbf{U}$,
   (b) associative: $(\mathbf{U} + \mathbf{V}) + \mathbf{W} = \mathbf{U} + (\mathbf{V} + \mathbf{W})$.
3. **(a)** There is a zero vector $\mathbf{0}$ such that $\mathbf{0} + \mathbf{V} = \mathbf{V} + \mathbf{0} = \mathbf{V}$ for every element $\mathbf{V}$ in the space.
   (b) Every element $\mathbf{V}$ has an additive inverse $(-\mathbf{V})$ such that $\mathbf{V} + (-\mathbf{V}) = \mathbf{0}$.
4. Multiplication of vectors by scalars has the expected properties:
   (a) $k(\mathbf{U} + \mathbf{V}) = k\mathbf{U} + k\mathbf{V}$;
   (b) $(k_1 + k_2)\mathbf{V} = k_1\mathbf{V} + k_2\mathbf{V}$;
   (c) $(k_1 k_2)\mathbf{V} = k_1(k_2\mathbf{V})$;
   (d) $0 \cdot \mathbf{V} = \mathbf{0}$, and $1 \cdot \mathbf{V} = \mathbf{V}$.

You should go over these and satisfy yourself that they are all true for ordinary two and three dimensional vector spaces. Now let's look at some examples of things we don't usually think of as vectors which, nevertheless, satisfy the above requirements.

▶ **Example 1.** Consider the set of polynomials of the third degree or less, namely functions of the form $f(x) = a_0 + a_1x + a_2x^2 + a_3x^3$. Is this a vector space? If so, find a basis. What is the dimension of the space?

We go over the requirements listed above:

1. The sum of two polynomials of degree ≤ 3 is a polynomial of degree ≤ 3 and so is a member of the set.

2. Addition of algebraic expressions is commutative and associative.

3. The "zero vector" is the polynomial with all coefficients a_i equal to 0, and adding it to any other polynomial just gives that other polynomial. The additive inverse of a function $f(x)$ is just $-f(x)$, and $-f(x) + f(x) = 0$ as required for a vector space.

4. All the listed familiar rules are just what we do every time we work with algebraic expressions.

So we have a vector space! Now let's try to find a basis for it. Consider the set of functions: $\{1, x, x^2, x^3\}$. They span the space since any polynomial of degree ≤ 3 is a linear combination of them. You can easily show (Problem 1) by computing the Wronskian [equation (8.5)] that they are linearly independent. Therefore they are a basis, and since there are 4 basis vectors, the dimension of the space is 4.

▶ **Example 2.** Consider the set of linear combinations of the functions

$$\{e^{ix},\ e^{-ix},\ \sin x,\ \cos x,\ x\sin x\}.$$

It is straightforward to verify that all our requirements above are met (Problem 1). To find a basis, we must find a linearly independent set of functions which spans the space. We note that the given functions are not linearly independent since e^{ix} and e^{-ix} are linear combinations of $\sin x$ and $\cos x$ (Chapter 2, Section 4). However, the set $\{\sin x,\ \cos x,\ x\sin x\}$ is a linearly independent set and it spans the space. So this is a possible basis and the dimension of the space is 3. Another possible basis would be $\{e^{ix}, e^{-ix}, x\sin x\}$. You will meet sets of functions like these as solutions of differential equations (*see* Chapter 8, Problems 5.13 to 5.18).

▶ **Example 3.** Modify Example 1 to consider the set of polynomials of degree ≤ 3 with $f(1) = 1$. Is this a vector space? Suppose we add two of the polynomials; then the value of the sum at $x = 1$ is 2, so it is not an element of the set. Thus requirement 1 is not satisfied so this is not a vector space. Note that a subset of the vectors of a vector space is not necessarily a subspace. On the other hand, if we consider polynomials of degree ≤ 3 with $f(1) = 0$, then the sum of two of them is zero at $x = 1$; this is a vector space. You can easily verify (Problem 1) that it is a subspace of dimension 3 and a possible basis is $\{x - 1,\ x^2 - 1,\ x^3 - 1\}$.

▶ **Example 4.** Consider the set of all polynomials of any degree $\leq N$. The sum of two polynomials of degree $\leq N$ is another such polynomial, and you can easily verify (Problem 1) that the rest of the requirements are met, so this is a vector space. A simple choice of basis is the set of powers of x from $x^0 = 1$ to x^N. Thus we see that the dimension of this space is $N + 1$.

Example 5. Consider the set of all 2 by 3 matrices with matrix addition as the law of combination, and multiplication by scalars defined as in Section 6. Recall that you add matrices by adding corresponding elements. Thus a sum of two 2 by 3 matrices is another 2 by 3 matrix. For matrix addition and multiplication by scalars, it is straightforward to show that the other requirements listed above are satisfied (Problem 1). As a basis, we could use the six matrices:

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

Satisfy yourself that these are linearly independent and that they span the space (that is, that you could write any 2 by 3 matrix as a linear combination of these six). Since there are 6 basis vectors, the dimension of this space is 6.

Inner Product, Norm, Orthogonality The definitions of these terms need to be generalized when our "vectors" are functions, that is, we want to generalize equations (10.1) to (10.3). A natural generalization of a sum is an integral, so we might reasonably replace $\sum A_i B_i$ by $\int A(x)B(x)\,dx$, and $\sum A_i^2$ by $\int [A(x)]^2\,dx$. However, in applications we frequently want to consider complex functions of the real variable x (for example, e^{ix} as in Example 2). Thus, given functions $A(x)$ and $B(x)$ on $a \le x \le b$, we define

$$\text{(14.1)}\qquad [\text{Inner Product of } A(x) \text{ and } B(x)] = \int_a^b A^*(x)B(x)\,dx,$$

$$\text{(14.2)}\qquad [\text{Norm of } A(x)] = ||A(x)|| = \sqrt{\int_a^b A^*(x)A(x)\,dx},$$

$$\text{(14.3)}\qquad A(x) \text{ and } B(x) \text{ are orthogonal on } (a,b) \quad \text{if} \quad \int_a^b A^*(x)B(x)\,dx = 0.$$

Let's now generalize our definition (14.1) of inner product still further. Let A, B, C, $\cdots$ be elements of a vector space, and let a, b, c, $\cdots$ be scalars. We will use the bracket $\langle A|B\rangle$ to mean the inner product of A and B. This vector space is called an *inner product space* if an inner product is defined subject to the conditions:

$$\text{(14.4a)}\qquad \langle A|B\rangle^* = \langle B|A\rangle;$$
$$\text{(14.4b)}\qquad \langle A|A\rangle \ge 0, \quad \langle A|A\rangle = 0 \text{ if and only if } A = 0;$$
$$\text{(14.4c)}\qquad \langle C|aA + bB\rangle = a\langle C|A\rangle + b\langle C|B\rangle.$$

(See Problem 11.) It follows from (14.4) that (Problem 12)

$$\text{(14.5a)}\qquad \langle aA + bB|C\rangle = a^*\langle A|C\rangle + b^*\langle B|C\rangle, \quad \text{and}$$
$$\text{(14.5b)}\qquad \langle aA|bB\rangle = a^*b\langle A|B\rangle.$$

You will find various other notations for the inner product, such as (A, B) or $[A, B]$ or $\langle A, B\rangle$. The notation $\langle A|B\rangle$ is used in quantum mechanics. Most mathematics books put the complex conjugate on the second factor in (14.1) and make the corresponding changes in (14.4) and (14.5). Most physics and mathematical methods books handle the complex conjugate as we have. If you are confused by this notation and equations (14.4) and (14.5), keep going back to (14.1) where $\langle A|B\rangle = \int A^* B$ until you get used to the bracket notation. Also study carefully our use of the bracket notation in the next section and do Problems 11 to 14.

Schwarz's Inequality In Section 10 we proved the Schwarz inequality for n-dimensional Euclidean space. For an inner product space satisfying (14.4), it becomes [compare (10.9)]

$$|\langle A|B\rangle|^2 \le \langle A|A\rangle\langle B|B\rangle. \tag{14.6}$$

To prove this, we first note that it is true if $B = 0$. For $B \neq 0$, let $C = A - \mu B$, where $\mu = \langle B|A\rangle/\langle B|B\rangle$, and find $\langle C|C\rangle$ which is ≥ 0 by (14.4b). Using (14.4) and (14.5), we write

$$\langle A - \mu B|A - \mu B\rangle = \langle A|A\rangle - \mu^*\langle B|A\rangle - \mu\langle A|B\rangle + \mu^*\mu\langle B|B\rangle \ge 0. \tag{14.7}$$

Now substitute the values of μ and μ^* to get (see Problem 13)

$$\begin{aligned}\langle A|A\rangle - \frac{\langle A|B\rangle}{\langle B|B\rangle}\langle B|A\rangle - \frac{\langle B|A\rangle}{\langle B|B\rangle}\langle A|B\rangle + \frac{\langle A|B\rangle}{\langle B|B\rangle}\frac{\langle B|A\rangle}{\langle B|B\rangle}\langle B|B\rangle \\ = \langle A|A\rangle - \frac{\langle A|B\rangle\langle A|B\rangle^*}{\langle B|B\rangle} = \langle A|A\rangle - \frac{|\langle A|B\rangle|^2}{\langle B|B\rangle} \ge 0\end{aligned} \tag{14.8}$$

which gives (14.6).

For a function space as in (14.1) to (14.3), Schwarz's inequality becomes (see Problem 14):

$$\left|\int_a^b A^*(x)B(x)\,dx\right|^2 \le \left(\int_a^b A^*(x)A(x)\,dx\right)\left(\int_a^b B^*(x)B(x)\,dx\right). \tag{14.9}$$

Orthonormal Basis; Gram-Schmidt Method Two functions are called *orthogonal* if they satisfy (14.3); a function is *normalized* if its norm in (14.2) is 1. By a combination of the two words, we call a set of functions *orthonormal* if they are all mutually orthogonal and they all have norm 1. It is often convenient to write the functions of a vector space in terms of an orthonormal basis (compare writing ordinary vectors in three dimensions in terms of **i**, **j**, **k**). Let's see how the Gram-Schmidt method applies to a vector space of functions with inner product, norm, and orthogonality defined by (14.1) to (14.3). (Compare Section 10, Example 4 and the paragraph before it.)

▶ **Example 6.** In Example 1, we found that the set of all polynomials of degree ≤ 3 is a vector space of dimension 4 with basis 1, x, x^2, x^3. Let's consider these polynomials on the interval $-1 \le x \le 1$ and construct an orthonormal basis. To keep track of what we're doing, let $f_0, f_1, f_2, f_3 = 1, x, x^2, x^3$; let p_0, p_1, p_2, p_3 be a corresponding orthogonal basis (which we find by the Gram-Schmidt method); and let e_0, e_1, e_2, e_3, be the orthonormal basis (which we get by normalizing the functions p_i). Recall the Gram-Schmidt routine (see Section 10, Example 4): Normalize the first function to get e_0. Then for the rest of the functions, subtract from f_i each preceding e_j multiplied by the inner product of e_j and f_i, that is, find

$$p_i = f_i - \sum_{j<i} e_j \langle e_j | f_i \rangle = f_i - \sum_{j<i} e_j \int_{-1}^{1} e_j f_i \, dx. \tag{14.10}$$

Finally, normalize p_i to get e_i.

We can save effort by noting in advance that many of the inner products we need are going to be zero. You can easily show (Problem 15) that the integral of an odd power of x from $x = -1$ to 1 is zero, and consequently any even power of x is orthogonal to any odd power. Observe that the f_i are alternately even and odd powers of x. Then you can show that the corresponding p_i and e_i will also involve just even or just odd powers of x. The Gram-Schmidt method gives the following results (Problem 16).

$$f_0 = 1 = p_0, \quad ||p_0||^2 = \int_{-1}^{1} 1^2 \, dx = 2, \quad e_0 = \frac{1}{\sqrt{2}}.$$

$f_1 = x;\quad p_1 = x$ because x is orthogonal to e_0.

$$||p_1||^2 = \int_{-1}^{1} x^2 \, dx = \frac{2}{3}, \quad e_1 = x\sqrt{\frac{3}{2}}.$$

$f_2 = x^2$. Since x^2 is orthogonal to e_1 but not to e_0,

$$p_2 = x^2 - \frac{1}{\sqrt{2}} \int_{-1}^{1} \frac{1}{\sqrt{2}} x^2 \, dx = x^2 - \frac{1}{3}.$$

$$||p_2||^2 = \int_{-1}^{1} \left(x^2 - \frac{1}{3}\right)^2 dx = \frac{8}{45}, \quad e_2 = (3x^2 - 1)\sqrt{\frac{5}{8}}.$$

$f_3 = x^3$. Since x^3 is orthogonal to e_0 and e_2,

$$p_3 = x^3 - x\sqrt{\frac{3}{2}} \int_{-1}^{1} x\sqrt{\frac{3}{2}}\, x^3 \, dx = x^3 - \frac{3}{5}x,$$

$$||p_3||^2 = \int_{-1}^{1} \left(x^3 - \frac{3}{5}x\right)^2 dx = \frac{8}{175}, \quad e_3 = (5x^3 - 3x)\sqrt{\frac{7}{8}}.$$

This process could be continued for a vector space with basis $1, x, x^2, \cdots, x^N$ (but it is not very efficient). The orthonormal functions e_i are well-known functions called (normalized) *Legendre polynomials*. In Chapters 12 and 13, we will discover these functions as solutions of differential equations and see their applications in physics problems.

Infinite Dimensional Spaces If a vector space does not have a finite basis, it is called an infinite dimensional vector space. It is beyond our scope to go into a detailed mathematical study of such spaces. However, you should know that, by analogy with finite dimensional vector spaces, we still use the term basis functions for sets of functions (like x^n or $\sin nx$) in terms of which we can expand suitably restricted functions in infinite series. So far we have discussed only power series (Chapter 1). In later chapters you will discover many other sets of functions which provide useful bases in applications: sines and cosines in Chapter 7, various special functions in Chapters 12 and 13. When we introduce them, we will discuss questions of convergence of the infinite series, and of completeness of sets of basis functions.

PROBLEMS, SECTION 14

1. Verify the statements indicated in Examples 1 to 5 above.

For each of the following sets, either verify (as in Example 1) that it is a vector space, or show which requirements are not satisfied. If it is a vector space, find a basis and the dimension of the space.

2. Linear combinations of the set of functions $\{e^x, \sinh x, xe^x\}$.

3. Linear combinations of the set of functions $\{x, \cos x, x\cos x, e^x \cos x, (2-3e^x)\cos x, x(1+5\cos x)\}$.

4. Polynomials of degree ≤ 3 with $a_2 = 0$.

5. Polynomials of degree ≤ 5 with $a_1 = a_3$.

6. Polynomials of degree ≤ 6 with $a_3 = 3$.

7. Polynomials of degree ≤ 7 with all the even coefficients equal to each other and all the odd coefficients equal to each other.

8. Polynomials of degree ≤ 7 but with all odd powers missing.

9. Polynomials of degree ≤ 10 but with all even powers having positive coefficients.

10. Polynomials of degree ≤ 13, but with the coefficient of each odd power equal to half the preceding coefficient of an even power.

11. Verify that the definitions in (14.1) and (14.2) satisfy the requirements for an inner product listed in (14.4) and (14.5). *Hint:* Write out all the equations (14.4) and (14.5) in the integral notation of (14.1) and (14.2).

12. Verify that the relations in (14.5) follow from (14.4). *Hints:* For (14.5a), take the complex conjugate of (14.4c). To take the complex conjugate of a bracket, use (14.4a).

13. Verify (14.7) and (14.8) *Hints:* Remember that a norm squared, like $\langle B|B\rangle$, is a real and non-negative.scalar, so its complex conjugate is just itself. But $\langle B|A\rangle$ is a complex scalar and $\langle B|A\rangle = \langle A|B\rangle^*$ by (14.4). Show that $\mu^* = \langle A|B\rangle / \langle B|B\rangle$.

14. Verify that (14.9) is (14.6) with the definition of scalar product as in (14.1).

15. For Example 6, verify the claimed orthogonality on $(-1, 1)$ of an even power of x and an odd power of x. *Hint:* For example, consider $\int_{-1}^{1} x^2 x^3\,dx$.

16. For Example 6, verify the details of the terms omitted in the functions p_i because of orthogonality. *Hint:* See Problem 15. Also verify the calculations of inner products and norms and the orthonormal set e_i.

15. MISCELLANEOUS PROBLEMS

1. Show that if each element of one row (or column) of a determinant is the sum of two terms, the determinant can be written as a sum of two determinants; for example,

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

Use this result to verify Fact 4b of Section 3.

2. What is wrong with the following argument? "If we add the first row of a determinant to the second row and the second row to the first row, then the first two rows of the determinant are identical, and the value of the determinant is zero. Therefore all determinants have the value zero."

3. (a) Find the equations of the line through the points $(4,-1,2)$ and $(3,1,4)$.

(b) Find the equation of the plane through the points $(0,0,0)$, $(1,2,3)$ and $(2,1,1)$.

(c) Find the distance from the point $(1,1,1)$ to the plane $3x-2y+6z=12$.

(d) Find the distance from the point $(1,0,2)$ to the line $\mathbf{r}=2\mathbf{i}+\mathbf{j}-\mathbf{k}+(\mathbf{i}-2\mathbf{j}+2\mathbf{k})t$.

(e) Find the angle between the plane in (c) and the line in (d).

4. Given the line $\mathbf{r}=3\mathbf{i}-\mathbf{j}+(2\mathbf{i}+\mathbf{j}-2\mathbf{k})t$:

(a) Find the equation of the plane containing the line and the point $(2,1,0)$.

(b) Find the angle between the line and the (y,z) plane.

(c) Find the perpendicular distance between the line and the x axis.

(d) Find the equation of the plane through the point $(2,1,0)$ and perpendicular to the line.

(e) Find the equations of the line of intersection of the plane in (d) and the plane $y=2z$.

5. (a) Write the equations of a straight line through the points $(2,7,-1)$ and $(5,7,3)$.

(b) Find the equation of the plane determined by the two lines $\mathbf{r}=(\mathbf{i}-2\mathbf{j}+\mathbf{k})t$ and $\mathbf{r}=(6\mathbf{i}-3\mathbf{j}+2\mathbf{k})t$.

(c) Find the angle which the line in (a) makes with the plane in (b).

(d) Find the distance from $(1,1,1)$ to the plane in (b).

(e) Find the distance from $(1,6,-3)$ to the line in (a).

6. Derive the formula

$$D=\frac{|ax_0+by_0+cz_0-d|}{\sqrt{a^2+b^2+c^2}}$$

for the distance from (x_0,y_0,z_0) to $ax+by+cz=d$.

7. Given the matrices A, B, C below, find or mark as meaningless the matrices: A^{T}, A^{-1}, AB, $\bar{\mathrm{A}}$, $\mathrm{A}^{\mathrm{T}}\mathrm{B}^{\mathrm{T}}$, $\mathrm{B}^{\mathrm{T}}\mathrm{A}^{\mathrm{T}}$, BA^{T}, ABC, $\mathrm{AB}^{\mathrm{T}}\mathrm{C}$, $\mathrm{B}^{\mathrm{T}}\mathrm{AC}$, $\mathrm{A}^{\dagger}$, $\mathrm{B}^{\mathrm{T}}\mathrm{C}$, $\mathrm{B}^{-1}\mathrm{C}$, $\mathrm{C}^{-1}\mathrm{A}$, CB^{T}.

$$\mathrm{A}=\begin{pmatrix}1 & -1\\ 0 & i\end{pmatrix},\quad \mathrm{B}=\begin{pmatrix}2 & 1 & -1\\ 0 & 3 & 5\end{pmatrix},\quad \mathrm{C}=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}.$$

8. Given

$$\mathrm{A}=\begin{pmatrix}1 & 0 & 2i\\ i & -3 & 0\\ 1 & 0 & i\end{pmatrix},\quad \text{find } \mathrm{A}^{\mathrm{T}},\bar{\mathrm{A}},\mathrm{A}^{\dagger},\mathrm{A}^{-1}.$$

9. The following matrix product is used in discussing a thick lens in air:

$$\mathrm{A}=\begin{pmatrix}1 & (n-1)/R_2\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ d/n & 1\end{pmatrix}\begin{pmatrix}1 & -(n-1)/R_1\\ 0 & 1\end{pmatrix},$$

where d is the thickness of the lens, n is its index of refraction, and R_1 and R_2 are the radii of curvature of the lens surfaces. It can be shown that element A_{12} of A is $-1/f$ where f is the focal length of the lens. Evaluate A and $\det \mathrm{A}$ (which should equal 1) and find $1/f$. [See Am. J. Phys. **48**, 397–399 (1980).]

10. The following matrix product is used in discussing two thin lenses in air:

$$\mathrm{M} = \begin{pmatrix} 1 & -1/f_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ d & 1 \end{pmatrix} \begin{pmatrix} 1 & -1/f_1 \\ 0 & 1 \end{pmatrix},$$

where f_1 and f_2 are the focal lengths of the lenses and d is the distance between them. As in Problem 9, element M_{12} is $-1/f$ where f is the focal length of the combination. Find M, det M, and $1/f$.

11. There is a one-to-one correspondence between two-dimensional vectors and complex numbers. Show that the real and imaginary parts of the product $z_1 z_2^*$ (the star denotes complex conjugate) are respectively the scalar product and $\pm$ the magnitude of the vector product of the vectors corresponding to z_1 and z_2.

12. The vectors $\mathbf{A} = a\mathbf{i} + b\mathbf{j}$ and $\mathbf{B} = c\mathbf{i} + d\mathbf{j}$ form two sides of a parallelogram. Show that the area of the parallelogram is given by the absolute value of the following determinant. (Also see Chapter 6, Section 3.)

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

13. The plane $2x+3y+6z = 6$ intersects the coordinate axes at points P, Q, R, forming a triangle. Find the vectors $\overrightarrow{PQ}$ and $\overrightarrow{PR}$. Write a vector formula for the area of the triangle PQR, and find the area.

In Problems 14 to 17, multiply matrices to find the resultant transformation. *Caution:* Be sure you are multiplying the matrices in the right order.

14. $\begin{cases} x' = (x + y\sqrt{3})/2 \\ y' = (-x\sqrt{3} + y)/2 \end{cases}$ $\qquad \begin{cases} x'' = (-x' + y'\sqrt{3})/2 \\ y'' = -(x'\sqrt{3} + y')/2 \end{cases}$

15. $\begin{cases} x' = 2x + 5y \\ y' = x + 3y \end{cases}$ $\qquad \begin{cases} x'' = x' - 2y' \\ y'' = 3x' - 5y' \end{cases}$

16. $\begin{cases} x' = (x + y\sqrt{2} + z)/2 \\ y' = (x\sqrt{2} - z\sqrt{2})/2 \\ z' = (-x + y\sqrt{2} - z)/2 \end{cases}$ $\qquad \begin{cases} x'' = (x'\sqrt{2} + z'\sqrt{2})/2 \\ y'' = (-x' - y'\sqrt{2} + z')/2 \\ z'' = (x' - y'\sqrt{2} - z')/2 \end{cases}$

17. $\begin{cases} x' = (2x + y + 2z)/3 \\ y' = (x + 2y - 2z)/3 \\ z' = (2x - 2y - z)/3 \end{cases}$ $\qquad \begin{cases} x'' = (2x' + y' + 2z')/3 \\ y'' = (-x' - 2y' + 2z')/3 \\ z'' = (-2x' + 2y' + z')/3 \end{cases}$

Find the eigenvalues and eigenvectors of the matrices in the following problems.

18. $\begin{pmatrix} 1 & 0 \\ 3 & -2 \end{pmatrix}$ **19.** $\begin{pmatrix} 5 & 1 \\ 4 & 2 \end{pmatrix}$ **20.** $\begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix}$ **21.** $\begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}$

22. $\begin{pmatrix} 3 & 0 & -2 \\ 0 & 4 & 0 \\ -2 & 0 & 3 \end{pmatrix}$ **23.** $\begin{pmatrix} 3 & 0 & 1 \\ 0 & 3 & 1 \\ 1 & 1 & 2 \end{pmatrix}$ **24.** $\begin{pmatrix} 2 & -3 & 4 \\ -3 & 2 & 0 \\ 4 & 0 & 2 \end{pmatrix}$

25. Find the C matrix which diagonalizes the matrix M of Problem 18. Observe that M is not symmetric, and C is not orthogonal (see Section 11). However, C does have an inverse; find C^{-1} and show that $\mathrm{C}^{-1}\mathrm{MC} = \mathrm{D}$.

26. Repeat Problem 25 for Problem 19.

In Problems 27 to 30, rotate the given quadric surface to principal axes. What is the name of the surface? What is the shortest distance from the origin to the surface?

27. $x^2 + y^2 - 5z^2 + 4xy = 15$

28. $7x^2 + 4y^2 + z^2 - 8xz = 36$

29. $3x^2 + 5y^2 - 3z^2 + 6yz = 54$

30. $7x^2 + 7y^2 + 7z^2 + 10xz - 24yz = 20$

31. Find the characteristic vibration frequencies of a system of masses and springs as in Figure 12.1 if the spring constants are k, $3k$, k.

32. Do Problem 31 if the spring constants are $6k$, $2k$, $3k$.

33. Prove the Caley-Hamilton theorem (Problem 11.60) for any matrix M for which $\mathrm{D} = \mathrm{C}^{-1}\mathrm{MC}$ is diagonal. See hints in Problem 11.60.

34. In problems 6.30 and 6.31, you found the matrices e^{A} and e^{C} (put $k = 1$) where A and C are the Pauli matrices from Problem 6.6. Now find the matrix $(\mathrm{A}+\mathrm{C})$ and its powers and so find the matrix $e^{\mathrm{A}+\mathrm{C}}$ to show that $e^{\mathrm{A}+\mathrm{C}} \neq e^{\mathrm{A}}e^{\mathrm{C}}$. See Problem 6.29.

35. Show that a square matrix A has an inverse if and only if $\lambda = 0$ is not an eigenvalue of A. *Hint:* Write the condition for A to have an inverse (Section 6), and the condition for A to have the eigenvalue $\lambda = 0$ (Section 11).

36. Write the three 3 by 3 matrices for 180° rotations about the x, y, z axes. Show that these three matrices commute (contrary to what we usually expect—see Problems 7.30 and 7.31). By writing the multiplication table, show that these three matrices with the unit matrix form a group. To which order 4 group is it isomorphic? *Hint:* See Problem 13.5.

37. Show that for a given irreducible representation of a group, the character of the class consisting of the identity is always the dimension of the irreducible representation. *Hint:* What is the trace of a unit n-by-n matrix?

38. For a cyclic group, show that every element is a class by itself. Show this also for an Abelian group.