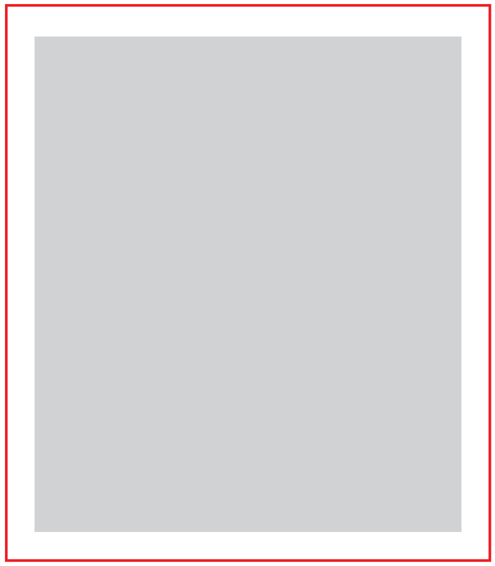

Carl Friedrich Gauss (1777–1855), who is often referred to as the Prince of Mathematics, was born on April 30, 1777, in Brunswick, Duchy of Brunswick (now Germany), into a laboring family. His talents were recognized at a very early age and were developed with the aid of a maternal uncle. In 1792, he received financial support, which continued long after he finished his schooling, from the Duke of Brunswick. In 1795, Gauss left home to study at Göttingen University, although he received his degree from the University of Helmstedt in 1799. He then returned to Brunswick, where he devoted himself to mathematical research. In 1801, he published his first major work on number theory. In the same year, he correctly predicted the orbit of the newly discovered “planet,” Ceres, using the least squares method without explaining his method. In 1805, he married Johanna Ostoff, with whom he had three children. In 1807, Gauss took the position of Director of the new observatory at Göttingen, partially because of the death of the Duke of Brunswick. He remained at Göttingen until his death 48 years later. His wife and infant son died in 1809, leaving him grief-stricken. He remarried less than a year later to a friend of his wife, but it would appear that it was more a marriage of convenience, even though they had three children together. One of his greatest contributions was the introduction of rigor to mathematical proofs, although he also worked on experimental problems of a practical nature for much of his life. Gauss did not seek fame or fortune, and he published only a portion of his extensive research. He often claimed that he had solved a problem years earlier when younger mathematicians published their results, which occasionally upset other mathematicians. However, after Gauss died, his unpublished work was discovered, and it vindicated his claims. Gauss died on February 23, 1855, when he was 77 years old.

CHAPTER 9

Linear Algebra and Vector Spaces

Although we learned how to solve simultaneous linear algebraic equations in high school, the fundamental theory behind the manipulations we learned constitutes the study of linear algebra, one of the most sophisticated and beautiful fields of mathematics. We can only touch on some of the more relevant topics for our purposes in this chapter. In Section 1, we introduce determinants and show how they can be used to solve simultaneous linear algebraic equations by Cramer's rule. As elegant as Cramer's rule may be, it is not well suited computationally, and in Section 2, we use Gaussian elimination to not only solve n linear equations in n unknowns, but other cases as well. The key quantity in Section 2 is the augmented matrix. In Section 3, we discuss matrices more fully, and learn how to multiply matrices together and to find their inverses, among other things. Section 4 deals with the idea of the rank of a matrix, one of the most important quantities for determining the nature of the solutions to sets of linear algebraic equations. Closely related to rank is the concept of linear independence of a set of vectors, which leads naturally to Section 5, where we introduce and discuss abstract vector spaces. After presenting the axioms of a vector space, we define a basis of a vector space and its dimension. When we define the operation of an inner product between the vectors in a vector space, we then have what is called an inner product space, which is the subject of Section 6. Introducing an inner product allows us to discuss the lengths of abstract vectors, the angle between them, the distance between them, orthogonality, and a number of other geometric quantities. Finally, in Section 7, we generalize the results of the previous two sections to include complex inner product spaces, which play a particularly important role in quantum mechanics.

9.1 Determinants

We frequently encounter simultaneous algebraic equations in physical applications. Let's start off with just two equations in two unknowns:

$$\begin{aligned} a_{11}x + a_{12}y &= h_1 \\ a_{21}x + a_{22}y &= h_2 \end{aligned} \tag{1}$$

If we multiply the first of these equations by a_{22} and the second by a_{12} and then subtract, we obtain

$$(a_{11}\,a_{22} - a_{12}\,a_{21})x = h_1\,a_{22} - h_2\,a_{12}$$

or

$$x = \frac{a_{22}h_1 - a_{12}h_2}{a_{11}\,a_{22} - a_{12}\,a_{21}} \tag{2}$$

Similarly, if we multiply the first equation by a_{21} and the second by a_{11} and then subtract, we get

$$y = \frac{a_{21}h_1 - a_{11}h_2}{a_{11}\,a_{22} - a_{12}\,a_{21}} \tag{3}$$

Notice that the denominators in both Equations 2 and 3 are the same. If we had started with three equations instead of two as in Equation 1, the denominators would have come out to be $a_{11}\,a_{22}\,a_{33} + a_{13}\,a_{21}\,a_{32} + a_{12}\,a_{23}\,a_{31} - a_{13}\,a_{22}\,a_{31} - a_{12}\,a_{21}\,a_{33} - a_{11}\,a_{23}\,a_{32}$.

We represent $a_{11}\,a_{22} - a_{12}\,a_{21}$ and the corresponding expression for three simultaneous equations by

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}\,a_{22} - a_{12}\,a_{21} \tag{4}$$

and

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{array}{c} a_{11}\,a_{22}\,a_{33} + a_{13}\,a_{21}\,a_{32} + a_{12}\,a_{23}\,a_{31} \\ - a_{13}\,a_{22}\,a_{31} - a_{12}\,a_{21}\,a_{33} - a_{11}\,a_{23}\,a_{32} \end{array} \tag{5}$$

The quantities introduced in Equations 4 and 5 are called a 2×2 *determinant* and a 3×3 *determinant*, respectively. The reason for introducing this notation is that it readily generalizes to n equations in n unknowns. An $n \times n$ determinant, called an *nth order determinant*, is a square array of n^2 elements arranged in n rows and n columns. We'll represent a determinant consisting of elements a_{ij} by $|\mathsf{A}|$. Note that the element a_{ij} occurs in the ith row and the jth column of $|\mathsf{A}|$. As of now, Equations 4 and 5 simply introduce symbols for the right-hand sides of these equations, but we shall develop convenient procedures for evaluating any size determinant.

Let's rearrange the right side of Equation 5 in the following way:

$$|\mathsf{A}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{array}{c} a_{11}(a_{22}\,a_{33} - a_{23}\,a_{32}) - a_{12}(a_{21}\,a_{33} - a_{23}\,a_{31}) \\ + a_{13}(a_{21}\,a_{32} - a_{22}\,a_{31}) \end{array} \tag{6}$$

Notice that each term in parentheses in Equation 6 is equal to the 2×2 determinant that is obtained by striking out the row and the column of the factor in front of each set of parentheses. Furthermore, these factors are the members of the first row of |A|. Thus, Equation 6 shows us that we can evaluate a 3×3 determinant in terms of three 2×2 determinants.

We can express Equation 6 in a systematic way by first introducing the *minor* of an element of a $n \times n$ determinant |A|. The minor M_{ij} of an element a_{ij} is a $(n-1) \times (n-1)$ determinant obtained by deleting the ith row and the jth column. We now define the *cofactor* A_{ij} of a_{ij} by $A_{ij} = (-1)^{i+j} M_{ij}$. For example, A_{12}, the cofactor of a_{12}, in Equation 6 is

$$A_{12} = (-1)^{1+2} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} = -(a_{21}a_{33} - a_{23}a_{31})$$

The introduction of cofactors allows us to write Equation 6 as

$$|\mathrm{A}| = a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13} \tag{7}$$

Equation 7 represents |A| as an *expansion in cofactors*. In particular, it is an expansion in cofactors about the first row of |A|. You can start with Equation 5 to show that |A| can be expressed in an expansion of cofactors about *any* row or *any* column (Problem 5).

Example 1:
Expand

$$|\mathrm{A}| = \begin{vmatrix} 2 & -1 & 1 \\ 0 & 3 & -1 \\ 2 & -2 & 1 \end{vmatrix}$$

in an expansion in cofactors about the second row and about the third column of |A|.

SOLUTION: We use Equation 7:

$$|\mathrm{A}| = -0(-1+2) + 3(2-2) - (-1)(-4+2) = -2$$

$$|\mathrm{A}| = 1(0-6) - (-1)(-4+2) + 1(6-0) = -2$$

We shall show below that Equation 7 readily generalizes to determinants of any order.

Example 2:
The *determinantal equation*

$$\begin{vmatrix} x & 1 & 0 & 0 \\ 1 & x & 1 & 0 \\ 0 & 1 & x & 1 \\ 0 & 0 & 1 & x \end{vmatrix} = 0$$

occurs in a quantum-mechanical treatment of a butadiene molecule. Expand this determinantal equation into a quartic equation for x.

SOLUTION: Expand about the first row of elements to obtain

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

Now expand about the first row of each of the 3×3 determinants to obtain

$$(x)(x)\begin{vmatrix} x & 1 \\ 1 & x \end{vmatrix} - (x)(1)\begin{vmatrix} 1 & 1 \\ 0 & x \end{vmatrix} - (1)\begin{vmatrix} x & 1 \\ 1 & x \end{vmatrix} - (-1)\begin{vmatrix} 0 & 1 \\ 0 & x \end{vmatrix} = 0$$

or

$$x^2(x^2 - 1) - x(x) - (1)(x^2 - 1) = 0$$

or

$$x^4 - 3x^2 + 1 = 0$$

Note that because we can choose any row or column to expand the determinant, it is easiest to take the one with the most zeroes.

Up to now we have used 3×3 determinants to illustrate how to evaluate determinants. We're going to discuss a number of general properties of determinants below, and so we need to discuss determinants more generally at this point. First, consider the product of elements of an $n \times n$ determinant

$$a_{1j_1}\, a_{2j_2}\, a_{3j_3} \cdots a_{nj_n}$$

where the j's are distinct and take on the values 1 through n. Note that there is only one element from each row and one element from each column in this product, and that the first subscripts are in their "natural order." Also note that there are $n!$ possible products because j_1 can take on one of n values, j_2 one of $n - 1$ values, and so on. Now consider the process of interchanging the elements in the above product successively until the second set of subscripts is in its "natural order." This will require either an even or an odd number of interchanges. For example,

$$a_{12}\, a_{21}\, a_{33} \longrightarrow a_{21}\, a_{12}\, a_{33}$$
$$a_{12}\, a_{23}\, a_{31} \longrightarrow a_{31}\, a_{23}\, a_{12} \longrightarrow a_{31}\, a_{12}\, a_{23}$$

require one and two interchanges, respectively. Now define the symbol

$$\epsilon_{j_1\, j_2\, j_3 \cdots j_n} = \pm 1$$

depending upon whether it takes an even or an odd number of interchanges to order the n j's into their "natural order." For example, $\epsilon_{12} = +1$, $\epsilon_{21} = -1$, $\epsilon_{132} = -1$, and so on. Finally, the general definition of an $n \times n$ determinant is

$$|\mathsf{A}| = \sum \epsilon_{j_1 j_2 j_3 \cdots j_n} a_{1j_1} a_{2j_2} a_{3j_3} \cdots a_{nj_n} \tag{8}$$

where the summation is over all the $n!$ permutations of the $j_1, j_2, \cdots, j_n$.

Equation 8 is the formal definition of a determinant. It may not be familiar to you and may be even a little formidable, but we won't have to use it to evaluate a determinant. Its primary use is to prove some important general properties of determinants. First, let's use Equation 8 to evaluate a 2×2 determinant. For a 2×2 determinant, Equation 8 gives

$$|\mathsf{A}| = \sum_{j_1} \sum_{j_2} \epsilon_{j_1 j_2} a_{1j_1} a_{2j_2}$$

In these summations, j_1 and j_2 take on the values 1 and 2, but there are no terms with $j_1 = j_2$ because the j's are distinct. Therefore,

$$|\mathsf{A}| = \epsilon_{12} a_{11} a_{22} + \epsilon_{21} a_{12} a_{21}$$

Using the fact that $\epsilon_{12} = +1$ and $\epsilon_{21} = -1$, we have

$$|\mathsf{A}| = a_{11} a_{22} - a_{12} a_{21}$$

in agreement with Equation 4. Problem 11 has you show that Equation 8 yields Equation 5 for 3×3 determinant.

The real utility of Equation 8 is that we can use it to derive general results for determinants. For example, suppose that we interchange two adjacent rows, r and $r + 1$. If we denote the resulting determinant by $|\mathsf{A}_{r \leftrightarrow r+1}|$, then

$$|\mathsf{A}_{r \leftrightarrow r+1}| = \sum \epsilon_{j_1 j_2 j_3 \cdots j_n} a_{1j_1} a_{2j_2} \cdots a_{(r+1)j_{r+1}} a_{rj_r} \cdots a_{nj_n}$$

We can get the order of the first subscripts back into natural order by one permutation, however, so $|\mathsf{A}_{r \leftrightarrow r+1}|$ differs from $|\mathsf{A}|$ by one inversion. Therefore we have that $|\mathsf{A}_{r \leftrightarrow r+1}| = -|\mathsf{A}|$. To analyze the interchange of two adjacent rows is fairly easy. Suppose now that we want to interchange two rows separated by one row. For concreteness, let these rows be rows 2, 3, and 4. We can interchange rows 2 and 4 by the process $(2, 3, 4) \rightarrow (2, 4, 3) \rightarrow (4, 2, 3) \rightarrow (4, 3, 2)$, which requires three steps. This is a general result and so we can write $|\mathsf{A}_{r \leftrightarrow r+2}| = -|\mathsf{A}|$. Problem 12 has you show that it requires $2k - 1$ steps to interchange rows r and $r + k$, so that we see that

$$|\mathsf{A}_{r \leftrightarrow r+k}| = (-1)^{2k-1} |\mathsf{A}| = -|\mathsf{A}| \tag{9}$$

or that $|\mathsf{A}|$ changes sign when *any* two rows are exchanged. (This is property 3 below.)

Equation 9 says that $|\mathsf{A}_{r \leftrightarrow r+k}| = -|\mathsf{A}|$ even if the two rows are identical. If the two rows are identical, however, then the value of $|\mathsf{A}|$ cannot change. The only way that $|A| = -|A|$ is if $|A| = 0$. Thus, we see that $|A| = 0$ if two rows (or two columns) are the same. This is property 2 below. The rest of the properties listed below can be proved using Equation 8 in a similar manner.

Example 3:
Show that

$$|A| = \begin{vmatrix} 1 & 2 & 2 \\ 0 & -1 & 0 \\ 3 & 1 & 6 \end{vmatrix} = 0$$

SOLUTION: Expand the cofactors about the first column to obtain

$$|A| = 1(-6) - 0(10) + 3(+2) = 0$$

Notice that the third column is twice the first column. We'll see below that $|A| = 0$ because of this.

Some properties of determinants that are useful to know are:

1. The value of a determinant is unchanged if the rows are made into columns in the same order; in other words, first row becomes first column, second row becomes second column, and so on. For example,

$$\begin{vmatrix} 1 & 2 & 5 \\ -1 & 0 & -1 \\ 3 & 1 & 2 \end{vmatrix} = \begin{vmatrix} 1 & -1 & 3 \\ 2 & 0 & 1 \\ 5 & -1 & 2 \end{vmatrix} = -6$$

2. If any two rows or columns are the same, the value of the determinant is zero. For example,

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

3. If any two rows or columns are interchanged, the sign of the determinant is changed. For example,

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

4. If every element in a single row or column is multiplied by a factor k, the value of the determinant is multiplied by k (Problem 13). For example,

$$\begin{vmatrix} 6 & 8 \\ -1 & 2 \end{vmatrix} = 2\begin{vmatrix} 3 & 4 \\ -1 & 2 \end{vmatrix} = 20$$

5. If any row or column is written as the sum or difference of two or more terms, the determinant can be written as the sum or difference of two or more determinants

according to (Problem 14)

$$\begin{vmatrix} a_{11} \pm a'_{11} & a_{12} & a_{13} \\ a_{21} \pm a'_{21} & a_{22} & a_{23} \\ a_{31} \pm a'_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \pm \begin{vmatrix} a'_{11} & a_{12} & a_{13} \\ a'_{21} & a_{22} & a_{23} \\ a'_{31} & a_{32} & a_{33} \end{vmatrix}$$

For example,

$$\begin{vmatrix} 3 & 3 \\ 2 & 6 \end{vmatrix} = \begin{vmatrix} 2+1 & 3 \\ -2+4 & 6 \end{vmatrix} = \begin{vmatrix} 2 & 3 \\ -2 & 6 \end{vmatrix} + \begin{vmatrix} 1 & 3 \\ 4 & 6 \end{vmatrix}$$

6. The value of a determinant is unchanged if one row or column is added or subtracted to another, as in

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

For example

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

where we added column 2 to column 1. This procedure may be repeated n times to obtain

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11}+na_{12} & a_{12} & a_{13} \\ a_{21}+na_{22} & a_{22} & a_{23} \\ a_{31}+na_{32} & a_{32} & a_{33} \end{vmatrix} \tag{10}$$

This result is easy to prove:

$$\begin{aligned} \begin{vmatrix} a_{11}+na_{12} & a_{12} & a_{13} \\ a_{21}+na_{22} & a_{22} & a_{23} \\ a_{31}+na_{32} & a_{32} & a_{33} \end{vmatrix} &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + n \begin{vmatrix} a_{12} & a_{12} & a_{13} \\ a_{22} & a_{22} & a_{23} \\ a_{32} & a_{32} & a_{33} \end{vmatrix} \\ &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + 0 \end{aligned}$$

where we used Rule 5 to write the first line. The second determinant on the right side equals zero because two columns are the same.

Example 4:
Show that the value of

$$|A| = \begin{vmatrix} 3 & -1 & 2 \\ -2 & 1 & 1 \\ 1 & 4 & 3 \end{vmatrix}$$

is unchanged if we add 2 times row 2 to row 3.

SOLUTION: First of all, $|\mathsf{A}| = 3(-1) - (-1)(-7) + 2(-9) = -28$. Now

$$\begin{vmatrix} 3 & -1 & 2 \\ -2 & 1 & 1 \\ -3 & 6 & 5 \end{vmatrix} = 3(-1) - (-1)(-7) + 2(-9) = -28$$

Simultaneous linear algebraic equations can be solved in terms of determinants. For simplicity, we will consider only a pair of equations, but the final result is easy to generalize. Consider the two equations

$$\begin{aligned} a_{11}x + a_{12}y &= h_1 \\ a_{21}x + a_{22}y &= h_2 \end{aligned} \tag{11}$$

The determinant of the coefficients of x and y is

$$|\mathsf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

According to Rule 4,

$$\begin{vmatrix} a_{11}x & a_{12} \\ a_{21}x & a_{22} \end{vmatrix} = x\,|\mathsf{A}|$$

Furthermore, according to Rule 6,

$$\begin{vmatrix} a_{11}x + a_{12}y & a_{12} \\ a_{21}x + a_{22}y & a_{22} \end{vmatrix} = x\,|\mathsf{A}| \tag{12}$$

If we substitute Equation 11 into Equation 12, then we have

$$\begin{vmatrix} h_1 & a_{12} \\ h_2 & a_{22} \end{vmatrix} = x\,|\mathsf{A}|$$

Solving for x gives

$$x = \frac{\begin{vmatrix} h_1 & a_{12} \\ h_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} \tag{13}$$

Similarly, we get

$$y = \frac{\begin{vmatrix} a_{11} & h_1 \\ a_{21} & h_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} \tag{14}$$

Notice that Equations 13 and 14 are identical to Equations 2 and 3. The solution for x and y in terms of determinants is called *Cramer's rule*. Note that the determinant in the numerator is obtained by replacing the column in |A| that is associated with the unknown quantity and replacing it with the column associated with the right sides of Equations 11. We shall show after the next Example that this result is readily extended to more than two simultaneous equations.

Example 5:
Use Cramer's rule to solve the equations

$$x + y + z = 2$$

$$2x - y - z = 1$$

$$x + 2y - z = -3$$

SOLUTION: The extension of Equations 13 and 14 is

$$x = \frac{\begin{vmatrix} 2 & 1 & 1 \\ 1 & -1 & -1 \\ -3 & 2 & -1 \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 1 \\ 2 & -1 & -1 \\ 1 & 2 & -1 \end{vmatrix}} = \frac{9}{9} = 1$$

Similarly,

$$y = \frac{\begin{vmatrix} 1 & 2 & 1 \\ 2 & 1 & -1 \\ 1 & -3 & -1 \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 1 \\ 2 & -1 & -1 \\ 1 & 2 & -1 \end{vmatrix}} = \frac{-9}{9} = -1$$

and

$$z = \frac{\begin{vmatrix} 1 & 1 & 2 \\ 2 & -1 & 1 \\ 1 & 2 & -3 \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 1 \\ 2 & -1 & -1 \\ 1 & 2 & -1 \end{vmatrix}} = \frac{18}{9} = 2$$

Before we finish this section, let's discuss the expansion of a determinant in terms of its cofactors and how it can be used to derive Cramer's rule for n simultaneous equations. Equation 7 says that

$$|\mathrm{A}| = a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13}$$

where A_{ij} is the cofactor of a_{ij}. We derived Equation 7 from the expansion of a 3×3 determinant, but it is a general result in the sense that if $|\mathsf{A}|$ is an $n \times n$ determinant, then

$$|\mathsf{A}| = a_{11}A_{11} + a_{12}A_{12} + \cdots + a_{1n}A_{1n}$$

or, in summation notation,

$$|\mathsf{A}| = \sum_{j=1}^{n} a_{1j}A_{1j} \tag{15}$$

Equation 15 can be derived directly from Equation 8, but the general proof is a little long. (See "Lipschutz" in the References.)

Equation 15 represents an expansion of $|\mathsf{A}|$ about its first row. More generally, we can expand about the ith row of $|\mathsf{A}|$, and so we also have

$$|\mathsf{A}| = \sum_{j=1}^{n} a_{ij}A_{ij} \tag{16}$$

If we had expanded about the ith column, Equation 16 would read

$$|\mathsf{A}| = \sum_{k=1}^{n} a_{ki}A_{ki} \tag{17}$$

Now if we replace the ith column by some other column, say the jth column, then $|\mathsf{A}| = 0$ because $|\mathsf{A}|$ now has two identical columns. The cofactors in Equation 17 don't change, however, so we have the result

$$\sum_{k=1}^{n} a_{kj}A_{ki} = 0 \qquad i \neq j \tag{18}$$

Notice that we can combine Equations 17 and 18 to read

$$\sum_{k=1}^{n} a_{kj}A_{ki} = |\mathsf{A}|\,\delta_{ij} \tag{19}$$

where δ_{ij}, the *Kronecker delta*

$$\delta_{ij} = \begin{matrix} 1 & \qquad i = j \\ 0 & \qquad i \neq j \end{matrix} \tag{20}$$

We will refer to Equation 19 several times in later chapters.

We can use Equation 19 to derive Cramer's rule. Start with the set of equations

$$\begin{matrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = h_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = h_2 \\ \vdots \qquad \vdots \qquad \ddots \qquad \vdots \qquad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = h_n \end{matrix} \tag{21}$$

Multiply the first equation by A_{11}, the second by A_{21}, and so on, and then add to obtain

$$\sum_{j=1}^{n} a_{j1}A_{j1}x_1 + \sum_{j=1}^{n} a_{j2}A_{j1}x_2 + \cdots + \sum_{j=1}^{n} a_{jn}A_{j1}x_n = \sum_{j=1}^{n} A_{j1}h_j \qquad (22)$$

According to Equation 19, all the terms on the left side except the first one vanish, and so the left side is equal to $|\mathrm{A}|\, x_1$. According to Equation 17, the right side is the original determinant $|\mathrm{A}|$ but with the first column replaced by $h_1, h_2, \ldots, h_n$. Thus, we see that

$$|\mathrm{A}|\, x_1 = \begin{vmatrix} h_1 & a_{12} & \cdots & a_{1n} \\ h_2 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_n & a_{n2} & \cdots & a_{nn} \end{vmatrix} \qquad (23)$$

Similar equations for x_2 through x_n can be obtained by multiplying Equation 21 by other cofactors (Problem 19).

9.1 Problems

1. Evaluate the determinant $|\mathrm{A}| = \begin{vmatrix} 2 & 1 & 1 \\ -1 & 3 & 2 \\ 2 & 0 & 1 \end{vmatrix}$. Add column 2 to column 1 to get $\begin{vmatrix} 3 & 1 & 1 \\ 2 & 3 & 2 \\ 2 & 0 & 1 \end{vmatrix}$ and evaluate it. Compare your result with the value of $|\mathrm{A}|$. Now add row 2 to row 1 of $|\mathrm{A}|$ to get $\begin{vmatrix} 1 & 4 & 3 \\ -1 & 3 & 2 \\ 2 & 0 & 1 \end{vmatrix}$ and evaluate it. Compare your result with the value of $|\mathrm{A}|$ above.

2. Interchange columns 1 and 3 in $|\mathrm{A}|$ in Problem 1 and evaluate the resulting determinant. Compare your result with the value of $|\mathrm{A}|$. Interchange rows 1 and 2 and do the same.

3. Evaluate the determinant $|\mathrm{A}| = \begin{vmatrix} 1 & 6 & 1 \\ -2 & 4 & -2 \\ 1 & -3 & 1 \end{vmatrix}$. Can you determine its value by inspection? What about $|\mathrm{A}| = \begin{vmatrix} 2 & 6 & 1 \\ -4 & 4 & -2 \\ 2 & -3 & 1 \end{vmatrix}$?

4. Starting with $|\mathrm{A}|$ in Problem 1, add two times the third row to the second row and evaluate the resulting determinant.

5. Use Equation 5 to derive an expansion in cofactors about the third column.

6. Evaluate $|\mathrm{A}| = \begin{vmatrix} 1 & \sin x & \cos x \\ 0 & \cos x & -\sin x \\ 0 & -\sin x & -\cos x \end{vmatrix}$.

7. Find the values of x that satisfy the determinantal equation, $\begin{vmatrix} x & 1 & 1 & 1 \\ 1 & x & 0 & 0 \\ 1 & 0 & x & 0 \\ 1 & 0 & 0 & x \end{vmatrix} = 0.$

8. Find the values of x that satisfy the determinantal equation, $\begin{vmatrix} x & 1 & 0 & 1 \\ 1 & x & 1 & 0 \\ 0 & 1 & x & 1 \\ 1 & 0 & 1 & x \end{vmatrix} = 0.$

9. Show that $\begin{vmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1.$

10. Evaluate (a) ϵ_{13245}, (b) ϵ_{32145}, and (c) ϵ_{54321}.

11. Show that Equation 8 yields Equation 5 for a 3×3 determinant.

12. Show that it requires $2k - 1$ steps to interchange rows r and $r + k$ in a determinant.

13. Use Equation 8 to prove property 4.

14. Use Equation 8 to prove property 5.

15. Solve the following set of equations using Cramer's rule:

$$x + y = 2$$
$$3x - 2y = 5$$

16. Solve the following set of equations using Cramer's rule:

$$x + 2y + 3z = -5$$
$$-x - 3y + z = -14$$
$$2x + y + z = 1$$

17. Use Cramer's rule to solve

$$x + 2y = 3$$
$$2x + 4y = 1$$

What goes wrong here? Why?

18. Verify Equation 19 for the determinant in Equation 5 for $i = 2$ and $j = 1$.

19. Derive an equation for x_2 starting with Equation 21.

20. Use any CAS to evaluate $\begin{vmatrix} 2 & 5 & 1 \\ 3 & 1 & 2 \\ -2 & 1 & 0 \end{vmatrix}$.

21. Use any CAS to evaluate $\begin{vmatrix} 1 & 0 & 3 & -2 \\ 6 & 1 & -1 & 3 \\ 2 & 0 & 1 & 1 \\ 4 & 3 & 2 & 5 \end{vmatrix}$.

9.2 Gaussian Elimination

Although Cramer's rule provides a systematic, compact approach to solving simultaneous linear algebraic equations, it is not a convenient computational procedure because of the necessity of evaluating numerous determinants. Nor does Cramer's rule apply if the number of equations does not equal the number of unknowns. In this section, we shall present an alternative method of solving simultaneous equations that is not only computationally convenient, but is not limited to $n \times n$ systems. Before we present this method, however, we shall discuss some general ideas about systems of linear algebraic equations.

Let's start off again with two equations in two unknowns.

$$a_{11}x_1 + a_{12}x_2 = h_1$$
$$a_{21}x_1 + a_{22}x_2 = h_2$$

Geometrically we have three possibilities: 1. the graphs of the two straight lines intersect and we have a unique solution; 2. the lines are parallel and we have no solution; and 3. the lines coincide and we have an infinite number of solutions (Figure 9.1). An example of the first case is

$$2x_1 + x_2 = 3$$
$$x_1 - 3x_2 = -2$$

with $x_1 = 1$ and $x_2 = 1$ as its unique solution. An example of the second case is

$$2x_1 + x_2 = 3$$
$$2x_1 + x_2 = 5$$

These two lines are parallel and have no point in common. An example of the third case is

$$2x_1 + x_2 = 3$$
$$4x_1 + 2x_2 = 6$$

These two lines actually coincide and the solution can be written as $x_2 = 3 - 2x_1$, where x_1 can take on any value.

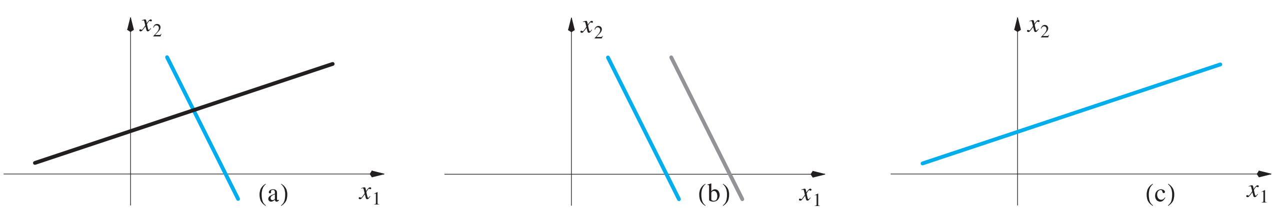

Figure 9.1
The three geometric possibilities of two linear equations in two unknowns, x_1 and x_2. (a) The colored line ($2x_1 + x_2 = 3$) and the black line ($x_1 - 3x_2 = -2$) have a unique point of intersection. (b) The colored line ($2x_1 + x_2 = 3$) and the black line ($2x_1 + x_2 = 5$) are parallel and have no point of intersection. (c) The two lines ($2x_1 - x_2 = 1$) and ($4x_1 - 2x_2 = 2$) superimpose, and so there is an infinite number of solutions.

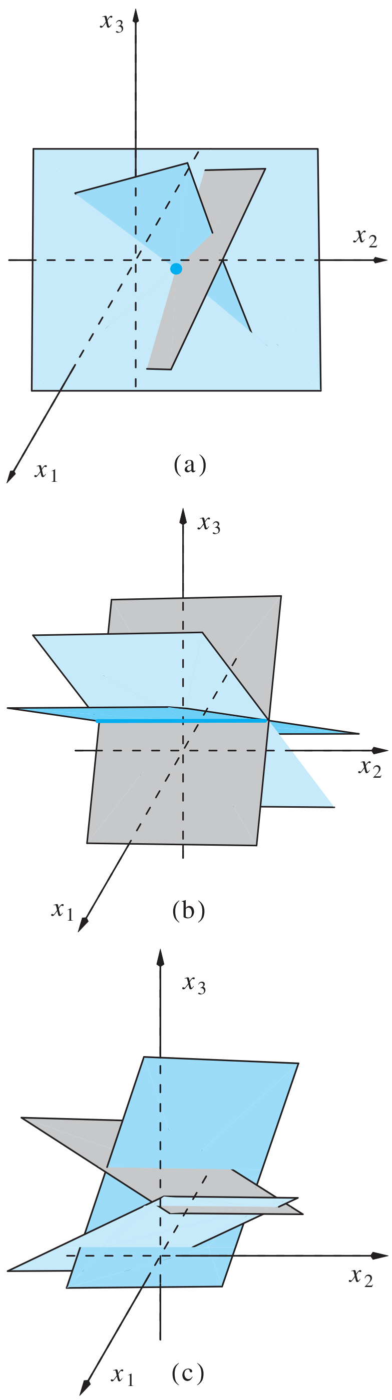

Figure 9.2
The three geometric possibilities of the graphs of three linear algebraic equations: (a) a unique solution; (b) an infinite number of solutions; and (c) no solution.

The geometric interpretation for a 3×3 set of equations involves planes. If the three planes intersect at one point, there is a unique solution (Figure 9.2a). If the three planes intersect as shown in Figure 9.2b, there is an infinite number of solutions. And if the three planes have no common point of intersection, as in Figure 9.2c, there is no solution.

Let's now consider a general $n \times n$ system:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= h_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= h_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= h_n \end{aligned} \tag{1}$$

If all the h_j in Equations 1 are equal to zero, the system of equations is called *homogeneous*. If at least one $h_j \neq 0$, the system is called *nonhomogeneous*. We may re-express Equations 1 as

$$\mathsf{AX} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{pmatrix} = \mathsf{H} \tag{2}$$

if we agree that the rth equation in Equations 1 is formed by multiplying each element of the rth row of A by the corresponding element of X, adding the results, and then equating the sum to the rth element in H. For example, the second line in Equations 1 is given by

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = h_2 \tag{3}$$

Thus, we can write Equations 1 as AX = H, where A is called the *coefficient matrix*, X is called the column vector of unknowns, and H is called the constant vector of the system. Note that the left side of Equation 3 can be viewed as the dot product of the rth row vector of A and the column vector X.

The quantity A in Equation 2 is an $n \times n$ matrix, which is an array of elements that obeys certain algebraic rules such as Equation 3. We shall discuss matrices and their algebraic rules in some detail in the next section. The important point here is that a matrix is *not* equal to a single number. However, we can associate a determinant with a matrix and write

$$\det \mathsf{A} = |\mathsf{A}| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \tag{4}$$

which is a single number.

Clearly, the coefficient matrix A must have a lot to say about the existence and nature of the solutions to Equations 1. Cramer's rule contains |A| in the denominators of the equations for the unknowns, so |A| cannot equal zero if there exists a unique solution to a nonhomogeneous system. If $|\mathsf{A}| = 0$, then A is said to be *singular*; if $|\mathsf{A}| \neq 0$, then A is said to be *nonsingular*. In fact, we have the following theorem, which we shall prove later:

The $n \times n$ system $\mathsf{AX} = \mathsf{H}$ *has a unique solution if and only if* A *is nonsingular.*

If $\mathsf{H} = 0$, that is, if the system is homogeneous, then $x_1 = x_2 = \cdots = x_n = 0$ (called the *trivial solution*) is always a solution. But the above theorem says that a solution is unique if A is nonsingular, so if A is nonsingular, there is *only* a trivial solution to a homogeneous system. To have a nontrivial solution to an $n \times n$ set of homogeneous equations, the coefficient matrix must be singular.

We shall now spend the rest of this section actually finding solutions to systems of linear equations, even if the coefficient matrix is not square. Let's consider the equations

$$\begin{aligned} 2x_1 + x_2 + 3x_3 &= 4 \\ 2x_1 - 2x_2 - x_3 &= 1 \\ -2x_1 + 4x_2 + x_3 &= 1 \end{aligned} \tag{5}$$

The coefficient matrix and the constant vector are

$$\mathsf{A} = \begin{pmatrix} 2 & 1 & 3 \\ 2 & -2 & -1 \\ -2 & 4 & 1 \end{pmatrix} \qquad \text{and} \qquad \mathsf{H} = \begin{pmatrix} 4 \\ 1 \\ 1 \end{pmatrix}$$

We now form a new matrix, called the *augmented matrix*, by adjoining H to A so that it is the last column

$$\mathsf{A|H} = \left(\begin{array}{rrr|r} 2 & 1 & 3 & 4 \\ 2 & -2 & -1 & 1 \\ -2 & 4 & 1 & 1 \end{array}\right) \tag{6}$$

Clearly this matrix contains *all* the information in Equations 5, and is just a succinct expression of them. Just as we may multiply any of the equations in Equations 5 by a nonzero constant without jeopardizing the solutions, we may multiply any row of A|H without altering its content. Similarly, we may interchange any two rows of either Equations 5 or A|H and replace any row by the sum of that row and a constant times another row. These three operations are called *elementary (row) operations*:

1. We may multiply any row by a nonzero constant.
2. We may interchange any pair of rows.
3. We may replace any row by the sum of that row and a constant times another row.

The key point is that these elementary operations produce an *equivalent system*, that is, a system with the same solution as the original system. Matrices that differ by a set of elementary operations are said to be *equivalent*.

We are now going to manipulate A|H by elementary operations so that there are zeros in the lower left positions of A|H. Add −1 times row 1 to row 2 and add row 1 to row 3 in Equation 5 to get

$$\left(\begin{array}{rrr|r} 2 & 1 & 3 & 4 \\ 0 & -3 & -4 & -3 \\ 0 & 5 & 4 & 5 \end{array}\right)$$

Now add 5/3 times row 2 to row 3 to get

$$\left(\begin{array}{rrr|r} 2 & 1 & 3 & 4 \\ 0 & -3 & -4 & -3 \\ 0 & 0 & -8/3 & 0 \end{array}\right)$$

Write out the corresponding system of equations:

$$\begin{aligned} 2x_1 + x_2 + 3x_3 &= 4 \\ -3x_2 - 4x_3 &= -3 \\ -8x_3/3 &= 0 \end{aligned}$$

and work your way from bottom to top to find $x_3 = 0$, $x_2 = 1$, and $x_1 = 3/2$.

This procedure is called *Gaussian elimination* and the final form of A|H is said to be in *echelon form*. The following Examples provide two other applications of Gaussian elimination.

Example 1:
Solve the equations

$$\begin{aligned} x_1 + x_2 - x_3 &= 2 \\ 2x_1 - x_2 + 3x_3 &= 5 \\ 3x_1 + 2x_2 - 2x_3 &= 5 \end{aligned}$$

SOLUTION: The augmented matrix is

$$\left(\begin{array}{rrr|r} 1 & 1 & -1 & 2 \\ 2 & -1 & 3 & 5 \\ 3 & 2 & -2 & 5 \end{array}\right)$$

Add −2 times row 1 to row 2 and −3 times row 1 to row 3 to obtain

$$\left(\begin{array}{rrr|r} 1 & 1 & -1 & 2 \\ 0 & -3 & 5 & 1 \\ 0 & -1 & 1 & -1 \end{array}\right)$$

To avoid introducing fractions, interchange rows 2 and 3 and then add −3 times the new row 2 to the new row 3 to get

$$\left(\begin{array}{rrr|r} 1 & 1 & -1 & 2 \\ 0 & -1 & 1 & -1 \\ 0 & 0 & 2 & 4 \end{array}\right)$$

The corresponding set of equations is

$$\begin{aligned} x_1 + x_2 - x_3 &= 2 \\ -x_2 + x_3 &= -1 \\ 2x_3 &= 4 \end{aligned}$$

Solving these equations from top to bottom gives $x_3 = 2$, $x_2 = 3$, and $x_1 = 1$.

Example 2:
Solve the equations

$$\begin{aligned} x_1 + x_2 + x_3 &= -2 \\ x_1 - x_2 + x_3 &= 2 \\ -x_1 + x_2 - x_3 &= -2 \end{aligned}$$

SOLUTION: The augmented matrix is

$$\left(\begin{array}{rrr|r} 1 & 1 & 1 & -2 \\ 1 & -1 & 1 & 2 \\ -1 & 1 & -1 & -2 \end{array}\right)$$

Add -1 times row 1 to row 2 and add row 1 to row 3 to obtain

$$\left(\begin{array}{rrr|r} 1 & 1 & 1 & -2 \\ 0 & -2 & 0 & 4 \\ 0 & 2 & 0 & -4 \end{array}\right)$$

Now add row 2 to row 3 to get

$$\left(\begin{array}{rrr|r} 1 & 1 & 1 & -2 \\ 0 & -2 & 0 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

In this case, the corresponding set of equations is

$$\begin{aligned} x_1 + x_2 + x_3 &= -2 \\ -2x_2 \qquad &= 4 \\ 0x_3 &= 0 \end{aligned}$$

The solutions are $x_3 =$ arbitrary, $x_2 = -2$, and $x_1 = -x_3$, so the solution is not unique. Note that $|A| = 0$ in this case, so we should not expect a unique solution.

Example 3:
Solve the equations

$$2x_1 - x_3 = -1$$
$$3x_1 + 2x_2 = 4$$
$$4x_2 + 3x_3 = 6$$

SOLUTION: The augmented matrix is

$$\left(\begin{array}{ccc|c} 2 & 0 & -1 & -1 \\ 3 & 2 & 0 & 4 \\ 0 & 4 & 3 & 6 \end{array}\right)$$

Add $-3/2$ times row 1 to row 2 to obtain

$$\left(\begin{array}{ccc|c} 2 & 0 & -1 & -1 \\ 0 & 2 & 3/2 & 11/2 \\ 0 & 4 & 3 & 6 \end{array}\right)$$

Now add -2 times row 2 to row 3 to get

$$\left(\begin{array}{ccc|c} 2 & 0 & -1 & -1 \\ 0 & 2 & 3/2 & 11/2 \\ 0 & 0 & 0 & -10/2 \end{array}\right)$$

This last line says that $-10/2 = 0$, meaning that there is no solution to the above equations. They are inconsistent.

Up to now we have considered only $n \times n$ systems of equations. Suppose we have a system with more equations than unknowns. (Such systems are called *overdetermined*.) For example, consider

$$\begin{aligned} x_1 + x_2 &= 4 \\ 3x_1 - 4x_2 &= 9 \\ 5x_1 - 2x_2 &= 17 \end{aligned} \tag{7}$$

The augmented matrix is

$$\left(\begin{array}{cc|c} 1 & 1 & 4 \\ 3 & -4 & 9 \\ 5 & -2 & 17 \end{array}\right)$$

Multiplying row 1 by -3 and adding to row 2, and then multiplying row 1 by -5 and adding to row 3 gives

$$\left(\begin{array}{cc|c} 1 & 1 & 4 \\ 0 & -7 & -3 \\ 0 & -7 & -3 \end{array}\right)$$

Now multiply row 2 by −1 and add to row 3:

$$\left(\begin{array}{rr|r} 1 & 1 & 4 \\ 0 & -7 & -3 \\ 0 & 0 & 0 \end{array}\right)$$

The corresponding algebraic equations are

$$\begin{aligned} x_1 + x_2 &= 4 \\ -7x_2 &= -3 \end{aligned}$$

and the solution is $x_2 = 3/7$ and $x_1 = 25/7$. The coefficient matrix of the final set of equations is $\begin{pmatrix} 1 & 1 \\ 0 & -7 \end{pmatrix}$ and is nonsingular; thus the solution is unique.

Example 4:
Solve the equations

$$\begin{aligned} 3x_1 - 2x_2 &= 2 \\ -6x_1 + 4x_2 &= -4 \\ -3x_1 + 2x_2 &= 2 \end{aligned}$$

SOLUTION: The augmented matrix is

$$\left(\begin{array}{rr|r} 3 & -2 & 2 \\ -6 & 4 & -4 \\ -3 & 2 & 2 \end{array}\right)$$

Adding 2 times row 1 to row 2, then adding row 1 to row 3, and then interchanging the resultant rows 2 and 3 gives

$$\left(\begin{array}{rr|r} 3 & -2 & 2 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{array}\right)$$

The second line here claims that $0 = 4$, so there is no solution.

Example 5:
Solve the equations

$$\begin{aligned} x_1 - 2x_2 &= 3 \\ 2x_1 - 4x_2 &= 6 \\ -3x_1 + 6x_2 &= -9 \end{aligned}$$

SOLUTION: The augmented matrix is

$$\left(\begin{array}{rr|r} 1 & -2 & 3 \\ 2 & -4 & 6 \\ -3 & 6 & -9 \end{array}\right)$$

Add -2 times row 1 to row 2 and 3 times row 1 to row 3 to get

$$\left(\begin{array}{rr|r} 1 & -2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

The corresponding algebraic equations are $x_1 - 2x_2 = 3$, or $x_1 = 2x_2 + 3$. Thus, there is an infinite number of solutions.

In summary, if we have more equations than unknowns, then there are three possible outcomes: 1. there is a unique solution; 2. there is no solution; and 3. there is an infinite number of solutions. In each case, Gaussian elimination leads us to the correct result.

If we have more unknowns than equations, as in

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 0 \\ x_1 + 3x_2 + 2x_3 + 4x_4 &= 0 \\ 2x_1 \qquad\quad + x_3 - x_4 &= 0 \end{aligned} \tag{8}$$

then the system is called *underdetermined*. The augmented matrix is

$$\left(\begin{array}{rrrr|r} 1 & 1 & 1 & 1 & 0 \\ 1 & 3 & 2 & 4 & 0 \\ 2 & 0 & 1 & -1 & 0 \end{array}\right) \sim \left(\begin{array}{rrrr|r} 1 & 1 & 1 & 1 & 0 \\ 0 & 2 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

where we have written $\sim$ to indicate that the first matrix is equivalent to the second; that is, it can be manipulated into the second by elementary operations. The corresponding algebraic equations are

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 0 \\ 2x_2 + x_3 + 3x_4 &= 0 \end{aligned}$$

Solving for x_1 and x_2 in terms of x_3 and x_4, we have $x_1 = (x_4 - x_3)/2$ and $x_2 = -(x_3 + 3x_4)/2$. Thus, there is an infinite number of solutions in this case.

Example 6:

Solve the equations

$$\begin{aligned} x_1 + x_2 + 2x_3 + x_4 &= 5 \\ 2x_1 + 3x_2 - x_3 - 2x_3 &= 2 \\ 4x_1 + 5x_2 + 3x_3 \qquad &= 7 \end{aligned}$$

SOLUTION: The augmented matrix is

$$\left(\begin{array}{cccc|c} 1 & 1 & 2 & 1 & 5 \\ 2 & 3 & -1 & -2 & 2 \\ 4 & 5 & 3 & 0 & 7 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 1 & 2 & 1 & 5 \\ 0 & 1 & -5 & -4 & -8 \\ 0 & 0 & 0 & 0 & -5 \end{array}\right)$$

The last line corresponds to $0 = -5$, so there are no solutions.

There are only two possibilities when there are more unknowns than equations. Either there is an infinite number of solutions, or there are no solutions. Either way, Gaussian elimination will give the correct answer.

Before we finish this section, we should point out that any CAS can be used to solve simultaneous linear equations (Problems 14 through 16).

9.2 Problems

1. Solve the equations

$$\begin{aligned} x_1 + 2x_2 - 3x_3 &= 4 \\ 2x_1 - x_2 + x_3 &= 1 \\ 3x_1 + 2x_2 - x_3 &= 5 \end{aligned}$$

2. Solve the equations

$$\begin{aligned} 2x + 5y + z &= 5 \\ x + 4y + 2z &= 1 \\ 4x + 10y - z &= 1 \end{aligned}$$

3. Solve the equations

$$\begin{aligned} x + y \phantom{{}+z} &= 1 \\ x \phantom{{}+y} + z &= 1 \\ 2x + y + z &= 0 \end{aligned}$$

4. Solve the equations

$$\begin{aligned} 2x_1 + x_2 - x_3 + x_4 &= -2 \\ x_1 - x_2 - x_3 + x_4 &= 1 \\ x_1 - 4x_2 - 2x_3 + 2x_4 &= 6 \\ 4x_1 + x_2 - 3x_3 + 3x_4 &= -1 \end{aligned}$$

5. Solve the equations

$$\begin{aligned} x + 2y - 6z &= 2 \\ x + 4y + 4z &= 1 \\ 3x + 10y + 2z &= -1 \end{aligned}$$

6. Solve the equations

$$\begin{aligned} x_1 \phantom{{}+2x_2} + 2x_3 - x_4 &= 3 \\ x_2 + x_3 \phantom{{}-x_4} &= 5 \\ 3x_1 + 2x_2 \phantom{{}+x_3} - 2x_4 &= -1 \\ -x_3 + 4x_4 &= 13 \\ 2x_1 \phantom{{}+x_2} - x_3 + 3x_4 &= 11 \end{aligned}$$

7. Solve the equations

$$\begin{aligned} 2x_1 - 4x_2 + x_3 - 3x_4 &= 6 \\ x_1 - 2x_2 + 3x_3 + 6x_4 &= 2 \end{aligned}$$

8. Solve the equations

$$\begin{aligned} x_1 - 2x_2 + 3x_3 - x_4 &= 0 \\ -x_1 \phantom{{}-2x_2} + 2x_3 + x_4 &= 0 \\ 2x_1 + x_2 \phantom{{}+3x_3} - 2x_4 &= 0 \end{aligned}$$

9. Solve the equations

$$x_1 - 2x_2 + x_3 - x_4 + 2x_5 = -7$$
$$x_2 + x_3 + 2x_4 - x_5 = 5$$
$$x_1 - x_2 + 2x_3 + 2x_4 + 2x_5 = -1$$

10. For what values of λ will the following equations have a unique solution?

$$x + y = \lambda x$$
$$-x + y = \lambda y$$

11. For what values of λ will the following equations have a unique solution?

$$x + y + z = 6$$
$$x + \lambda y + \lambda z = 2$$

12. For what values of λ will the equations in the previous problem have an infinite number of solutions?

13. For what values of λ will the following equations have a unique solution?

$$5x + \lambda y = 4$$
$$4x + 3y = 3$$
$$\lambda x - 6y = 3$$

14. Use any CAS to solve the equations in Problems 2 through 4.

15. Use any CAS to solve the equations in Problems 5 through 7.

16. Use any CAS to solve the equations in Problems 8 and 9.

9.3 Matrices

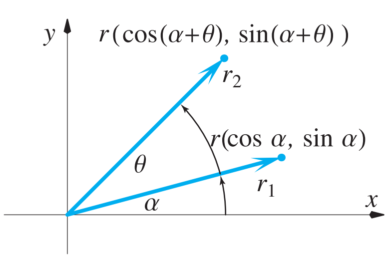

Figure 9.3
A pictorial representation of the rotation of the vector $\mathbf{r}_1$ through an angle θ in a counterclockwise direction. The result is the vector $\mathbf{r}_2$.

Up to now we have used matrices only as a representation of the coefficients in systems of linear algebraic equations. The utility of matrices far exceeds that use, however, and in this section we shall present some of the basic properties of matrices. Then, in Chapter 10, we will discuss a number of important physical applications of matrices.

Many physical operations such as magnification, rotation, and reflection through a plane can be represented mathematically by quantities called matrices. Consider the lower of the two vectors shown in Figure 9.3. The x and y components of the vector are given by $x_1 = r\cos\alpha$ and $y_1 = r\sin\alpha$, where r is the length of $\mathbf{r}_1$. Now, let's rotate the vector counterclockwise through an angle θ, so that $x_2 = r\cos(\alpha+\theta)$ and $y_2 = r\sin(\alpha+\theta)$ (see Figure 9.3). Using trigonometric formulas, we can write

$$x_2 = r\cos(\alpha+\theta) = r\cos\alpha\cos\theta - r\sin\alpha\sin\theta$$
$$y_2 = r\sin(\alpha+\theta) = r\cos\alpha\sin\theta + r\sin\alpha\cos\theta$$

or

$$x_2 = x_1 \cos\theta - y_1 \sin\theta$$
$$y_2 = x_1 \sin\theta + y_1 \cos\theta \tag{1}$$

We can display the set of coefficients here in the form

$$\mathsf{R} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{2}$$

We have expressed R in the form of a *matrix*, which is an array of numbers (or functions in this case) that obey a certain set of rules, called *matrix algebra*. Unlike determinants, matrices do not have to be square arrays. The matrix R in Equation 2 corresponds to a rotation of a vector through an angle θ.

Example 1:
Show that the matrices

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

correspond to reflections of a vector through the x axis and y axis, respectively.

SOLUTION: If we reflect the vector $\mathbf{r}_1 = x_1\,\mathbf{i} + y_1\,\mathbf{j}$ through the x axis, we obtain the vector $\mathbf{r}_2 = x_2\,\mathbf{i} + y_2\,\mathbf{j} = x_1\,\mathbf{i} - y_1\,\mathbf{j}$ (Figure 9.4a). Thus, we can write

$$x_2 = x_1 = x_1 + 0y_1$$
$$y_2 = -y_1 = 0x_1 - y_1$$

The set of coefficients can be expressed as

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

so we see that the matrix A corresponds to a reflection of a vector through the x axis. Similarly, for a reflection through the y axis

$$x_2 = -x_1 = -x_1 + 0y_1$$
$$y_2 = y_1 = 0x_1 + y_1$$

so that $\mathsf{B} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ corresponds to a reflection through the y axis (Figure 9.4b).

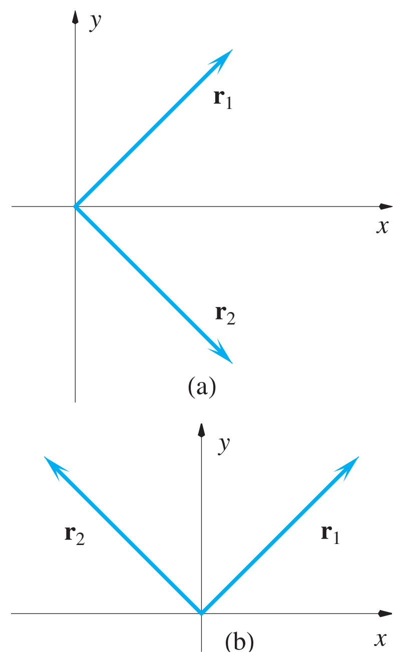

Figure 9.4
A pictorial representation of the reflection of a vector through (a) the x axis and (b) the y axis.

We shall see that matrices usually correspond to physical transformations.

The entries in a matrix A are called its *matrix elements* and are denoted by a_{ij} where, as in the case of determinants, i designates the row and j designates the column. Two matrices, A and B, are equal if and only if they are of the same dimension (that is, have the same number of rows and the same number of columns), and if and only if $a_{ij} = b_{ij}$ for all i and j. In other words, equal matrices are identical. Matrices can be added or subtracted only if they have the same number of rows and columns, in which case the elements of the resultant matrix are given by $a_{ij} + b_{ij}$. Thus, if

$$\mathsf{A} = \begin{pmatrix} -3 & 6 & 4 \\ 1 & 0 & 2 \end{pmatrix} \quad \text{and} \quad \mathsf{B} = \begin{pmatrix} 2 & 1 & 1 \\ -6 & 4 & 3 \end{pmatrix}$$

then

$$\mathsf{C} = \mathsf{A} + \mathsf{B} = \begin{pmatrix} -1 & 7 & 5 \\ -5 & 4 & 5 \end{pmatrix}$$

If we write

$$\mathsf{A} + \mathsf{A} = 2\mathsf{A} = \begin{pmatrix} -6 & 12 & 8 \\ 2 & 0 & 4 \end{pmatrix}$$

we see that scalar multiplication of a matrix means that each element is multiplied by the scalar. Thus,

$$c\,\mathsf{M} = \begin{pmatrix} cM_{11} & cM_{12} \\ cM_{21} & cM_{22} \end{pmatrix} \tag{3}$$

Example 2:
Using the matrices A and B above, form the matrix D = 3 A − 2 B.

SOLUTION:

$$\begin{aligned} \mathsf{D} &= 3\begin{pmatrix} -3 & 6 & 4 \\ 1 & 0 & 2 \end{pmatrix} - 2\begin{pmatrix} 2 & 1 & 1 \\ -6 & 4 & 3 \end{pmatrix} \\ &= \begin{pmatrix} -9 & 18 & 12 \\ 3 & 0 & 6 \end{pmatrix} - \begin{pmatrix} 4 & 2 & 2 \\ -12 & 8 & 6 \end{pmatrix} = \begin{pmatrix} -13 & 16 & 10 \\ 15 & -8 & 0 \end{pmatrix} \end{aligned}$$

One of the most important operations involving matrices is matrix multiplication. For simplicity, we will discuss the multiplication of square matrices first. Consider some linear transformation of $(x_1,\ y_1)$ into $(x_2,\ y_2)$:

$$\begin{aligned} x_2 &= a_{11}\,x_1 + a_{12}\,y_1 \\ y_2 &= a_{21}\,x_1 + a_{22}\,y_1 \end{aligned} \tag{4}$$

represented by the matrix equation

$$\mathsf{AV}_1 = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \mathsf{V}_2 \tag{5}$$

Now let's transform $(x_2,\ y_2)$ into $(x_3,\ y_3)$:

$$\begin{aligned} x_3 &= b_{11}\,x_2 + b_{12}\,y_2 \\ y_3 &= b_{21}\,x_2 + b_{22}\,y_2 \end{aligned} \tag{6}$$

represented by the matrix equation

$$\mathsf{BV}_2 = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_3 \\ y_3 \end{pmatrix} = \mathsf{V}_3 \tag{7}$$

Let the transformation of $(x_1,\ y_1)$ directly into $(x_3,\ y_3)$ be given by

$$\begin{aligned} x_3 &= c_{11}\,x_1 + c_{12}\,y_1 \\ y_3 &= c_{21}\,x_1 + c_{22}\,y_1 \end{aligned} \tag{8}$$

represented by the matrix equation

$$\mathsf{CV}_1 = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_3 \\ y_3 \end{pmatrix} = \mathsf{V}_3 \tag{9}$$

Symbolically, we can write that

$$\mathsf{V}_3 = \mathsf{CV}_1 = \mathsf{BV}_2 = \mathsf{BAV}_1$$

or that

$$\mathsf{C} = \mathsf{BA}$$

Let's find the relation between the elements of C and those of A and B. Substitute Equations 4 into 6 to obtain

$$\begin{aligned} x_3 &= b_{11}(a_{11}\,x_1 + a_{12}\,y_1) + b_{12}(a_{21}\,x_1 + a_{22}\,y_1) \\ y_3 &= b_{21}(a_{11}\,x_1 + a_{12}\,y_1) + b_{22}(a_{21}\,x_1 + a_{22}\,y_1) \end{aligned} \tag{10}$$

or

$$\begin{aligned} x_3 &= (b_{11}\,a_{11} + b_{12}\,a_{21})x_1 + (b_{11}\,a_{12} + b_{12}\,a_{22})y_1 \\ y_3 &= (b_{21}\,a_{11} + b_{22}\,a_{21})x_1 + (b_{21}\,a_{12} + b_{22}\,a_{22})y_1 \end{aligned}$$

Thus, we see that

$$\mathsf{C} = \mathsf{BA} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} b_{11}\,a_{11} + b_{12}\,a_{21} & b_{11}\,a_{12} + b_{12}\,a_{22} \\ b_{21}\,a_{11} + b_{22}\,a_{21} & b_{21}\,a_{12} + b_{22}\,a_{22} \end{pmatrix} \tag{11}$$

This result may look complicated, but it has a nice pattern which we will illustrate two ways. Mathematically, the ijth element of C is given by the formula

$$c_{ij} = \sum_k b_{ik}\, a_{kj} \tag{12}$$

Note that the right side of Equation 12 is the dot product of a row of B into a column of A. For example,

$$c_{11} = \sum_k b_{1k}\, a_{k1} = b_{11}\, a_{11} + b_{12}\, a_{21}$$

as in Equation 11. A more pictorial way is to notice that any element in C can be obtained by multiplying elements in any row in B by the corresponding elements in any column in A, adding them, and then placing them in C where the row and column intersect. For example, c_{11} is obtained by multiplying the elements of row 1 of B with the elements of column 1 of A, or by the scheme

$$\begin{matrix} \\ \rightarrow \end{matrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \overset{\downarrow\quad\;\;}{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}} = \begin{pmatrix} b_{11}\, a_{11} + b_{12}\, a_{21} & \cdot \\ \cdot & \cdot \end{pmatrix}$$

and c_{12} by

$$\begin{matrix} \\ \rightarrow \end{matrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \overset{\quad\;\;\downarrow}{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}} = \begin{pmatrix} \cdot & b_{11}\, a_{12} + b_{12}\, a_{22} \\ \cdot & \cdot \end{pmatrix}$$

Example 3:
Find C = BA if

$$\mathrm{B} = \begin{pmatrix} 1 & 2 & 1 \\ 3 & 0 & -1 \\ -1 & -1 & 2 \end{pmatrix} \quad \text{and} \quad \mathrm{A} = \begin{pmatrix} -3 & 0 & -1 \\ 1 & 4 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

SOLUTION:

$$\begin{aligned} \mathrm{C} &= \begin{pmatrix} 1 & 2 & 1 \\ 3 & 0 & -1 \\ -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} -3 & 0 & -1 \\ 1 & 4 & 0 \\ 1 & 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -3+2+1 & 0+8+1 & -1+0+1 \\ -9+0-1 & 0+0-1 & -3+0-1 \\ 3-1+2 & 0-4+2 & 1+0+2 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 9 & 0 \\ -10 & -1 & -4 \\ 4 & -2 & 3 \end{pmatrix} \end{aligned}$$

Example 4:
The matrix R given by Equation 2 represents a rotation through the angle θ. Show that R^2 represents a rotation through an angle 2θ.

SOLUTION:

$$R^2 = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$
$$= \begin{pmatrix} \cos^2\theta - \sin^2\theta & -2\sin\theta\cos\theta \\ 2\sin\theta\cos\theta & \cos^2\theta - \sin^2\theta \end{pmatrix}$$

Using standard trigonometric identities, we get

$$R^2 = \begin{pmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix}$$

which represents rotation through an angle 2θ.

Matrices do not have to be square to be multiplied together, but either Equation 11 or the pictorial method illustrated above suggests that the number of columns of B must be equal to the number of rows of A. When this is so, A and B are said to be *compatible*. We call a matrix having n rows and m columns an $n \times m$ matrix. Thus, an $n \times m$ matrix can multiply into only an $m \times p$ matrix and produces an $n \times p$ matrix.

For example, the product of a 2×3 matrix and a 3×3 matrix produces a 2×3 matrix:

$$\begin{pmatrix} 2 & -1 & 0 \\ 1 & 3 & -2 \end{pmatrix} \begin{pmatrix} 3 & 1 & 4 \\ -2 & 6 & -2 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 8 & -4 & 10 \\ -3 & 17 & -4 \end{pmatrix}$$

An important aspect of matrix multiplication is that BA does not usually equal AB. For example, if

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}$$

then

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

and

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

so $AB \neq BA$. If it does happen that $AB = BA$, then A and B are said to *commute*.

Example 5:
Do the matrices A and B commute if

$$\mathsf{A} = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathsf{B} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

SOLUTION:

$$\mathsf{AB} = \begin{pmatrix} 2 & 3 \\ 0 & 1 \end{pmatrix}$$

and

$$\mathsf{BA} = \begin{pmatrix} 2 & 2 \\ 0 & 1 \end{pmatrix}$$

so they do not commute.

Another property of matrix multiplication that differs from ordinary scalar multiplication is that the equation

$$\mathsf{AB} = 0$$

where 0 is the *zero matrix* or the *null matrix* (all elements equal to zero) does *not* imply that A or B necessarily is a zero matrix. For example,

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

Although matrices correspond to transformations and should not be confused with determinants, we can associate a determinant with a square matrix. In fact, if

$$\mathsf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

then

$$\det \mathsf{A} = |\mathsf{A}| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \tag{13}$$

If $\det \mathsf{A} \neq 0$, then A is said to be *nonsingular*. Conversely, if $\det \mathsf{A} = 0$, then A is said to be *singular*. A useful property of the determinants of matrices is that

$$\det \mathsf{AB} = (\det \mathsf{A})(\det \mathsf{B}) \tag{14}$$

(See Problem 21). Of course, A and B must both be square matrices and of the same dimension.

Example 6:
Given

$$\mathsf{A} = \begin{pmatrix} 2 & 0 & -1 \\ 1 & 1 & 3 \\ 2 & 0 & -1 \end{pmatrix} \quad \text{and} \quad \mathsf{B} = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix}$$

Show that $|\mathsf{AB}| = |\mathsf{A}|\ |\mathsf{B}|$.

SOLUTION: First calculate AB:

$$\mathsf{AB} = \begin{pmatrix} 4 & 0 & -1 \\ 1 & 1 & 3 \\ 2 & 0 & -1 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix} = \begin{pmatrix} -6 & 3 & 0 \\ 6 & 4 & 1 \\ -4 & 1 & 0 \end{pmatrix}$$

Then $|\mathsf{A}| = -2$, $|\mathsf{B}| = 3$, and $|\mathsf{AB}| = -6$.

The determinant of a matrix A is used in the construction of the inverse of A, which we define below.

A transformation that leaves $(x_1,\ y_1)$ unaltered is called the identity transformation, and the corresponding matrix is called the *identity matrix* or the *unit matrix*. All the elements of the identity matrix are equal to zero, except those along the main diagonal, which equal one:

$$\mathsf{I} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

A unit matrix is necessarily a square matrix. The elements of I are δ_{ij}, the Kronecker delta, which equals one when $i = j$ and zero when $i \neq j$. The unit matrix has the property that

$$\mathsf{I\,A} = \mathsf{A\,I} \tag{15}$$

The unit matrix is an example of a *diagonal matrix*. The only nonzero elements of a diagonal matrix are along its main diagonal. Generally, the elements on the main diagonal of a matrix are called *diagonal elements* and the others are called *off-diagonal elements*. Thus, we can say that all the off-diagonal elements of a diagonal matrix are zero. Diagonal matrices are necessarily square matrices. Also, any two $n \times n$ diagonal matrices commute with each other.

If $\mathsf{BA} = \mathsf{AB} = \mathsf{I}$, then B is said to be the *inverse* of A, and is denoted by A^{-1}. Thus, A^{-1} has the property that

$$\mathsf{AA}^{-1} = \mathsf{A}^{-1}\mathsf{A} = \mathsf{I} \tag{16}$$

If A represents some transformation, then A^{-1} undoes that transformation and restores the original state. The fact that $\mathsf{AA}^{-1} = \mathsf{A}^{-1}\mathsf{A}$ implies that A must be a square matrix. It should be clear on physical grounds that the inverse of R in Equation 2 is a rotation through the angle $-\theta$. Thus, we write

$$\mathsf{R}^{-1}(\theta) = \mathsf{R}(-\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{17}$$

which is obtained from R by replacing θ by $-\theta$. It's easy to show that $\mathsf{R}(\theta)\mathsf{R}^{-1}(\theta) = \mathsf{R}^{-1}(\theta)\mathsf{R}(\theta) = \mathsf{I}$.

We found the inverse of $\mathsf{R}(\theta)$ in Equation 2 by a physical argument, but how do we find the inverse of a (square) matrix in general? It turns out that we essentially derived the formula for the inverse of A in Section 1. Equation 19 of that section is

$$\sum_{k=1}^{n} \left(\frac{A_{ki}}{|\mathsf{A}|}\right) a_{kj} = \delta_{ij} \tag{18}$$

The quantities A_{ki} are the cofactors of the a_{ki} of A, and the right side of Equation 18 are the elements of a unit matrix.

Let's first define a *matrix of cofactors* of A by

$$\mathsf{A}_{\text{cof}} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix} \tag{19}$$

Now we define the *transpose* A^{T} of a matrix A to be the matrix that is obtained by interchanging the rows and columns of A. In terms of the matrix elements of a general matrix (a_{ij}), we have $a_{ij}^{\mathsf{T}} = a_{ji}$. For example, if

$$\mathsf{A} = \begin{pmatrix} 3 & 0 & 1 \\ -1 & 1 & 2 \\ -2 & -1 & -3 \end{pmatrix}, \quad \text{then} \quad \mathsf{A}^{\mathsf{T}} = \begin{pmatrix} 3 & -1 & -2 \\ 0 & 1 & -1 \\ 1 & 2 & -3 \end{pmatrix}$$

Notice that we can also form A^{T} from A by simply flipping A about its main diagonal.

We can now write the term in parentheses in Equation 18 as the ikth element of $\mathsf{A}_{\text{cof}}^{\mathsf{T}}/|\mathsf{A}|$, so that Equation 18 becomes, in matrix notation,

$$\frac{\mathsf{A}_{\text{cof}}^{\mathsf{T}}}{|\mathsf{A}|}\mathsf{A} = \mathsf{I} \tag{20}$$

where I is a unit matrix. Thus, we see that if $\det \mathsf{A} \neq 0$, then the inverse of A is given by

$$\mathsf{A}^{-1} = \frac{\mathsf{A}_{\text{cof}}^{\mathsf{T}}}{|\mathsf{A}|} \tag{21}$$

Some authors call $\mathsf{A}_{\text{cof}}^{\mathsf{T}}$ the *adjoint of* A, written as adj (A). One clear implication of Equation 21 is that A must be nonsingular. Singular matrices do not have inverses.

Equation 21 may look awkward to use, but it's pretty straightforward. Let's find the inverse of

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

The determinant of A is equal to −4 and the matrix of cofactors is

$$A_{cof} = \begin{pmatrix} -2 & 2 & -2 \\ 0 & 0 & -4 \\ -1 & -1 & 1 \end{pmatrix}$$

Using Equation 21, we have

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

It's readily verified that $A^{-1}A = AA^{-1} = I$.

Example 7:
Find the inverse of

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

SOLUTION: det A = 16 and

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

and so

$$A^{-1} = \frac{1}{16}\begin{pmatrix} 7 & -2 & -1 \\ 1 & 2 & -7 \\ 2 & 4 & 2 \end{pmatrix}$$

An example of a matrix that has no inverse, which occurs in a number of physical applications, is a matrix that corresponds to a projection of a vector onto a coordinate axis. For example, the matrix

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

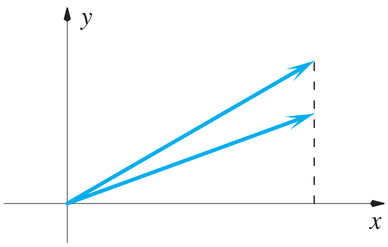

Figure 9.5
An illustration of why a matrix corresponding to a projection does not have an inverse. Both vectors have the same projection onto the x axis.

projects the vector $\mathbf{r} = x\,\mathbf{i} + y\,\mathbf{j}$ onto the x axis; that is,

$$\mathsf{P}\,\mathbf{r} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 0 \end{pmatrix}$$

Mathematically, P has no inverse because $|\mathsf{P}| = 0$. Physically, P has no inverse because any vector $\mathbf{r}$ with an x component x_0 will yield the same result, as you can see in Figure 9.5.

Finally, we mention that any CAS can readily find inverses of matrices (Problems 23 and 24).

9.3 Problems

1. Given the two matrices $\mathsf{A} = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix}$ and $\mathsf{B} = \begin{pmatrix} -1 & 1 & 0 \\ 3 & 0 & 2 \\ 1 & 1 & 1 \end{pmatrix}$, form the matrices $\mathsf{C} = 2\mathsf{A} - 3\mathsf{B}$ and $\mathsf{D} = 6\mathsf{B} - \mathsf{A}$.

2. Given the three matrices $\mathsf{A} = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\mathsf{B} = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, and $\mathsf{C} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, show that $\mathsf{A}^2 + \mathsf{B}^2 + \mathsf{C}^2 = \frac{3}{4}\mathsf{I}$, where I is a unit matrix. Also show that

$$\mathsf{AB} - \mathsf{BA} = i\mathsf{C}$$
$$\mathsf{BC} - \mathsf{CB} = i\mathsf{A}$$
$$\mathsf{CA} - \mathsf{AC} = i\mathsf{B}$$

3. Given the matrices $\mathsf{A} = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$, $\mathsf{B} = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}$, and $\mathsf{C} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$, show that

$$\mathsf{AB} - \mathsf{BA} = i\mathsf{C}$$
$$\mathsf{BC} - \mathsf{CB} = i\mathsf{A}$$
$$\mathsf{CA} - \mathsf{AC} = i\mathsf{B}$$
$$\mathsf{A}^2 + \mathsf{B}^2 + \mathsf{C}^2 = 2\mathsf{I}$$

where I is a unit matrix.

4. Does $(\mathsf{A} + \mathsf{B})^2$ always equal $\mathsf{A}^2 + 2\mathsf{AB} + \mathsf{B}^2$? Does $(\mathsf{AB})^2 = \mathsf{A}^2\mathsf{B}^2$?

5. A three-dimensional rotation of a vector about the z axis can be represented by the matrix

$\mathsf{R} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Show that $\det \mathsf{R} = |\mathsf{R}| = 1$ and that $\mathsf{R}^{-1} = \mathsf{R}(-\theta) = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

6. Show that (a) $(\mathsf{A}^{\mathsf{T}})^{\mathsf{T}} = \mathsf{A}$; (b) $(\mathsf{A} + \mathsf{B})^{\mathsf{T}} = \mathsf{A}^{\mathsf{T}} + \mathsf{B}^{\mathsf{T}}$; (c) $(\alpha\mathsf{A})^{\mathsf{T}} = \alpha\mathsf{A}^{\mathsf{T}}$; (d) $(\mathsf{AB})^{\mathsf{T}} = \mathsf{B}^{\mathsf{T}}\mathsf{A}^{\mathsf{T}}$.

7. Given the matrices $C_3 = \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}$, $\sigma_v = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, $\sigma_v' = \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}$, and $\sigma_v'' = \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}$, show that $\sigma_v C_3 = \sigma_v''$, $C_3 \sigma_v = \sigma_v'$, $\sigma_v'' \sigma_v' = C_3$, and $C_3 \sigma_v'' = \sigma_v$. Calculate the determinant associated with each matrix.

8. If $A^T = A^{-1}$, then A is said to be *orthogonal*. Which of the matrices in Problem 7 are orthogonal?

9. Find the matrix of cofactors, A_{cof}, and the adjoint of A, adj (A), for

(a) $\begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix}$ **(b)** $\begin{pmatrix} 3 & 1 & 2 \\ 1 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}$

10. Verify that $(AB)^T = B^T A^T$ if $A = \begin{pmatrix} 1 & 2 \\ 0 & -2 \end{pmatrix}$ and $B = \begin{pmatrix} 3 & 1 \\ -1 & 2 \end{pmatrix}$.

11. Prove that A^{-1} is unique.

12. Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}$. (a) Find a nonzero matrix B such that $AB = 0$. Does $BA = 0$? (b) Can you find B such that $AB = 0$ if $A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$?

13. Prove that (a) $(A^{-1})^{-1} = A$; (b) $(A^T)^{-1} = (A^{-1})^T$.

14. Prove that $\det(A^{-1}) = (\det A)^{-1}$. *Hint*: Use the relation $\det AB = (\det A)(\det B)$.

15. Prove that $(AB)^{-1} = B^{-1}A^{-1}$.

16. Use $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and $B = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix}$ to verify the relations in Problems 13 through 15.

17. Use $A = \begin{pmatrix} -1 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 4 & 1 & -1 \\ -1 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix}$ to verify the relations in Problems 13 through 15.

18. Find the inverse of (a) $\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$, (b) $\begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & 2 \\ 2 & 2 & 0 \end{pmatrix}$, and (c) $\begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}$.

19. Solve the equations

$$x + y = 3$$
$$4x - 3y = 5$$

by writing them as $AX = H$ and then $X = A^{-1}H$.

20. Show that two $n \times n$ diagonal matrices commute. Does an $n \times n$ diagonal matrix necessarily commute with any $n \times n$ matrix?

21. The general proof that $\det(AB) = (\det A)(\det B)$ is fairly long and so we shall not prove it here. Nevertheless, verify that it is true for 2×2 matrices.

22. A matrix that satisfies the relation $A^2 = A$ is called *idempotent*. Show that if A has an inverse, then it must be the identity matrix. Argue that a projection matrix must be idempotent. Does a projection matrix have an inverse?

23. Use any CAS to find the inverse of $\begin{pmatrix} 2 & 5 & 1 \\ 3 & 1 & 2 \\ -2 & 1 & 0 \end{pmatrix}$.

24. Use any CAS to find the inverse of $\begin{pmatrix} 1 & 0 & 3 & -2 \\ 6 & 1 & -1 & 3 \\ 2 & 0 & 1 & 1 \\ 4 & 3 & 2 & 5 \end{pmatrix}$.

9.4 Rank of a Matrix

In Section 2, we used Gaussian elimination to solve sets of linear algebraic equations and saw that we could have a unique solution, an infinity of solutions, or no solutions. Gaussian elimination leads directly to the correct result in each case, but it would be nice to have a general theory that tells us beforehand what to expect about the solutions. Surely, the nature of the solutions depends upon some property of the coefficient matrix and/or the augmented matrix since they describe the system of equations completely. This property is the *rank* of a matrix, which is the subject of this section.

There are several equivalent definitions of rank. One definition of rank is expressed in terms of square submatrices of A. A *square submatrix* of A is any square matrix obtained from A by deleting a certain number of rows and columns. If A happens to be square, then A is a submatrix of itself, obtained by deleting no rows and no columns. For example, consider the matrix

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

The possible square submatrices of A are the 2×2 matrices

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

and the 1×1 matrices (1), (2), (3), and (4).

The rank, $r(\mathsf{A})$, of A is the order of the largest square submatrix of A whose determinant is not equal to zero. The rank of the matrix A, above, is 2. The rank of

$$\mathsf{A} = \begin{pmatrix} 1 & 2 & 0 \\ 3 & 0 & 0 \end{pmatrix}$$

is also 2 because the submatrix $\begin{pmatrix} 1 & 2 \\ 3 & 0 \end{pmatrix}$ is nonsingular, even though the other two 2×2 submatrices are singular.

Example 1:
Determine the rank of

$$A = \begin{pmatrix} 2 & 1 & 0 & 5 \\ 3 & 6 & 1 & 1 \\ 5 & 7 & 1 & 8 \end{pmatrix}$$

SOLUTION: The largest that $r(A)$ can be is 3 since the largest possible square submatrix of A is 3×3, and there are four 3×3 square submatrices. The determinant of the 3×3 submatrix obtained by striking out the fourth column is zero, but the determinants of the other three 3×3 submatrices are not equal to zero. Therefore, $r(A) = 3$.

Another, perhaps more convenient but nevertheless equivalent, definition of rank is the number of nonzero rows in the matrix after it has been transformed into echelon form by elementary row operations. Since we are going to base the defintion of rank on the echelon form of a matrix, we should give a formal definition of what we mean by echelon form. We say that a matrix is in echelon if

1. All rows consisting of all zeros appear at the bottom.
2. If the first nonzero element of a row appears in column c, then all the elements in column c in lower rows are zero.
3. The first nonzero element of any nonzero row appears to the right of the first nonzero element in any higher row.

All the final versions of the augmented matrices in Section 2 were in echelon form.

Let's determine the rank of the matrix in Example 1 by this method. In obvious notation,

$$\begin{pmatrix} 2 & 1 & 0 & 5 \\ 3 & 6 & 1 & 1 \\ 5 & 7 & 1 & 8 \end{pmatrix} \sim \begin{pmatrix} 2 & 1 & 0 & 5 \\ 0 & 9/2 & 1 & -13/2 \\ 0 & 9/2 & 1 & -9/2 \end{pmatrix} \sim \begin{pmatrix} 2 & 1 & 0 & 5 \\ 0 & 9/2 & 1 & -13/2 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

There are three nonzero rows, so $r(A) = 3$.

Example 2:
Determine the rank of

$$A = \begin{pmatrix} 3 & 2 & 1 & -4 & 1 \\ 2 & 3 & 0 & -1 & -1 \\ 1 & -6 & 3 & -8 & 7 \end{pmatrix}$$

SOLUTION: Rearrange the rows so that the left-most column reads 1,2,3 (this avoids introducing fractions). Now add -2 times row 1 to row 2 and

-3 times row 1 to row 3 to obtain

$$\begin{pmatrix} 1 & -6 & 3 & -8 & 7 \\ 0 & 15 & -6 & 15 & -15 \\ 0 & 20 & -8 & 20 & -20 \end{pmatrix}$$

The last two rows are a constant multiple of each other, so if we multiply row 2 by $-4/3$ and add the result to row 3, we get

$$\begin{pmatrix} 1 & -6 & 3 & -8 & 7 \\ 0 & 15 & -6 & 15 & -15 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

There are two nonzero rows, so the rank of A is 2.

There are ten 3×3 submatrices of the matrix in Example 2 (can you show this?), so you would have to evaluate ten 3×3 determinants just to find out the rank of A is not equal to 3. This result suggests that the row echelon method is usually much easier to apply than our first definition of the rank of a matrix. Nevertheless, the definition of rank in terms of the largest nonsingular square submatrix of A is a standard definition.

There is another definition of rank that we will present here for completeness. We say that the m nonzero vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are *linearly dependent* if there exist constants $c_1, c_2, \ldots, c_m$ not all zero such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m = 0 \tag{1}$$

Linear dependence means that one of the vectors can be written as a linear combination of the others. If Equation 1 is satisfied only if all the $c_j = 0$, then the vectors are said to be *linearly independent*. In three dimensions, the unit vectors **i**, **j**, and **k** are linearly independent, but any other vector in three dimensions can be written as a linear combination of **i**, **j**, and **k** ($\mathbf{v} = v_x\,\mathbf{i} + v_y\,\mathbf{j} + v_z\,\mathbf{k}$).

We now define the rank of a matrix in terms of linear independence of vectors. Recall from Chapter 5 that a vector can be represented by an ordered n-tuple of numbers, $(v_1, v_2, \ldots, v_n)$, where we can think of the v_j as the components of $\mathbf{v}$ in some coordinate system. We think of the rows of the matrix A as vectors. If A is $n \times m$, then we have n m-dimensional vectors constituting the rows of A. The rank of A is the maximum number of linearly independent vectors that can be formed from these row vectors. In practice, this fundamental definition of rank isn't that useful because it often isn't easy to use Equation 1 to determine if a set of vectors is linearly independent or not. In fact, the easiest way to determine if a set of vectors is linearly independent or linearly dependent is to use the echelon matrix procedure to determine the rank of A, and hence the number of linearly independent vectors. Nevertheless, this definition is useful in theoretical discussions. Of course, the three definitions of rank that we have presented are equivalent.

To see more clearly the relation between rank and the number of linearly independent rows of a matrix, consider the following matrix in echelon form:

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

Row 1 cannot possibly be a linear combination of rows 2 and 3 because they have zeros in their first entries and so any linear combination of rows 2 and 3 must have a zero in its first position. Similarly, row 2 cannot be a multiple of row 3 because it has a zero in its first and second entries. Working from Row 3 upwards now, notice that no row can be a linear combination of higher rows because of the positions of the leading zeros in each row. Thus, there are three linearly independent vectors in this rank 3 matrix.

Example 3:
Determine whether the three vectors (3, 2, 1, −4, 1), (2, 3, 0, −1, −1), and (1, −6, 3, −8, 7) are linearly independent.

SOLUTION:

$$\mathsf{A} = \begin{pmatrix} 1 & -6 & 3 & -8 & 7 \\ 2 & 3 & 0 & -1 & -1 \\ 3 & 2 & 1 & -4 & 1 \end{pmatrix}$$

(We have arranged the rows in this way to avoid fractions.) This is the same matrix as in Example 2, where we determined that the rank of A is 2. Therefore, only two of the three vectors are linearly independent.

We now present a theorem on the existence of solutions to a set of m linear algebraic equations in n unknowns in terms of rank.

Let A *be the* $m \times n$ *coefficient matrix of the set of* m *linear algebraic equations* $\mathsf{AX} = \mathsf{H}$ *and let* $\mathsf{A}|\mathsf{H}$ *be the* $m \times (n+1)$ *augmented matrix of the system. If*

1. $r(\mathsf{A}) = r(\mathsf{A}|\mathsf{H}) = n$*, there is a unique solution.*
2. $r(\mathsf{A}) = r(\mathsf{A}|\mathsf{H}) < n$*, there are infinitely many solutions, expressible in terms of* $n - r(A)$ *parameters.*
3. $r(\mathsf{A}) < r(\mathsf{A}|\mathsf{H})$*, there are no solutions.*

This theorem summarizes all the possible cases for all linear systems, homogeneous or nonhomogeneous. Let's go back and examine each of the cases in

Section 2 in terms of the ranks of A and A|H. For Equations 5,

$$\mathsf{A} = \begin{pmatrix} 2 & 1 & 3 \\ 2 & -2 & -1 \\ -2 & 4 & 1 \end{pmatrix} \qquad \text{and} \qquad \mathsf{A|H} = \left(\begin{array}{ccc|c} 2 & 1 & 3 & 4 \\ 2 & -2 & -1 & 1 \\ -2 & 4 & 1 & 1 \end{array}\right)$$

Both $r(\mathsf{A})$ and $r(\mathsf{A|H}) = 3$, so Equations 5 have a unique solution.

For Example 1,

$$\mathsf{A} = \begin{pmatrix} 1 & 1 & -1 \\ 2 & -1 & 3 \\ 3 & 2 & -2 \end{pmatrix} \qquad \text{and} \qquad \mathsf{A|H} = \left(\begin{array}{ccc|c} 1 & 1 & -1 & 2 \\ 2 & -1 & 3 & 5 \\ 3 & 2 & -2 & 5 \end{array}\right)$$

In this case, $r(\mathsf{A}) = r(\mathsf{A|H}) = 3$, and so the solution is unique.

For Example 2,

$$\mathsf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ -1 & 1 & -1 \end{pmatrix} \qquad \text{and} \qquad \mathsf{A|H} = \left(\begin{array}{ccc|c} 1 & 1 & 1 & -2 \\ 1 & -1 & 1 & 2 \\ -1 & 1 & -1 & -2 \end{array}\right)$$

In this case, $r(\mathsf{A}) = r(\mathsf{A|H}) = 2 < 3$, and so there are infinitely many solutions, expressible in terms of one parameter.

For Equation 7,

$$\mathsf{A} = \begin{pmatrix} 1 & 1 \\ 3 & -4 \\ 5 & -2 \end{pmatrix} \qquad \text{and} \qquad \mathsf{A|H} = \left(\begin{array}{cc|c} 1 & 1 & 4 \\ 3 & -4 & 9 \\ 5 & -2 & 17 \end{array}\right)$$

In this case, $r(\mathsf{A}) = r(\mathsf{A|H}) = 2$, and so the solution is unique. The rest of the cases are left to the Problems.

Before leaving this section, we shall present a theorem regarding homogeneous sets of linear algebraic equations. Even though the above theorem applies to both homogeneous and nonhomogeneous systems, homogeneous systems occur quite often in physical problems, so we'll present the implications of the above general theorem to homogeneous systems.

> *The $m \times n$ homogeneous system* $\mathsf{AX} = 0$ *always has a trivial solution,* $x_1 = x_2 = \cdots = x_n = 0$. *[It is always consistent because* $r(\mathsf{A}) = r(\mathsf{A|H})$.*]*

If $r(\mathsf{A}) = n$, then the trivial solution is the only solution. If $r(\mathsf{A}) < n$, then the general theorem assures the existence of non-trivial solutions. In particular, these non-trivial solutions constitute an $[n - r(A)]$-parameter family of solutions.

An $n \times n$ homogeneous system has a property that we shall emphasize here by setting it off:

> *The $n \times n$ homogeneous system of linear algebraic equations* $\mathsf{AX} = 0$ *has a non-trivial solution if and only if* $\det \mathsf{A} = 0$.

This last theorem follows directly from everything above, but it is important enough to emphasize.

Example 4:
Determine the values of x such that the equations

$$\begin{aligned} xc_1 + c_2 \qquad\quad &= 0 \\ c_1 + xc_2 + c_3 \qquad &= 0 \\ c_2 + xc_3 + c_4 &= 0 \\ c_3 + xc_4 &= 0 \end{aligned}$$

have non-trivial solutions for the c_j. (This set of equations occurs in a quantum-mechanical calculation for a butadiene molecule.)

SOLUTION: To assure a non-trivial solution, the determinant of the coefficient matrix must vanish.

$$\begin{vmatrix} x & 1 & 0 & 0 \\ 1 & x & 1 & 0 \\ 0 & 1 & x & 1 \\ 0 & 0 & 1 & x \end{vmatrix} = 0$$

Expanding in cofactors about the first row gives

$$x(x^3 - 2x) - (x^2 - 1) = 0$$

or $x^4 - 3x^2 + 1 = 0$, or $x^2 = (3 \pm \sqrt{5})/2$, or $x = \pm 1.61804$ and ± 0.61804.

9.4 Problems

Use the concept of rank to investigate the nature of the solutions in Problems 1 through 12.

1. Example 4 of Section 2.
2. Equations 8 of Section 2.
3. Example 5 of Section 2.
4. Problem 1 of Section 2.
5. Problem 2 of Section 2.
6. Problem 3 of Section 2.
7. Problem 4 of Section 2.
8. Problem 5 of Section 2.
9. Problem 6 of Section 2.
10. Problem 7 of Section 2.
11. Problem 8 of Section 2.
12. Problem 9 of Section 2.

13. For what values of x will the following equations have non-trivial solutions?

$$\begin{aligned} xc_1 + c_2 + c_4 &= 0 \\ c_1 + xc_2 + c_3 &= 0 \\ c_2 + xc_3 + c_4 &= 0 \\ c_1 + c_3 + xc_4 &= 0 \end{aligned}$$

14. Determine the values of x for which the following equations will have a non-trivial solution:

$$\begin{aligned} xc_1 + c_2 + c_3 + c_4 &= 0 \\ c_1 + xc_2 + c_4 &= 0 \\ c_1 + xc_3 + c_4 &= 0 \\ c_1 + c_2 + c_3 + xc_4 &= 0 \end{aligned}$$

9.5 Vector Spaces

Although we didn't point it out explicitly in the previous section, matrices obey the following algebraic rules:

1. $\mathsf{A} + \mathsf{B} = \mathsf{B} + \mathsf{A}$
2. $\mathsf{A} + (\mathsf{B} + \mathsf{C}) = (\mathsf{A} + \mathsf{B}) + \mathsf{C}$
3. $a(\mathsf{A} + \mathsf{B}) = a\mathsf{A} + a\mathsf{B}$
4. $(a + b)\mathsf{A} = a\mathsf{A} + b\mathsf{A}$
5. $a(b\mathsf{A}) = (ab)\mathsf{A}$

where a and b are scalars. It so happens that many other mathematical quantities obey the same set of rules. For example, complex numbers, vectors, and functions obey these rules. There is a mathematical formalism that treats all these quantities in an abstract unified manner and allows us to see the similarities between them.

We define a *vector space* V to be a set of objects (which we'll call *vectors*) for which addition and multiplication by a scalar, either real or complex, are defined and satisfy the following requirements:

1. Two vectors, $\mathbf{x}$ and $\mathbf{y}$, may be added to give another vector, $\mathbf{x} + \mathbf{y}$, which is also in V. (We say that the set is *closed* under addition.)
2. Addition is commutative; in other words, $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
3. Addition is associative; in other words, $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z}) = \mathbf{x} + \mathbf{y} + \mathbf{z}$.
4. There exists in V a unique *zero vector*, $\mathbf{0}$, such that $\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x}$ for any $\mathbf{x}$ in V.
5. For every $\mathbf{x}$ in V, there is an additive inverse $-\mathbf{x}$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.
6. Any vector $\mathbf{x}$ may be multiplied by a scalar c such that $c\mathbf{x}$ is in V. In other words, the set is closed under scalar multiplication.

7. Scalar multiplication is associative; in other words, for any two numbers a and b, $a(b\mathbf{x}) = (ab)\mathbf{x}$.
8. Scalar multiplication is distributive over addition; in other words,

$$a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y} \text{ and } (a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}.$$

9. For the unit scalar, $1\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in V.

Properties 1 through 9 are called the *axioms of a vector space*. If the scalars are real numbers, V is called a *real vector space*; if they are complex, V is called a *complex vector space*.

The geometric vectors that we discussed in Chapter 5 satisfy all the above properties of a vector space, and form what is called a Euclidean vector space, in particular. The elements or members of a vector space, however, need not be geometric vectors. For example, the set of all nth order polynomials, P_n, with real or complex coefficients, forms a vector space, as long as we consider mth order polynomials ($m < n$) to be nth order polynomials with certain zero coefficients. Another vector space consists of all n-tuples of real numbers, where the sum of $(u_1, u_2, \ldots, u_n)$ and $(v_1, v_2, \ldots, v_n)$ is defined as $(u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n)$ and the product of an n-tuple by a scalar is defined as $a(u_1, u_2, \ldots, u_n) = (au_1, au_2, \ldots, au_n)$ (Problem 1). (We'll use ordered n-tuples fairly often to illustrate the properties of vector spaces, so we'll designate the space of all ordered n-tuples of real numbers by R^n and that of complex numbers by C^n.)

Example 1:

Show that the set of functions whose first derivatives are continuous in $[a, b]$ and satisfy

$$\frac{df}{dx} + 2f(x) = 0$$

form a vector space.

SOLUTION: To show that the set is closed under addition (1), let f and g be two elements of V (in other words, both f and g satisfy the above equation). Then,

$$\begin{aligned}\frac{d}{dx}(f + g) + 2(f + g) &= \frac{df}{dx} + \frac{dg}{dx} + 2f + 2g \\ &= \left(\frac{df}{dx} + 2f\right) + \left(\frac{dg}{dx} + 2g\right) = 0 + 0 = 0\end{aligned}$$

To show that the set is closed under scalar multiplication (6), consider

$$\frac{d}{dx}(af) + 2(af) = a\left(\frac{df}{dx} + 2f\right) = 0$$

The other axioms are satisfied by any continuous function.

Example 1 suggests that the set of solutions to any linear homogeneous differential equation forms a vector space (Chapter 11).

It often happens that a subset of the vectors in V forms a vector space with respect to the same addition and multiplication operations as V. In such a case, the set of vectors is said to form a *subspace* of V. A simple geometric example of a subspace is the xy-plane of a three-dimensional Euclidean space. The set of all vectors that lie in the xy-plane forms a vector space. To see another example of a subspace, consider the vector space R^n made up of ordered n-tuples of real numbers $(u_1, u_2, \ldots, u_n)$. The set of n-tuples $(a, a, \ldots, a)$ forms a subspace of R^n (Problem 5).

An important concept associated with vector spaces is the linear independence and linear dependence of vectors. We touched upon this idea in the previous section, but we shall study it more thoroughly here. Let $\{\mathbf{v}_j;\ \ j = 1, 2, \ldots, n\}$ be a set of nonzero vectors from a vector space V. We say that the set of vectors is *linearly independent* if the only way that

$$\sum_{j=1}^{n} c_j \mathbf{v}_j = 0$$

is for each and every $c_j = 0$. If the vectors are not independent, then they are *linearly dependent*. There are several convenient ways to determine if a set of vectors is independent or not. Let's test the three vectors (1, 0, 0), (1, −1, 1), and (1, 2, −1) for linear independence. Is there a set of c_j, not all zero, such that

$$\sum_{j=1}^{n} c_j \mathbf{v}_j = (c_1, 0, 0) + (c_2, -c_2, c_2) + (c_3, 2c_3, -c_3) = (0, 0, 0)$$

or is there a nontrivial solution to the equations

$$\begin{aligned} c_1 + c_2 + c_3 &= 0 \\ -c_2 + 2c_3 &= 0 \\ c_2 - c_3 &= 0 \end{aligned} \tag{1}$$

The determinant of the matrix of coefficients is nonzero, so the only solution is the solution $c_1 = c_2 = c_3 = 0$, and so the three vectors are linearly independent. We could also have arranged the three vectors as the rows of a matrix and then transformed it into echelon form:

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

which shows that the three vectors are linearly independent.

It's easy to see from the echelon form approach that a set of m n-tuples *must* be linearly dependent if $m > n$ because the bottom $m - n$ rows will always be zeros. For example, consider the three vectors (1, 1), (1, −1), and (−1, 1). Then,

$$\mathsf{A} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 \\ 0 & -2 \\ 0 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 \\ 0 & -2 \\ 0 & 0 \end{pmatrix}$$

and so only two of the vectors are linearly independent. We can also see that the rank of A can equal 2 at the most.

Example 2:
How many linearly independent vectors are there in the set {(1, 1, 0, 1), (−1, −1, 0, −1), (1, 0, 1, 1), (−1, 0, −1, −1)}?

SOLUTION:

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

where we placed the two zero rows at the bottom. Thus, there are two linearly independent vectors in the set.

If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are vectors in a vector space V and if any other vector $\mathbf{u}$ in V can be expressed as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ so that

$$\mathbf{u} = c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_n\,\mathbf{v}_n$$

where the c_j are constants, then we say that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ *span* V. For example, the three unit vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ span the three-dimensional space R^3, as does any three non-coplanar (linearly independent) vectors. If the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ in a vector space V are linearly independent and span V, then the set $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is called a *basis* or *basis set* for V. The unit vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$, or any three non-coplanar vectors, form a basis in R^3. The number of vectors in a basis is defined to be the *dimension* of the vector space. The dimension of a vector space V is equal to the maximum number of linearly independent vectors in V.

Suppose that $\{\mathbf{v}_j;\ j = 1, 2, \ldots, n\}$ is a set of linearly independent vectors in an n-dimensional vector space V. If the set composed of the $\mathbf{v}_j$ and any other (non-zero) vector $\mathbf{u}$ in V is linearly dependent, then the set $\{\mathbf{v}_j;\ j = 1, 2, \ldots, n\}$ is said to be *maximal*. Thus, the maximum number of linearly independent vectors in V is n. Because $\mathbf{u}$ and the $\mathbf{v}_j$ are linearly dependent, we can write

$$c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_n\,\mathbf{v}_n + c_u\,\mathbf{u} = 0$$

where the c's are not all zero. If $c_u = 0$, then the set $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is linearly dependent, contrary to our assertion. Therefore, $c_u \neq 0$ and we can write

$$\mathbf{u} = -\frac{c_1}{c_u}\mathbf{v}_1 - \frac{c_2}{c_u}\mathbf{v}_2 - \cdots - \frac{c_n}{c_u}\mathbf{v}_n$$

We see, then, that **u** can be written as a linear combination of the set of linearly independent vectors, and so the $\mathbf{v}_j$ constitute a basis.

Example 3:
Show that the vectors (1, 0, 0), (0, 1, 0), and (0, 0, 1) form a basis in R^3. What familiar vectors do they correspond to?

SOLUTION: They form a basis in R^3 because any vector $\mathbf{u} = (x, y, z)$ in R^3 can be written as

$$\mathbf{u} = x(1, 0, 0) + y(0, 1, 0) + z(0, 0, 1)$$

These vectors correspond to the unit vectors **i**, **j**, and **k**.

Example 4:
Show that the vector space V of ordered n-tuples is an n-dimensional space.

SOLUTION: The n vectors $(1, 0, \ldots, 0), (0, 1, \ldots, 0), \ldots, (0, 0, \ldots, 1)$ constitute a linearly independent set of vectors in V. Furthermore, they span V because any other vector in V can be written as a linear combination of these vectors according to

$$\mathbf{u} = (u_1, u_2, \ldots, u_n) = u_1(1, 0, \ldots, 0) + u_2(0, 1, \ldots, 0) + \cdots + u_n(0, 0, \ldots, 1)$$

Thus, the n vectors constitute a basis and the dimension of V is n.

Suppose that $\{\mathbf{v}_j;\ j = 1, 2, \ldots, n\}$ is a basis of V. Then any vector **u** of V can be written as a linear combination of the $\mathbf{v}_j$:

$$\mathbf{u} = \sum_{j=1}^{n} u_j \mathbf{v}_j$$

We say that u_j is the jth coordinate of **u** with respect to the given basis set. Problem 19 asks you to show that the coordinates of **u** with respect to a given basis in a given basis are unique.

When we study differential equations in Chapter 11, we'll see that the set of solutions to an nth order linear homogeneous differential equation forms a vector space. Consequently, it's not unusual to inquire about the linear independence of a set of functions over some interval I. In other words, we ask if the only way that

$$\sum_{j=1}^{n} c_j f_j(x) = 0 \tag{2}$$

is for $c_j = 0$ for $j = 1, 2, \ldots, n$. If that's the case, we say that the n functions $f_j(x)$, $j = 1, 2, \ldots, n$ are linearly independent. Otherwise, they are linearly dependent.

There is a convenient way to test for the linear independence of a set of functions. Start with Equation 2:

$$c_1 f_1(x) + c_2 f_2(x) + \cdots + c_n f_n(x) = 0 \tag{3}$$

If the $f_j(x)$ are differentiable up to $(n-1)$th order over the interval I, differentiate Equation 3 $n-1$ times to obtain

$$\begin{matrix} c_1 f_1'(x) & + c_2 f_2'(x) + & \cdots & + c_n f_n'(x) = 0 \\ \vdots & \vdots & \ddots & \vdots \\ c_1 f_1^{(n-1)}(x) & + c_2 f_2^{(n-1)}(x) + & \cdots & + c_n f_n^{(n-1)}(x) = 0 \end{matrix} \tag{4}$$

The coefficient matrix of these equations is

$$W(x) = \begin{vmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{vmatrix} \tag{5}$$

The determinant $W(x)$ in Equation 5 is called the *Wronskian* of the functions $f_j(x)$. If $W(x) \neq 0$ at any point in the interval I, then the $f_j(x)$ are linearly independent.

Example 5:
Test the functions $\sin x$ and $\cos x$ for linear independence over the interval $-\infty < x < \infty$.

SOLUTION: The Wronskian is

$$W(x) = \begin{vmatrix} \sin x & \cos x \\ \cos x & -\sin x \end{vmatrix} = -1$$

for all x, and so $\sin x$ and $\cos x$ are linearly independent for all values of x.

Unfortunately, the converse of the above result is not true. If $W(x) = 0$, the $f_j(x)$ may or may not be linearly independent, as the next Example shows. (See Problem 18 also.)

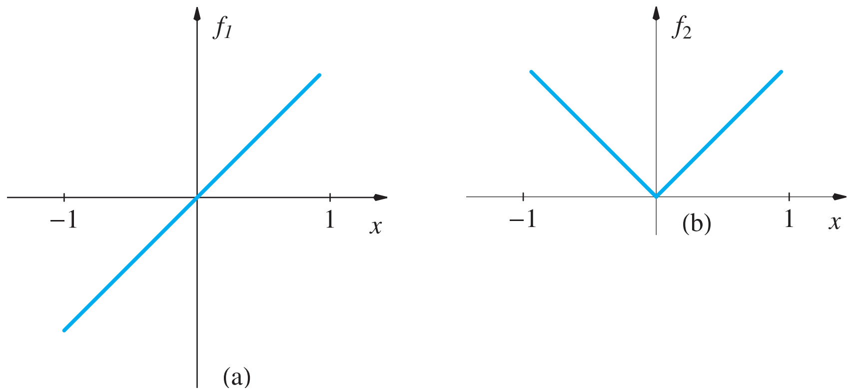

Figure 9.6
The functions $f_1(x) = x$ and $f_2(x) = |x|$ in Example 6 are linearly independent over the interval $-1 \leq x \leq 1$.

Example 6:
Show that $f_1(x) = x$ and $f_2(x) = |x|$ are linearly independent over the interval $-1 \leq x \leq 1$, but linearly dependent over the interval $0 \leq x \leq 1$.

SOLUTION: For the interval $[-1, 1]$,

$$W(x) = \begin{cases} \begin{vmatrix} x & x \\ 1 & -1 \end{vmatrix} = -2x & x \leq 0 \\ \begin{vmatrix} x & x \\ 1 & 1 \end{vmatrix} = 0 & x \geq 0 \end{cases}$$

$W(x) \neq 0$ for $-1 \leq x \leq 0$, so $f_1(x)$ and $f_2(x)$ are linearly independent over the interval $[-1, 1]$ (Figure 9.6).

For the interval $[0, 1]$

$$W(x) = \begin{vmatrix} x & x \\ 1 & 1 \end{vmatrix} = 0 \qquad 0 \leq x \leq 1$$

and so this test does not tell us anything. However, $f_1(x)$ and $f_2(x)$ are identical over the interval $[0, 1]$, so they are linearly dependent (Figure 9.7).

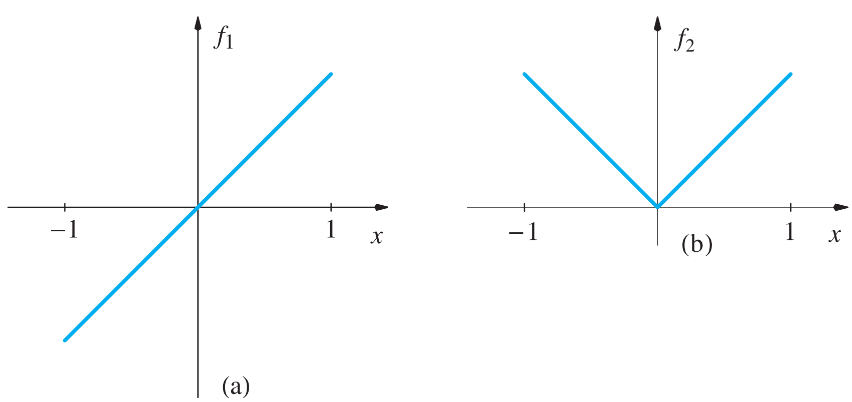

Figure 9.7
The functions $f_1(x) = x$ and $f_2(x) = |x|$ in Example 6 are linearly dependent over the interval $0 \leq x \leq 1$.

One final comment before we leave this section. It's easy to show (Problem 17) that if a set of differentiable functions $f_1(x)$, $f_2(x)$, . . . , $f_n(x)$ is linearly dependent on an interval I, then the Wronskian of these functions vanishes over the entire interval.

9.5 Problems

1. Show that the set of all ordered n-tuples of real numbers forms a vector space if addition of two n-tuples $(u_1, u_2, \ldots, u_n)$ and $(v_1, v_2, \ldots, v_n)$ is defined as $(u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n)$ and scalar multiplication is defined by $c(u_1, u_2, \ldots, u_n) = (cu_1, cu_2, \ldots, cu_n)$.

2. Show that the set of all two-dimensional geometric vectors forms a vector space.

3. Show that the set of all polynomials of degree less than or equal to 3 forms a vector space. What is its dimension?

4. Show that the set of functions that are continuous in the interval (a, b) forms a vector space.

5. Show that the set of n-tuples $(a, a, \ldots, a)$ is a subspace of R^n.

6. Let S be a subset of R^3 spanned by (1, 1, 0) and (1, 0, 1). Is S a subspace of V?

7. Test the following vectors for linear independence: (0, 1, 0, 0), (1, 1, 0, 0), (0, 1, 1, 0), and (0, 0, 0, 1).

8. Test the following vectors for linear independence: (1, 1, 1), (1, −1, 1), and (−1, 1, −1).

9. Is the vector (1, 0, 2) in the set spanned by (1, 1, 1), (1, −1, −1), (3, 1, 1)?

10. Show that the set of vectors {(1, 1, 1, 1), (1, −1, 1, −1), (1, 2, 3, 4), (1, 0, 2, 0)} is a basis for R^4.

11. Show that (1, 1, 0) and (1, 0, 1) are linearly independent in R^3 and find a third linearly independent vector.

12. Find the coordinates of (1, 2, 3) with respect to the basis (1, 1, 0), (1, 0, 1), (1, 1, 1) in R^3.

13. Use the Wronskian to test the three functions 1, $\sin x$, and $\cos x$ for linear independence.

14. Use the Wronskian to test the three functions e^x, e^{-x}, and $\sinh x$ for linear independence.

15. Evaluate the Wronskian of $f_1(x) = x^2$ and $f_2(x) = |x|x$ over the interval $[-1, 1]$. Are the functions linearly independent over the interval $[-1, 1]$? What about over the open interval $(0, 1]$?

16. Use the Wronskian to test for the linear independence of $f_1(x) = 1$, $f_2(x) = \sin^2 x$, and $f_3(x) = \cos^2 x$ for all x. If $W = 0$, can you check for linear independence by any other method?

17. We'll prove that if a set of $(n - 1)$ times differentiable functions $f_1(x)$, $f_2(x)$, . . . , $f_n(x)$ in an interval I is linearly dependent, their Wronskian vanishes identically on I. Argue that there must be nonzero constants in the expression $c_1 f_1(x) + c_2 f_2(x) + \cdots + c_n f_n(x) = 0$ for every x in I. Now form the set of n equations in the $f_j(x)$ and their first $n - 1$ derivatives. Why must the Wronskian equal zero?

18. This problem shows that the Wronskian of linearly independent functions may equal zero. First show that $f_1(x) = \begin{cases} 0 & x < 0 \\ x^2 & x \geq 0 \end{cases}$ and $f_2(x) = \begin{cases} -x^2 & x < 0 \\ 0 & x \geq 0 \end{cases}$ are linearly independent over $(-\infty, \infty)$. Now show that $W(x) = 0$.

19. Prove that the coordinates of a vector in a given basis are unique.

20. Show that the coefficient matrix of Equations 1 is made up of the three vectors in question arranged as columns. Using this observation, test the vectors (1, −1, 1, −1), (2, 3, −4, 1), and (0, −5, 6, −3) for linear independence in R^4.

9.6 Inner Product Spaces

The idea of a vector space generalizes the spaces of two- and three-dimensional geometric vectors that we discussed in Chapter 5. In those spaces we used a dot product to define lengths of vectors and the angle between two vectors. These concepts are so useful that it is desirable to introduce them into our general vector spaces. We shall now introduce a definition of an inner product for the vectors of a vector space.

A vector space is called an *inner product space* if in addition to the nine requirements that we listed in the previous section, there is a rule that associates with any two vectors **u** and **v** in V a real number, written as $\langle \mathbf{u}, \mathbf{v}\rangle$ [some authors use $(\mathbf{u}, \mathbf{v})$], that satisfies for all vectors in V

1. $$\langle a\,\mathbf{u_1} + b\,\mathbf{u_2}, \mathbf{u_3}\rangle = a\langle \mathbf{u_1}, \mathbf{u_3}\rangle + b\langle \mathbf{u_2}, \mathbf{u_3}\rangle \tag{1}$$

   where a and b are scalars. (The inner product is a linear property.)

2. $$\langle \mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{v}, \mathbf{u}\rangle \tag{2}$$

   (The inner product is commutative.)

3. $$\langle \mathbf{u}, \mathbf{u}\rangle \geq 0; \qquad \text{and} \qquad \langle \mathbf{u}, \mathbf{u}\rangle = 0 \qquad \text{if and only if} \qquad \mathbf{u} = 0 \tag{3}$$

   (This property is known as *positive definiteness*.)

Problem 1 has you prove that the dot product that we defined for geometric vectors is an inner product, so that the Euclidian space of two- or three-dimensional geometric vectors with a defined dot product is an inner product space.

Example 1:
Let $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$. Show that the product defined by

$$\langle \mathbf{u}, \mathbf{v}\rangle = u_1 v_1 + \cdots + u_n v_n$$

in the vector space R^n is an inner product.

SOLUTION: We shall verify each of the above three properties in turn:

1. $$\begin{aligned}\langle a\,\mathbf{u} + b\,\mathbf{v},\ \mathbf{w}\rangle &= (au_1 + bv_1)w_1 + (au_2 + bv_2)w_2 + \cdots + (au_n + bv_n)w_n \\ &= au_1w_1 + au_2w_2 + \cdots + au_nw_n + bv_1w_1 + \cdots + bv_nw_n \\ &= a\langle \mathbf{u}, \mathbf{w}\rangle + b\langle \mathbf{v}, \mathbf{w}\rangle\end{aligned}$$
2. $\langle \mathbf{u}, \mathbf{v}\rangle = u_1v_1 + \cdots + u_nv_n = v_1u_1 + \cdots + v_nu_n = \langle \mathbf{v}, \mathbf{u}\rangle$
3. $\langle \mathbf{u}, \mathbf{u}\rangle = u_1^2 + \cdots + u_n^2 > 0$ unless $u_1 = u_2 = \cdots = u_n = 0$.

Example 2:
Let V be the vector space of real-valued functions that are continuous on the interval $[\alpha, \beta]$. Show that

$$\langle f, g\rangle = \int_\alpha^\beta f(x)g(x)dx$$

is an inner product.

SOLUTION:

1. $\langle af_1 + bf_2, g\rangle = \int_\alpha^\beta (af_1 + bf_2)g dx = a\langle f_1, g\rangle + b\langle f_2, g\rangle$
2. $\langle f, g\rangle = \int_\alpha^\beta f(x)g(x)dx = \int_\alpha^\beta g(x)f(x)dx = \langle g, f\rangle$
3. $\langle f, f\rangle = \int_\alpha^\beta f^2(x)dx \geq 0$ and is equal to zero if and only if $f(x) = 0$ in $[\alpha, \beta]$.

Motivated by geometric vectors, we define the length of a vector in V by

$$\|\mathbf{u}\| = \langle \mathbf{u}, \mathbf{u}\rangle^{1/2} \tag{4}$$

We also call $\|\mathbf{u}\|$ the *norm* of $\mathbf{u}$. For the case of R^n, the norm is given by

$$\|\mathbf{u}\| = \langle \mathbf{u}, \mathbf{u}\rangle^{1/2} = (u_1^2 + u_2^2 + \cdots + u_n^2)^{1/2} \tag{5}$$

The inner product satisfies an important inequality called the *Schwarz inequality*:

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \leq \|\mathbf{u}\|\,\|\mathbf{v}\| \tag{6}$$

The proof of the Schwarz inequality goes as follows: Start with $\langle \mathbf{u} + \lambda\mathbf{v}, \mathbf{u} + \lambda\mathbf{v}\rangle \geq 0$, where λ is an arbitrary constant. Expand this inner product to write

$$\lambda^2\langle \mathbf{v}, \mathbf{v}\rangle + 2\lambda\langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{u}, \mathbf{u}\rangle \geq 0 \tag{7}$$

This inequality must be true for any value of λ, so we choose $\lambda = -\langle \mathbf{u}, \mathbf{v}\rangle / \langle \mathbf{v}, \mathbf{v}\rangle$. Substituting this choice of λ into Equation 7 gives

$$\langle \mathbf{u}, \mathbf{v}\rangle^2 \leq \langle \mathbf{u}, \mathbf{u}\rangle\langle \mathbf{v}, \mathbf{v}\rangle$$

The square root of this result gives Equation 6. Notice from Equation 6 that we can define the angle between $\mathbf{u}$ and $\mathbf{v}$ by $\cos\theta = \langle \mathbf{u}, \mathbf{v}\rangle / \|\mathbf{u}\|\,\|\mathbf{v}\|$, where $0 \leq \theta \leq \pi$ because $-1 \leq \langle \mathbf{u}, \mathbf{v}\rangle / \|\mathbf{u}\|\,\|\mathbf{v}\| \leq +1$.

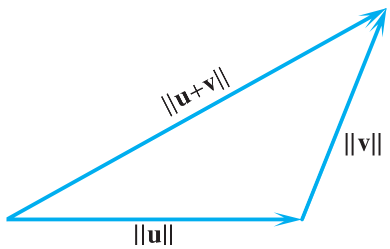

Figure 9.8
An illustration of the triangle inequality presented in Equation 10.

The norm in a vector space V satisfies the following properties:

$$\|\mathbf{v}\| \geq 0; \qquad \|\mathbf{v}\| = 0 \quad \text{if and only if} \quad \mathbf{v} = 0 \tag{8}$$

$$\|c\mathbf{v}\| = |c|\ \|\mathbf{v}\| \tag{9}$$

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\| \tag{10}$$

Equation 10 is called the *triangle inequality* (Figure 9.8). It can be readily proved using the Schwarz inequality (Problem 4).

Example 3:
Verify the Schwarz inequality for $\mathbf{u} = (2, 1, -1, 2, 0)$ and $\mathbf{v} = (-1, 0, 1, 2, -2)$.

SOLUTION: We use Equation 6:

$$\langle \mathbf{u}, \mathbf{v} \rangle = -2 + 0 - 1 + 4 + 0 = 1$$
$$\langle \mathbf{u}, \mathbf{u} \rangle = 4 + 1 + 1 + 4 + 0 = 10$$
$$\langle \mathbf{v}, \mathbf{v} \rangle = 1 + 0 + 1 + 4 + 4 = 10$$

The inequality reads $1 \leq 10$ in this case.

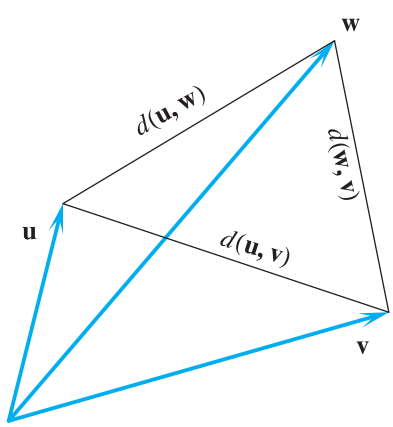

Figure 9.9
An illustration of the triangle inequality presented in Equation 13.

If $\mathbf{u}$ and $\mathbf{v}$ represent geometric vectors from an origin to points given by the tips of $\mathbf{u}$ and $\mathbf{v}$, then $|\mathbf{u} - \mathbf{v}|$ is the geometric distance between these points. We define the distance between vectors $\mathbf{u}$ and $\mathbf{v}$ in a vector space V by $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$, which you can show satisfies the following conditions (Problem 8):

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| \geq 0, \text{ which } = 0 \quad \text{if and only if} \quad \mathbf{u} = \mathbf{v} \tag{11}$$
$$= d(\mathbf{v}, \mathbf{u}) \tag{12}$$
$$\leq d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v}) \tag{13}$$

where $\mathbf{w}$ is a third vector (Figure 9.9). Equation 13 is another form of the triangle inequality.

If $\langle \mathbf{u}, \mathbf{v} \rangle = 0$, $\mathbf{u}$ and $\mathbf{v}$ are said to be *orthogonal*. For two- and three-dimensional geometric vectors, orthogonality means that the vectors are perpendicular to each other, but orthogonality is more general than that. Using the definition of the inner product of two functions as given in Example 2, we say the two functions, $f(x)$ and $g(x)$, are orthogonal over $[a, b]$ if

$$\langle f, g \rangle = \int_a^b f(x)g(x)dx = 0$$

Suppose that $\mathbf{v}_j$ for $j = 1, 2, \ldots, n$ is a set of orthogonal vectors in a vector space V, then we can express orthogonality by writing

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \|\mathbf{v}_j\|^2 \delta_{ij}$$

where δ_{ij} is the Kronecker delta. If the lengths of all the vectors are made to be unity by dividing each one by its length $\|\mathbf{v}_j\|$, then the new set is called *orthonormal*. An orthonormal set of vectors satisfies

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \delta_{ij} \tag{14}$$

Example 4:
Show that the geometric vectors $\mathbf{v}_1 = (\mathbf{i} + \mathbf{j})/\sqrt{2}$, $\mathbf{v}_2 = (\mathbf{i} - \mathbf{j})/\sqrt{2}$, and $\mathbf{v}_3 = \mathbf{k}$ form an orthonormal set of vectors.

SOLUTION:

$$\langle \mathbf{v}_1, \mathbf{v}_1 \rangle = \frac{(\mathbf{i}+\mathbf{j}) \cdot (\mathbf{i}+\mathbf{j})}{2} = 1 \qquad \langle \mathbf{v}_2, \mathbf{v}_2 \rangle = \frac{(\mathbf{i}-\mathbf{j}) \cdot (\mathbf{i}-\mathbf{j})}{2} = 1 \qquad \langle \mathbf{v}_3, \mathbf{v}_3 \rangle = \mathbf{k} \cdot \mathbf{k} = 1$$

$$\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \frac{1}{2}(\mathbf{i}+\mathbf{j}) \cdot (\mathbf{i}-\mathbf{j}) = 0 \qquad \langle \mathbf{v}_1, \mathbf{v}_3 \rangle = \frac{1}{\sqrt{2}}(\mathbf{i}+\mathbf{j}) \cdot \mathbf{k} = 0$$

$$\langle \mathbf{v}_2, \mathbf{v}_3 \rangle = \frac{1}{\sqrt{2}}(\mathbf{i}-\mathbf{j}) \cdot \mathbf{k} = 0$$

Therefore, $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \delta_{ij}$.

It's easy to show that an orthonormal set of vectors is linearly independent. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be an orthonormal set, and form

$$c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_n \mathbf{v}_n = 0 \tag{15}$$

where the c_j are to be determined. Now form the inner product of Equation 15 with each of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ in turn, and find that $c_j = 0$ for $j = 1, 2, \ldots, n$. Thus, the set of vectors is linearly independent.

We shall now show that every n-dimensional vector space V has an orthonormal basis by actually constructing one. Let $\mathbf{v}_j$ for $j = 1, 2, \ldots, n$ be any set of (nonzero) linearly independent vectors in V. Start with $\mathbf{v}_1$ and call it $\mathbf{u}_1$. Now take the second vector in the new set to be a linear combination of $\mathbf{v}_2$ and $\mathbf{u}_1$:

$$\mathbf{u}_2 = \mathbf{v}_2 + a_1 \mathbf{u}_1$$

such that $\langle \mathbf{u}_1, \mathbf{u}_2 \rangle = 0$. This condition gives $0 = \langle \mathbf{u}_1, \mathbf{v}_2 \rangle + a_1 \langle \mathbf{u}_1, \mathbf{u}_1 \rangle$ and so

$$\mathbf{u}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{u}_1, \mathbf{v}_2 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle} \mathbf{u}_1 \tag{16}$$

Now take the third member of the orthogonal set to be

$$\mathbf{u}_3 = \mathbf{v}_3 + b_1 \mathbf{u}_1 + b_2 \mathbf{u}_2$$

and require that $\langle \mathbf{u}_3, \mathbf{u}_1 \rangle = 0$ and $\langle \mathbf{u}_3, \mathbf{u}_2 \rangle = 0$. This gives (Problem 15)

$$\mathbf{u}_3 = \mathbf{v}_3 - \frac{\langle \mathbf{v}_3, \mathbf{u}_2 \rangle}{\langle \mathbf{u}_2, \mathbf{u}_2 \rangle}\,\mathbf{u}_2 - \frac{\langle \mathbf{v}_3, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle}\,\mathbf{u}_1 \tag{17}$$

The general pattern is evident now, and

$$\mathbf{u}_j = \mathbf{v}_j - \sum_{i=1}^{j-1} \frac{\langle \mathbf{v}_j, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle}\,\mathbf{u}_i \tag{18}$$

The $\mathbf{u}_j$ do not form an orthonormal set because they are not normalized, but it is easy to normalize each one simply by dividing by its length. This procedure for generating an orthonormal basis from a general basis is called *Gram-Schmidt orthogonalization*.

Example 5:
The three functions $f_1(x) = 1$, $f_2(x) = x$, and $f_3(x) = x^2$ form a basis for the vector space of all polynomials of degree equal to or less than 2. Using the definition of an inner product given in Example 2, find an orthonormal basis over the interval $[-1, 1]$.

SOLUTION: We start with $\mathbf{v}_1 = 1$, $\mathbf{v}_2 = x$, and $\mathbf{v}_3 = x^2$. Take $\mathbf{u}_1 = 1$ and write

$$\langle \mathbf{u}_1, \mathbf{v}_2 \rangle = \int_{-1}^{1} x\,dx = 0; \qquad \langle \mathbf{u}_1, \mathbf{u}_1 \rangle = \int_{-1}^{1} dx = 2$$

Using Equation 16, we find that $\mathbf{u}_2 = x$. Equation 17 requires that we evaluate

$$\langle \mathbf{u}_1, \mathbf{v}_3 \rangle = \int_{-1}^{1} x^2 dx = \frac{2}{3}$$

$$\langle \mathbf{u}_2, \mathbf{v}_3 \rangle = \int_{-1}^{1} x^3 dx = 0; \qquad \langle \mathbf{u}_2, \mathbf{u}_2 \rangle = \int_{-1}^{1} x^2 dx = \frac{2}{3}$$

which gives $\mathbf{u}_3 = x^2 - \frac{1}{3}$. We can normalize $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$:

$$\langle \mathbf{u}_1, \mathbf{u}_1 \rangle = 2 \qquad \langle \mathbf{u}_2, \mathbf{u}_2 \rangle = \int_{-1}^{1} x^2 dx = \frac{2}{3} \quad \text{and} \quad \langle \mathbf{u}_3, \mathbf{u}_3 \rangle = \int_{-1}^{1} \left(x^2 - \frac{1}{3} \right)^2 dx = \frac{8}{45}$$

and so the three orthonormal vectors are

$$\phi_1(x) = \left(\frac{1}{2} \right)^{1/2}; \quad \phi_2(x) = \left(\frac{3}{2} \right)^{1/2} x; \quad \phi_3(x) = \left(\frac{45}{8} \right)^{1/2} \left(x^2 - \frac{1}{3} \right)$$

9.6 Problems

1. Show that the dot product that we defined in Chapter 5 for geometric vectors is an inner product.
2. Show that the two geometric vectors $\mathbf{u} = \mathbf{i} + 2\,\mathbf{j} + \mathbf{k}$ and $\mathbf{v} = -\mathbf{i} + \mathbf{j} + \mathbf{k}$ satisfy the Schwarz inequality.
3. Show that the two functions $f_1(x) = 1 + x$ and $f_2(x) = x$ over the interval $[\,0, 1\,]$ satisfy the Schwarz inequality given the definition of the inner product in Example 2.
4. Prove the triangle inequality (Equation 10).
5. Show that $\|\mathbf{u}_1 + \mathbf{u}_2 + \cdots + \mathbf{u}_n\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{u}_2\|^2 + \cdots + \|\mathbf{u}_n\|^2$ if the $\mathbf{u}_j$ are orthogonal. This is the Pythagorean theorem in n-dimensions.
6. Let V be the vector space of real functions that are continuous on $[-\pi, \pi]$. Using the inner product defined in Example 2, show that 1, $\cos x$, $\cos 2x$, . . . , $\sin x$, $\sin 2x$, . . . form an orthogonal subset of V.
7. Suppose that $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ is an orthogonal basis of V. Show that
$$\mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{u}_1\rangle}{\langle \mathbf{u}_1, \mathbf{u}_1\rangle}\,\mathbf{u}_1 + \frac{\langle \mathbf{v}, \mathbf{u}_2\rangle}{\langle \mathbf{u}_2, \mathbf{u}_2\rangle}\,\mathbf{u}_2 + \cdots + \frac{\langle \mathbf{v}, \mathbf{u}_n\rangle}{\langle \mathbf{u}_n, \mathbf{u}_n\rangle}\,\mathbf{u}_n.$$
8. Show that the distance function $d(\mathbf{u}, \mathbf{v})$ satisfies Equations 11 to 13.
9. Using the inner product defined in Example 1, show that the norm of an ordered n-tuple of real numbers is $\|(u_1, u_2, \ldots, u_n)\| = (u_1^2 + u_2^2 + \cdots + u_n^2)^{1/2}$.
10. We defined the so-called Euclidean norm in Problem 9. We could also define the norm of an ordered n-tuple of real numbers by $\|\mathbf{u}\| = |u_1| + |u_2| + \cdots + |u_n|$. Show that this definition satisfies Equations 8 through 10.
11. Show that if we define the norm of an ordered n-tuple of real numbers by $\|\mathbf{u}\| = \max\,(|u_j|)$, then this definition satisfies Equations 8 through 10.
12. Calculate the norms of the vectors $(1, -2, 3)$ and $(2, 4, -1)$ according to the definition given in Problem 10.
13. Calculate the norms of the vectors $(1, -2, 3)$ and $(2, 4, -1)$ according to the definition given in Problem 11.
14. Calculate the distance functions of the vectors in Problem 12 using the three definitions of a norm that we have presented.
15. Derive Equation 17.
16. Construct an orthonormal basis from the three vectors $(1, -1, 0)$, $(1, 1, 0)$, and $(0, 1, 1)$.
17. The functions $f_1(x) = 1$, $f_2(x) = \sin 2x$, and $f_3(x) = \cos 2x$ over $[-\pi, \pi]$ are a basis for a three-dimensional vector space. Construct an orthonormal set from these three vectors.

9.7 Complex Inner Product Spaces

In our discussion of vector spaces, we have tacitly assumed that the scalars and vectors are real-valued quantities. It turns out that quantum mechanics can be formulated in terms of complex vector spaces with complex inner products, so in this brief section, we shall extend the results of the previous sections to include complex numbers.

The central notion of linear independence is not altered if we are dealing with a complex vector space instead of a real one. Nowhere in the previous section did we specify that the vectors or the set of constants in the definition of linear independence had to be real. Let's determine whether the vectors $(1, i, -1)$, $(1+i, 0, 1-i)$, and $(i, -1, -i)$ are linearly independent or linearly dependent. We form a matrix with the vectors as rows and then transform the matrix into echelon form:

$$\begin{pmatrix} 1 & i & -1 \\ 1+i & 0 & 1-i \\ i & -1 & -i \end{pmatrix} \sim \begin{pmatrix} 1 & i & -1 \\ 0 & 1-i & 1+i \\ 0 & 0 & 0 \end{pmatrix}$$

Thus we see that the vectors are not linearly independent. (See also Problem 1.)

The primary difference between a real and a complex inner product space is in our definition of an inner product. If we allow the scalars and vectors to be complex, Equations 1 through 3 of the previous section become

$$\langle \mathbf{u}|a\,\mathbf{v}_1 + b\,\mathbf{v}_2\rangle = a\langle \mathbf{u}|\mathbf{v}_1\rangle + b\langle \mathbf{u}|\mathbf{v}_2\rangle \tag{1}$$

$$\langle \mathbf{u}|\mathbf{v}\rangle = \langle \mathbf{v}|\mathbf{u}\rangle^* \tag{2}$$

$$\langle \mathbf{u}|\mathbf{u}\rangle \geq 0 \qquad \text{and} \qquad \langle \mathbf{u}|\mathbf{u}\rangle = 0 \quad \text{if and only if} \quad \mathbf{u} = 0 \tag{3}$$

We are using a vertical line rather than a comma to separate the two vectors enclosed by the angular brackets in Equations 1 through 3 to distinguish between real and complex product spaces. (This is standard notation in quantum mechanics.)

If we take the complex conjugate of Equation 1 and use Equation 2, we obtain

$$\langle a\,\mathbf{v}_1 + b\,\mathbf{v}_2|\mathbf{u}\rangle = a^*\langle \mathbf{v}_1|\mathbf{u}\rangle + b^*\langle \mathbf{v}_2|\mathbf{u}\rangle \tag{4}$$

In particular, if we let $\mathbf{v}_2 = 0$, then we have

$$\langle \mathbf{u}|a\,\mathbf{v}\rangle = a\langle \mathbf{u}|\mathbf{v}\rangle \tag{5}$$

from Equation 1 and

$$\langle a\,\mathbf{u}|\mathbf{v}\rangle = a^*\langle \mathbf{u}|\mathbf{v}\rangle \tag{6}$$

from Equation 4. Note that these two equations say that a scalar comes out of the inner product "as is" from the second position, but as its complex conjugate from the first position. This is standard notation in quantum mechanics texts, but not in all mathematics texts. Some mathematics texts define a complex inner product such that $\langle \mathbf{u}|a\,\mathbf{v}\rangle = a^*\langle \mathbf{u}|\mathbf{v}\rangle$ and $\langle a\,\mathbf{u}|\mathbf{v}\rangle = a\langle \mathbf{u}|\mathbf{v}\rangle$.

If $\mathbf{u}$ and $\mathbf{v}$ are ordered n-tuples of complex numbers, then we can define $\langle \mathbf{u}|\mathbf{v}\rangle$ by

$$\langle \mathbf{u}|\mathbf{v}\rangle = u_1^* v_1 + u_2^* v_2 + \cdots + u_n^* v_n \tag{7}$$

and the length of $\mathbf{u}$, $\|\mathbf{u}\|$ by

$$\|\mathbf{u}\| = \langle \mathbf{u}|\mathbf{u}\rangle = u_1^* u_1 + u_2^* u_2 + \cdots + u_n^* u_n \geq 0 \tag{8}$$

Equation 7 is sometimes called a *Hermitian inner product*. Problem 9 has you show that this definition satisfies Equations 1 through 3.

Example 1:
Given $\mathbf{u} = (1+i, 3, 4-i)$ and $\mathbf{v} = (3-4i, 1+i, 2i)$, find $\langle \mathbf{u}|\mathbf{v}\rangle$, $\langle \mathbf{v}|\mathbf{u}\rangle$, $\|\mathbf{u}\|$, and $\|\mathbf{v}\|$.

SOLUTION:

$$\langle \mathbf{u}|\mathbf{v}\rangle = (1-i)(3-4i) + 3(1+i) + (4+i)(2i) = 4i$$

$$\langle \mathbf{v}|\mathbf{u}\rangle = (3+4i)(1+i) + 3(1-i) - 2i(4-i) = -4i = \langle \mathbf{u}|\mathbf{v}\rangle^*$$

$$\|\mathbf{u}\| = [(1+i)(1-i) + 9 + (4+i)(4-i)]^{1/2} = \sqrt{28}$$

$$\|\mathbf{v}\| = [(3+4i)(3-4i) + (1+i)(1-i) + (2i)(-2i)]^{1/2} = \sqrt{31}$$

If a set of vectors $\mathbf{v}_j$ for $j = 1, 2, \ldots, n$ satisfies the condition $\langle \mathbf{v}_i|\mathbf{v}_j\rangle = \delta_{ij}$, we say that the set is orthonormal.

Example 2:
Show that the three 3-tuples $\mathbf{u}_1 = (1, i, 1+i)$, $\mathbf{u}_2 = (0, 1-i, i)$, and $\mathbf{u}_3 = (3i-3, 1+i, 2)$ form an orthogonal set. How would you make them orthonormal?

SOLUTION:

$$\langle \mathbf{u}_1|\mathbf{u}_2\rangle = 0 - i(1-i) + (1-i)i = 0$$

$$\langle \mathbf{u}_1|\mathbf{u}_3\rangle = (3i-3) - i(1+i) + 2(1-i) = 0$$

$$\langle \mathbf{u}_2|\mathbf{u}_3\rangle = 0 + (1+i)(1+i) - 2i = 0$$

To make them orthonormal, divide each one by its norm:

$$\|\mathbf{u}_1\| = \langle \mathbf{u}_1|\mathbf{u}_2\rangle^{1/2} = [(1)(1) - i(i) + (1-i)(1+i)]^{1/2} = 2$$

$$\|\mathbf{u}_2\| = \langle \mathbf{u}_2|\mathbf{u}_2\rangle^{1/2} = [(1+i)(1-i) + (-i)(i)]^{1/2} = 3^{1/2}$$

$$\|\mathbf{u}_3\| = \langle \mathbf{u}_3|\mathbf{u}_3\rangle^{1/2} = [(-3i-3)(3i-3) + (1-i)(1+i) + 4]^{1/2} = (24)^{1/2}$$

Example 3:
Show that the set of functions

$$\phi_m(\theta) = \frac{1}{\sqrt{2\pi}} e^{im\theta} \qquad 0 \le \theta \le 2\pi$$

for $m = 0, \pm 1, \pm 2, \ldots$ is orthonormal if we define the complex inner product by

$$\langle \phi_m | \phi_n \rangle = \int_0^{2\pi} \phi_m^*(\theta)\phi_n(\theta)\, d\theta$$

SOLUTION:

$$\langle \phi_m | \phi_n \rangle = \frac{1}{2\pi} \int_0^{2\pi} e^{-im\theta} e^{in\theta} d\theta$$
$$= \frac{1}{2\pi} \int_0^{2\pi} e^{i(n-m)\theta} d\theta$$

The integral here equals 1 if $m = n$ and 0 if $m \ne n$. If $m \ne n$, then it is an integral over complete cycles of the complex exponential function.

The Schwarz inequality takes on the same form for a complex vector space:

$$|\langle \mathbf{u} | \mathbf{v} \rangle| \le \|\mathbf{u}\| \, \|\mathbf{v}\| \tag{9}$$

and its proof parallels the one for a real vector space (Problem 11).

The Gram-Schmidt procedure is also valid for complex vector spaces. Let's construct an orthonormal basis from the two vectors $\mathbf{v}_1 = (-1, 1)$ and $\mathbf{v}_2 = (i, -1)$. We start with $\mathbf{u}_1 = \mathbf{v}_1 = (-1, 1)$ and write

$$\mathbf{u}_2 = \mathbf{v}_2 + a\, \mathbf{u}_1$$

Form the inner product with $\mathbf{u}_1$ from the right (see Problem 13) to get $\langle \mathbf{u}_2 | \mathbf{u}_1 \rangle = 0 = \langle \mathbf{v}_2 | \mathbf{u}_1 \rangle + \langle a\, \mathbf{u}_1 | \mathbf{u}_1 \rangle$, which in this case gives

$$0 = i - 1 + a^* \langle \mathbf{u}_1 | \mathbf{u}_1 \rangle = i - 1 + 2a^*$$

or $a = (1 + i)/2$. Thus, the two orthogonal vectors are $\mathbf{u}_1 = (-1, 1)$ and

$$\mathbf{u}_2 = (i, -1) + \frac{1+i}{2}(-1, 1) = \left(\frac{i-1}{2}, \frac{i-1}{2} \right) = \frac{1}{2}(i - 1, i - 1)$$

The two orthonormal vectors are

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}}(-1, 1) \qquad \text{and} \qquad \mathbf{u}_2 = \frac{1}{2}(i - 1, i - 1)$$

9.7 Problems

1. Show that the three vectors $(1, i, -1)$, $(1+i, 0, 1-i)$, and $(i, -1, -i)$ are not linearly independent by expressing one of them as a linear combination of the others.
2. Determine if the vectors $(1, 1, -i)$, $(0, i, i)$, and $(0, 1, -1)$ are linearly independent.
3. Determine if the vectors $(i, 0, 0)$, $(i, i, 0)$, and (i, i, i) are linearly independent.
4. Suppose that $\langle \mathbf{u} \mid \mathbf{v} \rangle = 2 + i$. Determine (a) $\langle (1-i)\mathbf{u} \mid \mathbf{v} \rangle$ and (b) $\langle \mathbf{u} \mid 2i\, \mathbf{v} \rangle$.
5. Show that the projection of a geometric vector **u** onto another geometric vector **v** is given by $\dfrac{\langle \mathbf{u}|\mathbf{v}\rangle \mathbf{v}}{\langle \mathbf{v}|\mathbf{v}\rangle}$. Calculate the projection of $\mathbf{u} = (1-i, 2i)$ onto $\mathbf{v} = (3+2i, -2+i)$ in a complex inner product space.
6. Determine if the four matrices $\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\boldsymbol{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\boldsymbol{\sigma}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, and $\boldsymbol{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ are linearly independent. These matrices are called the Pauli spin matrices.
7. Verify Equation 2 explicitly for (a) $\mathbf{u} = (1+i, 1)$; $\mathbf{v} = (-i, -1)$ and (b) $\mathbf{u} = (3, -i, 2i)$; $\mathbf{v} = (1, 3i, -1)$.
8. Let $\mathbf{u} = (1, 1)$ and $\mathbf{v} = (1, -2)$. Verify Equations 5 and 6 explicitly for $a = i$.
9. Show that the inner product defined by Equation 7 satisfies Equations 1 through 3.
10. Find the inner products and the lengths of the vectors in Problem 2.
11. Prove the Schwarz inequality for a complex vector space.
12. The two vectors $(1, i)$ and $(1, 1)$ are the basis of a two-dimensional complex vector space. Construct a pair of orthogonal vectors in this space.
13. Re-do the derivation at the end of the section by forming the inner product of $\mathbf{u}_1$ from the left.
14. We derive *Bessel's inequality* in this problem. Let $\boldsymbol{\phi}_j$, $j = 1, 2, \ldots, n$ be an orthonormal set of vectors in V and let **v** be any vector in V. If

$$\mathbf{v} = c_1 \boldsymbol{\phi}_1 + c_2 \boldsymbol{\phi}_j + \cdots + c_n \boldsymbol{\phi}_n \tag{10}$$

show that $c_j = \langle \mathbf{v} | \boldsymbol{\phi}_j \rangle$. Now form $\mathbf{u} = \mathbf{v} - \sum_{j=1}^{r} c_j \boldsymbol{\phi}_j$, where $r \le n$, and show that $\|\mathbf{v}\|^2 \ge \sum_{j=1}^{r} c_j^2$.

References

H. Anton, 2000, *Elementary Linear Algebra*, 8th ed., Wiley

H. Anton and C. Rorres, 1994, *Elementary Linear Algebra: Applications Version*, 7th ed., Wiley

P. Dennery and A. Kryzywicki, 1967, *Mathematics for Physicists*, Chapter 2, Dover Publications

F.B. Hildebrand, 1992, *Methods of Applied Mathematics*, Chapter 1, 2nd ed., Dover Publications

S. Lipschutz, 2000, *Linear Algebra*, 3rd ed., Schaum's Outline Series, McGraw-Hill

S. Lipschutz, 1989, *3000 Solved Problems in Linear Algebra*, Schaum's Outline Series, McGraw-Hill

C. Rorres and H. Anton, 1984, *Applications of Linear Algebra*, 3rd ed., Wiley