rocks and are measured by geophones attached to mile-long cables behind the ship.

- *Linear programming.* Many important management decisions today are made on the basis of linear programming models that use hundreds of variables. The airline industry, for instance, employs linear programs that schedule flight crews, monitor the locations of aircraft, or plan the varied schedules of support services such as maintenance and terminal operations.
- *Electrical networks.* Engineers use simulation software to design electrical circuits and microchips involving millions of transistors. Such software relies on linear algebra techniques and systems of linear equations.
- *Artificial intelligence.* Linear algebra plays a key role in everything from scrubbing data to facial recognition.
- *Signals and signal processing.* From a digital photograph to the daily price of a stock, important information is recorded as a signal and processed using linear transformations.
- *Machine learning.* Machines (specifically computers) use linear algebra to learn about anything from online shopping preferences to speech recognition.

Systems of linear equations lie at the heart of linear algebra, and this chapter uses them to introduce some of the central concepts of linear algebra in a simple and concrete setting. Sections 1.1 and 1.2 present a systematic method for solving systems of linear equations. This algorithm will be used for computations throughout the text. Sections 1.3 and 1.4 show how a system of linear equations is equivalent to a *vector equation* and to a *matrix equation*. This equivalence will reduce problems involving linear combinations of vectors to questions about systems of linear equations. The fundamental concepts of spanning, linear independence, and linear transformations, studied in the second half of the chapter, will play an essential role throughout the text as we explore the beauty and power of linear algebra.

1.1 Systems of Linear Equations

A **linear equation** in the variables $x_1, \dots, x_n$ is an equation that can be written in the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b \tag{1}$$

where b and the **coefficients** $a_1, \dots, a_n$ are real or complex numbers, usually known in advance. The subscript n may be any positive integer. In textbook examples and exercises, n is normally between 2 and 5. In real-life problems, n might be 50 or 5000, or even larger.

The equations

$$4x_1 - 5x_2 + 2 = x_1 \quad \text{and} \quad x_2 = 2\left(\sqrt{6} - x_1\right) + x_3$$

are both linear because they can be rearranged algebraically as in equation (1):

$$3x_1 - 5x_2 = -2 \quad \text{and} \quad 2x_1 + x_2 - x_3 = 2\sqrt{6}$$

The equations

$$4x_1 - 5x_2 = x_1x_2 \quad \text{and} \quad x_2 = 2\sqrt{x_1} - 6$$

are not linear because of the presence of x_1x_2 in the first equation and $\sqrt{x_1}$ in the second.

A **system of linear equations** (or a **linear system**) is a collection of one or more linear equations involving the same variables—say, $x_1, \ldots, x_n$. An example is

$$\begin{aligned} 2x_1 - x_2 + 1.5x_3 &= 8 \\ x_1 - 4x_3 &= -7 \end{aligned} \tag{2}$$

A **solution** of the system is a list $(s_1, s_2, \ldots, s_n)$ of numbers that makes each equation a true statement when the values $s_1, \ldots, s_n$ are substituted for $x_1, \ldots, x_n$, respectively. For instance, $(5, 6.5, 3)$ is a solution of system (2) because, when these values are substituted in (2) for x_1, x_2, x_3, respectively, the equations simplify to $8 = 8$ and $-7 = -7$.

The set of all possible solutions is called the **solution set** of the linear system. Two linear systems are called **equivalent** if they have the same solution set. That is, each solution of the first system is a solution of the second system, and each solution of the second system is a solution of the first.

Finding the solution set of a system of two linear equations in two variables is easy because it amounts to finding the intersection of two lines. A typical problem is

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

The graphs of these equations are lines, which we denote by ℓ_1 and ℓ_2. A pair of numbers (x_1, x_2) satisfies *both* equations in the system if and only if the point (x_1, x_2) lies on both ℓ_1 and ℓ_2. In the system above, the solution is the single point $(3, 2)$, as you can easily verify. See Figure 1.

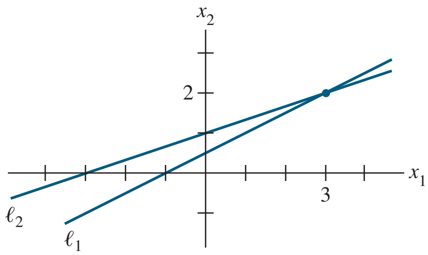
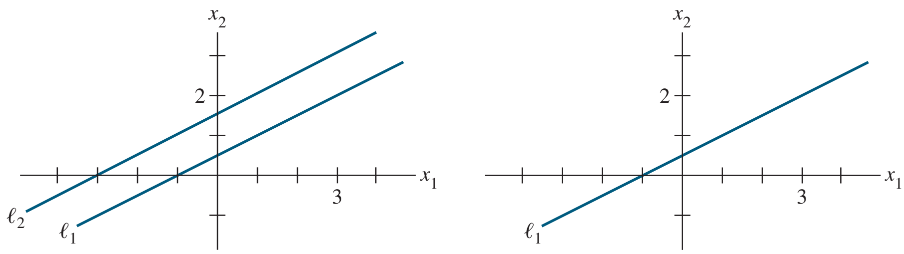


FIGURE 1 Exactly one solution.

Of course, two lines need not intersect in a single point—they could be parallel, or they could coincide and hence "intersect" at every point on the line. Figure 2 shows the graphs that correspond to the following systems:

$$\text{(a)}\quad \begin{aligned} x_1 - 2x_2 &= -1 \\ -x_1 + 2x_2 &= 3 \end{aligned} \qquad\qquad \text{(b)}\quad \begin{aligned} x_1 - 2x_2 &= -1 \\ -x_1 + 2x_2 &= 1 \end{aligned}$$

Figures 1 and 2 illustrate the following general fact about linear systems, to be verified in Section 1.2.



(a) (b)

FIGURE 2 (a) No solution. (b) Infinitely many solutions.

A system of linear equations has

1. no solution, or
2. exactly one solution, or
3. infinitely many solutions.

A system of linear equations is said to be **consistent** if it has either one solution or infinitely many solutions; a system is **inconsistent** if it has no solution.

Matrix Notation

The essential information of a linear system can be recorded compactly in a rectangular array called a **matrix**. Given the system

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_2 - 8x_3 &= 8 \\ 5x_1 - 5x_3 &= 10 \end{aligned} \tag{3}$$

with the coefficients of each variable aligned in columns, the matrix

$$\begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -8 \\ 5 & 0 & -5 \end{bmatrix}$$

is called the **coefficient matrix** (or **matrix of coefficients**) of the system (3), and the matrix

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix} \tag{4}$$

is called the **augmented matrix** of the system. (The second row here contains a zero because the second equation could be written as $0 \cdot x_1 + 2x_2 - 8x_3 = 8$.) An augmented matrix of a system consists of the coefficient matrix with an added column containing the constants from the respective right sides of the equations.

The **size** of a matrix tells how many rows and columns it has. The augmented matrix (4) above has 3 rows and 4 columns and is called a 3×4 (read "3 by 4") matrix. If m and n are positive integers, an $\boldsymbol{m \times n}$ **matrix** is a rectangular array of numbers with m rows and n columns. (The number of rows always comes first.) Matrix notation will simplify the calculations in the examples that follow.

Solving a Linear System

This section and the next describe an algorithm, or a systematic procedure, for solving linear systems. The basic strategy is *to replace one system with an equivalent system (that is one with the same solution set) that is easier to solve*.

Roughly speaking, use the x_1 term in the first equation of a system to eliminate the x_1 terms in the other equations. Then use the x_2 term in the second equation to eliminate the x_2 terms in the other equations, and so on, until you finally obtain a very simple equivalent system of equations.

Three basic operations are used to simplify a linear system: Replace one equation by the sum of itself and a multiple of another equation, interchange two equations, and multiply all the terms in an equation by a nonzero constant. After the first example, you will see why these three operations do not change the solution set of the system.

EXAMPLE 1 Solve system (3).

SOLUTION The elimination procedure is shown here with and without matrix notation, and the results are placed side by side for comparison:

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_2 - 8x_3 &= 8 \\ 5x_1 \qquad - 5x_3 &= 10 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix}$$

Keep x_1 in the first equation and eliminate it from the other equations. To do so, add -5 times equation 1 to equation 3. After some practice, this type of calculation is usually performed mentally:

$$\begin{array}{r r} -5\cdot[\text{equation 1}] & -5x_1 + 10x_2 - 5x_3 = 0 \\ +[\text{equation 3}] & 5x_1 \qquad\quad - 5x_3 = 10 \\ \hline [\text{new equation 3}] & 10x_2 - 10x_3 = 10 \end{array}$$

The result of this calculation is written in place of the original third equation:

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_2 - 8x_3 &= 8 \\ 10x_2 - 10x_3 &= 10 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & 10 & -10 & 10 \end{bmatrix}$$

Now, multiply equation 2 by $\frac{1}{2}$ in order to obtain 1 as the coefficient for x_2. (This calculation will simplify the arithmetic in the next step.)

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ x_2 - 4x_3 &= 4 \\ 10x_2 - 10x_3 &= 10 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 10 & -10 & 10 \end{bmatrix}$$

Use the x_2 in equation 2 to eliminate the $10x_2$ in equation 3. The "mental" computation is

$$\begin{array}{r r} -10\cdot[\text{equation 2}] & -10x_2 + 40x_3 = -40 \\ +[\text{equation 3}] & 10x_2 - 10x_3 = 10 \\ \hline [\text{new equation 3}] & 30x_3 = -30 \end{array}$$

The result of this calculation is written in place of the previous third equation (row):

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ x_2 - 4x_3 &= 4 \\ 30x_3 &= -30 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 30 & -30 \end{bmatrix}$$

Now, multiply equation 3 by $\frac{1}{30}$ in order to obtain 1 as the coefficient for x_3. (This calculation will simplify the arithmetic in the next step.)

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ x_2 - 4x_3 &= 4 \\ x_3 &= -1 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

The new system has a *triangular* form (the intuitive term *triangular* will be replaced by a precise term in the next section):

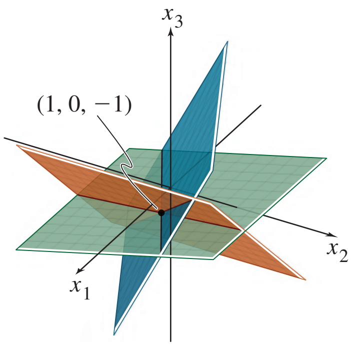
$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ x_2 - 4x_3 &= 4 \\ x_3 &= -1 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

Eventually, you want to eliminate the $-2x_2$ term from equation 1, but it is more efficient to use the x_3 in equation 3 first, to eliminate the $-4x_3$ and $+x_3$ terms in equations 2 and 1. The two "mental" calculations are

$$\begin{array}{rr} 4\cdot[\text{equation 3}] & 4x_3 = -4 \\ +\,[\text{equation 2}] & x_2 - 4x_3 = 4 \\ \hline [\text{new equation 2}] & x_2 = 0 \end{array} \qquad \begin{array}{rr} -1\cdot[\text{equation 3}] & -x_3 = 1 \\ +\,[\text{equation 1}] & x_1 - 2x_2 + x_3 = 0 \\ \hline [\text{new equation 1}] & x_1 - 2x_2 = 1 \end{array}$$

It is convenient to combine the results of these two operations:

$$\begin{aligned} x_1 - 2x_2 &= 1 \\ x_2 &= 0 \\ x_3 &= -1 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$



Each of the original equations determines a plane in three-dimensional space. The point $(1, 0, -1)$ lies in all three planes.

Now, having cleaned out the column above the x_3 in equation 3, move back to the x_2 in equation 2 and use it to eliminate the $-2x_2$ above it. Because of the previous work with x_3, there is now no arithmetic involving x_3 terms. Add 2 times equation 2 to equation 1 and obtain the system:

$$\begin{aligned} x_1 &= 1 \\ x_2 &= 0 \\ x_3 &= -1 \end{aligned} \qquad \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

The work is essentially done. It shows that the only solution of the original system is $(1, 0, -1)$. However, since there are so many calculations involved, it is a good practice to check the work. To verify that $(1, 0, -1)$ *is* a solution, substitute these values into the left side of the original system, and compute:

$$\begin{aligned} 1(1) - 2(0) + 1(-1) &= 1 - 0 - 1 = 0 \\ 2(0) - 8(-1) &= 0 + 8 = 8 \\ 5(1) - 5(-1) &= 5 + 5 = 10 \end{aligned}$$

The results agree with the right side of the original system, so $(1, 0, -1)$ is a solution of the system. ■

Example 1 illustrates how operations on equations in a linear system correspond to operations on the appropriate rows of the augmented matrix. The three basic operations listed earlier correspond to the following operations on the augmented matrix.

ELEMENTARY ROW OPERATIONS

1. (Replacement) Replace one row by the sum of itself and a multiple of another row.[1]
2. (Interchange) Interchange two rows.
3. (Scaling) Multiply all entries in a row by a nonzero constant.

[1] A common paraphrase of row replacement is "Add to one row a multiple of another row."

Row operations can be applied to any matrix, not merely to one that arises as the augmented matrix of a linear system. Two matrices are called **row equivalent** if there is a sequence of elementary row operations that transforms one matrix into the other.

It is important to note that row operations are *reversible*. If two rows are interchanged, they can be returned to their original positions by another interchange. If a row is scaled by a nonzero constant c, then multiplying the new row by $1/c$ produces the original row. Finally, consider a replacement operation involving two rows—say, rows 1 and 2—and suppose that c times row 1 is added to row 2 to produce a new row 2. To "reverse" this operation, add $-c$ times row 1 to (new) row 2 and obtain the original row 2. See Exercises 39–42 at the end of this section.

At the moment, we are interested in row operations on the augmented matrix of a system of linear equations. Suppose a system is changed to a new one via row operations. By considering each type of row operation, you can see that any solution of the original system remains a solution of the new system. Conversely, since the original system can be produced via row operations on the new system, each solution of the new system is also a solution of the original system. This discussion justifies the following statement.

If the augmented matrices of two linear systems are row equivalent, then the two systems have the same solution set.

Though Example 1 is lengthy, you will find that after some practice, the calculations go quickly. Row operations in the text and exercises will usually be extremely easy to perform, allowing you to focus on the underlying concepts. Still, you must learn to perform row operations accurately because they will be used throughout the text.

The rest of this section shows how to use row operations to determine the size of a solution set, without completely solving the linear system.

Existence and Uniqueness Questions

Section 1.2 will show why a solution set for a linear system contains either no solutions, one solution, or infinitely many solutions. Answers to the following two questions will determine the nature of the solution set for a linear system.

To determine which possibility is true for a particular system, we ask two questions.

TWO FUNDAMENTAL QUESTIONS ABOUT A LINEAR SYSTEM

1. Is the system consistent; that is, does at least one solution *exist*?
2. If a solution exists, is it the *only* one; that is, is the solution *unique*?

These two questions will appear throughout the text, in many different guises. This section and the next will show how to answer these questions via row operations on the augmented matrix.

EXAMPLE 2 Determine if the following system is consistent:

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_2 - 8x_3 &= 8 \\ 5x_1 \phantom{{}-2x_2} - 5x_3 &= 10 \end{aligned}$$

SOLUTION This is the system from Example 1. Suppose that we have performed the row operations necessary to obtain the triangular form

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ x_2 - 4x_3 &= 4 \\ x_3 &= -1 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

At this point, we know x_3. Were we to substitute the value of x_3 into equation 2, we could compute x_2 and hence could determine x_1 from equation 1. So a solution exists; the system is consistent. (In fact, x_2 is uniquely determined by equation 2 since x_3 has only one possible value, and x_1 is therefore uniquely determined by equation 1. So the solution is unique.) ■

EXAMPLE 3 Determine if the following system is consistent:

$$\begin{aligned} x_2 - 4x_3 &= 8 \\ 2x_1 - 3x_2 + 2x_3 &= 1 \\ 4x_1 - 8x_2 + 12x_3 &= 1 \end{aligned} \tag{5}$$

SOLUTION The augmented matrix is

$$\begin{bmatrix} 0 & 1 & -4 & 8 \\ 2 & -3 & 2 & 1 \\ 4 & -8 & 12 & 1 \end{bmatrix}$$

To obtain an x_1 in the first equation, interchange rows 1 and 2:

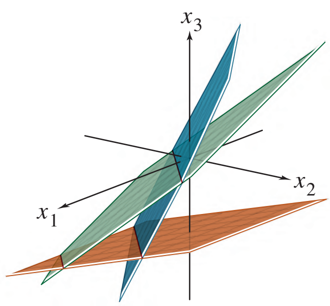
$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 4 & -8 & 12 & 1 \end{bmatrix}$$

To eliminate the $4x_1$ term in the third equation, add -2 times row 1 to row 3:

$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & -2 & 8 & -1 \end{bmatrix} \tag{6}$$

Next, use the x_2 term in the second equation to eliminate the $-2x_2$ term from the third equation. Add 2 times row 2 to row 3:

$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & 15 \end{bmatrix} \tag{7}$$



The system is inconsistent because there is no point that lies on all three planes.

The augmented matrix is now in triangular form. To interpret it correctly, go back to equation notation:

$$\begin{aligned} 2x_1 - 3x_2 + 2x_3 &= 1 \\ x_2 - 4x_3 &= 8 \\ 0 &= 15 \end{aligned} \tag{8}$$

The equation $0 = 15$ is a short form of $0x_1 + 0x_2 + 0x_3 = 15$. This system in triangular form obviously has a built-in contradiction. There are no values of x_1, x_2, x_3 that satisfy (8) because the equation $0 = 15$ is never true. Since (8) and (5) have the same solution set, the original system is inconsistent (it has no solution). ■

Pay close attention to the augmented matrix in (7). Its last row is typical of an inconsistent system in triangular form.

Reasonable Answers

Once you have one or more solutions to a system of equations, remember to check your answer by substituting the solution you found back into the original equation. For example, if you found $(2, 1, -1)$ was a solution to the system of equations

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

you could substitute your solution into the original equations to get

$$\begin{aligned} 2 - 2(1) + (-1) &= -1 \neq 2 \\ 2 \qquad\quad - 2(-1) &= 4 \neq -2 \\ 1 + (-1) &= 0 \neq 3 \end{aligned}$$

It is now clear that there must have been an error in your original calculations. If upon rechecking your arithmetic, you find the answer $(2, 1, 2)$, you can see that

$$\begin{aligned} 2 - 2(1) + (2) &= 2 = 2 \\ 2 \qquad\quad - 2(2) &= -2 = -2 \\ 1 + 2 &= 3 = 3 \end{aligned}$$

and you can now be confident you have a correct solution to the given system of equations.

Numerical Note

In real-world problems, systems of linear equations are solved by a computer. For a square coefficient matrix, computer programs nearly always use the elimination algorithm given here and in Section 1.2, modified slightly for improved accuracy.

The vast majority of linear algebra problems in business and industry are solved with programs that use *floating point arithmetic*. Numbers are represented as decimals $\pm .d_1 \cdots d_p \times 10^r$, where r is an integer and the number p of digits to the right of the decimal point is usually between 8 and 16. Arithmetic with such numbers typically is inexact, because the result must be rounded (or truncated) to the number of digits stored. "Roundoff error" is also introduced when a number such as $1/3$ is entered into the computer, since its decimal representation must be approximated by a finite number of digits. Fortunately, inaccuracies in floating point arithmetic seldom cause problems. The numerical notes in this book will occasionally warn of issues that you may need to consider later in your career.

Practice Problems

Throughout the text, practice problems should be attempted before working the exercises. Solutions appear after each exercise set.

1. State in words the next elementary row operation that should be performed on the system in order to solve it. [More than one answer is possible in (a).]

Practice Problems (Continued)

a. $$\begin{aligned} x_1 + 4x_2 - 2x_3 + 8x_4 &= 12 \\ x_2 - 7x_3 + 2x_4 &= -4 \\ 5x_3 - x_4 &= 7 \\ x_3 + 3x_4 &= -5 \end{aligned}$$

b. $$\begin{aligned} x_1 - 3x_2 + 5x_3 - 2x_4 &= 0 \\ x_2 + 8x_3 \qquad\quad &= -4 \\ 2x_3 \qquad\quad &= 3 \\ x_4 &= 1 \end{aligned}$$

2. The augmented matrix of a linear system has been transformed by row operations into the form below. Determine if the system is consistent.

$$\begin{bmatrix} 1 & 5 & 2 & -6 \\ 0 & 4 & -7 & 2 \\ 0 & 0 & 5 & 0 \end{bmatrix}$$

3. Is $(3, 4, -2)$ a solution of the following system?

$$\begin{aligned} 5x_1 - x_2 + 2x_3 &= 7 \\ -2x_1 + 6x_2 + 9x_3 &= 0 \\ -7x_1 + 5x_2 - 3x_3 &= -7 \end{aligned}$$

4. For what values of h and k is the following system consistent?

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

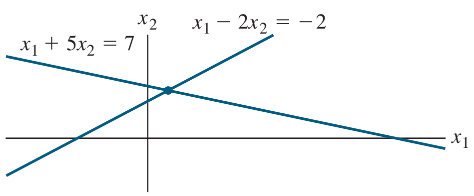
1.1 Exercises

Solve each system in Exercises 1–4 by using elementary row operations on the equations or on the augmented matrix. Follow the systematic elimination procedure described in this section.

1. $$\begin{aligned} x_1 + 5x_2 &= 7 \\ -2x_1 - 7x_2 &= -5 \end{aligned}$$

2. $$\begin{aligned} 2x_1 + 4x_2 &= -4 \\ 5x_1 + 7x_2 &= 11 \end{aligned}$$

3. Find the point (x_1, x_2) that lies on the line $x_1 + 5x_2 = 7$ and on the line $x_1 - 2x_2 = -2$. See the figure.



4. Find the point of intersection of the lines $x_1 - 5x_2 = 1$ and $3x_1 - 7x_2 = 5$.

Consider each matrix in Exercises 5 and 6 as the augmented matrix of a linear system. State in words the next two elementary row operations that should be performed in the process of solving the system.

5. $$\begin{bmatrix} 1 & 3 & -4 & 0 & 9 \\ 1 & 1 & 5 & 0 & -8 \\ 0 & 0 & 1 & 0 & 7 \\ 0 & 0 & 0 & 1 & -6 \end{bmatrix}$$

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

In Exercises 7–10, the augmented matrix of a linear system has been reduced by row operations to the form shown. In each case, continue the appropriate row operations and describe the solution set of the original system.

7. $$\begin{bmatrix} 1 & 7 & 3 & -4 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix}$$

8. $$\begin{bmatrix} 1 & 1 & 5 & 0 \\ 0 & 1 & 9 & 0 \\ 0 & 0 & 7 & -7 \end{bmatrix}$$

9. $$\begin{bmatrix} 1 & -1 & 0 & 0 & -4 \\ 0 & 1 & -3 & 0 & -7 \\ 0 & 0 & 1 & -3 & -1 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

10. $\begin{bmatrix} 1 & -2 & 0 & 3 & 0 \\ 0 & 1 & 0 & -4 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$

Solve the systems in Exercises 11–14.

11. $$\begin{aligned} x_2 + 4x_3 &= -4 \\ x_1 + 3x_2 + 3x_3 &= -2 \\ 3x_1 + 7x_2 + 5x_3 &= 6 \end{aligned}$$

12. $$\begin{aligned} x_1 - 3x_2 + 4x_3 &= -4 \\ 3x_1 - 7x_2 + 7x_3 &= -8 \\ -4x_1 + 6x_2 + 2x_3 &= 4 \end{aligned}$$

13. $$\begin{aligned} x_1 - 3x_3 &= 8 \\ 2x_1 + 2x_2 + 9x_3 &= 7 \\ x_2 + 5x_3 &= -2 \end{aligned}$$

14. $$\begin{aligned} x_1 - 3x_2 &= 5 \\ -x_1 + x_2 + 5x_3 &= 2 \\ x_2 + x_3 &= 0 \end{aligned}$$

15. Verify that the solution you found to Exercise 11 is correct by substituting the values you obtained back into the original equations.

16. Verify that the solution you found to Exercise 12 is correct by substituting the values you obtained back into the original equations.

17. Verify that the solution you found to Exercise 13 is correct by substituting the values you obtained back into the original equations.

18. Verify that the solution you found to Exercise 14 is correct by substituting the values you obtained back into the original equations.

Determine if the systems in Exercises 19 and 20 are consistent. Do not completely solve the systems.

19. $$\begin{aligned} x_1 + 3x_3 &= 2 \\ x_2 - 3x_4 &= 3 \\ -2x_2 + 3x_3 + 2x_4 &= 1 \\ 3x_1 + 7x_4 &= -5 \end{aligned}$$

20. $$\begin{aligned} x_1 - 2x_4 &= -3 \\ 2x_2 + 2x_3 &= 0 \\ x_3 + 3x_4 &= 1 \\ -2x_1 + 3x_2 + 2x_3 + x_4 &= 5 \end{aligned}$$

21. Do the three lines $x_1 - 4x_2 = 1$, $2x_1 - x_2 = -3$, and $-x_1 - 3x_2 = 4$ have a common point of intersection? Explain.

22. Do the three planes $x_1 + 2x_2 + x_3 = 4$, $x_2 - x_3 = 1$, and $x_1 + 3x_2 = 0$ have at least one common point of intersection? Explain.

In Exercises 23–26, determine the value(s) of h such that the matrix is the augmented matrix of a consistent linear system.

23. $\begin{bmatrix} 1 & h & 4 \\ 3 & 6 & 8 \end{bmatrix}$ **24.** $\begin{bmatrix} 1 & h & -3 \\ -2 & 4 & 6 \end{bmatrix}$

25. $\begin{bmatrix} 1 & 3 & -2 \\ -4 & h & 8 \end{bmatrix}$ **26.** $\begin{bmatrix} 3 & -4 & h \\ -6 & 8 & 9 \end{bmatrix}$

In Exercises 27–34, key statements from this section are either quoted directly, restated slightly (but still true), or altered in some way that makes them false in some cases. Mark each statement True or False, and *justify* your answer. (If true, give the approximate location where a similar statement appears, or refer to a definition or theorem. If false, give the location of a statement that has been quoted or used incorrectly, or cite an example that shows the statement is not true in all cases.) Similar true/false questions will appear in many sections of the text and will be flagged with a **(T/F)** at the beginning of the question.

27. **(T/F)** Every elementary row operation is reversible.

28. **(T/F)** Elementary row operations on an augmented matrix never change the solution set of the associated linear system.

29. **(T/F)** A 5×6 matrix has six rows.

30. **(T/F)** Two matrices are row equivalent if they have the same number of rows.

31. **(T/F)** The solution set of a linear system involving variables $x_1, \ldots, x_n$ is a list of numbers $(s_1, \ldots, s_n)$ that makes each equation in the system a true statement when the values $s_1, \ldots, s_n$ are substituted for $x_1, \ldots, x_n$, respectively.

32. **(T/F)** An inconsistent system has more than one solution.

33. **(T/F)** Two fundamental questions about a linear system involve existence and uniqueness.

34. **(T/F)** Two linear systems are equivalent if they have the same solution set.

35. Find an equation involving g, h, and k that makes this augmented matrix correspond to a consistent system:

$$\begin{bmatrix} 1 & -3 & 5 & g \\ 0 & 2 & -3 & h \\ -3 & 5 & -9 & k \end{bmatrix}$$

36. Construct three different augmented matrices for linear systems whose solution set is $x_1 = -2$, $x_2 = 1$, $x_3 = 0$.

37. Suppose the system below is consistent for all possible values of f and g. What can you say about the coefficients c and d? Justify your answer.

$$\begin{aligned} x_1 + 5x_2 &= f \\ cx_1 + dx_2 &= g \end{aligned}$$

38. Suppose a, b, c, and d are constants such that a is not zero and the system below is consistent for all possible values of f and g. What can you say about the numbers a, b, c, and d? Justify your answer.

$$\begin{aligned} ax_1 + bx_2 &= f \\ cx_1 + dx_2 &= g \end{aligned}$$

In Exercises 39–42, find the elementary row operation that transforms the first matrix into the second, and then find the reverse row operation that transforms the second matrix into the first.

39. $\begin{bmatrix} 0 & -2 & 5 \\ 1 & 4 & -7 \\ 3 & -1 & 6 \end{bmatrix}, \begin{bmatrix} 1 & 4 & -7 \\ 0 & -2 & 5 \\ 3 & -1 & 6 \end{bmatrix}$

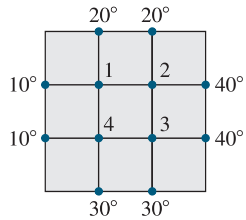
40. $\begin{bmatrix} 1 & 3 & -4 \\ 0 & -2 & 6 \\ 0 & -5 & 9 \end{bmatrix}, \begin{bmatrix} 1 & 3 & -4 \\ 0 & 1 & -3 \\ 0 & -5 & 9 \end{bmatrix}$

41. $\begin{bmatrix} 1 & -3 & 2 & 0 \\ 0 & 4 & -5 & 6 \\ 5 & -7 & 8 & -9 \end{bmatrix}, \begin{bmatrix} 1 & -3 & 2 & 0 \\ 0 & 4 & -5 & 6 \\ 0 & 8 & -2 & -9 \end{bmatrix}$

42. $\begin{bmatrix} 1 & 2 & -5 & 0 \\ 0 & 1 & -3 & -2 \\ 0 & -3 & 9 & 5 \end{bmatrix}, \begin{bmatrix} 1 & 2 & -5 & 0 \\ 0 & 1 & -3 & -2 \\ 0 & 0 & 0 & -1 \end{bmatrix}$

An important concern in the study of heat transfer is to determine the steady-state temperature distribution of a thin plate when the temperature around the boundary is known. Assume the plate shown in the figure represents a cross section of a metal beam, with negligible heat flow in the direction perpendicular to the plate. Let $T_1, \dots, T_4$ denote the temperatures at the four interior nodes of the mesh in the figure. The temperature at a node is approximately equal to the average of the four nearest nodes—to the left, above, to the right, and below.[2] For instance,

$$T_1 = (10 + 20 + T_2 + T_4)/4, \quad \text{or} \quad 4T_1 - T_2 - T_4 = 30$$



43. Write a system of four equations whose solution gives estimates for the temperatures $T_1, \dots, T_4$.

44. Solve the system of equations from Exercise 43. [*Hint:* To speed up the calculations, interchange rows 1 and 4 before starting "replace" operations.]

[2] See Frank M. White, *Heat and Mass Transfer* (Reading, MA: Addison-Wesley Publishing, 1991), pp. 145–149.

Solutions to Practice Problems

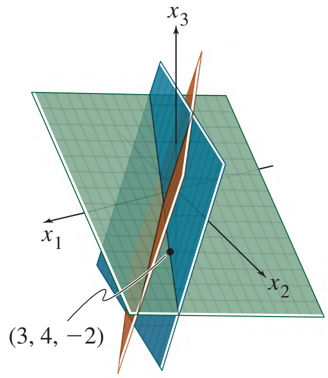
1. a. For "hand computation," the best choice is to interchange equations 3 and 4. Another possibility is to multiply equation 3 by $1/5$. Or, replace equation 4 by its sum with $-1/5$ times row 3. (In any case, do not use the x_2 in equation 2 to eliminate the $4x_2$ in equation 1. Wait until a triangular form has been reached and the x_3 terms and x_4 terms have been eliminated from the first two equations.)

b. The system is in triangular form. Further simplification begins with the x_4 in the fourth equation. Use the x_4 to eliminate all x_4 terms above it. The appropriate step now is to add 2 times equation 4 to equation 1. (After that, move to equation 3, multiply it by $1/2$, and then use the equation to eliminate the x_3 terms above it.)

2. The system corresponding to the augmented matrix is

$$\begin{aligned} x_1 + 5x_2 + 2x_3 &= -6 \\ 4x_2 - 7x_3 &= 2 \\ 5x_3 &= 0 \end{aligned}$$

The third equation makes $x_3 = 0$, which is certainly an allowable value for x_3. After eliminating the x_3 terms in equations 1 and 2, you could go on to solve for unique values for x_2 and x_1. Hence a solution exists, and it is unique. Contrast this situation with that in Example 3.



Since (3, 4, −2) satisfies the first two equations, it is on the line of the intersection of the first two planes. Since (3, 4, −2) does not satisfy all three equations, it does not lie on all three planes.

3. It is easy to check if a specific list of numbers is a solution. Set $x_1 = 3$, $x_2 = 4$, and $x_3 = -2$, and find that

$$\begin{aligned} 5(3) - (4) + 2(-2) &= 15 - 4 - 4 = 7 \\ -2(3) + 6(4) + 9(-2) &= -6 + 24 - 18 = 0 \\ -7(3) + 5(4) - 3(-2) &= -21 + 20 + 6 = 5 \end{aligned}$$

Although the first two equations are satisfied, the third is not, so $(3, 4, -2)$ is not a solution of the system. Notice the use of parentheses when making the substitutions. They are strongly recommended as a guard against arithmetic errors.

4. When the second equation is replaced by its sum with 3 times the first equation, the system becomes

$$\begin{aligned} 2x_1 - x_2 &= h \\ 0 &= k + 3h \end{aligned}$$

If $k + 3h$ is nonzero, the system has no solution. The system is consistent for any values of h and k that make $k + 3h = 0$.

1.2 Row Reduction and Echelon Forms

This section refines the method of Section 1.1 into a row reduction algorithm that will enable us to analyze any system of linear equations.[1] By using only the first part of the algorithm, we will be able to answer the fundamental existence and uniqueness questions posed in Section 1.1.

The algorithm applies to any matrix, whether or not the matrix is viewed as an augmented matrix for a linear system. So the first part of this section concerns an arbitrary rectangular matrix and begins by introducing two important classes of matrices that include the "triangular" matrices of Section 1.1. In the definitions that follow, a *nonzero* row or column in a matrix means a row or column that contains at least one nonzero entry; a **leading entry** of a row refers to the leftmost nonzero entry (in a nonzero row).

DEFINITION

A rectangular matrix is in **echelon form** (or **row echelon form**) if it has the following three properties:

1. All nonzero rows are above any rows of all zeros.
2. Each leading entry of a row is in a column to the right of the leading entry of the row above it.
3. All entries in a column below a leading entry are zeros.

If a matrix in echelon form satisfies the following additional conditions, then it is in **reduced echelon form** (or **reduced row echelon form**):

4. The leading entry in each nonzero row is 1.
5. Each leading 1 is the only nonzero entry in its column.

[1] The algorithm here is a variant of what is commonly called *Gaussian elimination*. A similar elimination method for linear systems was used by Chinese mathematicians in about 250 B.C. The process was unknown in Western culture until the nineteenth century, when a famous German mathematician, Carl Friedrich Gauss, discovered it. A German engineer, Wilhelm Jordan, popularized the algorithm in an 1888 text on geodesy.

An **echelon matrix** (respectively, **reduced echelon matrix**) is one that is in echelon form (respectively, reduced echelon form). Property 2 says that the leading entries form an *echelon* ("steplike") pattern that moves down and to the right through the matrix. Property 3 is a simple consequence of property 2, but we include it for emphasis.

The "triangular" matrices of Section 1.1, such as

$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & 5/2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 0 & 29 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

are in echelon form. In fact, the second matrix is in reduced echelon form. Here are additional examples.

EXAMPLE 1 The following matrices are in echelon form. The leading entries (■) may have any nonzero value; the starred entries ($*$) may have any value (including zero).

$$\begin{bmatrix} ■ & * & * & * \\ 0 & ■ & * & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & ■ & * & * & * & * & * & * & * & * \\ 0 & 0 & 0 & ■ & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & ■ & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & ■ & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & ■ & * \end{bmatrix}$$

The following matrices are in reduced echelon form because the leading entries are 1's, and there are 0's below *and above* each leading 1.

$$\begin{bmatrix} 1 & 0 & * & * \\ 0 & 1 & * & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & * & 0 & 0 & 0 & * & * & 0 & * \\ 0 & 0 & 0 & 1 & 0 & 0 & * & * & 0 & * \\ 0 & 0 & 0 & 0 & 1 & 0 & * & * & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 1 & * & * & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & * \end{bmatrix}$$

■

Any nonzero matrix may be **row reduced** (that is, transformed by elementary row operations) into more than one matrix in echelon form, using different sequences of row operations. However, the reduced echelon form one obtains from a matrix is unique. The following theorem is proved in Appendix A at the end of the text.

THEOREM 1

Uniqueness of the Reduced Echelon Form

Each matrix is row equivalent to one and only one reduced echelon matrix.

If a matrix A is row equivalent to an echelon matrix U, we call U **an echelon form** (or row echelon form) **of A**; if U is in reduced echelon form, we call U **the reduced echelon form of A**. [Most matrix programs and calculators with matrix capabilities use the abbreviation RREF for reduced (row) echelon form. Some use REF for (row) echelon form.]

Pivot Positions

When row operations on a matrix produce an echelon form, further row operations to obtain the reduced echelon form do not change the positions of the leading entries. Since the reduced echelon form is unique, *the leading entries are always in the same positions* *in any echelon form obtained from a given matrix.* These leading entries correspond to leading 1's in the reduced echelon form.

DEFINITION

A **pivot position** in a matrix A is a location in A that corresponds to a leading 1 in the reduced echelon form of A. A **pivot column** is a column of A that contains a pivot position.

In Example 1, the squares (■) identify the pivot positions. Many fundamental concepts in the first four chapters will be connected in one way or another with pivot positions in a matrix.

EXAMPLE 2 Row reduce the matrix A below to echelon form, and locate the pivot columns of A.

$$A = \begin{bmatrix} 0 & -3 & -6 & 4 & 9 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 1 & 4 & 5 & -9 & -7 \end{bmatrix}$$

SOLUTION Use the same basic strategy as in Section 1.1. The top of the leftmost nonzero column is the first pivot position. A nonzero entry, or *pivot*, must be placed in this position. A good choice is to interchange rows 1 and 4 (because the mental computations in the next step will not involve fractions).

Pivot

$$\begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix}$$

Pivot column

Create zeros below the pivot, 1, by adding multiples of the first row to the rows below, and obtain matrix (1) below. The pivot position in the second row must be as far left as possible—namely in the second column. Choose the 2 in this position as the next pivot.

Pivot

$$\begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 5 & 10 & -15 & -15 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix} \tag{1}$$

Next pivot column

Add $-5/2$ times row 2 to row 3, and add $3/2$ times row 2 to row 4.

$$\begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -5 & 0 \end{bmatrix} \tag{2}$$

The matrix in (2) is different from any encountered in Section 1.1. There is no way to create a leading entry in column 3! (We can't use row 1 or 2 because doing so would destroy the echelon arrangement of the leading entries already produced.) However, if we interchange rows 3 and 4, we can produce a leading entry in column 4.

Pivot

$$\begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \text{General form:} \begin{bmatrix} \blacksquare & * & * & * & * \\ 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & \blacksquare & * \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Pivot columns

The matrix is in echelon form and thus reveals that columns 1, 2, and 4 of A are pivot columns.

Pivot positions

$$A = \begin{bmatrix} 0 & -3 & -6 & 4 & 9 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 1 & 4 & 5 & -9 & -7 \end{bmatrix} \tag{3}$$

Pivot columns ■

A **pivot**, as illustrated in Example 2, is a nonzero number in a pivot position that is used as needed to create zeros via row operations. The pivots in Example 2 were 1, 2, and -5. Notice that these numbers are not the same as the actual elements of A in the highlighted pivot positions shown in (3).

With Example 2 as a guide, we are ready to describe an efficient procedure for transforming a matrix into an echelon or reduced echelon matrix. Careful study and mastery of this procedure now will pay rich dividends later in the course.

The Row Reduction Algorithm

The algorithm that follows consists of four steps, and it produces a matrix in echelon form. A fifth step produces a matrix in reduced echelon form. We illustrate the algorithm by an example.

EXAMPLE 3 Apply elementary row operations to transform the following matrix first into echelon form and then into reduced echelon form:

$$\begin{bmatrix} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{bmatrix}$$

SOLUTION

Step 1

Begin with the leftmost nonzero column. This is a pivot column. The pivot position is at the top.

$$\begin{bmatrix} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{bmatrix}$$

Pivot column

Step 2

Select a nonzero entry in the pivot column as a pivot. If necessary, interchange rows to move this entry into the pivot position.

Interchange rows 1 and 3. (We could have interchanged rows 1 and 2 instead.)

$$\begin{array}{l} \quad\text{Pivot} \\ \begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix} \end{array}$$

Step 3

Use row replacement operations to create zeros in all positions below the pivot.

As a preliminary step, we could divide the top row by the pivot, 3. But with two 3's in column 1, it is just as easy to add -1 times row 1 to row 2.

$$\begin{array}{l} \quad\text{Pivot} \\ \begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix} \end{array}$$

Step 4

Cover (or ignore) the row containing the pivot position and cover all rows, if any, above it. Apply steps 1–3 to the submatrix that remains. Repeat the process until there are no more nonzero rows to modify.

With row 1 covered, step 1 shows that column 2 is the next pivot column; for step 2, select as a pivot the "top" entry in that column.

$$\begin{array}{l} \quad\quad\text{Pivot} \\ \begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix} \\ \quad\quad\text{New pivot column} \end{array}$$

For step 3, we could insert an optional step of dividing the "top" row of the submatrix by the pivot, 2. Instead, we add $-3/2$ times the "top" row to the row below. This produces

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

When we cover the row containing the second pivot position for step 4, we are left with a new submatrix having only one row:

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

Pivot

Steps 1–3 require no work for this submatrix, and we have reached an echelon form of the full matrix. If we want the reduced echelon form, we perform one more step.

Step 5

Beginning with the rightmost pivot and working upward and to the left, create zeros above each pivot. If a pivot is not 1, make it 1 by a scaling operation.

The rightmost pivot is in row 3. Create zeros above it, adding suitable multiples of row 3 to rows 2 and 1.

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 0 & -9 \\ 0 & 2 & -4 & 4 & 0 & -14 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \begin{array}{l} \leftarrow \text{Row } 1 + (-6)\cdot \text{row } 3 \\ \leftarrow \text{Row } 2 + (-2)\cdot \text{row } 3 \\ \end{array}$$

The next pivot is in row 2. Scale this row, dividing by the pivot.

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 0 & -9 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \begin{array}{l} \\ \leftarrow \text{Row scaled by } \frac{1}{2} \\ \end{array}$$

Create a zero in column 2 by adding 9 times row 2 to row 1.

$$\begin{bmatrix} 3 & 0 & -6 & 9 & 0 & -72 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \begin{array}{l} \leftarrow \text{Row } 1 + (9)\cdot \text{row } 2 \\ \\ \end{array}$$

Finally, scale row 1, dividing by the pivot, 3.

$$\begin{bmatrix} 1 & 0 & -2 & 3 & 0 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \begin{array}{l} \leftarrow \text{Row scaled by } \frac{1}{3} \\ \\ \end{array}$$

This is the reduced echelon form of the original matrix. ■

The combination of steps 1–4 is called the **forward phase** of the row reduction algorithm. Step 5, which produces the unique reduced echelon form, is called the **backward phase**.

Numerical Note

In step 2 on page 41, a computer program usually selects as a pivot the entry in a column having the largest absolute value. This strategy, called **partial pivoting**, is used because it reduces roundoff errors in the calculations.

Solutions of Linear Systems

The row reduction algorithm leads directly to an explicit description of the solution set of a linear system when the algorithm is applied to the augmented matrix of the system.

Suppose, for example, that the augmented matrix of a linear system has been changed into the equivalent *reduced* echelon form

$$\begin{bmatrix} 1 & 0 & -5 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

There are three variables because the augmented matrix has four columns. The associated system of equations is

$$\begin{aligned} x_1 \qquad - 5x_3 &= 1 \\ x_2 + \;\; x_3 &= 4 \\ 0 &= 0 \end{aligned} \tag{4}$$

The variables x_1 and x_2 corresponding to pivot columns in the matrix are called **basic variables**.[2] The other variable, x_3, is called a **free variable**.

Whenever a system is consistent, as in (4), the solution set can be described explicitly by solving the *reduced* system of equations for the basic variables in terms of the free variables. This operation is possible because the reduced echelon form places each basic variable in one and only one equation. In (4), solve the first equation for x_1 and the second for x_2. (Ignore the third equation; it offers no restriction on the variables.)

$$\begin{cases} x_1 = 1 + 5x_3 \\ x_2 = 4 - x_3 \\ x_3 \text{ is free} \end{cases} \tag{5}$$

The statement "x_3 is free" means that you are free to choose any value for x_3. Once that is done, the formulas in (5) determine the values for x_1 and x_2. For instance, when $x_3 = 0$, the solution is $(1, 4, 0)$; when $x_3 = 1$, the solution is $(6, 3, 1)$. *Each different choice of x_3 determines a (different) solution of the system, and every solution of the system is determined by a choice of x_3.*

EXAMPLE 4 Find the general solution of the linear system whose augmented matrix has been reduced to

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

SOLUTION The matrix is in echelon form, but we want the reduced echelon form before solving for the basic variables. The row reduction is completed next. The symbol $\sim$ before a matrix indicates that the matrix is row equivalent to the preceding matrix.

$$\begin{bmatrix} 1 & 6 & 2 & -5 & -2 & -4 \\ 0 & 0 & 2 & -8 & -1 & 3 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & 6 & 2 & -5 & 0 & 10 \\ 0 & 0 & 2 & -8 & 0 & 10 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 6 & 2 & -5 & 0 & 10 \\ 0 & 0 & 1 & -4 & 0 & 5 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & 6 & 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & -4 & 0 & 5 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

---

[2] Some texts use the term *leading variables* because they correspond to the columns containing leading entries.

There are five variables because the augmented matrix has six columns. The associated system now is

$$\begin{aligned} x_1 + 6x_2 \quad\;\; + 3x_4 \quad\;\; &= 0 \\ x_3 - 4x_4 \quad\;\; &= 5 \\ x_5 &= 7 \end{aligned} \tag{6}$$

The pivot columns of the matrix are 1, 3, and 5, so the basic variables are x_1, x_3, and x_5. The remaining variables, x_2 and x_4, must be free. Solve for the basic variables to obtain the general solution:

$$\begin{cases} x_1 = -6x_2 - 3x_4 \\ x_2 \text{ is free} \\ x_3 = 5 + 4x_4 \\ x_4 \text{ is free} \\ x_5 = 7 \end{cases} \tag{7}$$

Note that the value of x_5 is already fixed by the third equation in system (6). ■

Parametric Descriptions of Solution Sets

The descriptions in (5) and (7) are *parametric descriptions* of solution sets in which the free variables act as parameters. *Solving a system* amounts to finding a parametric description of the solution set or determining that the solution set is empty.

Whenever a system is consistent and has free variables, the solution set has many parametric descriptions. For instance, in system (4), we may add 5 times equation 2 to equation 1 and obtain the equivalent system

$$\begin{aligned} x_1 + 5x_2 \quad\quad\; &= 21 \\ x_2 + x_3 &= \;\;4 \end{aligned}$$

We could treat x_2 as a parameter and solve for x_1 and x_3 in terms of x_2, and we would have an accurate description of the solution set. However, to be consistent, we make the (arbitrary) convention of always using the free variables as the parameters for describing a solution set. (The answer section at the end of the text also reflects this convention.)

Whenever a system is inconsistent, the solution set is empty, even when the system has free variables. In this case, the solution set has *no* parametric representation.

Back-Substitution

Consider the following system, whose augmented matrix is in echelon form but is *not* in reduced echelon form:

$$\begin{aligned} x_1 - 7x_2 + 2x_3 - 5x_4 + 8x_5 &= \;\;10 \\ x_2 - 3x_3 + 3x_4 + \;\;x_5 &= -5 \\ x_4 - \;\;x_5 &= \;\;\;4 \end{aligned}$$

A computer program would solve this system by back-substitution, rather than by computing the reduced echelon form. That is, the program would solve equation 3 for x_4 in terms of x_5 and substitute the expression for x_4 into equation 2, solve equation 2 for x_2, and then substitute the expressions for x_2 and x_4 into equation 1 and solve for x_1.

Our matrix format for the backward phase of row reduction, which produces the reduced echelon form, has the same number of arithmetic operations as back-substitution. But the discipline of the matrix format substantially reduces the likelihood of errors during hand computations. The best strategy is to use only the *reduced* echelon form to solve a system! The *Study Guide* that accompanies this text offers several helpful suggestions for performing row operations accurately and rapidly.

Numerical Note

In general, the forward phase of row reduction takes much longer than the backward phase. An algorithm for solving a system is usually measured in flops (or floating point operations). A **flop** is one arithmetic operation $(+, -, *, /)$ on two real floating point numbers.[3] For an $n \times (n+1)$ matrix, the reduction to echelon form can take $2n^3/3 + n^2/2 - 7n/6$ flops (which is approximately $2n^3/3$ flops when n is moderately large—say, $n \geq 30$). In contrast, further reduction to reduced echelon form needs at most n^2 flops.

Existence and Uniqueness Questions

Although a nonreduced echelon form is a poor tool for solving a system, this form is just the right device for answering two fundamental questions posed in Section 1.1.

EXAMPLE 5 Determine the existence and uniqueness of the solutions to the system

$$\begin{aligned} 3x_2 - 6x_3 + 6x_4 + 4x_5 &= -5 \\ 3x_1 - 7x_2 + 8x_3 - 5x_4 + 8x_5 &= 9 \\ 3x_1 - 9x_2 + 12x_3 - 9x_4 + 6x_5 &= 15 \end{aligned}$$

SOLUTION The augmented matrix of this system was row reduced in Example 3 to

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \tag{8}$$

The basic variables are x_1, x_2, and x_5; the free variables are x_3 and x_4. There is no equation such as $0 = 1$ that would indicate an inconsistent system, so we could use back-substitution to find a solution. But the *existence* of a solution is already clear in (8). Also, the solution is *not unique* because there are free variables. Each different choice of x_3 and x_4 determines a different solution. Thus the system has infinitely many solutions. ■

When a system is in echelon form and contains no equation of the form $0 = b$, with b nonzero, every nonzero equation contains a basic variable with a nonzero coefficient. Either the basic variables are completely determined (with no free variables) or at least one of the basic variables may be expressed in terms of one or more free variables. In the former case, there is a unique solution; in the latter case, there are infinitely many solutions (one for each choice of values for the free variables).

These remarks justify the following theorem.

---

[3] Traditionally, a *flop* was only a multiplication or division because addition and subtraction took much less time and could be ignored. The definition of *flop* given here is preferred now, as a result of advances in computer architecture. See Golub and Van Loan, *Matrix Computations*, 2nd ed. (Baltimore: The Johns Hopkins Press, 1989), pp. 19–20.

THEOREM 2

Existence and Uniqueness Theorem

A linear system is consistent if and only if the rightmost column of the augmented matrix is *not* a pivot column—that is, if and only if an echelon form of the augmented matrix has *no* row of the form

$$[\,0 \quad \cdots \quad 0 \quad b\,] \qquad \text{with } b \text{ nonzero}$$

If a linear system is consistent, then the solution set contains either (i) a unique solution, when there are no free variables, or (ii) infinitely many solutions, when there is at least one free variable.

The following procedure outlines how to find and describe all solutions of a linear system.

USING ROW REDUCTION TO SOLVE A LINEAR SYSTEM

1. Write the augmented matrix of the system.
2. Use the row reduction algorithm to obtain an equivalent augmented matrix in echelon form. Decide whether the system is consistent. If there is no solution, stop; otherwise, go to the next step.
3. Continue row reduction to obtain the reduced echelon form.
4. Write the system of equations corresponding to the matrix obtained in step 3.
5. Rewrite each nonzero equation from step 4 so that its one basic variable is expressed in terms of any free variables appearing in the equation.

Reasonable Answers

Remember that each augmented matrix corresponds to a system of equations. If you row reduce the augmented matrix $\begin{bmatrix} 1 & -2 & 1 & 2 \\ 1 & -1 & 2 & 5 \\ 0 & 1 & 1 & 3 \end{bmatrix}$ to get the matrix $\begin{bmatrix} 1 & 0 & 3 & 8 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, the solution set is

$$\begin{cases} x_1 = 8 - 3x_3 \\ x_2 = 3 - x_3 \\ x_3 \text{ is free} \end{cases}$$

The system of equations corresponding to the original augmented matrix is

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

You can now check whether your solution is correct by substituting it into the original equations. Notice that you can just leave the free variables in the solution.

$$\begin{aligned}
(8 - 3x_3) - 2(3 - x_3) + (x_3) &= 8 - 3x_3 - 6 + 2x_3 + x_3 = 2\\
(8 - 3x_3) - (3 - x_3) + 2(x_3) &= 8 - 3x_3 - 3 + x_3 + 2x_3 = 5\\
(3 - x_3) + (x_3) &= 3 - x_3 + x_3 = 3
\end{aligned}$$

You can now be confident you have a correct solution to the system of equations represented by the augmented matrix.

Practice Problems

1. Find the general solution of the linear system whose augmented matrix is

$$\begin{bmatrix} 1 & -3 & -5 & 0 \\ 0 & 1 & -1 & -1 \end{bmatrix}$$

2. Find the general solution of the system

$$\begin{aligned}
x_1 - 2x_2 - x_3 + 3x_4 &= 0\\
-2x_1 + 4x_2 + 5x_3 - 5x_4 &= 3\\
3x_1 - 6x_2 - 6x_3 + 8x_4 &= 2
\end{aligned}$$

3. Suppose a 4×7 coefficient matrix for a system of equations has 4 pivots. Is the system consistent? If the system is consistent, how many solutions are there?

1.2 Exercises

In Exercises 1 and 2, determine which matrices are in reduced echelon form and which others are only in echelon form.

1. a. $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ b. $\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

c. $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ d. $\begin{bmatrix} 1 & 1 & 0 & 1 & 1 \\ 0 & 2 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$

2. a. $\begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ b. $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$

c. $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ d. $\begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 2 & 2 & 2 \\ 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

Row reduce the matrices in Exercises 3 and 4 to reduced echelon form. Circle the pivot positions in the final matrix and in the original matrix, and list the pivot columns.

3. $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 5 & 6 & 7 \\ 6 & 7 & 8 & 9 \end{bmatrix}$ **4.** $\begin{bmatrix} 1 & 3 & 5 & 7 \\ 3 & 5 & 7 & 9 \\ 5 & 7 & 9 & 1 \end{bmatrix}$

5. Describe the possible echelon forms of a nonzero 2×2 matrix. Use the symbols ■, $*$, and 0, as in the first part of Example 1.

6. Repeat Exercise 5 for a nonzero 3×2 matrix.

Find the general solutions of the systems whose augmented matrices are given in Exercises 7–14.

7. $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 8 & 9 & 4 \end{bmatrix}$ **8.** $\begin{bmatrix} 1 & 4 & 0 & 7 \\ 2 & 7 & 0 & 11 \end{bmatrix}$

9. $\begin{bmatrix} 0 & 1 & -6 & 5 \\ 1 & -2 & 7 & -4 \end{bmatrix}$ **10.** $\begin{bmatrix} 1 & -2 & -1 & 3 \\ 3 & -6 & -2 & 2 \end{bmatrix}$

11. $\begin{bmatrix} 3 & -4 & 2 & 0 \\ -9 & 12 & -6 & 0 \\ -6 & 8 & -4 & 0 \end{bmatrix}$ **12.** $\begin{bmatrix} 1 & -7 & 0 & 6 & 5 \\ 0 & 0 & 1 & -2 & -3 \\ -1 & 7 & -4 & 2 & 7 \end{bmatrix}$

13. $\begin{bmatrix} 1 & -3 & 0 & -1 & 0 & -2 \\ 0 & 1 & 0 & 0 & -4 & 1 \\ 0 & 0 & 0 & 1 & 9 & -4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

14. $\begin{bmatrix} 1 & 2 & -5 & -4 & 0 & -5 \\ 0 & 1 & -6 & -4 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

You may find it helpful to review the information in the Reasonable Answers box from this section before answering Exercises 15–18.

15. Write down the equations corresponding to the augmented matrix in Exercise 9 and verify your answer to Exercise 9 is correct by substituting the solutions you obtained back into the original equations.

16. Write down the equations corresponding to the augmented matrix in Exercise 10 and verify your answer to Exercise 10 is correct by substituting the solutions you obtained back into the original equations.

17. Write down the equations corresponding to the augmented matrix in Exercise 11 and verify your answer to Exercise 11 is correct by substituting the solutions you obtained back into the original equations.

18. Write down the equations corresponding to the augmented matrix in Exercise 12 and verify your answer to Exercise 12 is correct by substituting the solutions you obtained back into the original equations.

Exercises 19 and 20 use the notation of Example 1 for matrices in echelon form. Suppose each matrix represents the augmented matrix for a system of linear equations. In each case, determine if the system is consistent. If the system is consistent, determine if the solution is unique.

19. a. $\begin{bmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & \blacksquare & 0 \end{bmatrix}$

b. $\begin{bmatrix} 0 & \blacksquare & * & * & * \\ 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 & \blacksquare \end{bmatrix}$

20. a. $\begin{bmatrix} \blacksquare & * & * \\ 0 & \blacksquare & * \\ 0 & 0 & 0 \end{bmatrix}$

b. $\begin{bmatrix} \blacksquare & * & * & * & * \\ 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & \blacksquare & * \end{bmatrix}$

In Exercises 21 and 22, determine the value(s) of h such that the matrix is the augmented matrix of a consistent linear system.

21. $\begin{bmatrix} 2 & 3 & h \\ 4 & 6 & 7 \end{bmatrix}$ **22.** $\begin{bmatrix} 1 & -4 & -3 \\ 6 & h & -9 \end{bmatrix}$

In Exercises 23 and 24, choose h and k such that the system has (a) no solution, (b) a unique solution, and (c) many solutions. Give separate answers for each part.

23. $\begin{aligned} x_1 + hx_2 &= 2 \\ 4x_1 + 8x_2 &= k \end{aligned}$ **24.** $\begin{aligned} x_1 + 4x_2 &= 5 \\ 2x_1 + hx_2 &= k \end{aligned}$

In Exercises 25–34, mark each statement True or False **(T/F)**. Justify each answer.[4]

25. **(T/F)** In some cases, a matrix may be row reduced to more than one matrix in reduced echelon form, using different sequences of row operations.

26. **(T/F)** The echelon form of a matrix is unique.

27. **(T/F)** The row reduction algorithm applies only to augmented matrices for a linear system.

28. **(T/F)** The pivot positions in a matrix depend on whether row interchanges are used in the row reduction process.

29. **(T/F)** A basic variable in a linear system is a variable that corresponds to a pivot column in the coefficient matrix.

30. **(T/F)** Reducing a matrix to echelon form is called the *forward phase* of the row reduction process.

31. **(T/F)** Finding a parametric description of the solution set of a linear system is the same as *solving* the system.

32. **(T/F)** Whenever a system has free variables, the solution set contains a unique solution.

33. **(T/F)** If one row in an echelon form of an augmented matrix is $[0 \quad 0 \quad 0 \quad 0 \quad 5]$, then the associated linear system is inconsistent.

34. **(T/F)** A general solution of a system is an explicit description of all solutions of the system.

35. Suppose a 3×5 *coefficient* matrix for a system has three pivot columns. Is the system consistent? Why or why not?

36. Suppose a system of linear equations has a 3×5 *augmented* matrix whose fifth column is a pivot column. Is the system consistent? Why (or why not)?

37. Suppose the coefficient matrix of a system of linear equations has a pivot position in every row. Explain why the system is consistent.

38. Suppose the coefficient matrix of a linear system of three equations in three variables has a pivot in each column. Explain why the system has a unique solution.

39. Restate the last sentence in Theorem 2 using the concept of pivot columns: "If a linear system is consistent, then the solution is unique if and only if ________."

40. What would you have to know about the pivot columns in an augmented matrix in order to know that the linear system is consistent and has a unique solution?

41. A system of linear equations with fewer equations than unknowns is sometimes called an *underdetermined system.*

[4] True/false questions of this type will appear in many sections. Methods for justifying your answers were described before the True or False exercises in Section 1.1.

Suppose that such a system happens to be consistent. Explain why there must be an infinite number of solutions.

42. Give an example of an inconsistent underdetermined system of two equations in three unknowns.

43. A system of linear equations with more equations than unknowns is sometimes called an *overdetermined system*. Can such a system be consistent? Illustrate your answer with a specific system of three equations in two unknowns.

44. Suppose an $n \times (n+1)$ matrix is row reduced to reduced echelon form. Approximately what fraction of the total number of operations (flops) is involved in the backward phase of the reduction when $n = 30$? when $n = 300$?

Suppose experimental data are represented by a set of points in the plane. An **interpolating polynomial** for the data is a polynomial whose graph passes through every point. In scientific work, such a polynomial can be used, for example, to estimate values between the known data points. Another use is to create curves for graphical images on a computer screen. One method for finding an interpolating polynomial is to solve a system of linear equations.

45. Find the interpolating polynomial $p(t) = a_0 + a_1 t + a_2 t^2$ for the data $(1, 11)$, $(2, 16)$, $(3, 19)$. That is, find a_0, a_1, and a_2 such that

$$\begin{aligned} a_0 + a_1(1) + a_2(1)^2 &= 11 \\ a_0 + a_1(2) + a_2(2)^2 &= 16 \\ a_0 + a_1(3) + a_2(3)^2 &= 19 \end{aligned}$$

T **46.** In a wind tunnel experiment, the force on a projectile due to air resistance was measured at different velocities:

Velocity (100 ft/sec)	0	2	4	6	8	10
Force (100 lb)	0	2.90	14.8	39.6	74.3	119

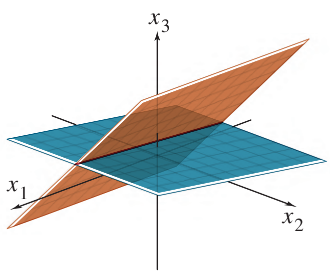
Find an interpolating polynomial for these data and estimate the force on the projectile when the projectile is traveling at 750 ft/sec. Use $p(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + a_5 t^5$. What happens if you try to use a polynomial of degree less than 5? (Try a cubic polynomial, for instance.)[5]

[5] Exercises marked with the symbol **T** are designed to be worked with the aid of a " Matrix program" (a computer program, such as MATLAB, Maple, Mathematica, MathCad, Octave, or Derive, or a programmable calculator with matrix capabilities, such as those manufactured by Texas Instruments or Hewlett-Packard).

Solutions to Practice Problems

1. The reduced echelon form of the augmented matrix and the corresponding system are

$$\begin{bmatrix} 1 & 0 & -8 & -3 \\ 0 & 1 & -1 & -1 \end{bmatrix} \quad \text{and} \quad \begin{aligned} x_1 \qquad - 8x_3 &= -3 \\ x_2 - \quad x_3 &= -1 \end{aligned}$$



The general solution of the system of equations is the line of intersection of the two planes.

The basic variables are x_1 and x_2, and the general solution is

$$\begin{cases} x_1 = -3 + 8x_3 \\ x_2 = -1 + x_3 \\ x_3 \text{ is free} \end{cases}$$

Note: It is essential that the general solution describe each variable, with any parameters clearly identified. The following statement does *not* describe the solution:

$$\begin{cases} x_1 = -3 + 8x_3 \\ x_2 = -1 + x_3 \\ x_3 = 1 + x_2 \quad \text{Incorrect solution} \end{cases}$$

This description implies that x_2 and x_3 are *both* free, which certainly is not the case.

2. Row reduce the system's augmented matrix:

$$\begin{bmatrix} 1 & -2 & -1 & 3 & 0 \\ -2 & 4 & 5 & -5 & 3 \\ 3 & -6 & -6 & 8 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & -1 & 3 & 0 \\ 0 & 0 & 3 & 1 & 3 \\ 0 & 0 & -3 & -1 & 2 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & -2 & -1 & 3 & 0 \\ 0 & 0 & 3 & 1 & 3 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix}$$

Solutions to Practice Problems (Continued)

This echelon matrix shows that the system is *inconsistent*, because its rightmost column is a pivot column; the third row corresponds to the equation 0 = 5. There is no need to perform any more row operations. Note that the presence of the free variables in this problem is irrelevant because the system is inconsistent.

3. Since the coefficient matrix has four pivots, there is a pivot in every row of the coefficient matrix. This means that when the coefficient matrix is row reduced, it will *not* have a row of zeros, thus the corresponding row reduced augmented matrix can never have a row of the form $[0\ \ 0\ \ \cdots\ \ 0\ \ b]$, where b is a nonzero number. By Theorem 2, the system is consistent. Moreover, since there are seven columns in the coefficient matrix and only four pivot columns, there will be three free variables resulting in infinitely many solutions.

1.3 Vector Equations

Important properties of linear systems can be described with the concept and notation of vectors. This section connects equations involving vectors to ordinary systems of equations. The term *vector* appears in a variety of mathematical and physical contexts, which we will discuss in Chapter 4, "Vector Spaces." Until then, *vector* will mean an *ordered list of numbers*. This simple idea enables us to get to interesting and important applications as quickly as possible.

Vectors in $\mathbb{R}^2$

A matrix with only one column is called a **column vector** or simply a **vector**. Examples of vectors with two entries are

$$\mathbf{u} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} .2 \\ .3 \end{bmatrix}, \qquad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

where w_1 and w_2 are any real numbers. The set of all vectors with two entries is denoted by $\mathbb{R}^2$ (read "r-two"). The $\mathbb{R}$ stands for the real numbers that appear as entries in the vectors, and the exponent 2 indicates that each vector contains two entries.[1]

Two vectors in $\mathbb{R}^2$ are **equal** if and only if their corresponding entries are equal. Thus $\begin{bmatrix} 4 \\ 7 \end{bmatrix}$ and $\begin{bmatrix} 7 \\ 4 \end{bmatrix}$ are *not* equal, because vectors in $\mathbb{R}^2$ are *ordered pairs* of real numbers.

Given two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^2$, their **sum** is the vector $\mathbf{u} + \mathbf{v}$ obtained by adding corresponding entries of $\mathbf{u}$ and $\mathbf{v}$. For example,

$$\begin{bmatrix} 1 \\ -2 \end{bmatrix} + \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 1+2 \\ -2+5 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

Given a vector $\mathbf{u}$ and a real number c, the **scalar multiple** of $\mathbf{u}$ by c is the vector $c\mathbf{u}$ obtained by multiplying each entry in $\mathbf{u}$ by c. For instance,

$$\text{if} \quad \mathbf{u} = \begin{bmatrix} 3 \\ -1 \end{bmatrix} \quad \text{and} \quad c = 5, \qquad \text{then} \quad c\mathbf{u} = 5\begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 15 \\ -5 \end{bmatrix}$$

[1] Most of the text concerns vectors and matrices that have only real entries. However, all definitions and theorems in Chapters 1–5, and in most of the rest of the text, remain valid if the entries are complex numbers. Complex vectors and matrices arise naturally, for example, in electrical engineering and physics.

The number c in $c\mathbf{u}$ is called a **scalar**; it is written in lightface type to distinguish it from the boldface vector $\mathbf{u}$.

The operations of scalar multiplication and vector addition can be combined, as in the following example.

EXAMPLE 1 Given $\mathbf{u} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 \\ -5 \end{bmatrix}$, find $4\mathbf{u}$, $(-3)\mathbf{v}$, and $4\mathbf{u} + (-3)\mathbf{v}$.

SOLUTION

$$4\mathbf{u} = \begin{bmatrix} 4 \\ -8 \end{bmatrix}, \qquad (-3)\mathbf{v} = \begin{bmatrix} -6 \\ 15 \end{bmatrix}$$

and

$$4\mathbf{u} + (-3)\mathbf{v} = \begin{bmatrix} 4 \\ -8 \end{bmatrix} + \begin{bmatrix} -6 \\ 15 \end{bmatrix} = \begin{bmatrix} -2 \\ 7 \end{bmatrix}$$ ■

Sometimes, for convenience (and also to save space), this text may write a column vector such as $\begin{bmatrix} 3 \\ -1 \end{bmatrix}$ in the form $(3, -1)$. In this case, the parentheses and the comma distinguish the vector $(3, -1)$ from the 1×2 row matrix $\begin{bmatrix} 3 & -1 \end{bmatrix}$, written with brackets and no comma. Thus

$$\begin{bmatrix} 3 \\ -1 \end{bmatrix} \neq \begin{bmatrix} 3 & -1 \end{bmatrix}$$

because the matrices have different shapes, even though they have the same entries.

Geometric Descriptions of $\mathbb{R}^2$

Consider a rectangular coordinate system in the plane. Because each point in the plane is determined by an ordered pair of numbers, *we can identify a geometric point* (a, b) *with the column vector* $\begin{bmatrix} a \\ b \end{bmatrix}$. So we may regard $\mathbb{R}^2$ as the set of all points in the plane. See Figure 1.

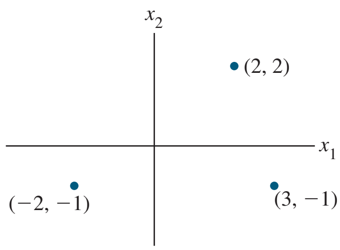


FIGURE 1 Vectors as points.

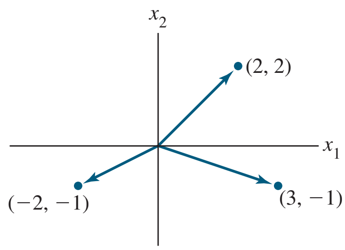


FIGURE 2 Vectors with arrows.

The geometric visualization of a vector such as $\begin{bmatrix} 3 \\ -1 \end{bmatrix}$ is often aided by including an arrow (directed line segment) from the origin $(0, 0)$ to the point $(3, -1)$, as in Figure 2. In this case, the individual points along the arrow itself have no special significance.[2]

The sum of two vectors has a useful geometric representation. The following rule can be verified by analytic geometry.

[2] In physics, arrows can represent forces and usually are free to move about in space. This interpretation of vectors will be discussed in Section 4.1.

Parallelogram Rule for Addition

If **u** and **v** in $\mathbb{R}^2$ are represented as points in the plane, then $\mathbf{u} + \mathbf{v}$ corresponds to the fourth vertex of the parallelogram whose other vertices are **u**, **0**, and **v**. See Figure 3.

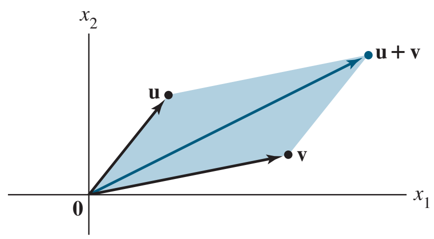


FIGURE 3 The parallelogram rule.

EXAMPLE 2 The vectors $\mathbf{u} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -6 \\ 1 \end{bmatrix}$, and $\mathbf{u} + \mathbf{v} = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$ are displayed in Figure 4. ■

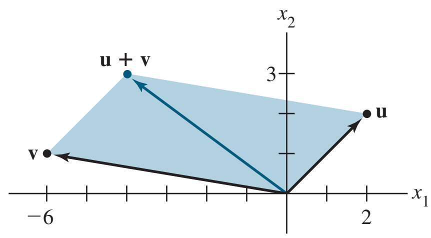
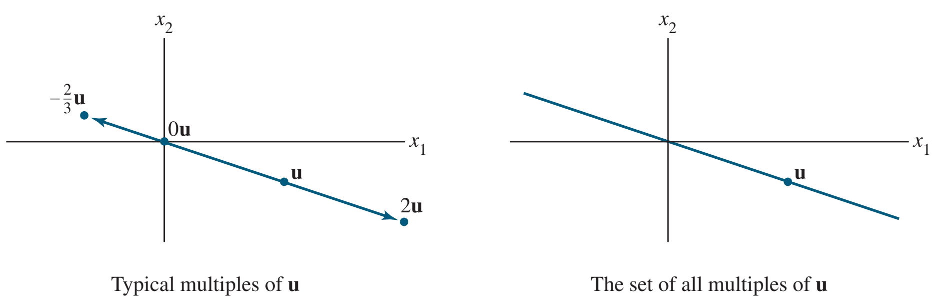


FIGURE 4

The next example illustrates the fact that the set of all scalar multiples of one fixed nonzero vector is a line through the origin, $(0, 0)$.

EXAMPLE 3 Let $\mathbf{u} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$. Display the vectors **u**, 2**u**, and $-\frac{2}{3}\mathbf{u}$ on a graph.

SOLUTION See Figure 5, where **u**, $2\mathbf{u} = \begin{bmatrix} 6 \\ -2 \end{bmatrix}$, and $-\frac{2}{3}\mathbf{u} = \begin{bmatrix} -2 \\ 2/3 \end{bmatrix}$ are displayed. The arrow for 2**u** is twice as long as the arrow for **u**, and the arrows point in the same direction. The arrow for $-\frac{2}{3}\mathbf{u}$ is two-thirds the length of the arrow for **u**, and the arrows point in opposite directions. In general, the length of the arrow for $c\mathbf{u}$ is $|c|$ times the length of the arrow for **u**. [Recall that the length of the line segment from $(0, 0)$ to (a, b) is $\sqrt{a^2 + b^2}$.



■

FIGURE 5

Vectors in $\mathbb{R}^3$

Vectors in $\mathbb{R}^3$ are 3×1 column matrices with three entries. They are represented geometrically by points in a three-dimensional coordinate space, with arrows from the origin sometimes included for visual clarity. The vectors $\mathbf{a} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$ and $2\mathbf{a}$ are displayed in Figure 6.

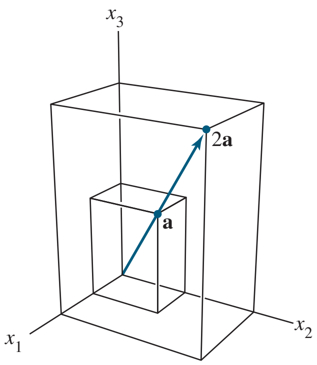


FIGURE 6
Scalar multiples.

Vectors in $\mathbb{R}^n$

If n is a positive integer, $\mathbb{R}^n$ (read "r-n") denotes the collection of all lists (or *ordered n-tuples*) of n real numbers, usually written as $n \times 1$ column matrices, such as

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

The vector whose entries are all zero is called the **zero vector** and is denoted by $\mathbf{0}$. (The number of entries in $\mathbf{0}$ will be clear from the context.)

Equality of vectors in $\mathbb{R}^n$ and the operations of scalar multiplication and vector addition in $\mathbb{R}^n$ are defined entry by entry just as in $\mathbb{R}^2$. These operations on vectors have the following properties, which can be verified directly from the corresponding properties for real numbers. See Practice Problem 1 and Exercises 41 and 42 at the end of this section.

Algebraic Properties of $\mathbb{R}^n$

For all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in $\mathbb{R}^n$ and all scalars c and d:

(i) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

(ii) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

(iii) $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$

(iv) $\mathbf{u} + (-\mathbf{u}) = -\mathbf{u} + \mathbf{u} = \mathbf{0}$,
where $-\mathbf{u}$ denotes $(-1)\mathbf{u}$

(v) $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$

(vi) $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$

(vii) $c(d\mathbf{u}) = (cd)\mathbf{u}$

(viii) $1\mathbf{u} = \mathbf{u}$

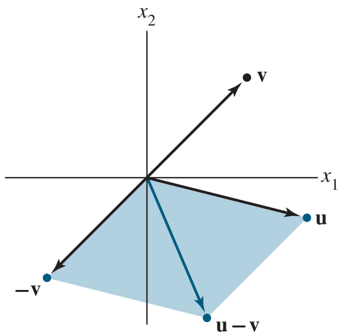


FIGURE 7
Vector subtraction.

For simplicity of notation, a vector such as $\mathbf{u} + (-1)\mathbf{v}$ is often written as $\mathbf{u} - \mathbf{v}$. Figure 7 shows $\mathbf{u} - \mathbf{v}$ as the sum of $\mathbf{u}$ and $-\mathbf{v}$.

Linear Combinations

Given vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ in $\mathbb{R}^n$ and given scalars $c_1, c_2, \dots, c_p$, the vector $\mathbf{y}$ defined by

$$\mathbf{y} = c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$$

is called a **linear combination** of $\mathbf{v}_1, \dots, \mathbf{v}_p$ with **weights** $c_1, \dots, c_p$. Algebraic Property (ii) above permits us to omit parentheses when forming such a linear combination. The weights in a linear combination can be any real numbers, including zero. For example, some linear combinations of vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are

$$\sqrt{3}\,\mathbf{v}_1 + \mathbf{v}_2, \quad \tfrac{1}{2}\mathbf{v}_1 \ (= \tfrac{1}{2}\mathbf{v}_1 + 0\mathbf{v}_2), \quad \text{and} \quad \mathbf{0} \ (= 0\mathbf{v}_1 + 0\mathbf{v}_2)$$

EXAMPLE 4 Figure 8 identifies selected linear combinations of $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. (Note that sets of parallel grid lines are drawn through integer multiples of $\mathbf{v}_1$ and $\mathbf{v}_2$.) Estimate the linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$ that generate the vectors $\mathbf{u}$ and $\mathbf{w}$.

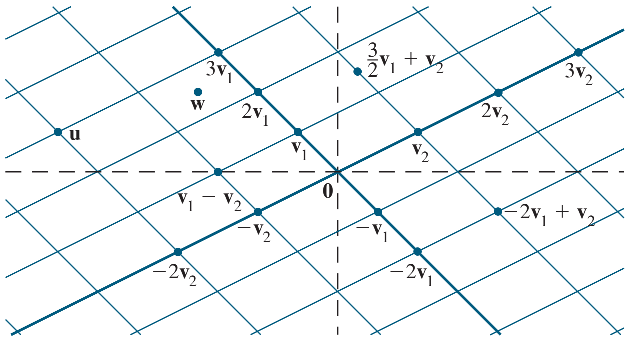


FIGURE 8 Linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$.

SOLUTION The parallelogram rule shows that $\mathbf{u}$ is the sum of $3\mathbf{v}_1$ and $-2\mathbf{v}_2$; that is,

$$\mathbf{u} = 3\mathbf{v}_1 - 2\mathbf{v}_2$$

This expression for $\mathbf{u}$ can be interpreted as instructions for traveling from the origin to $\mathbf{u}$ along two straight paths. First, travel 3 units in the $\mathbf{v}_1$ direction to $3\mathbf{v}_1$, and then travel -2 units in the $\mathbf{v}_2$ direction (parallel to the line through $\mathbf{v}_2$ and $\mathbf{0}$). Next, although the vector $\mathbf{w}$ is not on a grid line, $\mathbf{w}$ appears to be about halfway between two pairs of grid lines, at the vertex of a parallelogram determined by $(5/2)\mathbf{v}_1$ and $(-1/2)\mathbf{v}_2$. (See Figure 9.) Thus a reasonable estimate for $\mathbf{w}$ is

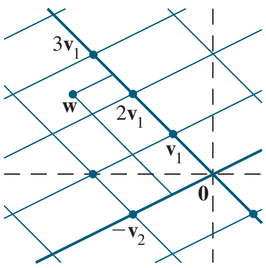


FIGURE 9

$$\mathbf{w} = \tfrac{5}{2}\mathbf{v}_1 - \tfrac{1}{2}\mathbf{v}_2$$ ■

The next example connects a problem about linear combinations to the fundamental existence question studied in Sections 1.1 and 1.2.

EXAMPLE 5 Let $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$. Determine whether $\mathbf{b}$ can be generated (or written) as a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$. That is, determine whether weights x_1 and x_2 exist such that

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{b} \qquad (1)$$

If vector equation (1) has a solution, find it.

SOLUTION Use the definitions of scalar multiplication and vector addition to rewrite the vector equation

$$x_1 \underset{\mathbf{a}_1}{\underset{\uparrow}{\begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}}} + x_2 \underset{\mathbf{a}_2}{\underset{\uparrow}{\begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}}} = \underset{\mathbf{b}}{\underset{\uparrow}{\begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}}}$$

which is the same as

$$\begin{bmatrix} x_1 \\ -2x_1 \\ -5x_1 \end{bmatrix} + \begin{bmatrix} 2x_2 \\ 5x_2 \\ 6x_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$$

and

$$\begin{bmatrix} x_1 + 2x_2 \\ -2x_1 + 5x_2 \\ -5x_1 + 6x_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix} \tag{2}$$

The vectors on the left and right sides of (2) are equal if and only if their corresponding entries are both equal. That is, x_1 and x_2 make the vector equation (1) true if and only if x_1 and x_2 satisfy the system

$$\begin{aligned} x_1 + 2x_2 &= 7 \\ -2x_1 + 5x_2 &= 4 \\ -5x_1 + 6x_2 &= -3 \end{aligned} \tag{3}$$

To solve this system, row reduce the augmented matrix of the system as follows:[3]

$$\begin{bmatrix} 1 & 2 & 7 \\ -2 & 5 & 4 \\ -5 & 6 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 7 \\ 0 & 9 & 18 \\ 0 & 16 & 32 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 7 \\ 0 & 1 & 2 \\ 0 & 16 & 32 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

The solution of (3) is $x_1 = 3$ and $x_2 = 2$. Hence $\mathbf{b}$ is a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$, with weights $x_1 = 3$ and $x_2 = 2$. That is,

$$3\begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix} + 2\begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$$ ■

Observe in Example 5 that the original vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{b}$ are the columns of the augmented matrix that we row reduced:

$$\begin{array}{c} \begin{bmatrix} 1 & 2 & 7 \\ -2 & 5 & 4 \\ -5 & 6 & -3 \end{bmatrix} \\ \begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{b} \end{array} \end{array}$$

For brevity, write this matrix in a way that identifies its columns—namely

$$[\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{b}\,] \tag{4}$$

It is clear how to write this augmented matrix immediately from vector equation (1), without going through the intermediate steps of Example 5. Take the vectors in the order in which they appear in (1) and put them into the columns of a matrix as in (4).

The discussion above is easily modified to establish the following fundamental fact.

---

[3] The symbol $\sim$ between matrices denotes row equivalence (Section 1.2).

A vector equation

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$$

has the same solution set as the linear system whose augmented matrix is

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n & \mathbf{b} \end{bmatrix} \tag{5}$$

In particular, $\mathbf{b}$ can be generated by a linear combination of $\mathbf{a}_1, \dots, \mathbf{a}_n$ if and only if there exists a solution to the linear system corresponding to the matrix (5).

One of the key ideas in linear algebra is to study the set of all vectors that can be generated or written as a linear combination of a fixed set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ of vectors.

DEFINITION

If $\mathbf{v}_1, \dots, \mathbf{v}_p$ are in $\mathbb{R}^n$, then the set of all linear combinations of $\mathbf{v}_1, \dots, \mathbf{v}_p$ is denoted by Span $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ and is called the **subset of $\mathbb{R}^n$ spanned** (or **generated**) **by $\mathbf{v}_1, \dots, \mathbf{v}_p$**. That is, Span $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is the collection of all vectors that can be written in the form

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p$$

with $c_1, \dots, c_p$ scalars.

Asking whether a vector $\mathbf{b}$ is in Span $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ amounts to asking whether the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_p\mathbf{v}_p = \mathbf{b}$$

has a solution, or, equivalently, asking whether the linear system with augmented matrix $[\,\mathbf{v}_1 \ \cdots \ \mathbf{v}_p \ \ \mathbf{b}\,]$ has a solution.

Note that Span $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ contains every scalar multiple of $\mathbf{v}_1$ (for example), since $c\mathbf{v}_1 = c\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_p$. In particular, the zero vector must be in Span $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$.

A Geometric Description of Span{v} and Span{u, v}

Let $\mathbf{v}$ be a nonzero vector in $\mathbb{R}^3$. Then Span $\{\mathbf{v}\}$ is the set of all scalar multiples of $\mathbf{v}$, which is the set of points on the line in $\mathbb{R}^3$ through $\mathbf{v}$ and $\mathbf{0}$. See Figure 10.

If $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors in $\mathbb{R}^3$, with $\mathbf{v}$ not a multiple of $\mathbf{u}$, then Span $\{\mathbf{u}, \mathbf{v}\}$ is the plane in $\mathbb{R}^3$ that contains $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{0}$. In particular, Span $\{\mathbf{u}, \mathbf{v}\}$ contains the line in $\mathbb{R}^3$ through $\mathbf{u}$ and $\mathbf{0}$ and the line through $\mathbf{v}$ and $\mathbf{0}$. See Figure 11.

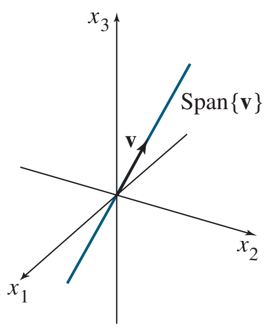


FIGURE 10 Span $\{\mathbf{v}\}$ as a line through the origin.

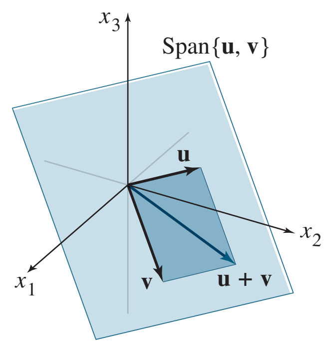


FIGURE 11 Span $\{\mathbf{u}, \mathbf{v}\}$ as a plane through the origin.

EXAMPLE 6 Let $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 5 \\ -13 \\ -3 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} -3 \\ 8 \\ 1 \end{bmatrix}$. Then Span $\{\mathbf{a}_1, \mathbf{a}_2\}$ is a plane through the origin in $\mathbb{R}^3$. Is $\mathbf{b}$ in that plane?

SOLUTION Does the equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{b}$ have a solution? To answer this, row reduce the augmented matrix $[\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{b}\,]$:

$$\begin{bmatrix} 1 & 5 & -3 \\ -2 & -13 & 8 \\ 3 & -3 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & -3 \\ 0 & -3 & 2 \\ 0 & -18 & 10 \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & -3 \\ 0 & -3 & 2 \\ 0 & 0 & -2 \end{bmatrix}$$

The third equation is $0 = -2$, which shows that the system has no solution. The vector equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{b}$ has no solution, and so $\mathbf{b}$ is *not* in Span $\{\mathbf{a}_1, \mathbf{a}_2\}$. ■

Linear Combinations in Applications

The final example shows how scalar multiples and linear combinations can arise when a quantity such as "cost" is broken down into several categories. The basic principle for the example concerns the cost of producing several units of an item when the cost per unit is known:

$$\begin{Bmatrix} \text{number} \\ \text{of units} \end{Bmatrix} \cdot \begin{Bmatrix} \text{cost} \\ \text{per unit} \end{Bmatrix} = \begin{Bmatrix} \text{total} \\ \text{cost} \end{Bmatrix}$$

EXAMPLE 7 A company manufactures two products. For \$1.00 worth of product B, the company spends \$.45 on materials, \$.25 on labor, and \$.15 on overhead. For \$1.00 worth of product C, the company spends \$.40 on materials, \$.30 on labor, and \$.15 on overhead. Let

$$\mathbf{b} = \begin{bmatrix} .45 \\ .25 \\ .15 \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} .40 \\ .30 \\ .15 \end{bmatrix}$$

Then $\mathbf{b}$ and $\mathbf{c}$ represent the "costs per dollar of income" for the two products.

a. What economic interpretation can be given to the vector $100\mathbf{b}$?

b. Suppose the company wishes to manufacture x_1 dollars worth of product B and x_2 dollars worth of product C. Give a vector that describes the various costs the company will have (for materials, labor, and overhead).

SOLUTION

a. Compute

$$100\mathbf{b} = 100\begin{bmatrix} .45 \\ .25 \\ .15 \end{bmatrix} = \begin{bmatrix} 45 \\ 25 \\ 15 \end{bmatrix}$$

The vector $100\mathbf{b}$ lists the various costs for producing \$100 worth of product B—namely \$45 for materials, \$25 for labor, and \$15 for overhead.

b. The costs of manufacturing x_1 dollars worth of B are given by the vector $x_1\mathbf{b}$, and the costs of manufacturing x_2 dollars worth of C are given by $x_2\mathbf{c}$. Hence the total costs for both products are given by the vector $x_1\mathbf{b} + x_2\mathbf{c}$. ■

Practice Problems

1. Prove that $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ for any $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$.
2. For what value(s) of h will $\mathbf{y}$ be in Span$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ if

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 5 \\ -4 \\ -7 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} -4 \\ 3 \\ h \end{bmatrix}$$

3. Let $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$, $\mathbf{u}$, and $\mathbf{v}$ be vectors in $\mathbb{R}^n$. Suppose the vectors $\mathbf{u}$ and $\mathbf{v}$ are in Span $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$. Show that $\mathbf{u} + \mathbf{v}$ is also in Span $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$. [*Hint:* The solution requires the use of the definition of the span of a set of vectors. It is useful to review this definition before starting this exercise.]

1.3 Exercises

In Exercises 1 and 2, compute $\mathbf{u} + \mathbf{v}$ and $\mathbf{u} - 2\mathbf{v}$.

1. $\mathbf{u} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} -3 \\ 3 \end{bmatrix}$

2. $\mathbf{u} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

In Exercises 3 and 4, display the following vectors using arrows on an xy-graph: $\mathbf{u}$, $\mathbf{v}$, $-\mathbf{v}$, $-2\mathbf{v}$, $\mathbf{u} + \mathbf{v}$, $\mathbf{u} - \mathbf{v}$, and $\mathbf{u} - 2\mathbf{v}$. Notice that $\mathbf{u} - \mathbf{v}$ is the vertex of a parallelogram whose other vertices are $\mathbf{u}$, $\mathbf{0}$, and $-\mathbf{v}$.

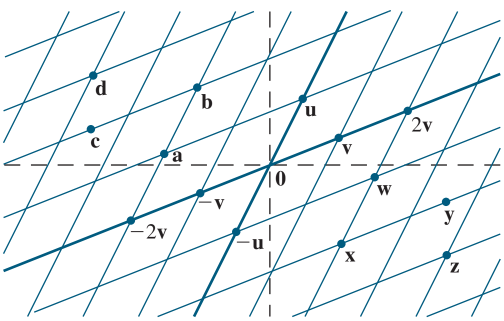
3. $\mathbf{u}$ and $\mathbf{v}$ as in Exercise 1
4. $\mathbf{u}$ and $\mathbf{v}$ as in Exercise 2

In Exercises 5 and 6, write a system of equations that is equivalent to the given vector equation.

5. $x_1 \begin{bmatrix} 4 \\ -3 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} -8 \\ 7 \\ 0 \end{bmatrix} = \begin{bmatrix} 9 \\ -6 \\ -5 \end{bmatrix}$

6. $x_1 \begin{bmatrix} -2 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 8 \\ 5 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ -6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Use the accompanying figure to write each vector listed in Exercises 7 and 8 as a linear combination of $\mathbf{u}$ and $\mathbf{v}$. Is every vector in $\mathbb{R}^2$ a linear combination of $\mathbf{u}$ and $\mathbf{v}$?



7. Vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$
8. Vectors $\mathbf{w}$, $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$

In Exercises 9 and 10, write a vector equation that is equivalent to the given system of equations.

9.
$$\begin{aligned} x_2 + 5x_3 &= 0 \\ 4x_1 + 6x_2 - x_3 &= 0 \\ -x_1 + 3x_2 - 8x_3 &= 0 \end{aligned}$$

10.
$$\begin{aligned} 4x_1 + x_2 + 3x_3 &= 9 \\ x_1 - 7x_2 - 2x_3 &= 2 \\ 8x_1 + 6x_2 - 5x_3 &= 15 \end{aligned}$$

In Exercises 11 and 12, determine if $\mathbf{b}$ is a linear combination of $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$.

11. $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} 5 \\ -6 \\ 8 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix}$

12. $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 0 \\ 5 \\ 5 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} 2 \\ 0 \\ 8 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -5 \\ 11 \\ -7 \end{bmatrix}$

In Exercises 13 and 14, determine if $\mathbf{b}$ is a linear combination of the vectors formed from the columns of the matrix A.

13. $A = \begin{bmatrix} 1 & -4 & 2 \\ 0 & 3 & 5 \\ -2 & 8 & -4 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 3 \\ -7 \\ -3 \end{bmatrix}$

14. $A = \begin{bmatrix} 1 & -2 & -6 \\ 0 & 3 & 7 \\ 1 & -2 & 5 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 11 \\ -5 \\ 9 \end{bmatrix}$

In Exercises 15 and 16, list five vectors in Span $\{\mathbf{v}_1, \mathbf{v}_2\}$. For each vector, show the weights on $\mathbf{v}_1$ and $\mathbf{v}_2$ used to generate the vector and list the three entries of the vector. Do not make a sketch.

15. $\mathbf{v}_1 = \begin{bmatrix} 7 \\ 1 \\ -6 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -5 \\ 3 \\ 0 \end{bmatrix}$

16. $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -2 \\ 0 \\ 3 \end{bmatrix}$

17. Let $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 4 \\ -2 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} -2 \\ -3 \\ 7 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ h \end{bmatrix}$. For what value(s) of h is $\mathbf{b}$ in the plane spanned by $\mathbf{a}_1$ and $\mathbf{a}_2$?

18. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ -4 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -5 \\ 1 \\ 7 \end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix} h \\ -1 \\ -5 \end{bmatrix}$. For what value(s) of h is $\mathbf{y}$ in the plane generated by $\mathbf{v}_1$ and $\mathbf{v}_2$?

19. Give a geometric description of Span $\{\mathbf{v}_1, \mathbf{v}_2\}$ for the vectors $\mathbf{v}_1 = \begin{bmatrix} 8 \\ 2 \\ -6 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 12 \\ 3 \\ -9 \end{bmatrix}$.

20. Give a geometric description of Span $\{\mathbf{v}_1, \mathbf{v}_2\}$ for the vectors in Exercise 16.

21. Let $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Show that $\begin{bmatrix} h \\ k \end{bmatrix}$ is in Span $\{\mathbf{u}, \mathbf{v}\}$ for all h and k.

22. Construct a 3×3 matrix A, with nonzero entries, and a vector $\mathbf{b}$ in $\mathbb{R}^3$ such that $\mathbf{b}$ is *not* in the set spanned by the columns of A.

In Exercises 23–32, mark each statement True or False **(T/F)**. Justify each answer.

23. **(T/F)** Another notation for the vector $\begin{bmatrix} -4 \\ 3 \end{bmatrix}$ is $[-4 \;\; 3]$.

24. **(T/F)** Any list of five real numbers is a vector in $\mathbb{R}^5$.

25. **(T/F)** The points in the plane corresponding to $\begin{bmatrix} -2 \\ 5 \end{bmatrix}$ and $\begin{bmatrix} -5 \\ 2 \end{bmatrix}$ lie on a line through the origin.

26. **(T/F)** The vector $\mathbf{u}$ results when a vector $\mathbf{u} - \mathbf{v}$ is added to the vector $\mathbf{v}$.

27. **(T/F)** An example of a linear combination of vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ is the vector $\frac{1}{2}\mathbf{v}_1$.

28. **(T/F)** The weights $c_1, \dots, c_p$ in a linear combination $c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$ cannot all be zero.

29. **(T/F)** The solution set of the linear system whose augmented matrix is $[\, \mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \mathbf{a}_3 \;\; \mathbf{b} \,]$ is the same as the solution set of the equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 = \mathbf{b}$.

30. **(T/F)** When $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors, Span $\{\mathbf{u}, \mathbf{v}\}$ contains the line through $\mathbf{u}$ and the origin.

31. **(T/F)** The set Span $\{\mathbf{u}, \mathbf{v}\}$ is always visualized as a plane through the origin.

32. **(T/F)** Asking whether the linear system corresponding to an augmented matrix $[\, \mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \mathbf{a}_3 \;\; \mathbf{b} \,]$ has a solution amounts to asking whether $\mathbf{b}$ is in Span $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$.

33. Let $A = \begin{bmatrix} 1 & 0 & -4 \\ 0 & 3 & -2 \\ -2 & 6 & 3 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ -4 \end{bmatrix}$. Denote the columns of A by $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$, and let $W = \text{Span}\,\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$.

a. Is $\mathbf{b}$ in $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$? How many vectors are in $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$?

b. Is $\mathbf{b}$ in W? How many vectors are in W?

c. Show that $\mathbf{a}_1$ is in W. [*Hint:* Row operations are unnecessary.]

34. Let $A = \begin{bmatrix} 2 & 0 & 6 \\ -1 & 8 & 5 \\ 1 & -2 & 1 \end{bmatrix}$, let $\mathbf{b} = \begin{bmatrix} 10 \\ 3 \\ 3 \end{bmatrix}$, and let W be the set of all linear combinations of the columns of A.

a. Is $\mathbf{b}$ in W?

b. Show that the third column of A is in W.

35. A mining company has two mines. One day's operation at mine 1 produces ore that contains 20 metric tons of copper and 550 kilograms of silver, while one day's operation at mine 2 produces ore that contains 30 metric tons of copper and 500 kilograms of silver. Let $\mathbf{v}_1 = \begin{bmatrix} 20 \\ 550 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 30 \\ 500 \end{bmatrix}$. Then $\mathbf{v}_1$ and $\mathbf{v}_2$ represent the "output per day" of mine 1 and mine 2, respectively.

a. What physical interpretation can be given to the vector $5\mathbf{v}_1$?

b. Suppose the company operates mine 1 for x_1 days and mine 2 for x_2 days. Write a vector equation whose solution gives the number of days each mine should operate in order to produce 150 tons of copper and 2825 kilograms of silver. Do not solve the equation.

T c. Solve the equation in (b).

36. A steam plant burns two types of coal: anthracite (A) and bituminous (B). For each ton of A burned, the plant produces 27.6 million Btu of heat, 3100 grams (g) of sulfur dioxide, and 250 g of particulate matter (solid-particle pollutants). For each ton of B burned, the plant produces 30.2 million Btu, 6400 g of sulfur dioxide, and 360 g of particulate matter.

a. How much heat does the steam plant produce when it burns x_1 tons of A and x_2 tons of B?

b. Suppose the output of the steam plant is described by a vector that lists the amounts of heat, sulfur dioxide, and particulate matter. Express this output as a linear combination of two vectors, assuming that the plant burns x_1 tons of A and x_2 tons of B.

T c. Over a certain time period, the steam plant produced 162 million Btu of heat, 23,610 g of sulfur dioxide, and 1623 g of particulate matter. Determine how many tons of each type of coal the steam plant must have burned. Include a vector equation as part of your solution.

37. Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be points in $\mathbb{R}^3$ and suppose that for $j = 1, \ldots, k$ an object with mass m_j is located at point $\mathbf{v}_j$. Physicists call such objects *point masses*. The total mass of the system of point masses is

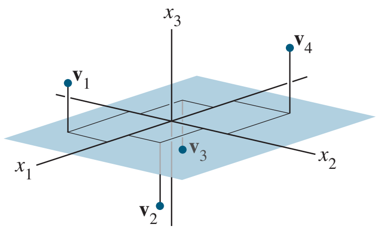
$$m = m_1 + \cdots + m_k$$

The *center of mass* (or *center of gravity*) of the system is

$$\bar{\mathbf{v}} = \frac{1}{m}[m_1\mathbf{v}_1 + \cdots + m_k\mathbf{v}_k]$$

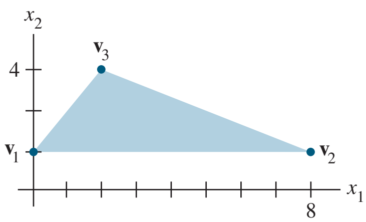
Compute the center of gravity of the system consisting of the following point masses (see the figure):

Point	Mass
$\mathbf{v}_1 = (5, -4, 3)$	2 g
$\mathbf{v}_2 = (4, 3, -2)$	5 g
$\mathbf{v}_3 = (-4, -3, -1)$	2 g
$\mathbf{v}_4 = (-9, 8, 6)$	1 g



38. Let $\mathbf{v}$ be the center of mass of a system of point masses located at $\mathbf{v}_1, \ldots, \mathbf{v}_k$ as in Exercise 37. Is $\mathbf{v}$ in Span $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$? Explain.

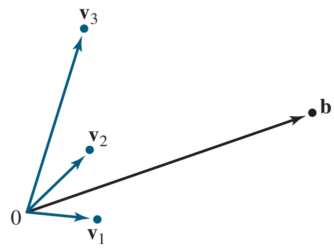
39. A thin triangular plate of uniform density and thickness has vertices at $\mathbf{v}_1 = (0, 1)$, $\mathbf{v}_2 = (8, 1)$, and $\mathbf{v}_3 = (2, 4)$, as in the figure below, and the mass of the plate is 3 g.



a. Find the (x, y)-coordinates of the center of mass of the plate. This "balance point" of the plate coincides with the center of mass of a system consisting of three 1-gram point masses located at the vertices of the plate.

b. Determine how to distribute an additional mass of 6 g at the three vertices of the plate to move the balance point of the plate to $(2, 2)$. [*Hint:* Let w_1, w_2, and w_3 denote the masses added at the three vertices, so that $w_1 + w_2 + w_3 = 6$.]

40. Consider the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, and $\mathbf{b}$ in $\mathbb{R}^2$, shown in the figure. Does the equation $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{b}$ have a solution? Is the solution unique? Use the figure to explain your answers.



41. Use the vectors $\mathbf{u} = (u_1, \ldots, u_n)$, $\mathbf{v} = (v_1, \ldots, v_n)$, and $\mathbf{w} = (w_1, \ldots, w_n)$ to verify the following algebraic properties of $\mathbb{R}^n$.

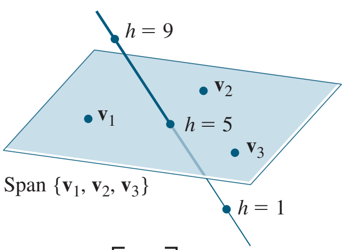
a. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

b. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$ for each scalar c

42. Use the vector $\mathbf{u} = (u_1, \ldots, u_n)$ to verify the following algebraic properties of $\mathbb{R}^n$.

a. $\mathbf{u} + (-\mathbf{u}) = (-\mathbf{u}) + \mathbf{u} = \mathbf{0}$

b. $c(d\mathbf{u}) = (cd)\mathbf{u}$ for all scalars c and d



The points $\begin{bmatrix} -4 \\ 3 \\ h \end{bmatrix}$ lie on a line that intersects the plane when $h = 5$.

Solutions to Practice Problems

1. Take arbitrary vectors $\mathbf{u} = (u_1, \ldots, u_n)$ and $\mathbf{v} = (v_1, \ldots, v_n)$ in $\mathbb{R}^n$, and compute

$$\begin{aligned} \mathbf{u} + \mathbf{v} &= (u_1 + v_1, \ldots, u_n + v_n) && \text{Definition of vector addition} \\ &= (v_1 + u_1, \ldots, v_n + u_n) && \text{Commutativity of addition in } \mathbb{R} \\ &= \mathbf{v} + \mathbf{u} && \text{Definition of vector addition} \end{aligned}$$

2. The vector $\mathbf{y}$ belongs to Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ if and only if there exist scalars x_1, x_2, x_3 such that

$$x_1\begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix} + x_2\begin{bmatrix} 5 \\ -4 \\ -7 \end{bmatrix} + x_3\begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -4 \\ 3 \\ h \end{bmatrix}$$

This vector equation is equivalent to a system of three linear equations in three unknowns. If you row reduce the augmented matrix for this system, you find that

$$\begin{bmatrix} 1 & 5 & -3 & -4 \\ -1 & -4 & 1 & 3 \\ -2 & -7 & 0 & h \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & -3 & -4 \\ 0 & 1 & -2 & -1 \\ 0 & 3 & -6 & h-8 \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & -3 & -4 \\ 0 & 1 & -2 & -1 \\ 0 & 0 & 0 & h-5 \end{bmatrix}$$

The system is consistent if and only if there is no pivot in the fourth column. That is, $h - 5$ must be 0. So $\mathbf{y}$ is in Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ if and only if $h = 5$.

Remember: The presence of a free variable in a system does not guarantee that the system is consistent.

3. Since the vectors $\mathbf{u}$ and $\mathbf{v}$ are in Span $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$, there exist scalars c_1, c_2, c_3 and d_1, d_2, d_3 such that

$$\mathbf{u} = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + c_3\mathbf{w}_3 \quad \text{and} \quad \mathbf{v} = d_1\mathbf{w}_1 + d_2\mathbf{w}_2 + d_3\mathbf{w}_3.$$

Notice

$$\begin{aligned} \mathbf{u} + \mathbf{v} &= c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + c_3\mathbf{w}_3 + d_1\mathbf{w}_1 + d_2\mathbf{w}_2 + d_3\mathbf{w}_3 \\ &= (c_1 + d_1)\mathbf{w}_1 + (c_2 + d_2)\mathbf{w}_2 + (c_3 + d_3)\mathbf{w}_3 \end{aligned}$$

Since $c_1 + d_1, c_2 + d_2$, and $c_3 + d_3$ are also scalars, the vector $\mathbf{u} + \mathbf{v}$ is in Span $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$.

1.4 The Matrix Equation $A\mathbf{x} = \mathbf{b}$

A fundamental idea in linear algebra is to view a linear combination of vectors as the product of a matrix and a vector. The following definition permits us to rephrase some of the concepts of Section 1.3 in new ways.

DEFINITION

If A is an $m \times n$ matrix, with columns $\mathbf{a}_1, \dots, \mathbf{a}_n$, and if $\mathbf{x}$ is in $\mathbb{R}^n$, then the **product of A and $\mathbf{x}$**, denoted by $A\mathbf{x}$, is **the linear combination of the columns of A using the corresponding entries in $\mathbf{x}$ as weights**; that is,

$$A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$$

Note that $A\mathbf{x}$ is defined only if the number of columns of A equals the number of entries in $\mathbf{x}$.

EXAMPLE 1

a. $$\begin{bmatrix} 1 & 2 & -1 \\ 0 & -5 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \\ 7 \end{bmatrix} = 4\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3\begin{bmatrix} 2 \\ -5 \end{bmatrix} + 7\begin{bmatrix} -1 \\ 3 \end{bmatrix}$$
$$= \begin{bmatrix} 4 \\ 0 \end{bmatrix} + \begin{bmatrix} 6 \\ -15 \end{bmatrix} + \begin{bmatrix} -7 \\ 21 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

b. $$\begin{bmatrix} 2 & -3 \\ 8 & 0 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 7 \end{bmatrix} = 4\begin{bmatrix} 2 \\ 8 \\ -5 \end{bmatrix} + 7\begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 32 \\ -20 \end{bmatrix} + \begin{bmatrix} -21 \\ 0 \\ 14 \end{bmatrix} = \begin{bmatrix} -13 \\ 32 \\ -6 \end{bmatrix}$$ ■

EXAMPLE 2 For $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ in $\mathbb{R}^m$, write the linear combination $3\mathbf{v}_1 - 5\mathbf{v}_2 + 7\mathbf{v}_3$ as a matrix times a vector.

SOLUTION Place $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ into the columns of a matrix A and place the weights 3, -5, and 7 into a vector $\mathbf{x}$. That is,

$$3\mathbf{v}_1 - 5\mathbf{v}_2 + 7\mathbf{v}_3 = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} \begin{bmatrix} 3 \\ -5 \\ 7 \end{bmatrix} = A\mathbf{x}$$ ■

Section 1.3 showed how to write a system of linear equations as a vector equation involving a linear combination of vectors. For example, the system

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

is equivalent to

$$x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ -5 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix} \tag{2}$$

As in Example 2, the linear combination on the left side is a matrix times a vector, so that (2) becomes

$$\begin{bmatrix} 1 & 2 & -1 \\ 0 & -5 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix} \tag{3}$$

Equation (3) has the form $A\mathbf{x} = \mathbf{b}$. Such an equation is called a **matrix equation**, to distinguish it from a vector equation such as is shown in (2).

Notice how the matrix in (3) is just the matrix of coefficients of the system (1). Similar calculations show that any system of linear equations, or any vector equation such as (2), can be written as an equivalent matrix equation in the form $A\mathbf{x} = \mathbf{b}$. This simple observation will be used repeatedly throughout the text.

Here is the formal result.

THEOREM 3

If A is an $m \times n$ matrix, with columns $\mathbf{a}_1, \dots, \mathbf{a}_n$, and if $\mathbf{b}$ is in $\mathbb{R}^m$, the matrix equation

$$A\mathbf{x} = \mathbf{b} \tag{4}$$

has the same solution set as the vector equation

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b} \tag{5}$$

which, in turn, has the same solution set as the system of linear equations whose augmented matrix is

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n & \mathbf{b} \end{bmatrix} \tag{6}$$

Theorem 3 provides a powerful tool for gaining insight into problems in linear algebra, because a system of linear equations may now be viewed in three different but equivalent ways: as a matrix equation, as a vector equation, or as a system of linear equations. Whenever you construct a mathematical model of a problem in real life, you are free to choose whichever viewpoint is most natural. Then you may switch from one formulation of a problem to another whenever it is convenient. In any case, the matrix equation (4), the vector equation (5), and the system of equations are all solved in the same way—by row reducing the augmented matrix (6). Other methods of solution will be discussed later.

Existence of Solutions

The definition of $A\mathbf{x}$ leads directly to the following useful fact.

> The equation $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is a linear combination of the columns of A.

Section 1.3 considered the existence question, "Is $\mathbf{b}$ in Span $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$?" Equivalently, "Is $A\mathbf{x} = \mathbf{b}$ consistent?" A harder existence problem is to determine whether the equation $A\mathbf{x} = \mathbf{b}$ is consistent *for all* possible $\mathbf{b}$.

EXAMPLE 3 Let $A = \begin{bmatrix} 1 & 3 & 4 \\ -4 & 2 & -6 \\ -3 & -2 & -7 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$. Is the equation $A\mathbf{x} = \mathbf{b}$ consistent for all possible b_1, b_2, b_3?

SOLUTION Row reduce the augmented matrix for $A\mathbf{x} = \mathbf{b}$:

$$\begin{bmatrix} 1 & 3 & 4 & b_1 \\ -4 & 2 & -6 & b_2 \\ -3 & -2 & -7 & b_3 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 4 & b_1 \\ 0 & 14 & 10 & b_2 + 4b_1 \\ 0 & 7 & 5 & b_3 + 3b_1 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 3 & 4 & b_1 \\ 0 & 14 & 10 & b_2 + 4b_1 \\ 0 & 0 & 0 & b_3 + 3b_1 - \frac{1}{2}(b_2 + 4b_1) \end{bmatrix}$$

The third entry in column 4 equals $b_1 - \frac{1}{2}b_2 + b_3$. The equation $A\mathbf{x} = \mathbf{b}$ is *not* consistent for every $\mathbf{b}$ because some choices of $\mathbf{b}$ can make $b_1 - \frac{1}{2}b_2 + b_3$ nonzero. ■

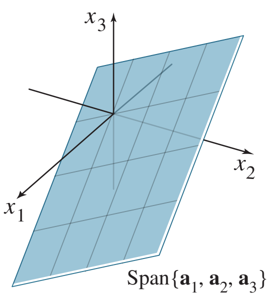


FIGURE 1
The columns of $A = [\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3\,]$ span a plane through $\mathbf{0}$.

The reduced matrix in Example 3 provides a description of all $\mathbf{b}$ for which the equation $A\mathbf{x} = \mathbf{b}$ *is* consistent: The entries in $\mathbf{b}$ must satisfy

$$b_1 - \tfrac{1}{2}b_2 + b_3 = 0$$

This is the equation of a plane through the origin in $\mathbb{R}^3$. The plane is the set of all linear combinations of the three columns of A. See Figure 1.

The equation $A\mathbf{x} = \mathbf{b}$ in Example 3 fails to be consistent for all $\mathbf{b}$ because the echelon form of A has a row of zeros. If A had a pivot in all three rows, we would not care about the calculations in the augmented column because in this case an echelon form of the augmented matrix could not have a row such as $[\,0 \quad 0 \quad 0 \quad 1\,]$.

In the next theorem, the sentence "The columns of A span $\mathbb{R}^m$" means that *every* $\mathbf{b}$ in $\mathbb{R}^m$ is a linear combination of the columns of A. In general, a set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ in $\mathbb{R}^m$ **spans** (or **generates**) $\mathbb{R}^m$ if every vector in $\mathbb{R}^m$ is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_p$—that is, if Span $\{\mathbf{v}_1, \dots, \mathbf{v}_p\} = \mathbb{R}^m$.

THEOREM 4

Let A be an $m \times n$ matrix. Then the following statements are logically equivalent. That is, for a particular A, either they are all true statements or they are all false.

a. For each $\mathbf{b}$ in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution.

b. Each $\mathbf{b}$ in $\mathbb{R}^m$ is a linear combination of the columns of A.

c. The columns of A span $\mathbb{R}^m$.

d. A has a pivot position in every row.

Theorem 4 is one of the most useful theorems in this chapter. Statements (a), (b), and (c) are equivalent because of the definition of $A\mathbf{x}$ and what it means for a set of vectors to span $\mathbb{R}^m$. The discussion after Example 3 suggests why (a) and (d) are equivalent; a proof is given at the end of the section. The exercises will provide examples of how Theorem 4 is used.

Warning: Theorem 4 is about a *coefficient matrix*, not an augmented matrix. If an augmented matrix $[\,A \quad \mathbf{b}\,]$ has a pivot position in every row, then the equation $A\mathbf{x} = \mathbf{b}$ may or may not be consistent.

Computation of $A\mathbf{x}$

The calculations in Example 1 were based on the definition of the product of a matrix A and a vector $\mathbf{x}$. The following simple example will lead to a more efficient method for calculating the entries in $A\mathbf{x}$ when working problems by hand.

EXAMPLE 4 Compute $A\mathbf{x}$, where $A = \begin{bmatrix} 2 & 3 & 4 \\ -1 & 5 & -3 \\ 6 & -2 & 8 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$.

SOLUTION From the definition,

$$\begin{aligned}
\begin{bmatrix} 2 & 3 & 4 \\ -1 & 5 & -3 \\ 6 & -2 & 8 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} &= x_1\begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix} + x_2\begin{bmatrix} 3 \\ 5 \\ -2 \end{bmatrix} + x_3\begin{bmatrix} 4 \\ -3 \\ 8 \end{bmatrix} \\
&= \begin{bmatrix} 2x_1 \\ -x_1 \\ 6x_1 \end{bmatrix} + \begin{bmatrix} 3x_2 \\ 5x_2 \\ -2x_2 \end{bmatrix} + \begin{bmatrix} 4x_3 \\ -3x_3 \\ 8x_3 \end{bmatrix} \qquad (7) \\
&= \begin{bmatrix} 2x_1 + 3x_2 + 4x_3 \\ -x_1 + 5x_2 - 3x_3 \\ 6x_1 - 2x_2 + 8x_3 \end{bmatrix}
\end{aligned}$$

The first entry in the product $A\mathbf{x}$ is a sum of products (sometimes called a *dot product*), using the first row of A and the entries in $\mathbf{x}$. That is,

$$\begin{bmatrix} 2 & 3 & 4 \\ & & \\ & & \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_1 + 3x_2 + 4x_3 \\ \\ \\ \end{bmatrix}$$

This matrix shows how to compute the first entry in $A\mathbf{x}$ directly, without writing down all the calculations shown in (7). Similarly, the second entry in $A\mathbf{x}$ can be calculated at once by multiplying the entries in the second row of A by the corresponding entries in $\mathbf{x}$ and then summing the resulting products:

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

Likewise, the third entry in $A\mathbf{x}$ can be calculated from the third row of A and the entries in $\mathbf{x}$. ■

Row–Vector Rule for Computing $A\mathbf{x}$

If the product $A\mathbf{x}$ is defined, then the ith entry in $A\mathbf{x}$ is the sum of the products of corresponding entries from row i of A and from the vector $\mathbf{x}$.

EXAMPLE 5

a. $$\begin{bmatrix} 1 & 2 & -1 \\ 0 & -5 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \\ 7 \end{bmatrix} = \begin{bmatrix} 1 \cdot 4 + 2 \cdot 3 + (-1) \cdot 7 \\ 0 \cdot 4 + (-5) \cdot 3 + 3 \cdot 7 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

b. $$\begin{bmatrix} 2 & -3 \\ 8 & 0 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 7 \end{bmatrix} = \begin{bmatrix} 2 \cdot 4 + (-3) \cdot 7 \\ 8 \cdot 4 + 0 \cdot 7 \\ (-5) \cdot 4 + 2 \cdot 7 \end{bmatrix} = \begin{bmatrix} -13 \\ 32 \\ -6 \end{bmatrix}$$

c. $$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r \\ s \\ t \end{bmatrix} = \begin{bmatrix} 1 \cdot r + 0 \cdot s + 0 \cdot t \\ 0 \cdot r + 1 \cdot s + 0 \cdot t \\ 0 \cdot r + 0 \cdot s + 1 \cdot t \end{bmatrix} = \begin{bmatrix} r \\ s \\ t \end{bmatrix}$$ ■

By definition, the matrix in Example 5(c) with 1's on the diagonal and 0's elsewhere is called an **identity matrix** and is denoted by I. The calculation in part (c) shows that $I\mathbf{x} = \mathbf{x}$ for every $\mathbf{x}$ in $\mathbb{R}^3$. There is an analogous $n \times n$ identity matrix, sometimes written as I_n. As in part (c), $I_n\mathbf{x} = \mathbf{x}$ for every $\mathbf{x}$ in $\mathbb{R}^n$.

Properties of the Matrix–Vector Product $A\mathbf{x}$

The facts in the next theorem are important and will be used throughout the text. The proof relies on the definition of $A\mathbf{x}$ and the algebraic properties of $\mathbb{R}^n$.

THEOREM 5

If A is an $m \times n$ matrix, $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^n$, and c is a scalar, then:

a. $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$;
b. $A(c\mathbf{u}) = c(A\mathbf{u})$.

PROOF For simplicity, take $n = 3$, $A = [\,\mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \mathbf{a}_3\,]$, and $\mathbf{u}$, $\mathbf{v}$ in $\mathbb{R}^3$. (The proof of the general case is similar.) For $i = 1, 2, 3$, let u_i and v_i be the ith entries in $\mathbf{u}$ and $\mathbf{v}$, respectively. To prove statement (a), compute $A(\mathbf{u} + \mathbf{v})$ as a linear combination of the columns of A using the entries in $\mathbf{u} + \mathbf{v}$ as weights.

$$A(\mathbf{u} + \mathbf{v}) = [\,\mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \mathbf{a}_3\,] \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ u_3 + v_3 \end{bmatrix}$$

Entries in $\mathbf{u} + \mathbf{v}$

$$= (u_1 + v_1)\mathbf{a}_1 + (u_2 + v_2)\mathbf{a}_2 + (u_3 + v_3)\mathbf{a}_3$$

Columns of A

$$= (u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + u_3\mathbf{a}_3) + (v_1\mathbf{a}_1 + v_2\mathbf{a}_2 + v_3\mathbf{a}_3)$$
$$= A\mathbf{u} + A\mathbf{v}$$

To prove statement (b), compute $A(c\mathbf{u})$ as a linear combination of the columns of A using the entries in $c\mathbf{u}$ as weights.

$$A(c\mathbf{u}) = [\,\mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \mathbf{a}_3\,] \begin{bmatrix} cu_1 \\ cu_2 \\ cu_3 \end{bmatrix} = (cu_1)\mathbf{a}_1 + (cu_2)\mathbf{a}_2 + (cu_3)\mathbf{a}_3$$
$$= c(u_1\mathbf{a}_1) + c(u_2\mathbf{a}_2) + c(u_3\mathbf{a}_3)$$
$$= c(u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + u_3\mathbf{a}_3)$$
$$= c(A\mathbf{u})$$ ■

Numerical Note

To optimize a computer algorithm to compute $A\mathbf{x}$, the sequence of calculations should involve data stored in contiguous memory locations. The most widely used professional algorithms for matrix computations are written in Fortran, a language that stores a matrix as a set of columns. Such algorithms compute $A\mathbf{x}$ as a linear combination of the columns of A. In contrast, if a program is written in the popular language C, which stores matrices by rows, $A\mathbf{x}$ should be computed via the alternative rule that uses the rows of A.

PROOF OF THEOREM 4 As was pointed out after Theorem 4, statements (a), (b), and (c) are logically equivalent. So, it suffices to show (for an arbitrary matrix A) that (a) and (d) are either both true or both false. This will tie all four statements together.

Let U be an echelon form of A. Given $\mathbf{b}$ in $\mathbb{R}^m$, we can row reduce the augmented matrix $[\,A \quad \mathbf{b}\,]$ to an augmented matrix $[\,U \quad \mathbf{d}\,]$ for some $\mathbf{d}$ in $\mathbb{R}^m$:

$$[\,A \quad \mathbf{b}\,] \sim \cdots \sim [\,U \quad \mathbf{d}\,]$$

If statement (d) is true, then each row of U contains a pivot position and there can be no pivot in the augmented column. So $A\mathbf{x} = \mathbf{b}$ has a solution for any $\mathbf{b}$, and (a) is true. If (d) is false, the last row of U is all zeros. Let $\mathbf{d}$ be any vector with a 1 in its last entry. Then $[\,U \quad \mathbf{d}\,]$ represents an *inconsistent* system. Since row operations are reversible, $[\,U \quad \mathbf{d}\,]$ can be transformed into the form $[\,A \quad \mathbf{b}\,]$. The new system $A\mathbf{x} = \mathbf{b}$ is also inconsistent, and (a) is false. ■

Practice Problems

1. Let $A = \begin{bmatrix} 1 & 5 & -2 & 0 \\ -3 & 1 & 9 & -5 \\ 4 & -8 & -1 & 7 \end{bmatrix}$, $\mathbf{p} = \begin{bmatrix} 3 \\ -2 \\ 0 \\ -4 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} -7 \\ 9 \\ 0 \end{bmatrix}$. It can be shown that $\mathbf{p}$ is a solution of $A\mathbf{x} = \mathbf{b}$. Use this fact to exhibit $\mathbf{b}$ as a specific linear combination of the columns of A.

2. Let $A = \begin{bmatrix} 2 & 5 \\ 3 & 1 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} -3 \\ 5 \end{bmatrix}$. Verify Theorem 5(a) in this case by computing $A(\mathbf{u} + \mathbf{v})$ and $A\mathbf{u} + A\mathbf{v}$.

3. Construct a 3×3 matrix A and vectors $\mathbf{b}$ and $\mathbf{c}$ in $\mathbb{R}^3$ so that $A\mathbf{x} = \mathbf{b}$ has a solution, but $A\mathbf{x} = \mathbf{c}$ does not.

1.4 Exercises

Compute the products in Exercises 1–4 using (a) the definition, as in Example 1, and (b) the row–vector rule for computing $A\mathbf{x}$. If a product is undefined, explain why.

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

2. $\begin{bmatrix} 2 \\ 6 \\ -1 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$

3. $\begin{bmatrix} 6 & 5 \\ -4 & -3 \\ 7 & 6 \end{bmatrix}\begin{bmatrix} 1 \\ -3 \end{bmatrix}$

4. $\begin{bmatrix} 8 & 3 & 1 \\ 5 & 1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

In Exercises 5–8, use the definition of $A\mathbf{x}$ to write the matrix equation as a vector equation, or vice versa.

5. $\begin{bmatrix} 7 & 2 & -9 & 3 \\ -4 & -5 & 7 & -2 \end{bmatrix}\begin{bmatrix} 6 \\ -9 \\ 1 \\ -8 \end{bmatrix} = \begin{bmatrix} -9 \\ 44 \end{bmatrix}$

6. $\begin{bmatrix} 7 & -3 \\ 2 & 1 \\ 9 & -6 \\ -3 & 2 \end{bmatrix}\begin{bmatrix} -2 \\ -5 \end{bmatrix} = \begin{bmatrix} 1 \\ -9 \\ 12 \\ -4 \end{bmatrix}$

7. $x_1\begin{bmatrix} 4 \\ -1 \\ 7 \\ -4 \end{bmatrix} + x_2\begin{bmatrix} -5 \\ 3 \\ -5 \\ 1 \end{bmatrix} + x_3\begin{bmatrix} 7 \\ -8 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ -8 \\ 0 \\ -7 \end{bmatrix}$

8. $z_1\begin{bmatrix} 4 \\ -2 \end{bmatrix} + z_2\begin{bmatrix} -4 \\ 5 \end{bmatrix} + z_3\begin{bmatrix} -5 \\ 4 \end{bmatrix} + z_4\begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 13 \end{bmatrix}$

In Exercises 9 and 10, write the system first as a vector equation and then as a matrix equation.

9. $\begin{aligned} 4x_1 + x_2 - 7x_3 &= 8 \\ x_2 + 6x_3 &= 0 \end{aligned}$

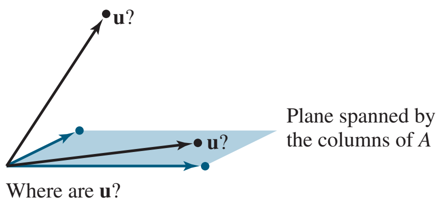
10. $\begin{aligned} 8x_1 - x_2 &= 4 \\ 5x_1 + 4x_2 &= 1 \\ x_1 - 3x_2 &= 2 \end{aligned}$

Given A and $\mathbf{b}$ in Exercises 11 and 12, write the augmented matrix for the linear system that corresponds to the matrix equation $A\mathbf{x} = \mathbf{b}$. Then solve the system and write the solution as a vector.

11. $A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & 5 \\ -2 & -4 & -3 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -2 \\ 2 \\ 9 \end{bmatrix}$

12. $A = \begin{bmatrix} 1 & 2 & 1 \\ -3 & -1 & 2 \\ 0 & 5 & 3 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$

13. Let $\mathbf{u} = \begin{bmatrix} 0 \\ 4 \\ 4 \end{bmatrix}$ and $A = \begin{bmatrix} 3 & -5 \\ -2 & 6 \\ 1 & 1 \end{bmatrix}$. Is $\mathbf{u}$ in the plane in $\mathbb{R}^3$ spanned by the columns of A? (See the figure.) Why or why not?



14. Let $\mathbf{u} = \begin{bmatrix} 2 \\ -3 \\ 2 \end{bmatrix}$ and $A = \begin{bmatrix} 5 & 8 & 7 \\ 0 & 1 & -1 \\ 1 & 3 & 0 \end{bmatrix}$. Is $\mathbf{u}$ in the subset of $\mathbb{R}^3$ spanned by the columns of A? Why or why not?

15. Let $A = \begin{bmatrix} 3 & -4 \\ -6 & 8 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$. Show that the equation $A\mathbf{x} = \mathbf{b}$ does not have a solution for all possible $\mathbf{b}$, and describe the set of all $\mathbf{b}$ for which $A\mathbf{x} = \mathbf{b}$ *does* have a solution.

16. Repeat Exercise 15: $A = \begin{bmatrix} 1 & -3 & -4 \\ -3 & 2 & 6 \\ 5 & -1 & -8 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$.

Exercises 17–20 refer to the matrices A and B below. Make appropriate calculations that justify your answers and mention an appropriate theorem.

$$A = \begin{bmatrix} 1 & 3 & 0 & 3 \\ -1 & -1 & -1 & 1 \\ 0 & -4 & 2 & -8 \\ 2 & 0 & 3 & -1 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 3 & -2 & 2 \\ 0 & 1 & 1 & -5 \\ 1 & 2 & -3 & 7 \\ -2 & -8 & 2 & -1 \end{bmatrix}$$

17. How many rows of A contain a pivot position? Does the equation $A\mathbf{x} = \mathbf{b}$ have a solution for each $\mathbf{b}$ in $\mathbb{R}^4$?

18. Do the columns of B span $\mathbb{R}^4$? Does the equation $B\mathbf{x} = \mathbf{y}$ have a solution for each $\mathbf{y}$ in $\mathbb{R}^4$?

19. Can each vector in $\mathbb{R}^4$ be written as a linear combination of the columns of the matrix A above? Do the columns of A span $\mathbb{R}^4$?

20. Can every vector in $\mathbb{R}^4$ be written as a linear combination of the columns of the matrix B above? Do the columns of B span $\mathbb{R}^3$?

21. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}$.

Does $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ span $\mathbb{R}^4$? Why or why not?

22. Let $\mathbf{v}_1 = \begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 0 \\ -3 \\ 8 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 4 \\ -1 \\ -5 \end{bmatrix}$.

Does $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ span $\mathbb{R}^3$? Why or why not?

In Exercises 23–34, mark each statement True or False **(T/F)**. Justify each answer.

23. **(T/F)** The equation $A\mathbf{x} = \mathbf{b}$ is referred to as a *vector equation*.

24. **(T/F)** Every matrix equation $A\mathbf{x} = \mathbf{b}$ corresponds to a vector equation with the same solution set.

25. **(T/F)** If the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent, then $\mathbf{b}$ is not in the set spanned by the columns of A.

26. **(T/F)** A vector $\mathbf{b}$ is a linear combination of the columns of a matrix A if and only if the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution.

27. **(T/F)** The equation $A\mathbf{x} = \mathbf{b}$ is consistent if the augmented matrix $[\, A \;\; \mathbf{b} \,]$ has a pivot position in every row.

28. **(T/F)** If A is an $m \times n$ matrix whose columns do not span $\mathbb{R}^m$, then the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent for some $\mathbf{b}$ in $\mathbb{R}^m$.

29. **(T/F)** The first entry in the product $A\mathbf{x}$ is a sum of products.

30. **(T/F)** Any linear combination vectors can always be written in the form $A\mathbf{x}$ for a suitable matrix A and vector $\mathbf{x}$.

31. **(T/F)** If the columns of an $m \times n$ matrix A span $\mathbb{R}^m$, then the equation $A\mathbf{x} = \mathbf{b}$ is consistent for each $\mathbf{b}$ in $\mathbb{R}^m$.

32. **(T/F)** The solution set of a linear system whose augmented matrix is $[\, \mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \mathbf{a}_3 \;\; \mathbf{b} \,]$ is the same as the solution set of $A\mathbf{x} = \mathbf{b}$, if $A = [\, \mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \mathbf{a}_3 \,]$.

33. **(T/F)** If A is an $m \times n$ matrix and if the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent for some $\mathbf{b}$ in $\mathbb{R}^m$, then A cannot have a pivot position in every row.

34. **(T/F)** If the augmented matrix $[\ A \quad \mathbf{b}\ ]$ has a pivot position in every row, then the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent.

35. Note that $\begin{bmatrix} 3 & -4 & 2 \\ 6 & -3 & 4 \\ -8 & 9 & -5 \end{bmatrix}\begin{bmatrix} -4 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ -9 \\ 8 \end{bmatrix}$. Use this fact (and no row operations) to find scalars c_1, c_2, c_3 such that $\begin{bmatrix} -2 \\ -9 \\ 8 \end{bmatrix} = c_1\begin{bmatrix} 3 \\ 6 \\ -8 \end{bmatrix} + c_2\begin{bmatrix} -4 \\ -3 \\ 9 \end{bmatrix} + c_3\begin{bmatrix} 2 \\ 4 \\ -5 \end{bmatrix}$.

36. Let $\mathbf{u} = \begin{bmatrix} 7 \\ 2 \\ 5 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 3 \\ 1 \\ 3 \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} 6 \\ 1 \\ 0 \end{bmatrix}$.

It can be shown that $3\mathbf{u} - 5\mathbf{v} - \mathbf{w} = \mathbf{0}$. Use this fact (and no row operations) to find x_1 and x_2 that satisfy the equation

$$\begin{bmatrix} 7 & 3 \\ 2 & 1 \\ 5 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 1 \\ 0 \end{bmatrix}.$$

37. Let $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$, and $\mathbf{v}$ represent vectors in $\mathbb{R}^5$, and let x_1, x_2, and x_3 denote scalars. Write the following vector equation as a matrix equation. Identify any symbols you choose to use.

$$x_1\mathbf{q}_1 + x_2\mathbf{q}_2 + x_3\mathbf{q}_3 = \mathbf{v}$$

38. Rewrite the (numerical) matrix equation below in symbolic form as a vector equation, using symbols $\mathbf{v}_1, \mathbf{v}_2, \ldots$ for the vectors and $c_1, c_2, \ldots$ for scalars. Define what each symbol represents, using the data given in the matrix equation.

$$\begin{bmatrix} -3 & 5 & -4 & 9 & 7 \\ 5 & 8 & 1 & -2 & -4 \end{bmatrix}\begin{bmatrix} -3 \\ 2 \\ 4 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ -1 \end{bmatrix}$$

39. Construct a 3×3 matrix, not in echelon form, whose columns span $\mathbb{R}^3$. Show that the matrix you construct has the desired property.

40. Construct a 3×3 matrix, not in echelon form, whose columns do *not* span $\mathbb{R}^3$. Show that the matrix you construct has the desired property.

STUDY GUIDE offers additional resources for mastering the concept of span.

41. Let A be a 3×2 matrix. Explain why the equation $A\mathbf{x} = \mathbf{b}$ cannot be consistent for all $\mathbf{b}$ in $\mathbb{R}^3$. Generalize your argument to the case of an arbitrary A with more rows than columns.

42. Could a set of three vectors in $\mathbb{R}^4$ span all of $\mathbb{R}^4$? Explain. What about n vectors in $\mathbb{R}^m$ when n is less than m?

43. Suppose A is a 4×3 matrix and $\mathbf{b}$ is a vector in $\mathbb{R}^4$ with the property that $A\mathbf{x} = \mathbf{b}$ has a unique solution. What can you say about the reduced echelon form of A? Justify your answer.

44. Suppose A is a 3×3 matrix and $\mathbf{b}$ is a vector in $\mathbb{R}^3$ with the property that $A\mathbf{x} = \mathbf{b}$ has a unique solution. Explain why the columns of A must span $\mathbb{R}^3$.

45. Let A be a 3×4 matrix, let $\mathbf{y}_1$ and $\mathbf{y}_2$ be vectors in $\mathbb{R}^3$, and let $\mathbf{w} = \mathbf{y}_1 + \mathbf{y}_2$. Suppose $\mathbf{y}_1 = A\mathbf{x}_1$ and $\mathbf{y}_2 = A\mathbf{x}_2$ for some vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathbb{R}^4$. What fact allows you to conclude that the system $A\mathbf{x} = \mathbf{w}$ is consistent? (*Note:* $\mathbf{x}_1$ and $\mathbf{x}_2$ denote vectors, not scalar entries in vectors.)

46. Let A be a 5×3 matrix, let $\mathbf{y}$ be a vector in $\mathbb{R}^3$, and let $\mathbf{z}$ be a vector in $\mathbb{R}^5$. Suppose $A\mathbf{y} = \mathbf{z}$. What fact allows you to conclude that the system $A\mathbf{x} = 4\mathbf{z}$ is consistent?

T In Exercises 47–50, determine if the columns of the matrix span $\mathbb{R}^4$.

47. $\begin{bmatrix} 7 & 2 & -5 & 8 \\ -5 & -3 & 4 & -9 \\ 6 & 10 & -2 & 7 \\ -7 & 9 & 2 & 15 \end{bmatrix}$

48. $\begin{bmatrix} 5 & -7 & -4 & 9 \\ 6 & -8 & -7 & 5 \\ 4 & -4 & -9 & -9 \\ -9 & 11 & 16 & 7 \end{bmatrix}$

49. $\begin{bmatrix} 12 & -7 & 11 & -9 & 5 \\ -9 & 4 & -8 & 7 & -3 \\ -6 & 11 & -7 & 3 & -9 \\ 4 & -6 & 10 & -5 & 12 \end{bmatrix}$

50. $\begin{bmatrix} 8 & 11 & -6 & -7 & 13 \\ -7 & -8 & 5 & 6 & -9 \\ 11 & 7 & -7 & -9 & -6 \\ -3 & 4 & 1 & 8 & 7 \end{bmatrix}$

T **51.** Find a column of the matrix in Exercise 49 that can be deleted and yet have the remaining matrix columns still span $\mathbb{R}^4$.

T **52.** Find a column of the matrix in Exercise 50 that can be deleted and yet have the remaining matrix columns still span $\mathbb{R}^4$. Can you delete more than one column?

Solutions to Practice Problems

1. The matrix equation

$$\begin{bmatrix} 1 & 5 & -2 & 0 \\ -3 & 1 & 9 & -5 \\ 4 & -8 & -1 & 7 \end{bmatrix}\begin{bmatrix} 3 \\ -2 \\ 0 \\ -4 \end{bmatrix} = \begin{bmatrix} -7 \\ 9 \\ 0 \end{bmatrix}$$

is equivalent to the vector equation

$$3\begin{bmatrix}1\\-3\\4\end{bmatrix}-2\begin{bmatrix}5\\1\\-8\end{bmatrix}+0\begin{bmatrix}-2\\9\\-1\end{bmatrix}-4\begin{bmatrix}0\\-5\\7\end{bmatrix}=\begin{bmatrix}-7\\9\\0\end{bmatrix},$$

which expresses $\mathbf{b}$ as a linear combination of the columns of A.

2. $$\mathbf{u}+\mathbf{v}=\begin{bmatrix}4\\-1\end{bmatrix}+\begin{bmatrix}-3\\5\end{bmatrix}=\begin{bmatrix}1\\4\end{bmatrix}$$

$$A(\mathbf{u}+\mathbf{v})=\begin{bmatrix}2&5\\3&1\end{bmatrix}\begin{bmatrix}1\\4\end{bmatrix}=\begin{bmatrix}2+20\\3+4\end{bmatrix}=\begin{bmatrix}22\\7\end{bmatrix}$$

$$A\mathbf{u}+A\mathbf{v}=\begin{bmatrix}2&5\\3&1\end{bmatrix}\begin{bmatrix}4\\-1\end{bmatrix}+\begin{bmatrix}2&5\\3&1\end{bmatrix}\begin{bmatrix}-3\\5\end{bmatrix}$$

$$=\begin{bmatrix}3\\11\end{bmatrix}+\begin{bmatrix}19\\-4\end{bmatrix}=\begin{bmatrix}22\\7\end{bmatrix}$$

Remark: There are, in fact, infinitely many correct solutions to Practice Problem 3. When creating matrices to satisfy specified criteria, it is often useful to create matrices that are straightforward, such as those already in reduced echelon form. Here is one possible solution:

3. Let

$$A=\begin{bmatrix}1&0&1\\0&1&1\\0&0&0\end{bmatrix},\ \mathbf{b}=\begin{bmatrix}3\\2\\0\end{bmatrix},\text{ and }\mathbf{c}=\begin{bmatrix}3\\2\\1\end{bmatrix}.$$

Notice the reduced echelon form of the augmented matrix corresponding to $A\mathbf{x}=\mathbf{b}$ is

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

which corresponds to a consistent system, and hence $A\mathbf{x}=\mathbf{b}$ has solutions. The reduced echelon form of the augmented matrix corresponding to $A\mathbf{x}=\mathbf{c}$ is

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

which corresponds to an inconsistent system, and hence $A\mathbf{x}=\mathbf{c}$ does not have any solutions.

1.5 Solution Sets of Linear Systems

Solution sets of linear systems are important objects of study in linear algebra. They will appear later in several different contexts. This section uses vector notation to give explicit and geometric descriptions of such solution sets.

Homogeneous Linear Systems

A system of linear equations is said to be **homogeneous** if it can be written in the form $A\mathbf{x} = \mathbf{0}$, where A is an $m \times n$ matrix and $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$. Such a system $A\mathbf{x} = \mathbf{0}$ *always* has at least one solution, namely $\mathbf{x} = \mathbf{0}$ (the zero vector in $\mathbb{R}^n$). This zero solution is usually called the **trivial solution**. For a given equation $A\mathbf{x} = \mathbf{0}$, the important question is whether there exists a **nontrivial solution**, that is, a nonzero vector $\mathbf{x}$ that satisfies $A\mathbf{x} = \mathbf{0}$. The Existence and Uniqueness Theorem in Section 1.2 (Theorem 2) leads immediately to the following fact.

> The homogeneous equation $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution if and only if the equation has at least one free variable.

EXAMPLE 1 Determine if the following homogeneous system has a nontrivial solution. Then describe the solution set.

$$\begin{aligned} 3x_1 + 5x_2 - 4x_3 &= 0 \\ -3x_1 - 2x_2 + 4x_3 &= 0 \\ 6x_1 + x_2 - 8x_3 &= 0 \end{aligned}$$

SOLUTION Let A be the matrix of coefficients of the system and row reduce the augmented matrix $[\, A \quad \mathbf{0} \,]$ to echelon form:

$$\begin{bmatrix} 3 & 5 & -4 & 0 \\ -3 & -2 & 4 & 0 \\ 6 & 1 & -8 & 0 \end{bmatrix} \sim \begin{bmatrix} 3 & 5 & -4 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & -9 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 3 & 5 & -4 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Since x_3 is a free variable, $A\mathbf{x} = \mathbf{0}$ has nontrivial solutions (one for each nonzero choice of x_3). To describe the solution set, continue the row reduction of $[\, A \quad \mathbf{0} \,]$ to *reduced* echelon form:

$$\begin{bmatrix} 1 & 0 & -\frac{4}{3} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{aligned} x_1 \qquad - \tfrac{4}{3}x_3 &= 0 \\ x_2 \qquad\quad &= 0 \\ 0 &= 0 \end{aligned}$$

Solve for the basic variables x_1 and x_2 and obtain $x_1 = \frac{4}{3}x_3$, $x_2 = 0$, with x_3 free. As a vector, the general solution of $A\mathbf{x} = \mathbf{0}$ has the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{4}{3}x_3 \\ 0 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix} = x_3\mathbf{v}, \quad \text{where } \mathbf{v} = \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix}$$

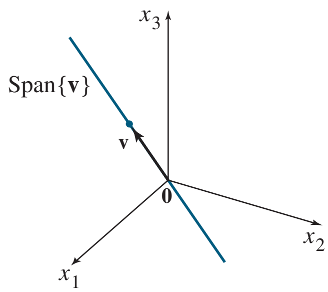


FIGURE 1

Here x_3 is factored out of the expression for the general solution vector. This shows that every solution of $A\mathbf{x} = \mathbf{0}$ in this case is a scalar multiple of $\mathbf{v}$. The trivial solution is obtained by choosing $x_3 = 0$. Geometrically, the solution set is a line through $\mathbf{0}$ in $\mathbb{R}^3$. See Figure 1. ■

Notice that a nontrivial solution $\mathbf{x}$ can have some zero entries so long as not all of its entries are zero.

EXAMPLE 2 A single linear equation can be treated as a very simple system of equations. Describe all solutions of the homogeneous "system"

$$10x_1 - 3x_2 - 2x_3 = 0 \tag{1}$$

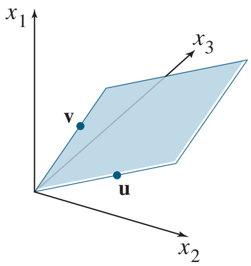


FIGURE 2

SOLUTION There is no need for matrix notation. Solve for the basic variable x_1 in terms of the free variables. The general solution is $x_1 = .3x_2 + .2x_3$, with x_2 and x_3 free. As a vector, the general solution is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} .3x_2 + .2x_3 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} .3x_2 \\ x_2 \\ 0 \end{bmatrix} + \begin{bmatrix} .2x_3 \\ 0 \\ x_3 \end{bmatrix}$$

$$= x_2 \underbrace{\begin{bmatrix} .3 \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{u}} + x_3 \underbrace{\begin{bmatrix} .2 \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{v}} \quad \text{(with } x_2, x_3 \text{ free)} \tag{2}$$

This calculation shows that every solution of (1) is a linear combination of the vectors **u** and **v**, shown in (2). That is, the solution set is Span $\{\mathbf{u}, \mathbf{v}\}$. Since neither **u** nor **v** is a scalar multiple of the other, the solution set is a plane through the origin. See Figure 2. ■

Examples 1 and 2, along with the exercises, illustrate the fact that the solution set of a homogeneous equation $A\mathbf{x} = \mathbf{0}$ can always be expressed explicitly as Span $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ for suitable vectors $\mathbf{v}_1, \dots, \mathbf{v}_p$. If the only solution is the zero vector, then the solution set is Span $\{\mathbf{0}\}$. If the equation $A\mathbf{x} = \mathbf{0}$ has only one free variable, the solution set is a line through the origin, as in Figure 1. A plane through the origin, as in Figure 2, provides a good mental image for the solution set of $A\mathbf{x} = \mathbf{0}$ when there are two or more free variables. Note, however, that a similar figure can be used to visualize Span $\{\mathbf{u}, \mathbf{v}\}$ even when **u** and **v** do not arise as solutions of $A\mathbf{x} = \mathbf{0}$. See Figure 11 in Section 1.3.

Parametric Vector Form

The original equation (1) for the plane in Example 2 is an *implicit* description of the plane. Solving this equation amounts to finding an *explicit* description of the plane as the set spanned by **u** and **v**. Equation (2) is called a **parametric vector equation** of the plane. Sometimes such an equation is written as

$$\mathbf{x} = s\mathbf{u} + t\mathbf{v} \quad (s, t \text{ in } \mathbb{R})$$

to emphasize that the parameters vary over all real numbers. In Example 1, the equation $\mathbf{x} = x_3\mathbf{v}$ (with x_3 free), or $\mathbf{x} = t\mathbf{v}$ (with t in $\mathbb{R}$), is a parametric vector equation of a line. Whenever a solution set is described explicitly with vectors as in Examples 1 and 2, we say that the solution is in **parametric vector form**.

Solutions of Nonhomogeneous Systems

When a nonhomogeneous linear system has many solutions, the general solution can be written in parametric vector form as one vector plus an arbitrary linear combination of vectors that satisfy the corresponding homogeneous system.

EXAMPLE 3 Describe all solutions of $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 3 & 5 & -4 \\ -3 & -2 & 4 \\ 6 & 1 & -8 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 7 \\ -1 \\ -4 \end{bmatrix}$$

SOLUTION Here A is the matrix of coefficients from Example 1. Row operations on $[\,A \quad \mathbf{b}\,]$ produce

$$\begin{bmatrix} 3 & 5 & -4 & 7 \\ -3 & -2 & 4 & -1 \\ 6 & 1 & -8 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -\frac{4}{3} & -1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \begin{aligned} x_1 \quad\quad - \tfrac{4}{3}x_3 &= -1 \\ x_2 \quad\quad\quad &= 2 \\ 0 &= 0 \end{aligned}$$

Thus $x_1 = -1 + \frac{4}{3}x_3$, $x_2 = 2$, and x_3 is free. As a vector, the general solution of $A\mathbf{x} = \mathbf{b}$ has the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 + \frac{4}{3}x_3 \\ 2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{4}{3}x_3 \\ 0 \\ x_3 \end{bmatrix} = \underset{\substack{\uparrow \\ \mathbf{p}}}{\begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}} + x_3 \underset{\substack{\uparrow \\ \mathbf{v}}}{\begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix}}$$

The equation $\mathbf{x} = \mathbf{p} + x_3\mathbf{v}$, or, writing t as a general parameter,

$$\mathbf{x} = \mathbf{p} + t\mathbf{v} \quad (t \text{ in } \mathbb{R}) \tag{3}$$

describes the solution set of $A\mathbf{x} = \mathbf{b}$ in parametric vector form. Recall from Example 1 that the solution set of $A\mathbf{x} = \mathbf{0}$ has the parametric vector equation

$$\mathbf{x} = t\mathbf{v} \quad (t \text{ in } \mathbb{R}) \tag{4}$$

[with the same $\mathbf{v}$ that appears in (3)]. Thus the solutions of $A\mathbf{x} = \mathbf{b}$ are obtained by adding the vector $\mathbf{p}$ to the solutions of $A\mathbf{x} = \mathbf{0}$. The vector $\mathbf{p}$ itself is just one particular solution of $A\mathbf{x} = \mathbf{b}$ [corresponding to $t = 0$ in (3)]. ■

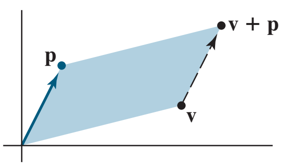


FIGURE 3
Adding $\mathbf{p}$ to $\mathbf{v}$ translates $\mathbf{v}$ to $\mathbf{v} + \mathbf{p}$.

To describe the solution set of $A\mathbf{x} = \mathbf{b}$ geometrically, we can think of vector addition as a *translation*. Given $\mathbf{v}$ and $\mathbf{p}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$, the effect of adding $\mathbf{p}$ to $\mathbf{v}$ is to *move* $\mathbf{v}$ in a direction parallel to the line through $\mathbf{p}$ and $\mathbf{0}$. We say that $\mathbf{v}$ is **translated by p** to $\mathbf{v} + \mathbf{p}$. See Figure 3. If each point on a line L in $\mathbb{R}^2$ or $\mathbb{R}^3$ is translated by a vector $\mathbf{p}$, the result is a line parallel to L. See Figure 4.

Suppose L is the line through $\mathbf{0}$ and $\mathbf{v}$, described by equation (4). Adding $\mathbf{p}$ to each point on L produces the translated line described by equation (3). Note that $\mathbf{p}$ is on the line in equation (3). We call (3) **the equation of the line through p parallel to v**. Thus *the solution set of* $A\mathbf{x} = \mathbf{b}$ *is a line through* $\mathbf{p}$ *parallel to the solution set of* $A\mathbf{x} = \mathbf{0}$. Figure 5 illustrates this case.

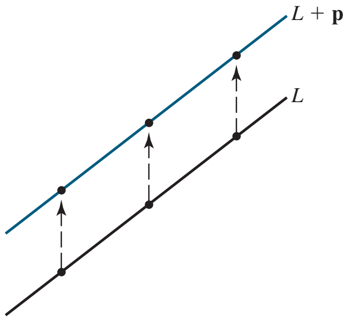


FIGURE 4
Translated line.

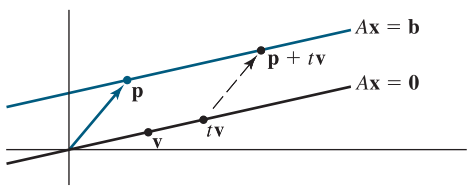


FIGURE 5 Parallel solution sets of $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{0}$.

The relation between the solution sets of $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{0}$ shown in Figure 5 generalizes to any *consistent* equation $A\mathbf{x} = \mathbf{b}$, although the solution set will be larger than a line when there are several free variables. The following theorem gives the precise statement. See Exercise 37 at the end of this section for a proof.

THEOREM 6

Suppose the equation $A\mathbf{x} = \mathbf{b}$ is consistent for some given $\mathbf{b}$, and let $\mathbf{p}$ be a solution. Then the solution set of $A\mathbf{x} = \mathbf{b}$ is the set of all vectors of the form $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$, where $\mathbf{v}_h$ is any solution of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

Theorem 6 says that if $A\mathbf{x} = \mathbf{b}$ has a solution, then the solution set is obtained by translating the solution set of $A\mathbf{x} = \mathbf{0}$, using any particular solution $\mathbf{p}$ of $A\mathbf{x} = \mathbf{b}$ for the translation. Figure 6 illustrates the case in which there are two free variables. Even when $n > 3$, our mental image of the solution set of a consistent system $A\mathbf{x} = \mathbf{b}$ (with $\mathbf{b} \neq \mathbf{0}$) is either a single nonzero point or a line or plane not passing through the origin.

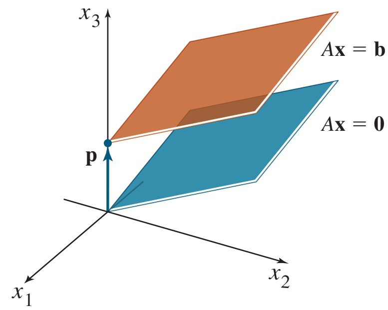


FIGURE 6 Parallel solution sets of $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{0}$.

Warning: Theorem 6 and Figure 6 apply only to an equation $A\mathbf{x} = \mathbf{b}$ that has at least one nonzero solution $\mathbf{p}$. When $A\mathbf{x} = \mathbf{b}$ has no solution, the solution set is empty.

The following algorithm outlines the calculations shown in Examples 1, 2, and 3.

WRITING A SOLUTION SET (OF A CONSISTENT SYSTEM) IN PARAMETRIC VECTOR FORM

1. Row reduce the augmented matrix to reduced echelon form.
2. Express each basic variable in terms of any free variables appearing in an equation.
3. Write a typical solution $\mathbf{x}$ as a vector whose entries depend on the free variables, if any.
4. Decompose $\mathbf{x}$ into a linear combination of vectors (with numeric entries) using the free variables as parameters.

Reasonable Answers

To verify that the solutions you found are indeed solutions to the homogeneous equation $A\mathbf{x} = \mathbf{0}$, simply multiply the matrix by each vector in your solution and check that the result is the zero vector. For example, if $A = \begin{bmatrix} 1 & -2 & 1 & 2 \\ 1 & -1 & 2 & 5 \\ 0 & 1 & 1 & 3 \end{bmatrix}$, and you found the homogeneous solutions to

Reasonable Answers (Continued)

be $x_3\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}+x_4\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}$, check $\begin{bmatrix}1&-2&1&2\\1&-1&2&5\\0&1&1&3\end{bmatrix}\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$ and $\begin{bmatrix}1&-2&1&2\\1&-1&2&5\\0&1&1&3\end{bmatrix}\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$. Then $A\left(x_3\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}+x_4\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}\right)$ $=x_3A\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}+x_4A\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}$, which is equal to $x_3\begin{bmatrix}0\\0\\0\end{bmatrix}+x_4\begin{bmatrix}0\\0\\0\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$, as desired.

If you are solving $A\mathbf{x}=\mathbf{b}$, then you can again verify that you have correct solutions by multiplying the matrix by each vector in your solutions. The product of A with the first vector (the one that is *not* part of the solution to the homogeneous equation) should be $\mathbf{b}$. The product of A with the remaining vectors (the ones that are part of the solution to the homogeneous equation) should of course be $\mathbf{0}$.

For example, to verify that $\begin{bmatrix}2\\1\\1\\2\end{bmatrix}+x_3\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}+x_4\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}$ are solutions to $A\mathbf{x}=\begin{bmatrix}5\\13\\8\end{bmatrix}$, check $\begin{bmatrix}1&-2&1&2\\1&-1&2&5\\0&1&1&3\end{bmatrix}\begin{bmatrix}2\\1\\1\\2\end{bmatrix}=\begin{bmatrix}5\\13\\8\end{bmatrix}$, and use the calculations from above. Notice $A\left(\begin{bmatrix}2\\1\\1\\2\end{bmatrix}+x_3\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}+x_4\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}\right)$ $=A\begin{bmatrix}2\\1\\1\\2\end{bmatrix}+x_3A\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}+x_4A\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}$, which is equal to $\begin{bmatrix}5\\13\\8\end{bmatrix}+x_3\begin{bmatrix}0\\0\\0\end{bmatrix}$ $+x_4\begin{bmatrix}0\\0\\0\end{bmatrix}=\begin{bmatrix}5\\13\\8\end{bmatrix}$, as desired.

Practice Problems

1. Each of the following equations determines a plane in $\mathbb{R}^3$. Do the two planes intersect? If so, describe their intersection.

$$\begin{aligned} x_1 + 4x_2 - 5x_3 &= 0 \\ 2x_1 - x_2 + 8x_3 &= 9 \end{aligned}$$

2. Write the general solution of $10x_1 - 3x_2 - 2x_3 = 7$ in parametric vector form, and relate the solution set to the one found in Example 2.
3. Prove the first part of Theorem 6: Suppose that $\mathbf{p}$ is a solution of $A\mathbf{x} = \mathbf{b}$, so that $A\mathbf{p} = \mathbf{b}$. Let $\mathbf{v}_h$ be any solution to the homogeneous equation $A\mathbf{x} = \mathbf{0}$, and let $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$. Show that $\mathbf{w}$ is a solution to $A\mathbf{x} = \mathbf{b}$.

1.5 Exercises

In Exercises 1–4, determine if the system has a nontrivial solution. Try to use as few row operations as possible.

1. $$\begin{aligned} 2x_1 - 5x_2 + 8x_3 &= 0 \\ -2x_1 - 7x_2 + x_3 &= 0 \\ 4x_1 + 2x_2 + 7x_3 &= 0 \end{aligned}$$

2. $$\begin{aligned} x_1 - 3x_2 + 7x_3 &= 0 \\ -2x_1 + x_2 - 4x_3 &= 0 \\ x_1 + 2x_2 + 9x_3 &= 0 \end{aligned}$$

3. $$\begin{aligned} -3x_1 + 5x_2 - 7x_3 &= 0 \\ -6x_1 + 7x_2 + x_3 &= 0 \end{aligned}$$

4. $$\begin{aligned} -5x_1 + 7x_2 + 9x_3 &= 0 \\ x_1 - 2x_2 + 6x_3 &= 0 \end{aligned}$$

In Exercises 5 and 6, follow the method of Examples 1 and 2 to write the solution set of the given homogeneous system in parametric vector form.

5. $$\begin{aligned} x_1 + 3x_2 + x_3 &= 0 \\ -4x_1 - 9x_2 + 2x_3 &= 0 \\ -3x_2 - 6x_3 &= 0 \end{aligned}$$

6. $$\begin{aligned} x_1 + 3x_2 - 5x_3 &= 0 \\ x_1 + 4x_2 - 8x_3 &= 0 \\ -3x_1 - 7x_2 + 9x_3 &= 0 \end{aligned}$$

In Exercises 7–12, describe all solutions of $A\mathbf{x} = \mathbf{0}$ in parametric vector form, where A is row equivalent to the given matrix.

7. $\begin{bmatrix} 1 & 3 & -3 & 7 \\ 0 & 1 & -4 & 5 \end{bmatrix}$

8. $\begin{bmatrix} 1 & -2 & -9 & 5 \\ 0 & 1 & 2 & -6 \end{bmatrix}$

9. $\begin{bmatrix} 2 & -8 & 6 \\ -1 & 4 & -3 \end{bmatrix}$

10. $\begin{bmatrix} 1 & 3 & 0 & -4 \\ 2 & 6 & 0 & -8 \end{bmatrix}$

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

12. $\begin{bmatrix} 1 & 5 & 2 & -6 & 9 & 0 \\ 0 & 0 & 1 & -7 & 4 & -8 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

You may find it helpful to review the information in the Reasonable Answers box from this section before answering Exercises 13–16.

13. Verify that the solutions you found to Exercise 9 are indeed homogeneous solutions.

14. Verify that the solutions you found to Exercise 10 are indeed homogeneous solutions.

15. Verify that the solutions you found to Exercise 11 are indeed homogeneous solutions.

16. Verify that the solutions you found to Exercise 12 are indeed homogeneous solutions.

17. Suppose the solution set of a certain system of linear equations can be described as $x_1 = 5 + 4x_3$, $x_2 = -2 - 7x_3$, with x_3 free. Use vectors to describe this set as a line in $\mathbb{R}^3$.

18. Suppose the solution set of a certain system of linear equations can be described as $x_1 = 3x_4$, $x_2 = 8 + x_4$, $x_3 = 2 - 5x_4$, with x_4 free. Use vectors to describe this set as a line in $\mathbb{R}^4$.

19. Follow the method of Example 3 to describe the solutions of the following system in parametric vector form. Also, give a geometric description of the solution set and compare it to that in Exercise 5.

$$\begin{aligned} x_1 + 3x_2 + x_3 &= 1 \\ -4x_1 - 9x_2 + 2x_3 &= -1 \\ -3x_2 - 6x_3 &= -3 \end{aligned}$$

20. As in Exercise 19, describe the solutions of the following system in parametric vector form, and provide a geometric comparison with the solution set in Exercise 6.

$$\begin{aligned} x_1 + 3x_2 - 5x_3 &= 4 \\ x_1 + 4x_2 - 8x_3 &= 7 \\ -3x_1 - 7x_2 + 9x_3 &= -6 \end{aligned}$$

21. Describe and compare the solution sets of $x_1 + 9x_2 - 4x_3 = 0$ and $x_1 + 9x_2 - 4x_3 = -2$.

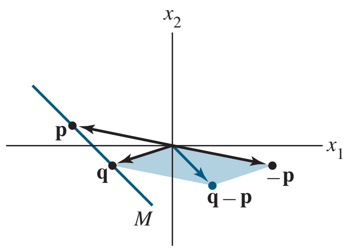
22. Describe and compare the solution sets of $x_1 - 3x_2 + 5x_3 = 0$ and $x_1 - 3x_2 + 5x_3 = 4$.

In Exercises 23 and 24, find the parametric equation of the line through **a** parallel to **b**.

23. $\mathbf{a} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -5 \\ 3 \end{bmatrix}$ **24.** $\mathbf{a} = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -4 \\ 9 \end{bmatrix}$

In Exercises 25 and 26, find a parametric equation of the line M through **p** and **q**. [*Hint:* M is parallel to the vector $\mathbf{q} - \mathbf{p}$. See the figure below.]

25. $\mathbf{p} = \begin{bmatrix} 2 \\ -5 \end{bmatrix}$, $\mathbf{q} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$ **26.** $\mathbf{p} = \begin{bmatrix} -6 \\ 3 \end{bmatrix}$, $\mathbf{q} = \begin{bmatrix} 0 \\ -4 \end{bmatrix}$



The line through **p** and **q**.

In Exercises 27–36, mark each statement True or False **(T/F)**. Justify each answer.

27. (T/F) A homogeneous equation is always consistent.

28. (T/F) If **x** is a nontrivial solution of $A\mathbf{x} = \mathbf{0}$, then every entry in **x** is nonzero.

29. (T/F) The equation $A\mathbf{x} = \mathbf{0}$ gives an explicit description of its solution set.

30. (T/F) The equation $\mathbf{x} = x_2\mathbf{u} + x_3\mathbf{v}$, with x_2 and x_3 free (and neither **u** nor **v** a multiple of the other), describes a plane through the origin.

31. (T/F) The homogeneous equation $A\mathbf{x} = \mathbf{0}$ has the trivial solution if and only if the equation has at least one free variable.

32. (T/F) The equation $A\mathbf{x} = \mathbf{b}$ is homogeneous if the zero vector is a solution.

33. (T/F) The equation $\mathbf{x} = \mathbf{p} + t\mathbf{v}$ describes a line through **v** parallel to **p**.

34. (T/F) The effect of adding **p** to a vector is to move the vector in a direction parallel to **p**.

35. (T/F) The solution set of $A\mathbf{x} = \mathbf{b}$ is the set of all vectors of the form $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$, where $\mathbf{v}_h$ is any solution of the equation $A\mathbf{x} = \mathbf{0}$.

36. (T/F) The solution set of $A\mathbf{x} = \mathbf{b}$ is obtained by translating the solution set of $A\mathbf{x} = \mathbf{0}$.

37. Prove the second part of Theorem 6: Let **w** be any solution of $A\mathbf{x} = \mathbf{b}$, and define $\mathbf{v}_h = \mathbf{w} - \mathbf{p}$. Show that $\mathbf{v}_h$ is a solution of $A\mathbf{x} = \mathbf{0}$. This shows that every solution of $A\mathbf{x} = \mathbf{b}$ has the form $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$, with **p** a particular solution of $A\mathbf{x} = \mathbf{b}$ and $\mathbf{v}_h$ a solution of $A\mathbf{x} = \mathbf{0}$.

38. Suppose $A\mathbf{x} = \mathbf{b}$ has a solution. Explain why the solution is unique precisely when $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

39. Suppose A is the 3×3 *zero* matrix (with all zero entries). Describe the solution set of the equation $A\mathbf{x} = \mathbf{0}$.

40. If $\mathbf{b} \neq \mathbf{0}$, can the solution set of $A\mathbf{x} = \mathbf{b}$ be a plane through the origin? Explain.

In Exercises 41–44, (a) does the equation $A\mathbf{x} = \mathbf{0}$ have a nontrivial solution and (b) does the equation $A\mathbf{x} = \mathbf{b}$ have at least one solution for every possible **b**?

41. A is a 3×3 matrix with three pivot positions.

42. A is a 3×3 matrix with two pivot positions.

43. A is a 3×2 matrix with two pivot positions.

44. A is a 2×4 matrix with two pivot positions.

45. Given $A = \begin{bmatrix} -2 & -6 \\ 7 & 21 \\ -3 & -9 \end{bmatrix}$, find one nontrivial solution of $A\mathbf{x} = \mathbf{0}$ by inspection. [*Hint:* Think of the equation $A\mathbf{x} = \mathbf{0}$ written as a vector equation.]

46. Given $A = \begin{bmatrix} 4 & -6 \\ -8 & 12 \\ 6 & -9 \end{bmatrix}$, find one nontrivial solution of $A\mathbf{x} = \mathbf{0}$ by inspection.

47. Construct a 3×3 nonzero matrix A such that the vector $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is a solution of $A\mathbf{x} = \mathbf{0}$.

48. Construct a 3×3 nonzero matrix A such that the vector $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ is a solution of $A\mathbf{x} = \mathbf{0}$.

49. Construct a 2×2 matrix A such that the solution set of the equation $A\mathbf{x} = \mathbf{0}$ is the line in $\mathbb{R}^2$ through $(4, 1)$ and the origin. Then, find a vector **b** in $\mathbb{R}^2$ such that the solution set of $A\mathbf{x} = \mathbf{b}$ is *not* a line in $\mathbb{R}^2$ parallel to the solution set of $A\mathbf{x} = \mathbf{0}$. Why does this *not* contradict Theorem 6?

50. Suppose A is a 3×3 matrix and **y** is a vector in $\mathbb{R}^3$ such that the equation $A\mathbf{x} = \mathbf{y}$ does *not* have a solution. Does there exist a vector **z** in $\mathbb{R}^3$ such that the equation $A\mathbf{x} = \mathbf{z}$ has a unique solution? Discuss.

51. Let A be an $m \times n$ matrix and let **u** be a vector in $\mathbb{R}^n$ that satisfies the equation $A\mathbf{x} = \mathbf{0}$. Show that for any scalar c, the vector $c\mathbf{u}$ also satisfies $A\mathbf{x} = \mathbf{0}$. [That is, show that $A(c\mathbf{u}) = \mathbf{0}$.]

52. Let A be an $m \times n$ matrix, and let **u** and **v** be vectors in $\mathbb{R}^n$ with the property that $A\mathbf{u} = \mathbf{0}$ and $A\mathbf{v} = \mathbf{0}$. Explain why $A(\mathbf{u} + \mathbf{v})$ must be the zero vector. Then explain why $A(c\mathbf{u} + d\mathbf{v}) = \mathbf{0}$ for each pair of scalars c and d.

Solutions to Practice Problems

1. Row reduce the augmented matrix:

$$\begin{bmatrix} 1 & 4 & -5 & 0 \\ 2 & -1 & 8 & 9 \end{bmatrix} \sim \begin{bmatrix} 1 & 4 & -5 & 0 \\ 0 & -9 & 18 & 9 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 3 & 4 \\ 0 & 1 & -2 & -1 \end{bmatrix}$$

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

Thus $x_1 = 4 - 3x_3$, $x_2 = -1 + 2x_3$, with x_3 free. The general solution in parametric vector form is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 - 3x_3 \\ -1 + 2x_3 \\ x_3 \end{bmatrix} = \underset{\mathbf{p}}{\begin{bmatrix} 4 \\ -1 \\ 0 \end{bmatrix}} + x_3 \underset{\mathbf{v}}{\begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}}$$

The intersection of the two planes is the line through $\mathbf{p}$ in the direction of $\mathbf{v}$.

2. The augmented matrix $\begin{bmatrix} 10 & -3 & -2 & 7 \end{bmatrix}$ is row equivalent to $\begin{bmatrix} 1 & -.3 & -.2 & .7 \end{bmatrix}$, and the general solution is $x_1 = .7 + .3x_2 + .2x_3$, with x_2 and x_3 free. That is,

$$\begin{aligned} \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} .7 + .3x_2 + .2x_3 \\ x_2 \\ x_3 \end{bmatrix} &= \begin{bmatrix} .7 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} .3 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} .2 \\ 0 \\ 1 \end{bmatrix} \\ &= \mathbf{p} + x_2\mathbf{u} + x_3\mathbf{v} \end{aligned}$$

The solution set of the nonhomogeneous equation $A\mathbf{x} = \mathbf{b}$ is the translated plane $\mathbf{p} + \text{Span}\,\{\mathbf{u}, \mathbf{v}\}$, which passes through $\mathbf{p}$ and is parallel to the solution set of the homogeneous equation in Example 2.

3. Using Theorem 5 from Section 1.4, notice

$$A(\mathbf{p} + \mathbf{v}_h) = A\mathbf{p} + A\mathbf{v}_h = \mathbf{b} + \mathbf{0} = \mathbf{b},$$

hence $\mathbf{p} + \mathbf{v}_h$ is a solution to $A\mathbf{x} = \mathbf{b}$.

1.6 Applications of Linear Systems

You might expect that a real-life problem involving linear algebra would have only one solution, or perhaps no solution. The purpose of this section is to show how linear systems with many solutions can arise naturally. The applications here come from economics, chemistry, and network flow.

A Homogeneous System in Economics

The system of 500 equations in 500 variables, mentioned in this chapter's introduction, is now known as a Leontief "input–output" (or "production") model.[1] Section 2.6 will examine this model in more detail, when more theory and better notation are available. For now, we look at a simpler "exchange model," also due to Leontief.

[1] See Wassily W. Leontief, "Input–Output Economics," *Scientific American*, October 1951, pp. 15–21.

Suppose a nation's economy is divided into many sectors, such as various manufacturing, communication, entertainment, and service industries. Suppose that for each sector we know its total output for one year and we know exactly how this output is divided or "exchanged" among the other sectors of the economy. Let the total dollar value of a sector's output be called the **price** of that output. Leontief proved the following result.

> There exist *equilibrium prices* that can be assigned to the total outputs of the various sectors in such a way that the income of each sector exactly balances its expenses.

The following example shows how to find the equilibrium prices.

EXAMPLE 1 Suppose an economy consists of the Coal, Electric (power), and Steel sectors, and the output of each sector is distributed among the various sectors as shown in Table 1, where the entries in a column represent the fractional parts of a sector's total output.

The second column of Table 1, for instance, says that the total output of the Electric sector is divided as follows: 40% to Coal, 50% to Steel, and the remaining 10% to Electric. (Electric treats this 10% as an expense it incurs in order to operate its business.) Since all output must be taken into account, the decimal fractions in each column must sum to 1.

Denote the prices (in dollar values) of the total annual outputs of the Coal, Electric, and Steel sectors by p_C, p_E, and p_S, respectively. If possible, find equilibrium prices that make each sector's income match its expenditures.

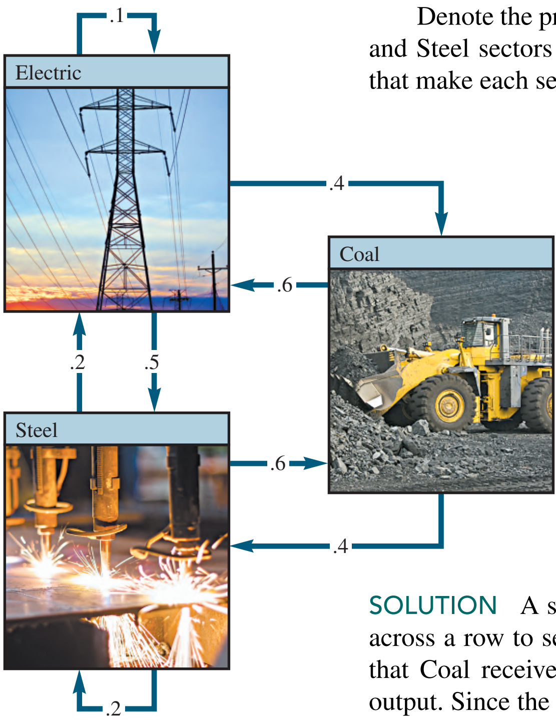


TABLE 1 A Simple Economy

Distribution of Output from			
Coal	**Electric**	**Steel**	**Purchased by**
.0	.4	.6	Coal
.6	.1	.2	Electric
.4	.5	.2	Steel

SOLUTION A sector looks down a column to see where its output goes, and it looks across a row to see what it needs as inputs. For instance, the first row of Table 1 says that Coal receives (and pays for) 40% of the Electric output and 60% of the Steel output. Since the respective values of the total outputs are p_E and p_S, Coal must spend $.4p_E$ dollars for its share of Electric's output and $.6p_S$ for its share of Steel's output. Thus Coal's total expenses are $.4p_E + .6p_S$. To make Coal's income, p_C, equal to its expenses, we want

$$p_C = .4p_E + .6p_S \tag{1}$$

The second row of the exchange table shows that the Electric sector spends $.6p_C$ for coal, $.1p_E$ for electricity, and $.2p_S$ for steel. Hence the income/expense requirement for Electric is

$$p_E = .6p_C + .1p_E + .2p_S \tag{2}$$

Finally, the third row of the exchange table leads to the final requirement:

$$p_S = .4p_C + .5p_E + .2p_S \tag{3}$$

To solve the system of equations (1), (2), and (3), move all the unknowns to the left sides of the equations and combine like terms. [For instance, on the left side of (2), write $p_E - .1p_E$ as $.9p_E$.]

$$\begin{aligned} p_C - .4p_E - .6p_S &= 0 \\ -.6p_C + .9p_E - .2p_S &= 0 \\ -.4p_C - .5p_E + .8p_S &= 0 \end{aligned}$$

Row reduction is next. For simplicity here, decimals are rounded to two places.

$$\begin{bmatrix} 1 & -.4 & -.6 & 0 \\ -.6 & .9 & -.2 & 0 \\ -.4 & -.5 & .8 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -.4 & -.6 & 0 \\ 0 & .66 & -.56 & 0 \\ 0 & -.66 & .56 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -.4 & -.6 & 0 \\ 0 & .66 & -.56 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & -.4 & -.6 & 0 \\ 0 & 1 & -.85 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -.94 & 0 \\ 0 & 1 & -.85 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The general solution is $p_C = .94p_S$, $p_E = .85p_S$, and p_S is free. The equilibrium price vector for the economy has the form

$$\mathbf{p} = \begin{bmatrix} p_C \\ p_E \\ p_S \end{bmatrix} = \begin{bmatrix} .94p_S \\ .85p_S \\ p_S \end{bmatrix} = p_S \begin{bmatrix} .94 \\ .85 \\ 1 \end{bmatrix}$$

Any (nonnegative) choice for p_S results in a choice of equilibrium prices. For instance, if we take p_S to be 100 (or \$100 million), then $p_C = 94$ and $p_E = 85$. The incomes and expenditures of each sector will be equal if the output of Coal is priced at \$94 million, that of Electric at \$85 million, and that of Steel at \$100 million. ■

Balancing Chemical Equations

Chemical equations describe the quantities of substances consumed and produced by chemical reactions. For instance, when propane gas burns, the propane (C_3H_8) combines with oxygen (O_2) to form carbon dioxide (CO_2) and water (H_2O), according to an equation of the form

$$(x_1)C_3H_8 + (x_2)O_2 \to (x_3)CO_2 + (x_4)H_2O \tag{4}$$

To "balance" this equation, a chemist must find whole numbers $x_1, \dots, x_4$ such that the total numbers of carbon (C), hydrogen (H), and oxygen (O) atoms on the left match the corresponding numbers of atoms on the right (because atoms are neither destroyed nor created in the reaction).

A systematic method for balancing chemical equations is to set up a vector equation that describes the numbers of atoms of each type present in a reaction. Since equation (4) involves three types of atoms (carbon, hydrogen, and oxygen), construct a vector in $\mathbb{R}^3$ for each reactant and product in (4) that lists the numbers of "atoms per molecule," as follows:

$$C_3H_8: \begin{bmatrix} 3 \\ 8 \\ 0 \end{bmatrix}, \quad O_2: \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}, \quad CO_2: \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \quad H_2O: \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \begin{matrix} \leftarrow \text{Carbon} \\ \leftarrow \text{Hydrogen} \\ \leftarrow \text{Oxygen} \end{matrix}$$

To balance equation (4), the coefficients $x_1, \dots, x_4$ must satisfy

$$x_1\begin{bmatrix}3\\8\\0\end{bmatrix} + x_2\begin{bmatrix}0\\0\\2\end{bmatrix} = x_3\begin{bmatrix}1\\0\\2\end{bmatrix} + x_4\begin{bmatrix}0\\2\\1\end{bmatrix}$$

To solve, move all the terms to the left (changing the signs in the third and fourth vectors):

$$x_1\begin{bmatrix}3\\8\\0\end{bmatrix} + x_2\begin{bmatrix}0\\0\\2\end{bmatrix} + x_3\begin{bmatrix}-1\\0\\-2\end{bmatrix} + x_4\begin{bmatrix}0\\-2\\-1\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$$

Row reduction of the augmented matrix for this equation leads to the general solution

$$x_1 = \tfrac{1}{4}x_4,\ x_2 = \tfrac{5}{4}x_4,\ x_3 = \tfrac{3}{4}x_4,\ \text{ with } x_4 \text{ free}$$

Since the coefficients in a chemical equation must be integers, take $x_4 = 4$, in which case $x_1 = 1$, $x_2 = 5$, and $x_3 = 3$. The balanced equation is

$$C_3H_8 + 5O_2 \rightarrow 3CO_2 + 4H_2O$$

The equation would also be balanced if, for example, each coefficient were doubled. For most purposes, however, chemists prefer to use a balanced equation whose coefficients are the smallest possible whole numbers.

Network Flow

Systems of linear equations arise naturally when scientists, engineers, or economists study the flow of some quantity through a network. For instance, urban planners and traffic engineers monitor the pattern of traffic flow in a grid of city streets. Electrical engineers calculate current flow through electrical circuits. Economists analyze the distribution of products from manufacturers to consumers through a network of wholesalers and retailers. For many networks, the systems of equations involve hundreds or even thousands of variables and equations.

A *network* consists of a set of points called *junctions*, or *nodes*, with lines or arcs called *branches* connecting some or all of the junctions. The direction of flow in each branch is indicated, and the flow amount (or rate) is either shown or is denoted by a variable.

The basic assumption of network flow is that the total flow into the network equals the total flow out of the network and that the total flow into a junction equals the total flow out of the junction. For example, Figure 1 shows 30 units flowing into a junction through one branch, with x_1 and x_2 denoting the flows out of the junction through other branches. Since the flow is "conserved" at each junction, we must have $x_1 + x_2 = 30$. In a similar fashion, the flow at each junction is described by a linear equation. The problem of network analysis is to determine the flow in each branch when partial information (such as the flow into and out of the network) is known.

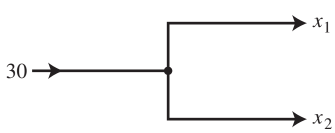


FIGURE 1
A junction or node.

EXAMPLE 2 The network in Figure 2 shows the traffic flow (in vehicles per hour) over several one-way streets in downtown Baltimore during a typical early afternoon. Determine the general flow pattern for the network.

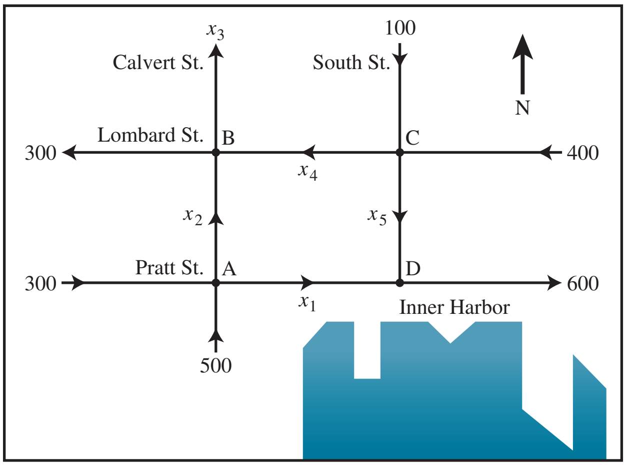


FIGURE 2 Baltimore streets.

SOLUTION Write equations that describe the flow, and then find the general solution of the system. Label the street intersections (junctions) and the unknown flows in the branches, as shown in Figure 2. At each intersection, set the flow in equal to the flow out.

Intersection	Flow in		Flow out
A	$300 + 500$	$=$	$x_1 + x_2$
B	$x_2 + x_4$	$=$	$300 + x_3$
C	$100 + 400$	$=$	$x_4 + x_5$
D	$x_1 + x_5$	$=$	600

Also, the total flow into the network $(500 + 300 + 100 + 400)$ equals the total flow out of the network $(300 + x_3 + 600)$, which simplifies to $x_3 = 400$. Combine this equation with a rearrangement of the first four equations to obtain the following system of equations:

$$\begin{aligned} x_1 + x_2 \qquad\qquad\qquad &= 800 \\ x_2 - x_3 + x_4 \qquad &= 300 \\ x_4 + x_5 &= 500 \\ x_1 \qquad\qquad\qquad\quad + x_5 &= 600 \\ x_3 \qquad\qquad &= 400 \end{aligned}$$

Row reduction of the associated augmented matrix leads to

$$\begin{aligned} x_1 \qquad\qquad + x_5 &= 600 \\ x_2 \qquad - x_5 &= 200 \\ x_3 \qquad &= 400 \\ x_4 + x_5 &= 500 \end{aligned}$$

The general flow pattern for the network is described by

$$\begin{cases} x_1 = 600 - x_5 \\ x_2 = 200 + x_5 \\ x_3 = 400 \\ x_4 = 500 - x_5 \\ x_5 \text{ is free} \end{cases}$$

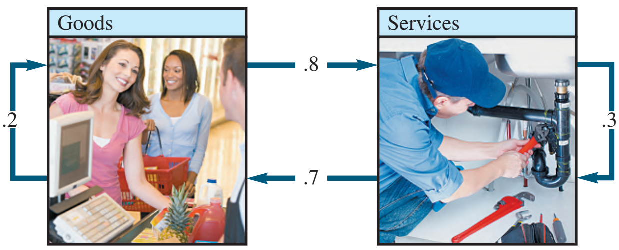
A negative flow in a network branch corresponds to flow in the direction opposite to that shown on the model. Since the streets in this problem are one way, none of the variables here can be negative. This fact leads to certain limitations on the possible values of the variables. For instance, $x_5 \le 500$ because x_4 cannot be negative. Other constraints on the variables are considered in Practice Problem 2. ■

Practice Problems

1. Suppose an economy has three sectors: Agriculture, Mining, and Manufacturing. Agriculture sells 5% of its output to Mining and 30% to Manufacturing, and retains the rest. Mining sells 20% of its output to Agriculture and 70% to Manufacturing, and retains the rest. Manufacturing sells 20% of its output to Agriculture and 30% to Mining, and retains the rest. Determine the exchange table for this economy, where the columns describe how the output of each sector is exchanged among the three sectors.
2. Consider the network flow studied in Example 2. Determine the possible range of values of x_1 and x_2. [*Hint:* The example showed that $x_5 \le 500$. What does this imply about x_1 and x_2? Also, use the fact that $x_5 \ge 0$.]

1.6 Exercises

1. Suppose an economy has only two sectors, Goods and Services. Each year, Goods sells 80% of its output to Services and keeps the rest, while Services sells 70% of its output to Goods and retains the rest. Find equilibrium prices for the annual outputs of the Goods and Services sectors that make each sector's income match its expenditures.



2. Find another set of equilibrium prices for the economy in Example 1. Suppose the same economy used Japanese yen instead of dollars to measure the value of the various sectors' outputs. Would this change the problem in any way? Discuss.

3. Consider an economy with three sectors, Chemicals & Metals, Fuels & Power, and Machinery. Chemicals sells 30% of its output to Fuels and 50% to Machinery and retains the rest. Fuels sells 80% of its output to Chemicals and 10% to Machinery and retains the rest. Machinery sells 40% to Chemicals and 40% to Fuels and retains the rest.

   a. Construct the exchange table for this economy.

   b. Develop a system of equations that leads to prices at which each sector's income matches its expenses. Then write the augmented matrix that can be row reduced to find these prices.

   c. Find a set of equilibrium prices when the price for the Machinery output is 100 units.

4. Suppose an economy has four sectors, Agriculture (A), Energy (E), Manufacturing (M), and Transportation (T). Sector A sells 10% of its output to E and 25% to M and retains the rest. Sector E sells 30% of its output to A, 35% to M, and 25% to T and retains the rest. Sector M sells 30% of its output to A, 15% to E, and 40% to T and retains the rest. Sector T sells 20% of its output to A, 10% to E, and 30% to M and retains the rest.

   a. Construct the exchange table for this economy.

   b. Find a set of equilibrium prices for the economy.

Balance the chemical equations in Exercises 5–10 using the vector equation approach discussed in this section.

5. Boron sulfide reacts violently with water to form boric acid and hydrogen sulfide gas (the smell of rotten eggs). The unbalanced equation is

   $B_2S_3 + H_2O \rightarrow H_3BO_3 + H_2S$

   [For each compound, construct a vector that lists the numbers of atoms of boron, sulfur, hydrogen, and oxygen.]

6. When solutions of sodium phosphate and barium nitrate are mixed, the result is barium phosphate (as a precipitate) and sodium nitrate. The unbalanced equation is

   $Na_3PO_4 + Ba(NO_3)_2 \rightarrow Ba_3(PO_4)_2 + NaNO_3$

   [For each compound, construct a vector that lists the numbers of atoms of sodium (Na), phosphorus, oxygen, barium, and nitrogen. For instance, barium nitrate corresponds to $(0, 0, 6, 1, 2)$.]

7. Alka-Seltzer contains sodium bicarbonate ($NaHCO_3$) and citric acid ($H_3C_6H_5O_7$). When a tablet is dissolved in water, the following reaction produces sodium citrate, water, and carbon dioxide (gas):

$$NaHCO_3 + H_3C_6H_5O_7 \rightarrow Na_3C_6H_5O_7 + H_2O + CO_2$$

8. The following reaction between potassium permanganate ($KMnO_4$) and manganese sulfate in water produces manganese dioxide, potassium sulfate, and sulfuric acid:

$$KMnO_4 + MnSO_4 + H_2O \rightarrow MnO_2 + K_2SO_4 + H_2SO_4$$

[For each compound, construct a vector that lists the numbers of atoms of potassium (K), manganese, oxygen, sulfur, and hydrogen.]

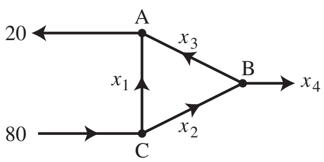
T 9. If possible, use exact arithmetic or rational format for calculations in balancing the following chemical reaction:

$$PbN_6 + CrMn_2O_8 \rightarrow Pb_3O_4 + Cr_2O_3 + MnO_2 + NO$$

T 10. The chemical reaction below can be used in some industrial processes, such as the production of arsene (AsH_3). Use exact arithmetic or rational format for calculations to balance this equation.

$$MnS + As_2Cr_{10}O_{35} + H_2SO_4 \rightarrow HMnO_4 + AsH_3 + CrS_3O_{12} + H_2O$$

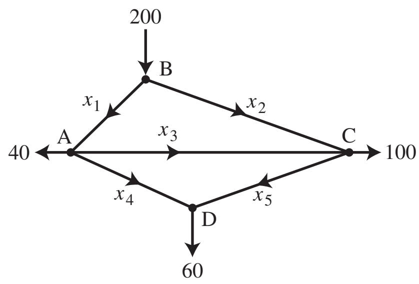
11. Find the general flow pattern of the network shown in the figure. Assuming that the flows are all nonnegative, what is the largest possible value for x_3?



12. a. Find the general traffic pattern in the freeway network shown in the figure. (Flow rates are in cars/minute.)

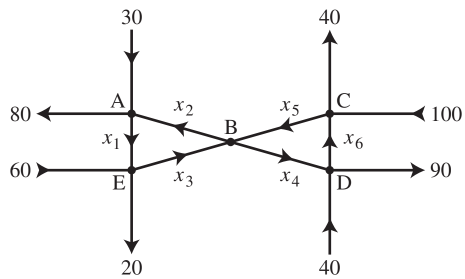
b. Describe the general traffic pattern when the road whose flow is x_4 is closed.

c. When $x_4 = 0$, what is the minimum value of x_1?

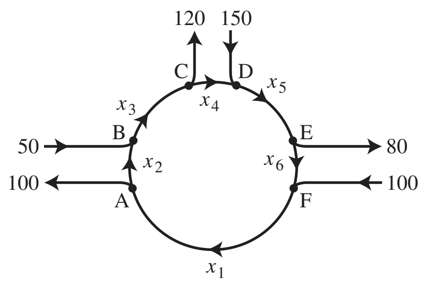


13. a. Find the general flow pattern in the network shown in the figure.

b. Assuming that the flow must be in the directions indicated, find the minimum flows in the branches denoted by x_2, x_3, x_4, and x_5.



14. Intersections in England are often constructed as one-way "roundabouts," such as the one shown in the figure. Assume that traffic must travel in the directions shown. Find the general solution of the network flow. Find the smallest possible value for x_6.



Solutions to Practice Problems

1. Write the percentages as decimals. Since all output must be taken into account, each column must sum to 1. This fact helps to fill in any missing entries.

Distribution of Output from			
Agriculture	**Mining**	**Manufacturing**	**Purchased by**
.65	.20	.20	Agriculture
.05	.10	.30	Mining
.30	.70	.50	Manufacturing

Solutions to Practice Problems (Continued)

2. Since $x_5 \leq 500$, the equations D and A for x_1 and x_2 imply that $x_1 \geq 100$ and $x_2 \leq 700$. The fact that $x_5 \geq 0$ implies that $x_1 \leq 600$ and $x_2 \geq 200$. So, $100 \leq x_1 \leq 600$, and $200 \leq x_2 \leq 700$.

1.7 Linear Independence

The homogeneous equations in Section 1.5 can be studied from a different perspective by writing them as vector equations. In this way, the focus shifts from the unknown solutions of $A\mathbf{x} = \mathbf{0}$ to the vectors that appear in the vector equations.

For instance, consider the equation

$$x_1\begin{bmatrix}1\\2\\3\end{bmatrix} + x_2\begin{bmatrix}4\\5\\6\end{bmatrix} + x_3\begin{bmatrix}2\\1\\0\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix} \tag{1}$$

This equation has a trivial solution, of course, where $x_1 = x_2 = x_3 = 0$. As in Section 1.5, the main issue is whether the trivial solution is the *only one*.

DEFINITION

An indexed set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ is said to be **linearly independent** if the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_p\mathbf{v}_p = \mathbf{0}$$

has only the trivial solution. The set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is said to be **linearly dependent** if there exist weights $c_1, \dots, c_p$, not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0} \tag{2}$$

Equation (2) is called a **linear dependence relation** among $\mathbf{v}_1, \dots, \mathbf{v}_p$ when the weights are not all zero. An indexed set is linearly dependent if and only if it is not linearly independent. For brevity, we may say that $\mathbf{v}_1, \dots, \mathbf{v}_p$ are linearly dependent when we mean that $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is a linearly dependent set. We use analogous terminology for linearly independent sets.

EXAMPLE 1 Let $\mathbf{v}_1 = \begin{bmatrix}1\\2\\3\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}4\\5\\6\end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix}2\\1\\0\end{bmatrix}$.

a. Determine if the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent.

b. If possible, find a linear dependence relation among $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$.

SOLUTION

a. We must determine if there is a nontrivial solution of equation (1) above. Row operations on the associated augmented matrix show that

$$\begin{bmatrix}1 & 4 & 2 & 0\\2 & 5 & 1 & 0\\3 & 6 & 0 & 0\end{bmatrix} \sim \begin{bmatrix}1 & 4 & 2 & 0\\0 & -3 & -3 & 0\\0 & 0 & 0 & 0\end{bmatrix}$$

Clearly, x_1 and x_2 are basic variables, and x_3 is free. Each nonzero value of x_3 determines a nontrivial solution of (1). Hence $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are linearly dependent (and not linearly independent).

b. To find a linear dependence relation among $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$, completely row reduce the augmented matrix and write the new system:

$$\begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{aligned} x_1 \qquad - 2x_3 &= 0 \\ x_2 + x_3 &= 0 \\ 0 &= 0 \end{aligned}$$

Thus $x_1 = 2x_3$, $x_2 = -x_3$, and x_3 is free. Choose any nonzero value for x_3—say, $x_3 = 5$. Then $x_1 = 10$ and $x_2 = -5$. Substitute these values into equation (1) and obtain

$$10\mathbf{v}_1 - 5\mathbf{v}_2 + 5\mathbf{v}_3 = \mathbf{0}$$

This is one (out of infinitely many) possible linear dependence relations among $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. ■

Linear Independence of Matrix Columns

Suppose that we begin with a matrix $A = [\,\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_n\,]$ instead of a set of vectors. The matrix equation $A\mathbf{x} = \mathbf{0}$ can be written as

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{0}$$

Each linear dependence relation among the columns of A corresponds to a nontrivial solution of $A\mathbf{x} = \mathbf{0}$. Thus we have the following important fact.

> The columns of a matrix A are linearly independent if and only if the equation $A\mathbf{x} = \mathbf{0}$ has *only* the trivial solution. (3)

EXAMPLE 2 Determine if the columns of the matrix $A = \begin{bmatrix} 0 & 1 & 4 \\ 1 & 2 & -1 \\ 5 & 8 & 0 \end{bmatrix}$ are linearly independent.

SOLUTION To study $A\mathbf{x} = \mathbf{0}$, row reduce the augmented matrix:

$$\begin{bmatrix} 0 & 1 & 4 & 0 \\ 1 & 2 & -1 & 0 \\ 5 & 8 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 4 & 0 \\ 0 & -2 & 5 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 4 & 0 \\ 0 & 0 & 13 & 0 \end{bmatrix}$$

At this point, it is clear that there are three basic variables and no free variables. So the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, and the columns of A are linearly independent. ■

Sets of One or Two Vectors

A set containing only one vector—say, $\mathbf{v}$—is linearly independent if and only if $\mathbf{v}$ is not the zero vector. This is because the vector equation $x_1\mathbf{v} = \mathbf{0}$ has only the trivial solution when $\mathbf{v} \neq \mathbf{0}$. The zero vector is linearly dependent because $x_1\mathbf{0} = \mathbf{0}$ has many nontrivial solutions.

The next example will explain the nature of a linearly dependent set of two vectors.

EXAMPLE 3 Determine if the following sets of vectors are linearly independent.

a. $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$ b. $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$

SOLUTION

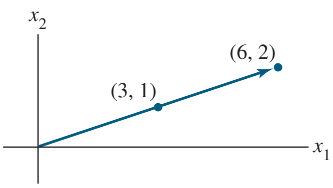
a. Notice that $\mathbf{v}_2$ is a multiple of $\mathbf{v}_1$, namely $\mathbf{v}_2 = 2\mathbf{v}_1$. Hence $-2\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{0}$, which shows that $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly dependent.

b. The vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are certainly *not* multiples of one another. Could they be linearly dependent? Suppose c and d satisfy

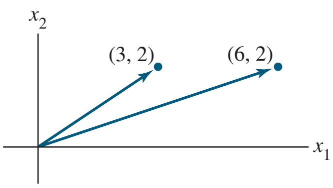
$$c\mathbf{v}_1 + d\mathbf{v}_2 = \mathbf{0}$$

If $c \neq 0$, then we can solve for $\mathbf{v}_1$ in terms of $\mathbf{v}_2$, namely $\mathbf{v}_1 = (-d/c)\mathbf{v}_2$. This result is impossible because $\mathbf{v}_1$ is *not* a multiple of $\mathbf{v}_2$. So c must be zero. Similarly, d must also be zero. Thus $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a linearly independent set. ■

The arguments in Example 3 show that you can always decide *by inspection* when a set of two vectors is linearly dependent. Row operations are unnecessary. Simply check whether at least one of the vectors is a scalar times the other. (The test applies only to sets of *two* vectors.)



Linearly dependent



Linearly independent

FIGURE 1

A set of two vectors $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly dependent if at least one of the vectors is a multiple of the other. The set is linearly independent if and only if neither of the vectors is a multiple of the other.

In geometric terms, two vectors are linearly dependent if and only if they lie on the same line through the origin. Figure 1 shows the vectors from Example 3.

Sets of Two or More Vectors

The proof of the next theorem is similar to the solution of Example 3. Details are given at the end of this section.

THEOREM 7

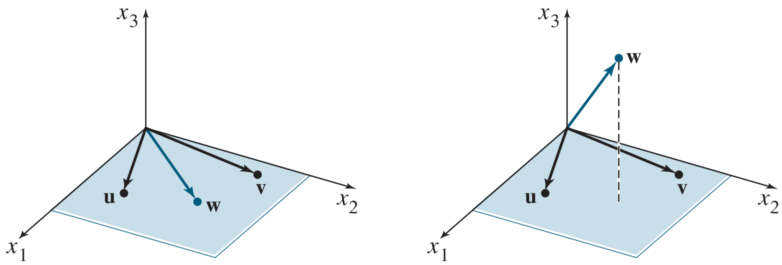
Characterization of Linearly Dependent Sets

An indexed set $S = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ of two or more vectors is linearly dependent if and only if at least one of the vectors in S is a linear combination of the others. In fact, if S is linearly dependent and $\mathbf{v}_1 \neq \mathbf{0}$, then some $\mathbf{v}_j$ (with $j > 1$) is a linear combination of the preceding vectors, $\mathbf{v}_1, \dots, \mathbf{v}_{j-1}$.

Warning: Theorem 7 does *not* say that *every* vector in a linearly dependent set is a linear combination of the preceding vectors. A vector in a linearly dependent set may fail to be a linear combination of the other vectors. See Practice Problem 1(c).

EXAMPLE 4 Let $\mathbf{u} = \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix}$. Describe the set spanned by $\mathbf{u}$ and $\mathbf{v}$, and explain why a vector $\mathbf{w}$ is in Span $\{\mathbf{u}, \mathbf{v}\}$ if and only if $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent.

SOLUTION The vectors $\mathbf{u}$ and $\mathbf{v}$ are linearly independent because neither vector is a multiple of the other, and so they span a plane in $\mathbb{R}^3$. (See Section 1.3.) In fact, Span $\{\mathbf{u}, \mathbf{v}\}$ is the x_1x_2-plane (with $x_3 = 0$). If $\mathbf{w}$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$, then $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent, by Theorem 7. Conversely, suppose that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent. By Theorem 7, some vector in $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a linear combination of the preceding vectors (since $\mathbf{u} \neq \mathbf{0}$). That vector must be $\mathbf{w}$, since $\mathbf{v}$ is not a multiple of $\mathbf{u}$. So $\mathbf{w}$ is in Span $\{\mathbf{u}, \mathbf{v}\}$. See Figure 2. ■



Linearly dependent, $\mathbf{w}$ in Span $\{\mathbf{u}, \mathbf{v}\}$

Linearly independent, $\mathbf{w}$ *not* in Span $\{\mathbf{u}, \mathbf{v}\}$

FIGURE 2 Linear dependence in $\mathbb{R}^3$.

Example 4 generalizes to any set $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ in $\mathbb{R}^3$ with $\mathbf{u}$ and $\mathbf{v}$ linearly independent. The set $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ will be linearly dependent if and only if $\mathbf{w}$ is in the plane spanned by $\mathbf{u}$ and $\mathbf{v}$.

The next two theorems describe special cases in which the linear dependence of a set is automatic. Moreover, Theorem 8 will be a key result for work in later chapters.

THEOREM 8

If a set contains more vectors than there are entries in each vector, then the set is linearly dependent. That is, any set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ is linearly dependent if $p > n$.

$$\begin{array}{cc} & p \\ n & \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix} \end{array}$$

FIGURE 3
If $p > n$, the columns are linearly dependent.

PROOF Let $A = [\,\mathbf{v}_1 \ \cdots \ \mathbf{v}_p\,]$. Then A is $n \times p$, and the equation $A\mathbf{x} = \mathbf{0}$ corresponds to a system of n equations in p unknowns. If $p > n$, there are more variables than equations, so there must be a free variable. Hence $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution, and the columns of A are linearly dependent. See Figure 3 for a matrix version of this theorem. ■

Warning: Theorem 8 says nothing about the case in which the number of vectors in the set does *not* exceed the number of entries in each vector.

EXAMPLE 5 The vectors $\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ -1 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \end{bmatrix}$ are linearly dependent by Theorem 8, because there are three vectors in the set and there are only two entries in each vector. Notice, however, that none of the vectors is a multiple of one of the other vectors. See Figure 4. ■

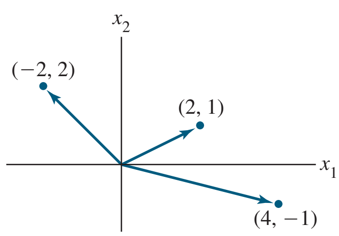


FIGURE 4
A linearly dependent set in $\mathbb{R}^2$.

THEOREM 9

If a set $S = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ contains the zero vector, then the set is linearly dependent.

PROOF By renumbering the vectors, we may suppose $\mathbf{v}_1 = \mathbf{0}$. Then the equation $1\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_p = \mathbf{0}$ shows that S is linearly dependent. ■

EXAMPLE 6 Determine by inspection if the given set is linearly dependent.

a. $\begin{bmatrix} 1 \\ 7 \\ 6 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 9 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \\ 8 \end{bmatrix}$ b. $\begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 8 \end{bmatrix}$ c. $\begin{bmatrix} -2 \\ 4 \\ 6 \\ 10 \end{bmatrix}, \begin{bmatrix} 3 \\ -6 \\ -9 \\ 15 \end{bmatrix}$

SOLUTION

a. The set contains four vectors, each of which has only three entries. So the set is linearly dependent by Theorem 8.

b. Theorem 8 does not apply here because the number of vectors does not exceed the number of entries in each vector. Since the zero vector is in the set, the set is linearly dependent by Theorem 9.

c. Compare the corresponding entries of the two vectors. The second vector seems to be $-3/2$ times the first vector. This relation holds for the first three pairs of entries, but fails for the fourth pair. Thus neither of the vectors is a multiple of the other, and hence they are linearly independent. ■

In general, you should read a section thoroughly *several* times to absorb an important concept such as linear independence. The notes in the *Study Guide* for this section will help you learn to form mental images of key ideas in linear algebra. For instance, the following proof is worth reading carefully because it shows how the definition of linear independence can be *used*.

PROOF OF THEOREM 7 (Characterization of Linearly Dependent Sets)
If some $\mathbf{v}_j$ in S equals a linear combination of the other vectors, then $\mathbf{v}_j$ can be subtracted from both sides of the equation, producing a linear dependence relation with a nonzero weight (-1) on $\mathbf{v}_j$. [For instance, if $\mathbf{v}_1 = c_2\mathbf{v}_2 + c_3\mathbf{v}_3$, then $\mathbf{0} = (-1)\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + 0\mathbf{v}_4 + \cdots + 0\mathbf{v}_p$.] Thus S is linearly dependent.

Conversely, suppose S is linearly dependent. If $\mathbf{v}_1$ is zero, then it is a (trivial) linear combination of the other vectors in S. Otherwise, $\mathbf{v}_1 \neq \mathbf{0}$, and there exist weights $c_1, \ldots, c_p$, not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$$

Let j be the largest subscript for which $c_j \neq 0$. If $j = 1$, then $c_1\mathbf{v}_1 = \mathbf{0}$, which is impossible because $\mathbf{v}_1 \neq \mathbf{0}$. So $j > 1$, and

$$c_1\mathbf{v}_1 + \cdots + c_j\mathbf{v}_j + 0\mathbf{v}_{j+1} + \cdots + 0\mathbf{v}_p = \mathbf{0}$$

$$c_j\mathbf{v}_j = -c_1\mathbf{v}_1 - \cdots - c_{j-1}\mathbf{v}_{j-1}$$

$$\mathbf{v}_j = \left(-\frac{c_1}{c_j}\right)\mathbf{v}_1 + \cdots + \left(-\frac{c_{j-1}}{c_j}\right)\mathbf{v}_{j-1} \quad ■$$

Practice Problems

1. Let $\mathbf{u} = \begin{bmatrix} 3 \\ 2 \\ -4 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -6 \\ 1 \\ 7 \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} 0 \\ -5 \\ 2 \end{bmatrix}$, and $\mathbf{z} = \begin{bmatrix} 3 \\ 7 \\ -5 \end{bmatrix}$.

a. Are the sets $\{\mathbf{u}, \mathbf{v}\}$, $\{\mathbf{u}, \mathbf{w}\}$, $\{\mathbf{u}, \mathbf{z}\}$, $\{\mathbf{v}, \mathbf{w}\}$, $\{\mathbf{v}, \mathbf{z}\}$, and $\{\mathbf{w}, \mathbf{z}\}$ each linearly independent? Why or why not?

b. Does the answer to Part (a) imply that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z}\}$ is linearly independent?

c. To determine if $\{\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z}\}$ is linearly dependent, is it wise to check if, say, $\mathbf{w}$ is a linear combination of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{z}$?

d. Is $\{\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z}\}$ linearly dependent?

2. Suppose that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly dependent set of vectors in $\mathbb{R}^n$ and $\mathbf{v}_4$ is a vector in $\mathbb{R}^n$. Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is also a linearly dependent set.

1.7 Exercises

In Exercises 1–4, determine if the vectors are linearly independent. Justify each answer.

1. $\begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 7 \\ 2 \\ -6 \end{bmatrix}, \begin{bmatrix} -2 \\ -1 \\ 6 \end{bmatrix}$

2. $\begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \\ -8 \end{bmatrix}, \begin{bmatrix} -3 \\ 4 \\ 1 \end{bmatrix}$

3. $\begin{bmatrix} 1 \\ -3 \end{bmatrix}, \begin{bmatrix} -3 \\ 6 \end{bmatrix}$

4. $\begin{bmatrix} -1 \\ 4 \end{bmatrix}, \begin{bmatrix} -2 \\ 8 \end{bmatrix}$

In Exercises 5–8, determine if the columns of the matrix form a linearly independent set. Justify each answer.

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

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

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

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

In Exercises 9 and 10, (a) for what values of h is $\mathbf{v}_3$ in Span $\{\mathbf{v}_1, \mathbf{v}_2\}$, and (b) for what values of h is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ linearly *dependent*? Justify each answer.

9. $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -3 \\ 10 \\ -6 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 2 \\ -7 \\ h \end{bmatrix}$

10. $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -5 \\ -3 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -2 \\ 10 \\ 6 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 2 \\ -10 \\ h \end{bmatrix}$

In Exercises 11–14, find the value(s) of h for which the vectors are linearly *dependent*. Justify each answer.

11. $\begin{bmatrix} 1 \\ -1 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ -5 \\ 7 \end{bmatrix}, \begin{bmatrix} -1 \\ 5 \\ h \end{bmatrix}$

12. $\begin{bmatrix} 2 \\ -4 \\ 1 \end{bmatrix}, \begin{bmatrix} -6 \\ 7 \\ -3 \end{bmatrix}, \begin{bmatrix} 8 \\ h \\ 4 \end{bmatrix}$

13. $\begin{bmatrix} 1 \\ 5 \\ -3 \end{bmatrix}, \begin{bmatrix} -2 \\ -9 \\ 6 \end{bmatrix}, \begin{bmatrix} 3 \\ h \\ -9 \end{bmatrix}$

14. $\begin{bmatrix} 1 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} -6 \\ 8 \\ 7 \end{bmatrix}, \begin{bmatrix} 4 \\ -2 \\ h \end{bmatrix}$

Determine by inspection whether the vectors in Exercises 15–20 are linearly *independent*. Justify each answer.

15. $\begin{bmatrix} 5 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 8 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 7 \end{bmatrix}$

16. $\begin{bmatrix} 4 \\ -2 \\ 6 \end{bmatrix}, \begin{bmatrix} 6 \\ -3 \\ 9 \end{bmatrix}$

17. $\begin{bmatrix} 3 \\ 5 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -6 \\ 5 \\ 4 \end{bmatrix}$

18. $\begin{bmatrix} 4 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 8 \\ 1 \end{bmatrix}$

19. $\begin{bmatrix} -8 \\ 12 \\ -4 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$

20. $\begin{bmatrix} 1 \\ 4 \\ -7 \end{bmatrix}, \begin{bmatrix} -2 \\ 5 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

In Exercises 21–28, mark each statement True or False **(T/F)**. Justify each answer on the basis of a careful reading of the text.

21. **(T/F)** The columns of a matrix A are linearly independent if the equation $A\mathbf{x} = \mathbf{0}$ has the trivial solution.

22. **(T/F)** Two vectors are linearly dependent if and only if they lie on a line through the origin.

23. **(T/F)** If S is a linearly dependent set, then each vector is a linear combination of the other vectors in S.

24. **(T/F)** If a set contains fewer vectors than there are entries in the vectors, then the set is linearly independent.

25. **(T/F)** The columns of any 4×5 matrix are linearly dependent.

26. **(T/F)** If $\mathbf{x}$ and $\mathbf{y}$ are linearly independent, and if $\mathbf{z}$ is in Span $\{\mathbf{x}, \mathbf{y}\}$, then $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ is linearly dependent.

27. **(T/F)** If $\mathbf{x}$ and $\mathbf{y}$ are linearly independent, and if $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ is linearly dependent, then $\mathbf{z}$ is in Span $\{\mathbf{x}, \mathbf{y}\}$.

28. **(T/F)** If a set in $\mathbb{R}^n$ is linearly dependent, then the set contains more vectors than there are entries in each vector.

In Exercises 29–32, describe the possible echelon forms of the matrix. Use the notation of Example 1 in Section 1.2.

29. A is a 3×3 matrix with linearly independent columns.

30. A is a 2×2 matrix with linearly dependent columns.

31. A is a 4×2 matrix, $A = [\mathbf{a}_1 \quad \mathbf{a}_2]$, and $\mathbf{a}_2$ is not a multiple of $\mathbf{a}_1$.

32. A is a 4×3 matrix, $A = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3]$, such that $\{\mathbf{a}_1, \mathbf{a}_2\}$ is linearly independent and $\mathbf{a}_3$ is not in Span $\{\mathbf{a}_1, \mathbf{a}_2\}$.

33. How many pivot columns must a 7×5 matrix have if its columns are linearly independent? Why?

34. How many pivot columns must a 5×7 matrix have if its columns span $\mathbb{R}^5$? Why?

35. Construct 3×2 matrices A and B such that $A\mathbf{x} = \mathbf{0}$ has only the trivial solution and $B\mathbf{x} = \mathbf{0}$ has a nontrivial solution.

36. a. Fill in the blank in the following statement: "If A is an $m \times n$ matrix, then the columns of A are linearly independent if and only if A has ______ pivot columns."

b. Explain why the statement in (a) is true.

Exercises 37 and 38 should be solved *without performing row operations*. [*Hint:* Write $A\mathbf{x} = \mathbf{0}$ as a vector equation.]

37. Given $A = \begin{bmatrix} 2 & 3 & 5 \\ -5 & 1 & -4 \\ -3 & -1 & -4 \\ 1 & 0 & 1 \end{bmatrix}$, observe that the third column is the sum of the first two columns. Find a nontrivial solution of $A\mathbf{x} = \mathbf{0}$.

38. Given $A = \begin{bmatrix} 5 & 1 & 8 \\ -9 & 5 & 6 \\ 6 & -5 & -9 \end{bmatrix}$, observe that the first column plus three times the second column equals the third column. Find a nontrivial solution of $A\mathbf{x} = \mathbf{0}$.

Each statement in Exercises 39–44 is either true (in all cases) or false (for at least one example). If false, construct a specific example to show that the statement is not always true. Such an example is called a *counterexample* to the statement. If a statement is true, give a justification. (One specific example cannot explain why a statement is always true. You will have to do more work here than in Exercises 21–28.)

39. **(T/F-C)** If $\mathbf{v}_1, \ldots, \mathbf{v}_4$ are in $\mathbb{R}^4$ and $\mathbf{v}_3 = 2\mathbf{v}_1 + \mathbf{v}_2$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is linearly dependent.

40. **(T/F-C)** If $\mathbf{v}_1, \ldots, \mathbf{v}_4$ are in $\mathbb{R}^4$ and $\mathbf{v}_3 = \mathbf{0}$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is linearly dependent.

41. **(T/F-C)** If $\mathbf{v}_1$ and $\mathbf{v}_2$ are in $\mathbb{R}^4$ and $\mathbf{v}_2$ is not a scalar multiple of $\mathbf{v}_1$, then $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent.

42. **(T/F-C)** If $\mathbf{v}_1, \ldots, \mathbf{v}_4$ are in $\mathbb{R}^4$ and $\mathbf{v}_3$ is *not* a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is linearly independent.

43. **(T/F-C)** If $\mathbf{v}_1, \ldots, \mathbf{v}_4$ are in $\mathbb{R}^4$ and $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is also linearly dependent.

44. **(T/F-C)** If $\mathbf{v}_1, \ldots, \mathbf{v}_4$ are linearly independent vectors in $\mathbb{R}^4$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is also linearly independent. [*Hint:* Think about $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 + 0 \cdot \mathbf{v}_4 = \mathbf{0}$.]

45. Suppose A is an $m \times n$ matrix with the property that for all $\mathbf{b}$ in $\mathbb{R}^m$ the equation $A\mathbf{x} = \mathbf{b}$ has at most one solution. Use the definition of linear independence to explain why the columns of A must be linearly independent.

46. Suppose an $m \times n$ matrix A has n pivot columns. Explain why for each $\mathbf{b}$ in $\mathbb{R}^m$ the equation $A\mathbf{x} = \mathbf{b}$ has at most one solution. [*Hint:* Explain why $A\mathbf{x} = \mathbf{b}$ cannot have infinitely many solutions.]

T In Exercises 47 and 48, use as many columns of A as possible to construct a matrix B with the property that the equation $B\mathbf{x} = \mathbf{0}$ has only the trivial solution. Solve $B\mathbf{x} = \mathbf{0}$ to verify your work.

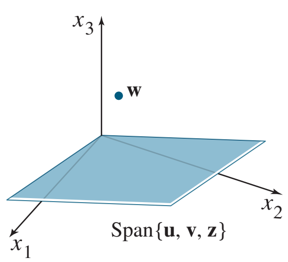
47. $A = \begin{bmatrix} 8 & -3 & 0 & -7 & 2 \\ -9 & 4 & 5 & 11 & -7 \\ 6 & -2 & 2 & -4 & 4 \\ 5 & -1 & 7 & 0 & 10 \end{bmatrix}$

48. $A = \begin{bmatrix} 12 & 10 & -6 & -3 & 7 & 10 \\ -7 & -6 & 4 & 7 & -9 & 5 \\ 9 & 9 & -9 & -5 & 5 & -1 \\ -4 & -3 & 1 & 6 & -8 & 9 \\ 8 & 7 & -5 & -9 & 11 & -8 \end{bmatrix}$

T 49. With A and B as in Exercise 47 select a column $\mathbf{v}$ of A that was not used in the construction of B and determine if $\mathbf{v}$ is in the set spanned by the columns of B. (Describe your calculations.)

T 50. Repeat Exercise 49 with the matrices A and B from Exercise 48. Then give an explanation for what you discover, assuming that B was constructed as specified.

STUDY GUIDE offers additional resources for mastering the concept of linear independence.



Solutions to Practice Problems

1. a. Yes. In each case, neither vector is a multiple of the other. Thus each set is linearly independent.

b. No. The observation in Part (a), by itself, says nothing about the linear independence of $\{\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z}\}$.

c. No. When testing for linear independence, it is usually a poor idea to check if one selected vector is a linear combination of the others. It may happen that the selected vector is not a linear combination of the others and yet the whole set of vectors is linearly dependent. In this practice problem, $\mathbf{w}$ is not a linear combination of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{z}$.

d. Yes, by Theorem 8. There are more vectors (four) than entries (three) in them.

2. Applying the definition of linearly dependent to $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ implies that there exist scalars c_1, c_2, and c_3, not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = \mathbf{0}.$$

Adding $0\,\mathbf{v}_4 = \mathbf{0}$ to both sides of this equation results in

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + 0\,\mathbf{v}_4 = \mathbf{0}.$$

Since c_1, c_2, c_3 and 0 are not *all* zero, the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ satisfies the definition of a linearly dependent set.

1.8 Introduction to Linear Transformations

The difference between a matrix equation $A\mathbf{x} = \mathbf{b}$ and the associated vector equation $x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$ is merely a matter of notation. However, a matrix equation $A\mathbf{x} = \mathbf{b}$ can arise in linear algebra (and in applications such as computer graphics and signal processing) in a way that is not directly connected with linear combinations of vectors. This happens when we think of the matrix A as an object that "acts" on a vector $\mathbf{x}$ by multiplication to produce a new vector called $A\mathbf{x}$.

For instance, the equations

$$\underset{\textstyle\uparrow \atop \textstyle A}{\begin{bmatrix} 4 & -3 & 1 & 3 \\ 2 & 0 & 5 & 1 \end{bmatrix}} \underset{\textstyle\uparrow \atop \textstyle \mathbf{x}}{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}} = \underset{\textstyle\uparrow \atop \textstyle \mathbf{b}}{\begin{bmatrix} 5 \\ 8 \end{bmatrix}} \quad \text{and} \quad \underset{\textstyle\uparrow \atop \textstyle A}{\begin{bmatrix} 4 & -3 & 1 & 3 \\ 2 & 0 & 5 & 1 \end{bmatrix}} \underset{\textstyle\uparrow \atop \textstyle \mathbf{u}}{\begin{bmatrix} 1 \\ 4 \\ -1 \\ 3 \end{bmatrix}} = \underset{\textstyle\uparrow \atop \textstyle \mathbf{0}}{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}$$

say that multiplication by A transforms $\mathbf{x}$ into $\mathbf{b}$ and transforms $\mathbf{u}$ into the zero vector. See Figure 1.

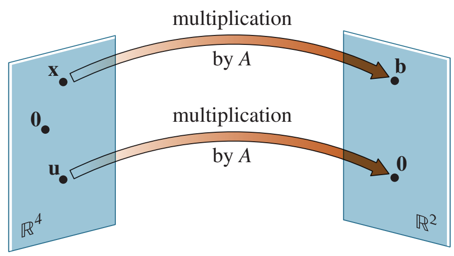


FIGURE 1 Transforming vectors via matrix multiplication.

From this new point of view, solving the equation $A\mathbf{x} = \mathbf{b}$ amounts to finding all vectors $\mathbf{x}$ in $\mathbb{R}^4$ that are transformed into the vector $\mathbf{b}$ in $\mathbb{R}^2$ under the "action" of multiplication by A.

The correspondence from $\mathbf{x}$ to $A\mathbf{x}$ is a *function* from one set of vectors to another. This concept generalizes the common notion of a function as a rule that transforms one real number into another.

A **transformation** (or **function** or **mapping**) T from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that assigns to each vector $\mathbf{x}$ in $\mathbb{R}^n$ a vector $T(\mathbf{x})$ in $\mathbb{R}^m$. The set $\mathbb{R}^n$ is called the **domain** of T, and $\mathbb{R}^m$

is called the **codomain** of T. The notation $T : \mathbb{R}^n \to \mathbb{R}^m$ indicates that the domain of T is $\mathbb{R}^n$ and the codomain is $\mathbb{R}^m$. For $\mathbf{x}$ in $\mathbb{R}^n$, the vector $T(\mathbf{x})$ in $\mathbb{R}^m$ is called the **image** of $\mathbf{x}$ (under the action of T). The set of all images $T(\mathbf{x})$ is called the **range** of T. See Figure 2.

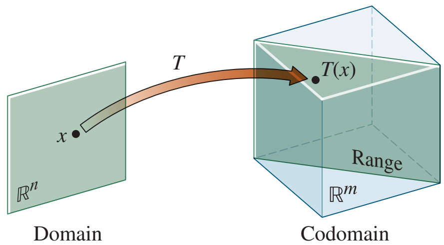
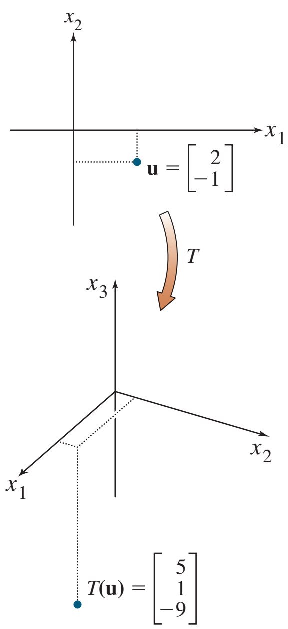


FIGURE 2 Domain, codomain, and range of $T : \mathbb{R}^n \to \mathbb{R}^m$.

The new terminology in this section is important because a dynamic view of matrix–vector multiplication is the key to understanding several ideas in linear algebra and to building mathematical models of physical systems that evolve over time. Such *dynamical systems* will be discussed in Sections 1.10, 4.8, and throughout Chapter 5.

Matrix Transformations

The rest of this section focuses on mappings associated with matrix multiplication. For each $\mathbf{x}$ in $\mathbb{R}^n$, $T(\mathbf{x})$ is computed as $A\mathbf{x}$, where A is an $m \times n$ matrix. For simplicity, we sometimes denote such a *matrix transformation* by $\mathbf{x} \mapsto A\mathbf{x}$. Observe that the domain of T is $\mathbb{R}^n$ when A has n columns and the codomain of T is $\mathbb{R}^m$ when each column of A has m entries. The range of T is the set of all linear combinations of the columns of A, because each image $T(\mathbf{x})$ is of the form $A\mathbf{x}$.



EXAMPLE 1 Let $A = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}$, $\mathbf{c} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}$, and define a transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ by $T(\mathbf{x}) = A\mathbf{x}$, so that

$$T(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - 3x_2 \\ 3x_1 + 5x_2 \\ -x_1 + 7x_2 \end{bmatrix}$$

a. Find $T(\mathbf{u})$, the image of $\mathbf{u}$ under the transformation T.

b. Find an $\mathbf{x}$ in $\mathbb{R}^2$ whose image under T is $\mathbf{b}$.

c. Is there more than one $\mathbf{x}$ whose image under T is $\mathbf{b}$?

d. Determine if $\mathbf{c}$ is in the range of the transformation T.

SOLUTION

a. Compute

$$T(\mathbf{u}) = A\mathbf{u} = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ -9 \end{bmatrix}$$

b. Solve $T(\mathbf{x}) = \mathbf{b}$ for $\mathbf{x}$. That is, solve $A\mathbf{x} = \mathbf{b}$, or

$$\begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix} \tag{1}$$

Using the method discussed in Section 1.4, row reduce the augmented matrix:

$$\begin{bmatrix} 1 & -3 & 3 \\ 3 & 5 & 2 \\ -1 & 7 & -5 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 14 & -7 \\ 0 & 4 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 1 & -.5 \\ 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1.5 \\ 0 & 1 & -.5 \\ 0 & 0 & 0 \end{bmatrix} \qquad (2)$$

Hence $x_1 = 1.5$, $x_2 = -.5$, and $\mathbf{x} = \begin{bmatrix} 1.5 \\ -.5 \end{bmatrix}$. The image of this $\mathbf{x}$ under T is the given vector $\mathbf{b}$.

c. Any $\mathbf{x}$ whose image under T is $\mathbf{b}$ must satisfy equation (1). From (2), it is clear that equation (1) has a unique solution. So there is exactly one $\mathbf{x}$ whose image is $\mathbf{b}$.

d. The vector $\mathbf{c}$ is in the range of T if $\mathbf{c}$ is the image of some $\mathbf{x}$ in $\mathbb{R}^2$, that is, if $\mathbf{c} = T(\mathbf{x})$ for some $\mathbf{x}$. This is just another way of asking if the system $A\mathbf{x} = \mathbf{c}$ is consistent. To find the answer, row reduce the augmented matrix:

$$\begin{bmatrix} 1 & -3 & 3 \\ 3 & 5 & 2 \\ -1 & 7 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 14 & -7 \\ 0 & 4 & 8 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 1 & 2 \\ 0 & 14 & -7 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & -35 \end{bmatrix}$$

The third equation, $0 = -35$, shows that the system is inconsistent. So $\mathbf{c}$ is *not* in the range of T. ■

The question in Example 1(c) is a *uniqueness* problem for a system of linear equations, translated here into the language of matrix transformations: Is $\mathbf{b}$ the image of a *unique* $\mathbf{x}$ in $\mathbb{R}^n$? Similarly, Example 1(d) is an *existence* problem: Does there *exist* an $\mathbf{x}$ whose image is $\mathbf{c}$?

The next two matrix transformations can be viewed geometrically. They reinforce the dynamic view of a matrix as something that transforms vectors into other vectors. Section 2.7 contains other interesting examples connected with computer graphics.

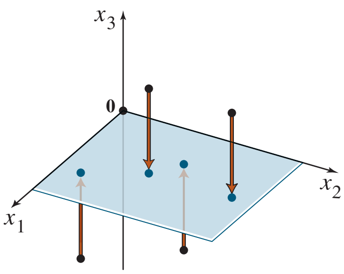


FIGURE 3
A projection transformation.

EXAMPLE 2 If $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, then the transformation $\mathbf{x} \mapsto A\mathbf{x}$ *projects* points in $\mathbb{R}^3$ onto the x_1x_2-plane because

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix}$$

See Figure 3. ■

sheep

sheared sheep

EXAMPLE 3 Let $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$. The transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(\mathbf{x}) = A\mathbf{x}$ is called a **shear transformation**. It can be shown that if T acts on each point in the 2×2 square shown in Figure 4, then the set of images forms the sheared parallelogram. The key idea is to show that T maps line segments onto line segments (as shown in Exercise 35) and then to check that the corners of the square map onto the vertices of the parallelogram. For instance, the image of the point $\mathbf{u} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$ is $T(\mathbf{u}) = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$, and the image of $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$ is $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$. T deforms the square as if the top of the square were pushed to the right while the base is held fixed. Shear transformations appear in physics, geology, and crystallography. ■

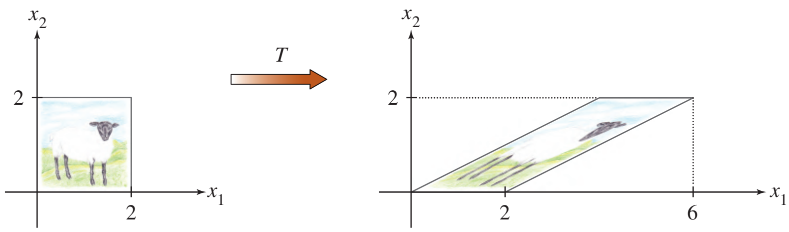


FIGURE 4 A shear transformation.

Linear Transformations

Theorem 5 in Section 1.4 shows that if A is $m \times n$, then the transformation $\mathbf{x} \mapsto A\mathbf{x}$ has the properties

$$A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} \quad \text{and} \quad A(c\mathbf{u}) = cA\mathbf{u}$$

for all $\mathbf{u}$, $\mathbf{v}$ in $\mathbb{R}^n$ and all scalars c. These properties, written in function notation, identify the most important class of transformations in linear algebra.

DEFINITION

A transformation (or mapping) T is **linear** if

(i) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}$, $\mathbf{v}$ in the domain of T;

(ii) $T(c\mathbf{u}) = cT(\mathbf{u})$ for all scalars c and all $\mathbf{u}$ in the domain of T.

Every matrix transformation is a linear transformation. Important examples of linear transformations that are not matrix transformations will be discussed in Chapters 4 and 5.

Linear transformations *preserve the operations of vector addition and scalar multiplication*. Property (i) says that the result $T(\mathbf{u} + \mathbf{v})$ of first adding $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ and then applying T is the same as first applying T to $\mathbf{u}$ and to $\mathbf{v}$ and then adding $T(\mathbf{u})$ and $T(\mathbf{v})$ in $\mathbb{R}^m$. These two properties lead easily to the following useful facts.

If T is a linear transformation, then

$$T(\mathbf{0}) = \mathbf{0} \tag{3}$$

and

$$T(c\mathbf{u} + d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v}) \tag{4}$$

for all vectors $\mathbf{u}$, $\mathbf{v}$ in the domain of T and all scalars c, d.

Property (3) follows from condition (ii) in the definition, because $T(\mathbf{0}) = T(0\mathbf{u}) = 0T(\mathbf{u}) = \mathbf{0}$. Property (4) requires both (i) and (ii):

$$T(c\mathbf{u} + d\mathbf{v}) = T(c\mathbf{u}) + T(d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v})$$

Observe that *if a transformation satisfies* (4) *for all* $\mathbf{u}$, $\mathbf{v}$ *and* c, d, *it must be linear*. (Set $c = d = 1$ for preservation of addition, and set $d = 0$ for preservation of scalar multiplication.) Repeated application of (4) produces a useful generalization:

$$\boxed{T(c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p) = c_1T(\mathbf{v}_1) + \cdots + c_pT(\mathbf{v}_p)} \tag{5}$$

In engineering and physics, (5) is referred to as a *superposition principle*. Think of $\mathbf{v}_1, \dots, \mathbf{v}_p$ as signals that go into a system and $T(\mathbf{v}_1), \dots, T(\mathbf{v}_p)$ as the responses of that system to the signals. The system satisfies the superposition principle if whenever an input is expressed as a linear combination of such signals, the system's response is *the same* linear combination of the responses to the individual signals. We will return to this idea in Chapter 4.

EXAMPLE 4 Given a scalar r, define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\mathbf{x}) = r\mathbf{x}$. T is called a **contraction** when $0 \le r \le 1$ and a **dilation** when $r > 1$. Let $r = 3$, and show that T is a linear transformation.

SOLUTION Let $\mathbf{u}$, $\mathbf{v}$ be in $\mathbb{R}^2$ and let c, d be scalars. Then

$$\begin{aligned} T(c\mathbf{u} + d\mathbf{v}) &= 3(c\mathbf{u} + d\mathbf{v}) && \text{Definition of } T \\ &= 3c\mathbf{u} + 3d\mathbf{v} && \Big\} \text{ Vector arithmetic} \\ &= c(3\mathbf{u}) + d(3\mathbf{v}) && \\ &= cT(\mathbf{u}) + dT(\mathbf{v}) \end{aligned}$$

Thus T is a linear transformation because it satisfies (4). See Figure 5. ■

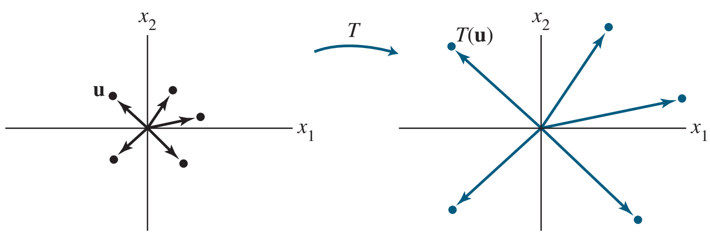


FIGURE 5 A dilation transformation.

EXAMPLE 5 Define a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$T(\mathbf{x}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}$$

Find the images under T of $\mathbf{u} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, and $\mathbf{u} + \mathbf{v} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$.

SOLUTION

$$T(\mathbf{u}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 4 \end{bmatrix}, \qquad T(\mathbf{v}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -3 \\ 2 \end{bmatrix},$$

$$T(\mathbf{u} + \mathbf{v}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 6 \\ 4 \end{bmatrix} = \begin{bmatrix} -4 \\ 6 \end{bmatrix}$$

Note that $T(\mathbf{u} + \mathbf{v})$ is obviously equal to $T(\mathbf{u}) + T(\mathbf{v})$. It appears from Figure 6 that T rotates $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u} + \mathbf{v}$ counterclockwise about the origin through 90°. In fact, T transforms the entire parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$ into the one determined by $T(\mathbf{u})$ and $T(\mathbf{v})$. (See Exercise 36.) ■

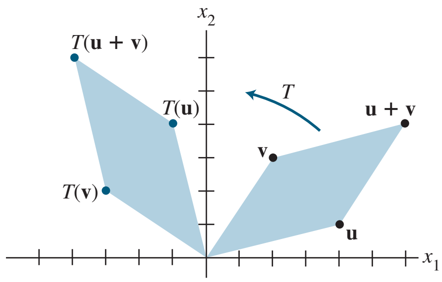


FIGURE 6 A rotation transformation.

The final example is not geometrical; instead, it shows how a linear mapping can transform one type of data into another.

EXAMPLE 6 A company manufactures two products, B and C. Using data from Example 7 in Section 1.3, we construct a "unit cost" matrix, $U = [\,\mathbf{b} \quad \mathbf{c}\,]$, whose columns describe the "costs per dollar of output" for the products:

$$\begin{array}{cccc} & \multicolumn{2}{c}{\text{Product}} & \\ & \text{B} & \text{C} & \\ U = & \left[\begin{array}{c} .45 \\ .25 \\ .15 \end{array}\right. & \left.\begin{array}{c} .40 \\ .30 \\ .15 \end{array}\right] & \begin{array}{l} \text{Materials} \\ \text{Labor} \\ \text{Overhead} \end{array} \end{array}$$

Let $\mathbf{x} = (x_1, x_2)$ be a "production" vector, corresponding to x_1 dollars of product B and x_2 dollars of product C, and define $T : \mathbb{R}^2 \to \mathbb{R}^3$ by

$$T(\mathbf{x}) = U\mathbf{x} = x_1 \begin{bmatrix} .45 \\ .25 \\ .15 \end{bmatrix} + x_2 \begin{bmatrix} .40 \\ .30 \\ .15 \end{bmatrix} = \begin{bmatrix} \text{Total cost of materials} \\ \text{Total cost of labor} \\ \text{Total cost of overhead} \end{bmatrix}$$

The mapping T transforms a list of production quantities (measured in dollars) into a list of total costs. The linearity of this mapping is reflected in two ways:

1. If production is increased by a factor of, say, 4, from $\mathbf{x}$ to $4\mathbf{x}$, then the costs will increase by the same factor, from $T(\mathbf{x})$ to $4T(\mathbf{x})$.
2. If $\mathbf{x}$ and $\mathbf{y}$ are production vectors, then the total cost vector associated with the combined production $\mathbf{x} + \mathbf{y}$ is precisely the sum of the cost vectors $T(\mathbf{x})$ and $T(\mathbf{y})$. ■

Practice Problems

1. Suppose $T : \mathbb{R}^5 \to \mathbb{R}^2$ and $T(\mathbf{x}) = A\mathbf{x}$ for some matrix A and for each $\mathbf{x}$ in $\mathbb{R}^5$. How many rows and columns does A have?
2. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Give a geometric description of the transformation $\mathbf{x} \mapsto A\mathbf{x}$.
3. The line segment from $\mathbf{0}$ to a vector $\mathbf{u}$ is the set of points of the form $t\mathbf{u}$, where $0 \le t \le 1$. Show that a linear transformation T maps this segment into the segment between $\mathbf{0}$ and $T(\mathbf{u})$.

1.8 Exercises

1. Let $A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, and define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\mathbf{x}) = A\mathbf{x}$. Find the images under T of $\mathbf{u} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$.

2. Let $A = \begin{bmatrix} .5 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & .5 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ -4 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$. Define $T : \mathbb{R}^3 \to \mathbb{R}^3$ by $T(\mathbf{x}) = A\mathbf{x}$. Find $T(\mathbf{u})$ and $T(\mathbf{v})$.

In Exercises 3–6, with T defined by $T(\mathbf{x}) = A\mathbf{x}$, find a vector $\mathbf{x}$ whose image under T is $\mathbf{b}$, and determine whether $\mathbf{x}$ is unique.

3. $A = \begin{bmatrix} 1 & 0 & -2 \\ -2 & 1 & 6 \\ 3 & -2 & -5 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -1 \\ 7 \\ -3 \end{bmatrix}$

4. $A = \begin{bmatrix} 1 & -3 & 2 \\ 0 & 1 & -4 \\ 3 & -5 & -9 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 6 \\ -7 \\ -9 \end{bmatrix}$

5. $A = \begin{bmatrix} 1 & -5 & -7 \\ -3 & 7 & 5 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$

6. $A = \begin{bmatrix} 1 & -2 & 1 \\ 3 & -4 & 5 \\ 0 & 1 & 1 \\ -3 & 5 & -4 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ 9 \\ 3 \\ -6 \end{bmatrix}$

7. Let A be a 4×6 matrix. What must a and b be in order to define $T : \mathbb{R}^a \to \mathbb{R}^b$ by $T(\mathbf{x}) = A\mathbf{x}$?

8. How many rows and columns must a matrix A have in order to define a mapping from $\mathbb{R}^3$ into $\mathbb{R}^6$ by the rule $T(\mathbf{x}) = A\mathbf{x}$?

For Exercises 9 and 10, find all $\mathbf{x}$ in $\mathbb{R}^4$ that are mapped into the zero vector by the transformation $\mathbf{x} \mapsto A\mathbf{x}$ for the given matrix A.

9. $A = \begin{bmatrix} 1 & -4 & 7 & -5 \\ 0 & 1 & -4 & 3 \\ 2 & -6 & 6 & -4 \end{bmatrix}$

10. $A = \begin{bmatrix} 1 & 3 & 9 & 2 \\ 1 & 0 & 3 & -4 \\ 0 & 1 & 2 & 3 \\ -2 & 3 & 0 & 5 \end{bmatrix}$

11. Let $\mathbf{b} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$, and let A be the matrix in Exercise 9. Is $\mathbf{b}$ in the range of the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$? Why or why not?

12. Let $\mathbf{b} = \begin{bmatrix} -1 \\ 3 \\ -1 \\ 4 \end{bmatrix}$, and let A be the matrix in Exercise 10. Is $\mathbf{b}$ in the range of the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$? Why or why not?

In Exercises 13–16, use a rectangular coordinate system to plot $\mathbf{u} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$, and their images under the given transformation T. (Make a separate and reasonably large sketch for each exercise.) Describe geometrically what T does to each vector $\mathbf{x}$ in $\mathbb{R}^2$.

13. $T(\mathbf{x}) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

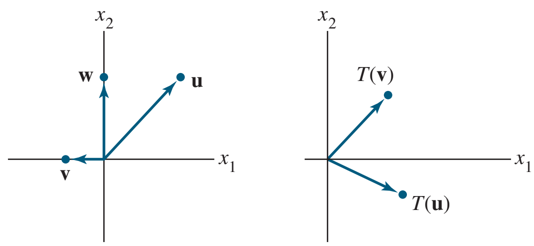
14. $T(\mathbf{x}) = \begin{bmatrix} .5 & 0 \\ 0 & .5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

15. $T(\mathbf{x}) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

16. $T(\mathbf{x}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

17. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation that maps $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ into $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and maps $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ into $\begin{bmatrix} 1 \\ -5 \end{bmatrix}$. Use the fact that T is linear to find the images under T of $5\mathbf{u}$, $4\mathbf{v}$, and $5\mathbf{u} + 4\mathbf{v}$.

18. The figure shows vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$, along with the images $T(\mathbf{u})$ and $T(\mathbf{v})$ under the action of a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$. Copy this figure carefully, and draw the image $T(\mathbf{w})$ as accurately as possible. [*Hint:* First, write $\mathbf{w}$ as a linear combination of $\mathbf{u}$ and $\mathbf{v}$.]



19. Let $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\mathbf{y}_1 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$, and $\mathbf{y}_2 = \begin{bmatrix} -1 \\ 6 \end{bmatrix}$, and let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation that maps $\mathbf{e}_1$ into $\mathbf{y}_1$ and maps $\mathbf{e}_2$ into $\mathbf{y}_2$. Find the images of $\begin{bmatrix} 5 \\ -3 \end{bmatrix}$ and $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

20. Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} -3 \\ 5 \end{bmatrix}$, and $\mathbf{v}_1 = \begin{bmatrix} 2 \\ -9 \end{bmatrix}$, and let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation that maps $\mathbf{x}$ into $x_1\mathbf{v}_1 + x_2\mathbf{v}_2$. Find a matrix A such that $T(\mathbf{x})$ is $A\mathbf{x}$ for each $\mathbf{x}$.

In Exercises 21–30, mark each statement True or False **(T/F)**. Justify each answer.

21. **(T/F)** A linear transformation is a special type of function.

22. **(T/F)** Every matrix transformation is a linear transformation.

23. **(T/F)** If A is a 3×5 matrix and T is a transformation defined by $T(\mathbf{x}) = A\mathbf{x}$, then the domain of T is $\mathbb{R}^3$.

24. **(T/F)** The codomain of the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is the set of all linear combinations of the columns of A.

25. **(T/F)** If A is an $m \times n$ matrix, then the range of the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is $\mathbb{R}^m$.

26. **(T/F)** If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation and if $\mathbf{c}$ is in $\mathbb{R}^m$, then a uniqueness question is "Is $\mathbf{c}$ in the range of T?"

27. **(T/F)** Every linear transformation is a matrix transformation.

28. **(T/F)** A linear transformation preserves the operations of vector addition and scalar multiplication.

29. **(T/F)** A transformation T is linear if and only if $T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2)$ for all $\mathbf{v}_1$ and $\mathbf{v}_2$ in the domain of T and for all scalars c_1 and c_2.

30. **(T/F)** The superposition principle is a physical description of a linear transformation.

31. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation that reflects each point through the x_1-axis. (See Practice Problem 2.) Make two sketches similar to Figure 6 that illustrate properties (i) and (ii) of a linear transformation.

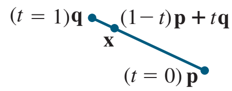
32. Suppose vectors $\mathbf{v}_1, \dots, \mathbf{v}_p$ span $\mathbb{R}^n$, and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Suppose $T(\mathbf{v}_i) = \mathbf{0}$ for $i = 1, \dots, p$. Show that T is the zero transformation. That is, show that if $\mathbf{x}$ is any vector in $\mathbb{R}^n$, then $T(\mathbf{x}) = \mathbf{0}$.

33. Given $\mathbf{v} \neq \mathbf{0}$ and $\mathbf{p}$ in $\mathbb{R}^n$, the line through $\mathbf{p}$ in the direction of $\mathbf{v}$ has the parametric equation $\mathbf{x} = \mathbf{p} + t\mathbf{v}$. Show that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ maps this line onto another line or onto a single point (a *degenerate line*).

34. Let $\mathbf{u}$ and $\mathbf{v}$ be linearly independent vectors in $\mathbb{R}^3$, and let P be the plane through $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{0}$. The parametric equation of P is $\mathbf{x} = s\mathbf{u} + t\mathbf{v}$ (with s, t in $\mathbb{R}$). Show that a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ maps P onto a plane through $\mathbf{0}$, or onto a line through $\mathbf{0}$, or onto just the origin in $\mathbb{R}^3$. What must be true about $T(\mathbf{u})$ and $T(\mathbf{v})$ in order for the image of the plane P to be a plane?

35. a. Show that the line through vectors $\mathbf{p}$ and $\mathbf{q}$ in $\mathbb{R}^n$ may be written in the parametric form $\mathbf{x} = (1 - t)\mathbf{p} + t\mathbf{q}$. (Refer to the figure with Exercises 25 and 26 in Section 1.5.)

b. The line segment from $\mathbf{p}$ to $\mathbf{q}$ is the set of points of the form $(1 - t)\mathbf{p} + t\mathbf{q}$ for $0 \leq t \leq 1$ (as shown in the figure below). Show that a linear transformation T maps this line segment onto a line segment or onto a single point.



36. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$. It can be shown that the set P of all points in the parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$ has the form $a\mathbf{u} + b\mathbf{v}$, for $0 \leq a \leq 1$, $0 \leq b \leq 1$. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Explain why the image of a point in P under the transformation T lies in the parallelogram determined by $T(\mathbf{u})$ and $T(\mathbf{v})$.

37. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = mx + b$.

a. Show that f is a linear transformation when $b = 0$.

b. Find a property of a linear transformation that is violated when $b \neq 0$.

c. Why is f called a linear function?

38. An *affine transformation* $T : \mathbb{R}^n \to \mathbb{R}^m$ has the form $T(x) = A\mathbf{x} + \mathbf{b}$, with A an $m \times n$ matrix and $\mathbf{b}$ in $\mathbb{R}^m$. Show that T is *not* a linear transformation when $\mathbf{b} \neq \mathbf{0}$. (Affine transformations are important in computer graphics.)

39. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and let $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ be a linearly dependent set in $\mathbb{R}^n$. Explain why the set $\{T(\mathbf{v}_1), T(\mathbf{v}_2), T(\mathbf{v}_3)\}$ is linearly dependent.

In Exercises 40–44, column vectors are written as rows, such as $\mathbf{x} = (x_1, x_2)$, and $T(\mathbf{x})$ is written as $T(x_1, x_2)$.

40. Show that the transformation T defined by $T(x_1, x_2) = (4x_1 - 2x_2, 3|x_2|)$ is not linear.

41. Show that the transformation T defined by $T(x_1, x_2) = (2x_1 - 3x_2, x_1 + 4, 5x_2)$ is not linear.

42. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Show that if T maps two linearly independent vectors onto a linearly dependent set, then the equation $T(\mathbf{x}) = \mathbf{0}$ has a nontrivial solution. [*Hint:* Suppose $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are linearly independent and yet $T(\mathbf{u})$ and $T(\mathbf{v})$ are linearly dependent. Then $c_1T(\mathbf{u}) + c_2T(\mathbf{v}) = \mathbf{0}$ for some weights c_1 and c_2, not both zero. Use this equation.]

43. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the transformation that reflects each vector $\mathbf{x} = (x_1, x_2, x_3)$ through the plane $x_3 = 0$ onto $T(\mathbf{x}) = (x_1, x_2, -x_3)$. Show that T is a linear transformation. [See Example 4 for ideas.]

44. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the transformation that projects each vector $\mathbf{x} = (x_1, x_2, x_3)$ onto the plane $x_2 = 0$, so $T(\mathbf{x}) = (x_1, 0, x_3)$. Show that T is a linear transformation.

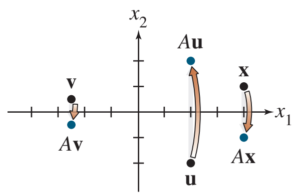
T In Exercises 45 and 46, the given matrix determines a linear transformation T. Find all $\mathbf{x}$ such that $T(\mathbf{x}) = \mathbf{0}$.

45. $\begin{bmatrix} 4 & -2 & 5 & -5 \\ -9 & 7 & -8 & 0 \\ -6 & 4 & 5 & 3 \\ 5 & -3 & 8 & -4 \end{bmatrix}$ **46.** $\begin{bmatrix} -9 & -4 & -9 & 4 \\ 5 & -8 & -7 & 6 \\ 7 & 11 & 16 & -9 \\ 9 & -7 & -4 & 5 \end{bmatrix}$

T 47. Let $\mathbf{b} = \begin{bmatrix} 7 \\ 5 \\ 9 \\ 7 \end{bmatrix}$ and let A be the matrix in Exercise 45. Is $\mathbf{b}$ in the range of the transformation $\mathbf{x} \mapsto A\mathbf{x}$? If so, find an $\mathbf{x}$ whose image under the transformation is $\mathbf{b}$.

T 48. Let $\mathbf{b} = \begin{bmatrix} -7 \\ -7 \\ 13 \\ -5 \end{bmatrix}$ and let A be the matrix in Exercise 46. Is $\mathbf{b}$ in the range of the transformation $\mathbf{x} \mapsto A\mathbf{x}$? If so, find an $\mathbf{x}$ whose image under the transformation is $\mathbf{b}$.

STUDY GUIDE offers additional resources for mastering linear transformations.



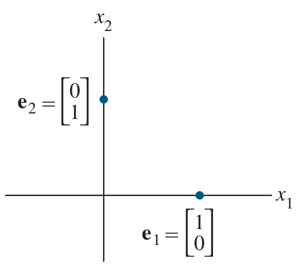
The transformation $\mathbf{x} \mapsto A\mathbf{x}$.

Solutions to Practice Problems

1. A must have five columns for $A\mathbf{x}$ to be defined. A must have two rows for the codomain of T to be $\mathbb{R}^2$.
2. Plot some random points (vectors) on graph paper to see what happens. A point such as $(4, 1)$ maps into $(4, -1)$. The transformation $\mathbf{x} \mapsto A\mathbf{x}$ reflects points through the x-axis (or x_1-axis).
3. Let $\mathbf{x} = t\mathbf{u}$ for some t such that $0 \le t \le 1$. Since T is linear, $T(t\mathbf{u}) = t\,T(\mathbf{u})$, which is a point on the line segment between $\mathbf{0}$ and $T(\mathbf{u})$.

1.9 The Matrix of a Linear Transformation

Whenever a linear transformation T arises geometrically or is described in words, we usually want a "formula" for $T(\mathbf{x})$. The discussion that follows shows that every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is actually a matrix transformation $\mathbf{x} \mapsto A\mathbf{x}$ and that important properties of T are intimately related to familiar properties of A. The key to finding A is to observe that T is completely determined by what it does to the columns of the $n \times n$ identity matrix I_n.



EXAMPLE 1 The columns of $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ are $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

Suppose T is a linear transformation from $\mathbb{R}^2$ into $\mathbb{R}^3$ such that

$$T(\mathbf{e}_1) = \begin{bmatrix} 5 \\ -7 \\ 2 \end{bmatrix} \quad \text{and} \quad T(\mathbf{e}_2) = \begin{bmatrix} -3 \\ 8 \\ 0 \end{bmatrix}$$

With no additional information, find a formula for the image of an arbitrary $\mathbf{x}$ in $\mathbb{R}^2$.

SOLUTION Write

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1\begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2\begin{bmatrix} 0 \\ 1 \end{bmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 \tag{1}$$

Since T is a *linear* transformation,

$$T(\mathbf{x}) = x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) \tag{2}$$

$$= x_1\begin{bmatrix} 5 \\ -7 \\ 2 \end{bmatrix} + x_2\begin{bmatrix} -3 \\ 8 \\ 0 \end{bmatrix} = \begin{bmatrix} 5x_1 - 3x_2 \\ -7x_1 + 8x_2 \\ 2x_1 + 0 \end{bmatrix} \quad \blacksquare$$

The step from equation (1) to equation (2) explains why knowledge of $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$ is sufficient to determine $T(\mathbf{x})$ for any $\mathbf{x}$. Moreover, since (2) expresses $T(\mathbf{x})$ as a linear combination of vectors, we can put these vectors into the columns of a matrix A and write (2) as

$$T(\mathbf{x}) = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A\mathbf{x}$$

THEOREM 10

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there exists a unique matrix A such that

$$T(\mathbf{x}) = A\mathbf{x} \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^n$$

In fact, A is the $m \times n$ matrix whose jth column is the vector $T(\mathbf{e}_j)$, where $\mathbf{e}_j$ is the jth column of the identity matrix in $\mathbb{R}^n$:

$$A = \begin{bmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \end{bmatrix} \tag{3}$$

PROOF Write $\mathbf{x} = I_n\mathbf{x} = [\,\mathbf{e}_1 \ \cdots \ \mathbf{e}_n\,]\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$, and use the linearity of T to compute

$$\begin{aligned} T(\mathbf{x}) &= T(x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n) = x_1T(\mathbf{e}_1) + \cdots + x_nT(\mathbf{e}_n) \\ &= \begin{bmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = A\mathbf{x} \end{aligned}$$

The uniqueness of A is treated in Exercise 41. ■

The matrix A in (3) is called the **standard matrix for the linear transformation** T.

We know now that every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be viewed as a matrix transformation, and vice versa. The term *linear transformation* focuses on a property of a mapping, while *matrix transformation* describes how such a mapping is implemented, as Examples 2 and 3 illustrate.

EXAMPLE 2 Find the standard matrix A for the dilation transformation $T(\mathbf{x}) = 3\mathbf{x}$, for $\mathbf{x}$ in $\mathbb{R}^2$.

SOLUTION Write

$$T(\mathbf{e}_1) = 3\mathbf{e}_1 = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \quad \text{and} \quad T(\mathbf{e}_2) = 3\mathbf{e}_2 = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$ ■

EXAMPLE 3 Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that rotates each point in $\mathbb{R}^2$ about the origin through an angle φ, with counterclockwise rotation for a positive angle. We could show geometrically that such a transformation is linear. (See Figure 6 in Section 1.8.) Find the standard matrix A of this transformation.

SOLUTION $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ rotates into $\begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix}$, and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ rotates into $\begin{bmatrix} -\sin\varphi \\ \cos\varphi \end{bmatrix}$. See Figure 1. By Theorem 10,

$$A = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$$

Example 5 in Section 1.8 is a special case of this transformation, with $\varphi = \pi/2$. ■

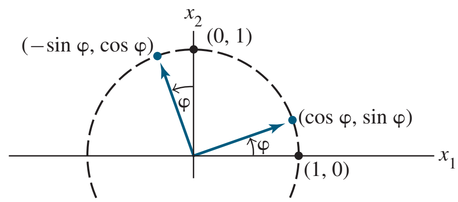


FIGURE 1 A rotation transformation.

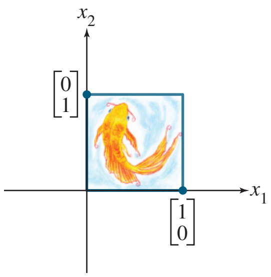


FIGURE 2
The unit square.

Geometric Linear Transformations of $\mathbb{R}^2$

Examples 2 and 3 illustrate linear transformations that are described geometrically. Tables 1–4 illustrate other common geometric linear transformations of the plane. Because the transformations are linear, they are determined completely by what they do to the columns of I_2. Instead of showing only the images of $\mathbf{e}_1$ and $\mathbf{e}_2$, the tables show what a transformation does to the unit square (Figure 2).

Other transformations can be constructed from those listed in Tables 1–4 by applying one transformation after another. For instance, a horizontal shear could be followed by a reflection in the x_2-axis. Section 2.1 will show that such a *composition* of linear transformations is linear. (Also, see Exercise 44.)

Existence and Uniqueness Questions

The concept of a linear transformation provides a new way to understand the existence and uniqueness questions asked earlier. The next two definitions give the appropriate terminology for transformations.

DEFINITION

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **onto** $\mathbb{R}^m$ if each $\mathbf{b}$ in $\mathbb{R}^m$ is the image of *at least one* $\mathbf{x}$ in $\mathbb{R}^n$.

Equivalently, T is onto $\mathbb{R}^m$ when the range of T is all of the codomain $\mathbb{R}^m$. That is, T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if, for each $\mathbf{b}$ in the codomain $\mathbb{R}^m$, there exists at least one solution of $T(\mathbf{x}) = \mathbf{b}$. "Does T map $\mathbb{R}^n$ onto $\mathbb{R}^m$?" is an existence question. The mapping T is *not* onto when there is some $\mathbf{b}$ in $\mathbb{R}^m$ for which the equation $T(\mathbf{x}) = \mathbf{b}$ has no solution. See Figure 3.

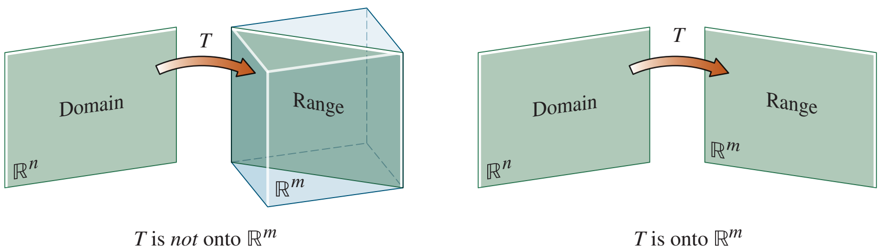


FIGURE 3 Is the range of T all of $\mathbb{R}^m$?

TABLE 1 Reflections

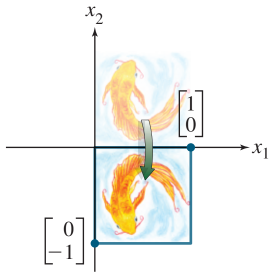
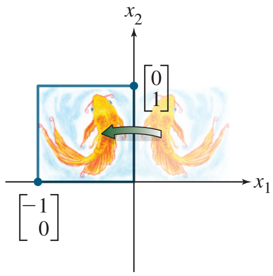
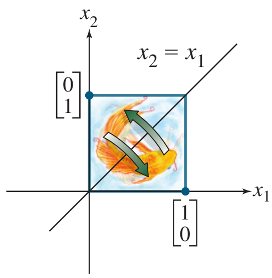
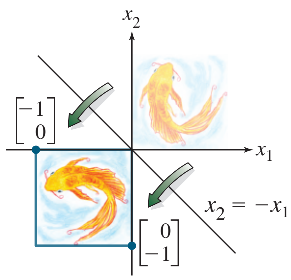
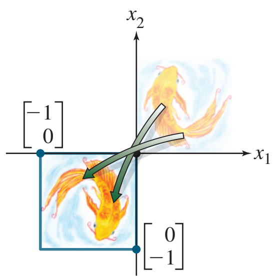
Transformation	Image of the Unit Square	Standard Matrix
Reflection through the x_1-axis		$\begin{bmatrix}1 & 0\\0 & -1\end{bmatrix}$
Reflection through the x_2-axis		$\begin{bmatrix}-1 & 0\\0 & 1\end{bmatrix}$
Reflection through the line $x_2 = x_1$		$\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$
Reflection through the line $x_2 = -x_1$		$\begin{bmatrix}0 & -1\\-1 & 0\end{bmatrix}$
Reflection through the origin		$\begin{bmatrix}-1 & 0\\0 & -1\end{bmatrix}$

TABLE 2 Contractions and Expansions

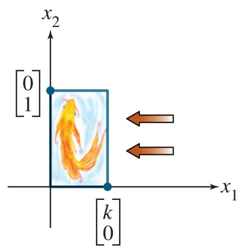
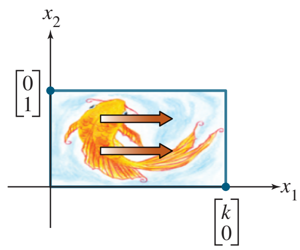
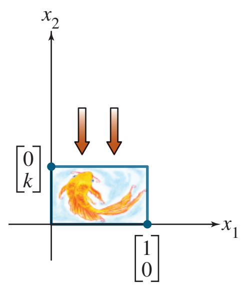
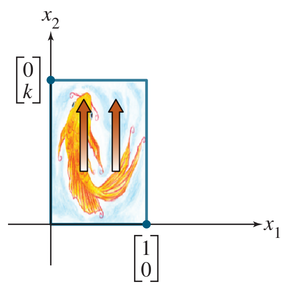
Transformation	Image of the Unit Square		Standard Matrix
Horizontal contraction and expansion	 $0 < k < 1$	 $k > 1$	$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$
Vertical contraction and expansion	 $0 < k < 1$	 $k > 1$	$\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$

TABLE 3 Shears

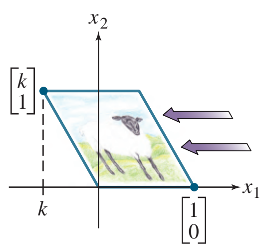
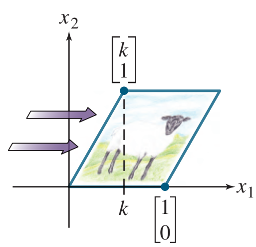
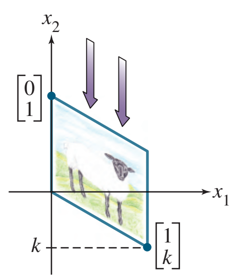
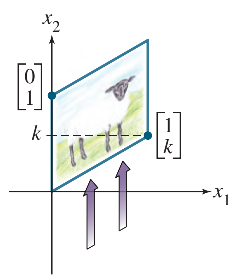
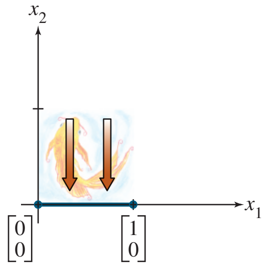
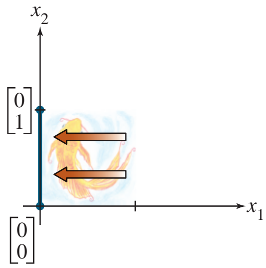
Transformation	Image of the Unit Square		Standard Matrix
Horizontal shear	 $k < 0$	 $k > 0$	$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$
Vertical shear	 $k < 0$	 $k > 0$	$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$

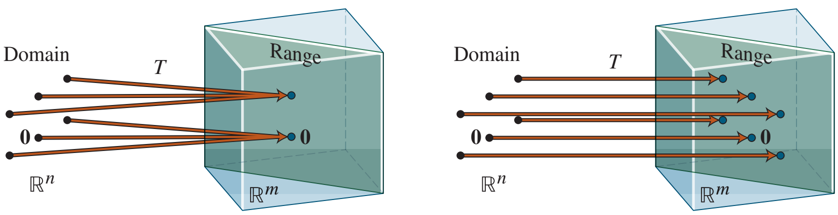
TABLE 4 Projections

Transformation	Image of the Unit Square	Standard Matrix
Projection onto the x_1-axis		$\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}$
Projection onto the x_2-axis		$\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}$

DEFINITION

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **one-to-one** if each $\mathbf{b}$ in $\mathbb{R}^m$ is the image of *at most one* $\mathbf{x}$ in $\mathbb{R}^n$.

Equivalently, T is one-to-one if, for each $\mathbf{b}$ in $\mathbb{R}^m$, the equation $T(\mathbf{x}) = \mathbf{b}$ has either a unique solution or none at all. "Is T one-to-one?" is a uniqueness question. The mapping T is *not* one-to-one when some $\mathbf{b}$ in $\mathbb{R}^m$ is the image of more than one vector in $\mathbb{R}^n$. If there is no such $\mathbf{b}$, then T is one-to-one. See Figure 4.



T is *not* one-to-one $\qquad$ T is one-to-one

FIGURE 4 Is every $\mathbf{b}$ the image of at most one vector?

The projection transformations shown in Table 4 are *not* one-to-one and do *not* map $\mathbb{R}^2$ onto $\mathbb{R}^2$. The transformations in Tables 1, 2, and 3 are one-to-one *and* do map $\mathbb{R}^2$ onto $\mathbb{R}^2$. Other possibilities are shown in the two examples below.

Example 4 and the theorems that follow show how the function properties of being one-to-one and mapping onto are related to important concepts studied earlier in this chapter.

EXAMPLE 4 Let T be the linear transformation whose standard matrix is

$$A = \begin{bmatrix} 1 & -4 & 8 & 1 \\ 0 & 2 & -1 & 3 \\ 0 & 0 & 0 & 5 \end{bmatrix}$$

Does T map $\mathbb{R}^4$ onto $\mathbb{R}^3$? Is T a one-to-one mapping?

SOLUTION Since A happens to be in echelon form, we can see at once that A has a pivot position in each row. By Theorem 4 in Section 1.4, for each $\mathbf{b}$ in $\mathbb{R}^3$, the equation $A\mathbf{x} = \mathbf{b}$ is consistent. In other words, the linear transformation T maps $\mathbb{R}^4$ (its domain) onto $\mathbb{R}^3$. However, since the equation $A\mathbf{x} = \mathbf{b}$ has a free variable (because there are four variables and only three basic variables), each $\mathbf{b}$ is the image of more than one $\mathbf{x}$. That is, T is *not* one-to-one. ■

THEOREM 11

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then T is one-to-one if and only if the equation $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution.

Remark: To prove a theorem that says "statement P is true if and only if statement Q is true," one must establish two things: (1) If P is true, then Q is true and (2) If Q is true, then P is true. The second requirement can also be established by showing (2a): If P is false, then Q is false. (This is called contrapositive reasoning.) This proof uses (1) and (2a) to show that P and Q are either both true or both false.

PROOF Since T is linear, $T(\mathbf{0}) = \mathbf{0}$. If T is one-to-one, then the equation $T(\mathbf{x}) = \mathbf{0}$ has at most one solution and hence only the trivial solution. If T is not one-to-one, then there is a $\mathbf{b}$ that is the image of at least two different vectors in $\mathbb{R}^n$—say, $\mathbf{u}$ and $\mathbf{v}$. That is, $T(\mathbf{u}) = \mathbf{b}$ and $T(\mathbf{v}) = \mathbf{b}$. But then, since T is linear,

$$T(\mathbf{u} - \mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v}) = \mathbf{b} - \mathbf{b} = \mathbf{0}$$

The vector $\mathbf{u} - \mathbf{v}$ is not zero, since $\mathbf{u} \neq \mathbf{v}$. Hence the equation $T(\mathbf{x}) = \mathbf{0}$ has more than one solution. So, either the two conditions in the theorem are both true or they are both false. ■

THEOREM 12

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and let A be the standard matrix for T. Then:

a. T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if and only if the columns of A span $\mathbb{R}^m$;

b. T is one-to-one if and only if the columns of A are linearly independent.

Remark: "If and only if" statements can be linked together. For example if "P if and only if Q" is known and "Q if and only if R" is known, then one can conclude "P if and only if R." This strategy is used repeatedly in this proof.

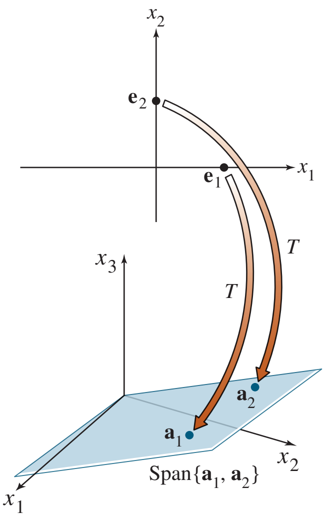
PROOF

a. By Theorem 4 in Section 1.4, the columns of A span $\mathbb{R}^m$ if and only if for each $\mathbf{b}$ in $\mathbb{R}^m$ the equation $A\mathbf{x} = \mathbf{b}$ is consistent—in other words, if and only if for every $\mathbf{b}$, the equation $T(\mathbf{x}) = \mathbf{b}$ has at least one solution. This is true if and only if T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$.

b. The equations $T(\mathbf{x}) = \mathbf{0}$ and $A\mathbf{x} = \mathbf{0}$ are the same except for notation. So, by Theorem 11, T is one-to-one if and only if $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. This happens if and only if the columns of A are linearly independent, as was already noted in the boxed statement (3) in Section 1.7. ■

Statement (a) in Theorem 12 is equivalent to the statement "T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if and only if every vector in $\mathbb{R}^m$ is a linear combination of the columns of A." See Theorem 4 in Section 1.4.

In the next example and in some exercises that follow, column vectors are written in rows, such as $\mathbf{x} = (x_1, x_2)$, and $T(\mathbf{x})$ is written as $T(x_1, x_2)$ instead of the more formal $T((x_1, x_2))$.



The transformation T is not onto $\mathbb{R}^3$.

EXAMPLE 5 Let $T(x_1, x_2) = (3x_1 + x_2,\ 5x_1 + 7x_2,\ x_1 + 3x_2)$. Show that T is a one-to-one linear transformation. Does T map $\mathbb{R}^2$ onto $\mathbb{R}^3$?

SOLUTION When $\mathbf{x}$ and $T(\mathbf{x})$ are written as column vectors, you can determine the standard matrix of T by inspection, visualizing the row–vector computation of each entry in $A\mathbf{x}$.

$$T(\mathbf{x}) = \begin{bmatrix} 3x_1 + x_2 \\ 5x_1 + 7x_2 \\ x_1 + 3x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} ? & ? \\ ? & ? \\ ? & ? \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 5 & 7 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{4}$$

So T is indeed a linear transformation, with its standard matrix A shown in (4). The columns of A are linearly independent because they are not multiples. By Theorem 12(b), T is one-to-one. To decide if T is onto $\mathbb{R}^3$, examine the span of the columns of A. Since A is 3×2, the columns of A span $\mathbb{R}^3$ if and only if A has 3 pivot positions, by Theorem 4. This is impossible, since A has only 2 columns. So the columns of A do not span $\mathbb{R}^3$, and the associated linear transformation is not onto $\mathbb{R}^3$. ■

Practice Problems

1. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that first performs a horizontal shear that maps $\mathbf{e}_2$ into $\mathbf{e}_2 - .5\mathbf{e}_1$ (but leaves $\mathbf{e}_1$ unchanged) and then reflects the result through the x_2-axis. Assuming that T is linear, find its standard matrix. [*Hint:* Determine the final location of the images of $\mathbf{e}_1$ and $\mathbf{e}_2$.]
2. Suppose A is a 7×5 matrix with 5 pivots. Let $T(\mathbf{x}) = A\mathbf{x}$ be a linear transformation from $\mathbb{R}^5$ into $\mathbb{R}^7$. Is T a one-to-one linear transformation? Is T onto $\mathbb{R}^7$?

1.9 Exercises

In Exercises 1–10, assume that T is a linear transformation. Find the standard matrix of T.

1. $T : \mathbb{R}^2 \to \mathbb{R}^4$, $T(\mathbf{e}_1) = (2, 1, 2, 1)$ and $T(\mathbf{e}_2) = (-5, 2, 0, 0)$, where $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$.
2. $T : \mathbb{R}^3 \to \mathbb{R}^2$, $T(\mathbf{e}_1) = (1, 3)$, $T(\mathbf{e}_2) = (4, 2)$, and $T(\mathbf{e}_3) = (-5, 4)$, where $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are the columns of the 3×3 identity matrix.
3. $T : \mathbb{R}^2 \to \mathbb{R}^2$ rotates points (about the origin) through $3\pi/2$ radians (in the counterclockwise direction).

4. $T: \mathbb{R}^2 \to \mathbb{R}^2$ rotates points (about the origin) through $-\pi/4$ radians (since the number is negative, the actual rotation is clockwise). [*Hint:* $T(\mathbf{e}_1) = (1/\sqrt{2}, -1/\sqrt{2})$.]

5. $T: \mathbb{R}^2 \to \mathbb{R}^2$ is a vertical shear transformation that maps $\mathbf{e}_1$ into $\mathbf{e}_1 - 2\mathbf{e}_2$ but leaves the vector $\mathbf{e}_2$ unchanged.

6. $T: \mathbb{R}^2 \to \mathbb{R}^2$ is a horizontal shear transformation that leaves $\mathbf{e}_1$ unchanged and maps $\mathbf{e}_2$ into $\mathbf{e}_2 + 5\mathbf{e}_1$.

7. $T: \mathbb{R}^2 \to \mathbb{R}^2$ first rotates points through $-3\pi/4$ radians (since the number is negative, the actual rotation is clockwise) and then reflects points through the horizontal x_1-axis. [*Hint:* $T(\mathbf{e}_1) = (-1/\sqrt{2}, 1/\sqrt{2})$.]

8. $T: \mathbb{R}^2 \to \mathbb{R}^2$ first reflects points through the vertical x_2-axis and then reflects points through the line $x_2 = x_1$.

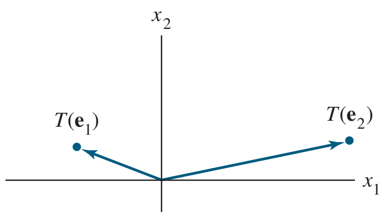
9. $T: \mathbb{R}^2 \to \mathbb{R}^2$ first performs a horizontal shear that transforms $\mathbf{e}_2$ into $\mathbf{e}_2 - 3\mathbf{e}_1$ (leaving $\mathbf{e}_1$ unchanged) and then reflects points through the line $x_2 = -x_1$.

10. $T: \mathbb{R}^2 \to \mathbb{R}^2$ first reflects points through the vertical x_2-axis and then rotates points $3\pi/2$ radians.

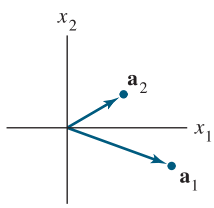
11. A linear transformation $T: \mathbb{R}^2 \to \mathbb{R}^2$ first reflects points through the x_1-axis and then reflects points through the x_2-axis. Show that T can also be described as a linear transformation that rotates points about the origin. What is the angle of that rotation?

12. Show that the transformation in Exercise 8 is merely a rotation about the origin. What is the angle of the rotation?

13. Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation such that $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$ are the vectors shown in the figure. Using the figure, sketch the vector $T(2, 1)$.



14. Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation with standard matrix $A = [\mathbf{a}_1 \ \ \mathbf{a}_2]$, where $\mathbf{a}_1$ and $\mathbf{a}_2$ are shown in the figure. Using the figure, draw the image of $\begin{bmatrix} -1 \\ 3 \end{bmatrix}$ under the transformation T.



In Exercises 15 and 16, fill in the missing entries of the matrix, assuming that the equation holds for all values of the variables.

15. $\begin{bmatrix} ? & ? & ? \\ ? & ? & ? \\ ? & ? & ? \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_1 - 3x_3 \\ 4x_1 \\ x_1 - x_2 + x_3 \end{bmatrix}$

16. $\begin{bmatrix} ? & ? \\ ? & ? \\ ? & ? \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - 3x_2 \\ -2x_1 + x_2 \\ x_1 \end{bmatrix}$

In Exercises 17–20, show that T is a linear transformation by finding a matrix that implements the mapping. Note that $x_1, x_2, \ldots$ are not vectors but are entries in vectors.

17. $T(x_1, x_2, x_3, x_4) = (0, x_1 + x_2, x_2 + x_3, x_3 + x_4)$

18. $T(x_1, x_2) = (2x_2 - 3x_1, x_1 - 4x_2, 0, x_2)$

19. $T(x_1, x_2, x_3) = (x_1 - 5x_2 + 4x_3, x_2 - 6x_3)$

20. $T(x_1, x_2, x_3, x_4) = 2x_1 + 3x_3 - 4x_4 \qquad (T: \mathbb{R}^4 \to \mathbb{R})$

21. Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation such that $T(x_1, x_2) = (x_1 + x_2, 4x_1 + 5x_2)$. Find $\mathbf{x}$ such that $T(\mathbf{x}) = (3, 8)$.

22. Let $T: \mathbb{R}^2 \to \mathbb{R}^3$ be a linear transformation such that $T(x_1, x_2) = (x_1 - 2x_2, -x_1 + 3x_2, 3x_1 - 2x_2)$. Find $\mathbf{x}$ such that $T(\mathbf{x}) = (-1, 4, 9)$.

In Exercises 23–32, mark each statement True or False **(T/F)**. Justify each answer.

23. **(T/F)** A linear transformation $T: \mathbb{R}^n \to \mathbb{R}^m$ is completely determined by its effect on the columns of the $n \times n$ identity matrix.

24. **(T/F)** A mapping $T: \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one if each vector in $\mathbb{R}^n$ maps onto a unique vector in $\mathbb{R}^m$.

25. **(T/F)** If $T: \mathbb{R}^2 \to \mathbb{R}^2$ rotates vectors about the origin through an angle ϕ, then T is a linear transformation.

26. **(T/F)** The columns of the standard matrix for a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ are the images of the columns of the $n \times n$ identity matrix.

27. **(T/F)** When two linear transformations are performed one after another, the combined effect may not always be a linear transformation.

28. **(T/F)** Not every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a matrix transformation.

29. **(T/F)** A mapping $T: \mathbb{R}^n \to \mathbb{R}^m$ is onto $\mathbb{R}^m$ if every vector $\mathbf{x}$ in $\mathbb{R}^n$ maps onto some vector in $\mathbb{R}^m$.

30. **(T/F)** The standard matrix of a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$ that reflects points through the horizontal axis, the vertical axis, or the origin has the form $\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}$, where a and d are ± 1.

31. **(T/F)** A is a 3×2 matrix, then the transformation $\mathbf{x} \mapsto A\mathbf{x}$ cannot be one-to-one.

32. **(T/F)** A is a 3×2 matrix, then the transformation $\mathbf{x} \mapsto A\mathbf{x}$ cannot map $\mathbb{R}^2$ onto $\mathbb{R}^3$.

In Exercises 33–36, determine if the specified linear transformation is (a) one-to-one and (b) onto. Justify each answer.

33. The transformation in Exercise 17

34. The transformation in Exercise 2

35. The transformation in Exercise 19

36. The transformation in Exercise 14

In Exercises 37 and 38, describe the possible echelon forms of the standard matrix for a linear transformation T. Use the notation of Example 1 in Section 1.2.

37. $T : \mathbb{R}^3 \to \mathbb{R}^4$ is one-to-one.

38. $T : \mathbb{R}^4 \to \mathbb{R}^3$ is onto.

39. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, with A its standard matrix. Complete the following statement to make it true: "T is one-to-one if and only if A has ____ pivot columns." Explain why the statement is true. [*Hint:* Look in the exercises for Section 1.7.]

40. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, with A its standard matrix. Complete the following statement to make it true: "T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if and only if A has ____ pivot columns." Find some theorems that explain why the statement is true.

41. Verify the uniqueness of A in Theorem 10. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation such that $T(\mathbf{x}) = B\mathbf{x}$ for some $m \times n$ matrix B. Show that if A is the standard matrix for T, then $A = B$. [*Hint:* Show that A and B have the same columns.]

42. Why is the question "Is the linear transformation T onto?" an existence question?

43. If a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ maps $\mathbb{R}^n$ *onto* $\mathbb{R}^m$, can you give a relation between m and n? If T is one-to-one, what can you say about m and n?

44. Let $S : \mathbb{R}^p \to \mathbb{R}^n$ and $T : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations. Show that the mapping $\mathbf{x} \mapsto T(S(\mathbf{x}))$ is a linear transformation (from $\mathbb{R}^p$ to $\mathbb{R}^m$). [*Hint:* Compute $T(S(c\mathbf{u} + d\mathbf{v}))$ for $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^p$ and scalars c and d. Justify each step of the computation, and explain why this computation gives the desired conclusion.]

T In Exercises 45–48, let T be the linear transformation whose standard matrix is given. In Exercises 45 and 46, decide if T is a one-to-one mapping. In Exercises 47 and 48, decide if T maps $\mathbb{R}^5$ onto $\mathbb{R}^5$. Justify your answers.

45. $\begin{bmatrix} -5 & 10 & -5 & 4 \\ 8 & 3 & -4 & 7 \\ 4 & -9 & 5 & -3 \\ -3 & -2 & 5 & 4 \end{bmatrix}$

46. $\begin{bmatrix} 7 & 5 & 4 & -9 \\ 10 & 6 & 16 & -4 \\ 12 & 8 & 12 & 7 \\ -8 & -6 & -2 & 5 \end{bmatrix}$

47. $\begin{bmatrix} 4 & -7 & 3 & 7 & 5 \\ 6 & -8 & 5 & 12 & -8 \\ -7 & 10 & -8 & -9 & 14 \\ 3 & -5 & 4 & 2 & -6 \\ -5 & 6 & -6 & -7 & 3 \end{bmatrix}$

48. $\begin{bmatrix} 9 & 13 & 5 & 6 & -1 \\ 14 & 15 & -7 & -6 & 4 \\ -8 & -9 & 12 & -5 & -9 \\ -5 & -6 & -8 & 9 & 8 \\ 13 & 14 & 15 & 2 & 11 \end{bmatrix}$

STUDY GUIDE offers additional resources for mastering existence and uniqueness.

Solution to Practice Problems

1. Follow what happens to $\mathbf{e}_1$ and $\mathbf{e}_2$. See Figure 5. First, $\mathbf{e}_1$ is unaffected by the shear and then is reflected into $-\mathbf{e}_1$. So $T(\mathbf{e}_1) = -\mathbf{e}_1$. Second, $\mathbf{e}_2$ goes to $\mathbf{e}_2 - .5\mathbf{e}_1$ by the shear transformation. Since reflection through the x_2-axis changes $\mathbf{e}_1$ into $-\mathbf{e}_1$ and leaves $\mathbf{e}_2$ unchanged, the vector $\mathbf{e}_2 - .5\mathbf{e}_1$ goes to $\mathbf{e}_2 + .5\mathbf{e}_1$. So $T(\mathbf{e}_2) = \mathbf{e}_2 + .5\mathbf{e}_1$.

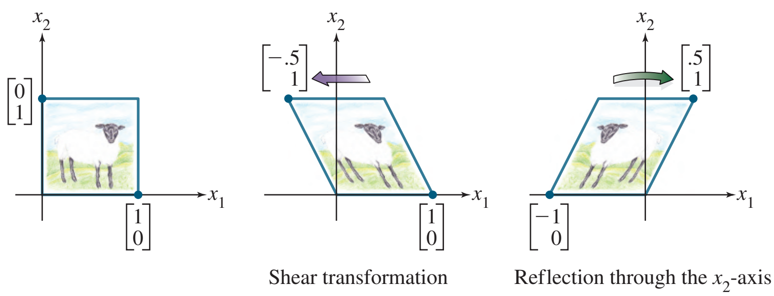


FIGURE 5 The composition of two transformations.

Thus the standard matrix of T is

$$\begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} -\mathbf{e}_1 & \mathbf{e}_2 + .5\mathbf{e}_1 \end{bmatrix} = \begin{bmatrix} -1 & .5 \\ 0 & 1 \end{bmatrix}$$

2. The standard matrix representation of T is the matrix A. Since A has 5 columns and 5 pivots, there is a pivot in every column so the columns are linearly independent. By Theorem 12, T is one-to-one. Since A has 7 rows and only 5 pivots, there is not a pivot in every row hence the columns of A do not span $\mathbb{R}^7$. By Theorem 12, and T is not onto.

1.10 Linear Models in Business, Science, and Engineering

The mathematical models in this section are all *linear*; that is, each describes a problem by means of a linear equation, usually in vector or matrix form. The first model concerns nutrition but actually is representative of a general technique in linear programming problems. The second model comes from electrical engineering. The third model introduces the concept of a *linear difference equation*, a powerful mathematical tool for studying dynamic processes in a wide variety of fields such as engineering, ecology, economics, telecommunications, and the management sciences. Linear models are important because natural phenomena are often linear or nearly linear when the variables involved are held within reasonable bounds. Also, linear models are more easily adapted for computer calculation than are complex nonlinear models.

As you read about each model, pay attention to how its linearity reflects some property of the system being modeled.

Constructing a Nutritious Weight-Loss Diet

The formula for the Cambridge Diet, a popular diet in the 1980s, was based on years of research. A team of scientists headed by Dr. Alan H. Howard developed this diet at Cambridge University after more than eight years of clinical work with obese patients.[1] The very low-calorie powdered formula diet combines a precise balance of carbohydrate, high-quality protein, and fat, together with vitamins, minerals, trace elements, and electrolytes. Millions of persons have used the diet to achieve rapid and substantial weight loss.

To achieve the desired amounts and proportions of nutrients, Dr. Howard had to incorporate a large variety of foodstuffs in the diet. Each foodstuff supplied several of the required ingredients, but not in the correct proportions. For instance, nonfat milk was a major source of protein but contained too much calcium. So soy flour was used for part of the protein because soy flour contains little calcium. However, soy flour contains proportionally too much fat, so whey was added since it supplies less fat in relation to calcium. Unfortunately, whey contains too much carbohydrate. . . .

The following example illustrates the problem on a small scale. Listed in Table 1 are three of the ingredients in the diet, together with the amounts of certain nutrients supplied by 100 grams (g) of each ingredient.[2]

[1] The first announcement of this rapid weight-loss regimen was given in the *International Journal of Obesity* (1978) **2**, 321–332.

[2] Ingredients in the diet as of 1984; nutrient data for ingredients adapted from USDA Agricultural Handbooks No. 8-1 and 8-6, 1976.

TABLE 1 The Cambridge Diet

Amounts (g) Supplied per 100 g of Ingredient				Amounts (g) Supplied by Cambridge Diet in One Day
Nutrient	**Nonfat milk**	**Soy flour**	**Whey**	
Protein	36	51	13	33
Carbohydrate	52	34	74	45
Fat	0	7	1.1	3

EXAMPLE 1 If possible, find some combination of nonfat milk, soy flour, and whey to provide the exact amounts of protein, carbohydrate, and fat supplied by the diet in one day (Table 1).

SOLUTION Let x_1, x_2, and x_3, respectively, denote the number of units (100 g) of these foodstuffs. One approach to the problem is to derive equations for each nutrient separately. For instance, the product

$$\left\{\begin{matrix} x_1 \text{ units of} \\ \text{nonfat milk} \end{matrix}\right\} \cdot \left\{\begin{matrix} \text{protein per unit} \\ \text{of nonfat milk} \end{matrix}\right\}$$

gives the amount of protein supplied by x_1 units of nonfat milk. To this amount, we would then add similar products for soy flour and whey and set the resulting sum equal to the amount of protein we need. Analogous calculations would have to be made for each nutrient.

A more efficient method, and one that is conceptually simpler, is to consider a "nutrient vector" for each foodstuff and build just one vector equation. The amount of nutrients supplied by x_1 units of nonfat milk is the scalar multiple

$$\overset{\text{Scalar}}{\left\{\begin{matrix} x_1 \text{ units of} \\ \text{nonfat milk} \end{matrix}\right\}} \cdot \overset{\text{Vector}}{\left\{\begin{matrix} \text{nutrients per unit} \\ \text{of nonfat milk} \end{matrix}\right\}} = x_1\mathbf{a}_1 \tag{1}$$

where $\mathbf{a}_1$ is the first column in Table 1. Let $\mathbf{a}_2$ and $\mathbf{a}_3$ be the corresponding vectors for soy flour and whey, respectively, and let $\mathbf{b}$ be the vector that lists the total nutrients required (the last column of the table). Then $x_2\mathbf{a}_2$ and $x_3\mathbf{a}_3$ give the nutrients supplied by x_2 units of soy flour and x_3 units of whey, respectively. So the relevant equation is

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 = \mathbf{b} \tag{2}$$

Row reduction of the augmented matrix for the corresponding system of equations shows that

$$\begin{bmatrix} 36 & 51 & 13 & 33 \\ 52 & 34 & 74 & 45 \\ 0 & 7 & 1.1 & 3 \end{bmatrix} \sim \cdots \sim \begin{bmatrix} 1 & 0 & 0 & .277 \\ 0 & 1 & 0 & .392 \\ 0 & 0 & 1 & .233 \end{bmatrix}$$

To three significant digits, the diet requires .277 units of nonfat milk, .392 units of soy flour, and .233 units of whey in order to provide the desired amounts of protein, carbohydrate, and fat. ■

It is important that the values of x_1, x_2, and x_3 found above are nonnegative. This is necessary for the solution to be physically feasible. (How could you use $-.233$ units of whey, for instance?) With a large number of nutrient requirements, it may be necessary to use a larger number of foodstuffs in order to produce a system of equations with a "nonnegative" solution. Thus many, many different combinations of foodstuffs may need to be examined in order to find a system of equations with such a solution. In fact, the manufacturer of the Cambridge Diet was able to supply 31 nutrients in precise amounts using only 33 ingredients.

The diet construction problem leads to the *linear* equation (2) because the amount of nutrients supplied by each foodstuff can be written as a scalar multiple of a vector, as in (1). That is, the nutrients supplied by a foodstuff are *proportional* to the amount of the foodstuff added to the diet mixture. Also, each nutrient in the mixture is the *sum* of the amounts from the various foodstuffs.

Problems of formulating specialized diets for humans and livestock occur frequently. Usually they are treated by linear programming techniques. Our method of constructing vector equations often simplifies the task of formulating such problems.

Linear Equations and Electrical Networks

Current flow in a simple electrical network can be described by a system of linear equations. A voltage source such as a battery forces a current of electrons to flow through the network. When the current passes through a resistor (such as a lightbulb or motor), some of the voltage is "used up"; by Ohm's law, this "voltage drop" across a resistor is given by

$$V = RI$$

where the voltage V is measured in *volts*, the resistance R in *ohms* (denoted by Ω), and the current flow I in *amperes* (*amps*, for short).

The network in Figure 1 contains three closed loops. The currents flowing in loops 1, 2, and 3 are denoted by I_1, I_2, and I_3, respectively. The designated directions of such *loop currents* are arbitrary. If a current turns out to be negative, then the actual direction of current flow is opposite to that chosen in the figure. If the current direction shown is away from the positive (longer) side of a battery (⊣⊢) around to the negative (shorter) side, the voltage is positive; otherwise, the voltage is negative.

Current flow in a loop is governed by the following rule.

KIRCHHOFF'S VOLTAGE LAW

The algebraic sum of the RI voltage drops in one direction around a loop equals the algebraic sum of the voltage sources in the same direction around the loop.

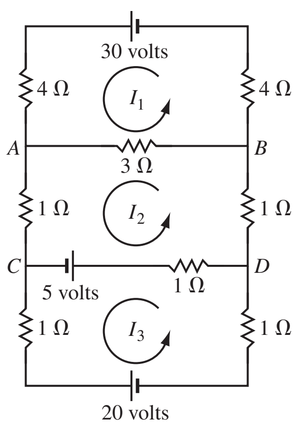


FIGURE 1

EXAMPLE 2 Determine the loop currents in the network in Figure 1.

SOLUTION For loop 1, the current I_1 flows through three resistors, and the sum of the RI voltage drops is

$$4I_1 + 4I_1 + 3I_1 = (4 + 4 + 3)I_1 = 11I_1$$

Current from loop 2 also flows in part of loop 1, through the short *branch* between A and B. The associated RI drop there is $3I_2$ volts. However, the current direction for the branch AB in loop 1 is opposite to that chosen for the flow in loop 2, so the algebraic sum of all RI drops for loop 1 is $11I_1 - 3I_2$. Since the voltage in loop 1 is $+30$ volts, Kirchhoff's voltage law implies that

$$11I_1 - 3I_2 = 30$$

The equation for loop 2 is

$$-3I_1 + 6I_2 - I_3 = 5$$

The term $-3I_1$ comes from the flow of the loop 1 current through the branch AB (with a negative voltage drop because the current flow there is opposite to the flow in loop 2). The term $6I_2$ is the sum of all resistances in loop 2, multiplied by the loop current. The term $-I_3 = -1 \cdot I_3$ comes from the loop 3 current flowing through the 1-ohm resistor in branch CD, in the direction opposite to the flow in loop 2. The loop 3 equation is

$$-I_2 + 3I_3 = -25$$

Note that the 5-volt battery in branch CD is counted as part of both loop 2 and loop 3, but it is -5 volts for loop 3 because of the direction chosen for the current in loop 3. The 20-volt battery is negative for the same reason.

The loop currents are found by solving the system

$$\begin{aligned} 11I_1 - 3I_2 \qquad\quad &= 30 \\ -3I_1 + 6I_2 - I_3 &= 5 \\ -I_2 + 3I_3 &= -25 \end{aligned} \tag{3}$$

Row operations on the augmented matrix lead to the solution: $I_1 = 3$ amps, $I_2 = 1$ amp, and $I_3 = -8$ amps. The negative value of I_3 indicates that the actual current in loop 3 flows in the direction opposite to that shown in Figure 1. ■

It is instructive to look at system (3) as a vector equation:

$$I_1 \underbrace{\begin{bmatrix} 11 \\ -3 \\ 0 \end{bmatrix}}_{\mathbf{r}_1} + I_2 \underbrace{\begin{bmatrix} -3 \\ 6 \\ -1 \end{bmatrix}}_{\mathbf{r}_2} + I_3 \underbrace{\begin{bmatrix} 0 \\ -1 \\ 3 \end{bmatrix}}_{\mathbf{r}_3} = \underbrace{\begin{bmatrix} 30 \\ 5 \\ -25 \end{bmatrix}}_{\mathbf{v}} \tag{4}$$

The first entry of each vector concerns the first loop, and similarly for the second and third entries. The first resistor vector $\mathbf{r}_1$ lists the resistance in the various loops through which current I_1 flows. A resistance is written negatively when I_1 flows against the flow direction in another loop. Examine Figure 1 and see how to compute the entries in $\mathbf{r}_1$; then do the same for $\mathbf{r}_2$ and $\mathbf{r}_3$. The matrix form of equation (4),

$$R\mathbf{i} = \mathbf{v}, \quad \text{where} \quad R = [\,\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{r}_3\,] \quad \text{and} \quad \mathbf{i} = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix}$$

provides a matrix version of Ohm's law. If all loop currents are chosen in the same direction (say, counterclockwise), then all entries off the main diagonal of R will be negative.

The matrix equation $R\mathbf{i} = \mathbf{v}$ makes the linearity of this model easy to see at a glance. For instance, if the voltage vector is doubled, then the current vector must double. Also, a *superposition principle* holds. That is, the solution of equation (4) is the sum of the solutions of the equations

$$R\mathbf{i} = \begin{bmatrix} 30 \\ 0 \\ 0 \end{bmatrix}, \qquad R\mathbf{i} = \begin{bmatrix} 0 \\ 5 \\ 0 \end{bmatrix}, \quad \text{and} \quad R\mathbf{i} = \begin{bmatrix} 0 \\ 0 \\ -25 \end{bmatrix}$$

Each equation here corresponds to the circuit with only one voltage source (the other sources being replaced by wires that close each loop). The model for current flow is *linear* precisely because Ohm's law and Kirchhoff's law are linear: The voltage drop across a resistor is *proportional* to the current flowing through it (Ohm), and the *sum* of the voltage drops in a loop equals the sum of the voltage sources in the loop (Kirchhoff).

Loop currents in a network can be used to determine the current in any branch of the network. If only one loop current passes through a branch, such as from B to D in Figure 1, the branch current equals the loop current. If more than one loop current passes through a branch, such as from A to B, the branch current is the algebraic sum of the loop currents in the branch (*Kirchhoff's current law*). For instance, the current in branch AB is $I_1 - I_2 = 3 - 1 = 2$ amps, in the direction of I_1. The current in branch CD is $I_2 - I_3 = 9$ amps.

Difference Equations

In many fields, such as ecology, economics, and engineering, a need arises to model mathematically a dynamic system that changes over time. Several features of the system are each measured at discrete time intervals, producing a sequence of vectors $\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{x}_2, \ldots$. The entries in $\mathbf{x}_k$ provide information about the *state* of the system at the time of the kth measurement.

If there is a matrix A such that $\mathbf{x}_1 = A\mathbf{x}_0$, $\mathbf{x}_2 = A\mathbf{x}_1$, and, in general,

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \quad \text{for } k = 0, 1, 2, \ldots \tag{5}$$

then (5) is called a **linear difference equation** (or **recurrence relation**). Given such an equation, one can compute $\mathbf{x}_1$, $\mathbf{x}_2$, and so on, provided $\mathbf{x}_0$ is known. Sections 4.8 and several sections in Chapter 5 will develop formulas for $\mathbf{x}_k$ and describe what can happen to $\mathbf{x}_k$ as k increases indefinitely. The discussion below illustrates how a difference equation might arise.

A subject of interest to demographers is the movement of populations or groups of people from one region to another. The simple model here considers the changes in the population of a certain city and its surrounding suburbs over a period of years.

Fix an initial year—say, 2020—and denote the populations of the city and suburbs that year by r_0 and s_0, respectively. Let $\mathbf{x}_0$ be the population vector

$$\mathbf{x}_0 = \begin{bmatrix} r_0 \\ s_0 \end{bmatrix} \begin{array}{l} \text{City population, 2020} \\ \text{Suburban population, 2020} \end{array}$$

For 2021 and subsequent years, denote the populations of the city and suburbs by the vectors

$$\mathbf{x}_1 = \begin{bmatrix} r_1 \\ s_1 \end{bmatrix}, \qquad \mathbf{x}_2 = \begin{bmatrix} r_2 \\ s_2 \end{bmatrix}, \qquad \mathbf{x}_3 = \begin{bmatrix} r_3 \\ s_3 \end{bmatrix}, \ldots$$

Our goal is to describe mathematically how these vectors might be related.

Suppose demographic studies show that each year about 5% of the city's population moves to the suburbs (and 95% remains in the city), while 3% of the suburban population moves to the city (and 97% remains in the suburbs). See Figure 2.

After 1 year, the original r_0 persons in the city are now distributed between city and suburbs as

$$\begin{bmatrix} .95r_0 \\ .05r_0 \end{bmatrix} = r_0 \begin{bmatrix} .95 \\ .05 \end{bmatrix} \begin{array}{l} \text{Remain in city} \\ \text{Move to suburbs} \end{array} \tag{6}$$

The s_0 persons in the suburbs in 2020 are distributed 1 year later as

$$s_0 \begin{bmatrix} .03 \\ .97 \end{bmatrix} \begin{array}{l} \text{Move to city} \\ \text{Remain in suburbs} \end{array} \tag{7}$$

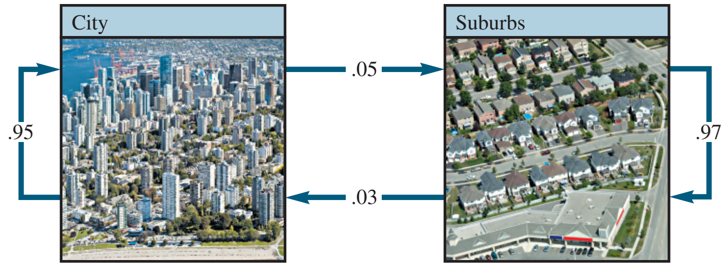


FIGURE 2 Annual percentage migration between city and suburbs.

The vectors in (6) and (7) account for all of the population in 2021.[3] Thus

$$\begin{bmatrix} r_1 \\ s_1 \end{bmatrix} = r_0 \begin{bmatrix} .95 \\ .05 \end{bmatrix} + s_0 \begin{bmatrix} .03 \\ .97 \end{bmatrix} = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix} \begin{bmatrix} r_0 \\ s_0 \end{bmatrix}$$

That is,

$$\mathbf{x}_1 = M\mathbf{x}_0 \tag{8}$$

where M is the **migration matrix** determined by the following table:

$$\begin{array}{cl} \text{From:} & \\ \begin{array}{cc} \text{City} & \text{Suburbs} \end{array} & \text{To:} \\ \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix} & \begin{array}{l} \text{City} \\ \text{Suburbs} \end{array} \end{array}$$

Equation (8) describes how the population changes from 2020 to 2021. If the migration percentages remain constant, then the change from 2021 to 2022 is given by

$$\mathbf{x}_2 = M\mathbf{x}_1$$

and similarly for 2022 to 2023 and subsequent years. In general,

$$\mathbf{x}_{k+1} = M\mathbf{x}_k \quad \text{for } k = 0, 1, 2, \dots \tag{9}$$

The sequence of vectors $\{\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots\}$ describes the population of the city/suburban region over a period of years.

EXAMPLE 3 Compute the population of the region just described for the years 2021 and 2022, given that the population in 2020 was 600,000 in the city and 400,000 in the suburbs.

SOLUTION The initial population in 2020 is $\mathbf{x}_0 = \begin{bmatrix} 600{,}000 \\ 400{,}000 \end{bmatrix}$. For 2021,

$$\mathbf{x}_1 = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix} \begin{bmatrix} 600{,}000 \\ 400{,}000 \end{bmatrix} = \begin{bmatrix} 582{,}000 \\ 418{,}000 \end{bmatrix}$$

For 2022,

$$\mathbf{x}_2 = M\mathbf{x}_1 = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix} \begin{bmatrix} 582{,}000 \\ 418{,}000 \end{bmatrix} = \begin{bmatrix} 565{,}440 \\ 434{,}560 \end{bmatrix}$$

[3] For simplicity, we ignore other influences on the population such as births, deaths, and migration into and out of the city/suburban region.

The model for population movement in (9) is *linear* because the correspondence $\mathbf{x}_k \mapsto \mathbf{x}_{k+1}$ is a linear transformation. The linearity depends on two facts: the number of people who chose to move from one area to another is *proportional* to the number of people in that area, as shown in (6) and (7), and the cumulative effect of these choices is found by *adding* the movement of people from the different areas.

Practice Problem

Find a matrix A and vectors $\mathbf{x}$ and $\mathbf{b}$ such that the problem in Example 1 amounts to solving the equation $A\mathbf{x} = \mathbf{b}$.

1.10 Exercises

1. The container of a breakfast cereal usually lists the number of calories and the amounts of protein, carbohydrate, and fat contained in one serving of the cereal. The amounts for two common cereals are given below. Suppose a mixture of these two cereals is to be prepared that contains exactly 295 calories, 9 g of protein, 48 g of carbohydrate, and 8 g of fat.

a. Set up a vector equation for this problem. Include a statement of what the variables in your equation represent.

b. Write an equivalent matrix equation, and then determine if the desired mixture of the two cereals can be prepared.

Nutrition Information per Serving

Nutrient	General Mills Cheerios®	Quaker® 100% Natural Cereal
Calories	110	130
Protein (g)	4	3
Carbohydrate (g)	20	18
Fat (g)	2	5

2. One serving of Post Shredded Wheat® supplies 160 calories, 5 g of protein, 6 g of fiber, and 1 g of fat. One serving of Crispix® supplies 110 calories, 2 g of protein, .1 g of fiber, and .4 g of fat.

a. Set up a matrix B and a vector $\mathbf{u}$ such that $B\mathbf{u}$ gives the amounts of calories, protein, fiber, and fat contained in a mixture of three servings of Shredded Wheat and two servings of Crispix.

T b. Suppose that you want a cereal with more fiber than Crispix but fewer calories than Shredded Wheat. Is it possible for a mixture of the two cereals to supply 130 calories, 3.20 g of protein, 2.46 g of fiber, and .64 g of fat? If so, what is the mixture?

3. After taking a nutrition class, a big Annie's® Mac and Cheese fan decides to improve the levels of protein and fiber in her favorite lunch by adding broccoli and canned chicken. The nutritional information for the foods referred to in this are given in the table.

Nutrition Information per Serving

Nutrient	Mac and Cheese	Broccoli	Chicken	Shells
Calories	270	51	70	260
Protein (g)	10	5.4	15	9
Fiber (g)	2	5.2	0	5

T a. If she wants to limit her lunch to 400 calories but get 30 g of protein and 10 g of fiber, what proportions of servings of Mac and Cheese, broccoli, and chicken should she use?

T b. She found that there was too much broccoli in the proportions from part (a), so she decided to switch from classical Mac and Cheese to Annie's® Whole Wheat Shells and White Cheddar. What proportions of servings of each food should she use to meet the same goals as in part (a)?

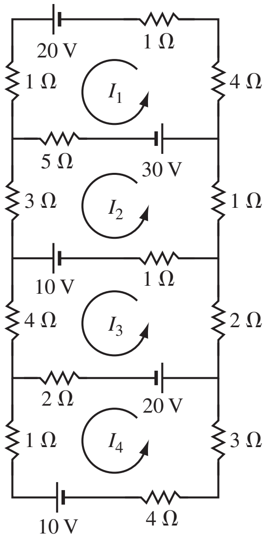
4. The Cambridge Diet supplies .8 g of calcium per day, in addition to the nutrients listed in Table 1 for Example 1. The amounts of calcium per unit (100 g) supplied by the three ingredients in the Cambridge Diet are as follows: 1.26 g from nonfat milk, .19 g from soy flour, and .8 g from whey. Another ingredient in the diet mixture is isolated soy protein, which provides the following nutrients in each unit: 80 g of protein, 0 g of carbohydrate, 3.4 g of fat, and .18 g of calcium.

a. Set up a matrix equation whose solution determines the amounts of nonfat milk, soy flour, whey, and isolated soy protein necessary to supply the precise amounts of protein, carbohydrate, fat, and calcium in the Cambridge Diet. State what the variables in the equation represent.

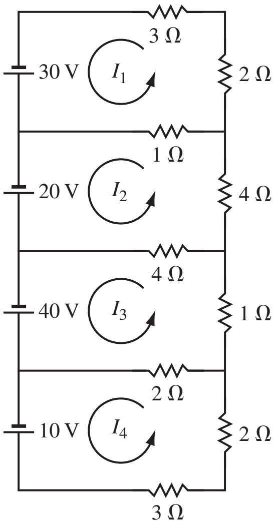
T b. Solve the equation in (a) and discuss your answer.

T In Exercises 5–8, write a matrix equation that determines the loop currents. If MATLAB or another matrix program is available, solve the system for the loop currents.

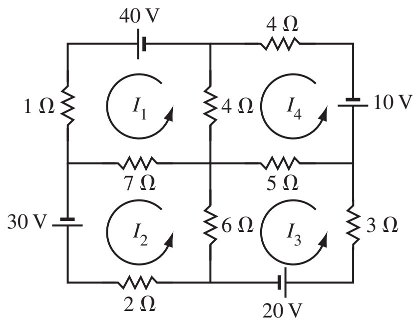
5.



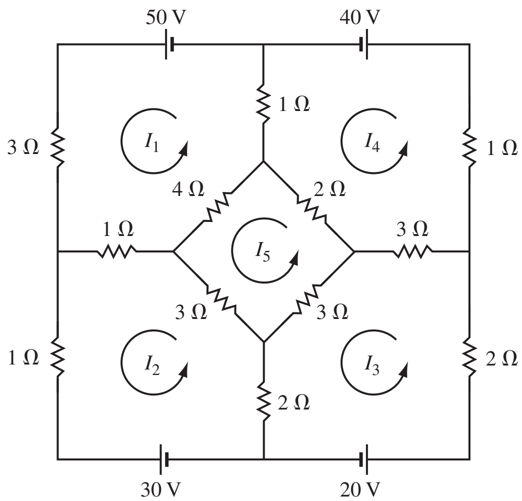
6.



7.



8.



9. In a certain region, about 7% of a city's population moves to the surrounding suburbs each year, and about 5% of the suburban population moves into the city. In 2020, there were 800,000 residents in the city and 500,000 in the suburbs. Set up a difference equation that describes this situation, where $\mathbf{x}_0$ is the initial population in 2020. Then estimate the populations in the city and in the suburbs two years later, in 2022. (Ignore other factors that might influence the population sizes.)

10. In a certain region, about 6% of a city's population moves to the surrounding suburbs each year, and about 4% of the suburban population moves into the city. In 2020, there were 10,000,000 residents in the city and 800,000 in the suburbs. Set up a difference equation that describes this situation, where $\mathbf{x}_0$ is the initial population in 2020. Then estimate the populations in the city and in the suburbs two years later, in 2022.

T 11. College Moving Truck Rental has a fleet of 20, 100, and 200 trucks in Pullman, Spokane, and Seattle, respectively. A truck rented at one location may be returned to any of the three locations. The various fractions of trucks returned to the three locations each month are shown in the matrix below. What will be the approximate distribution of the trucks after three months?

$$\begin{array}{cccl} & \text{Trucks Rented From:} & & \\ \text{Pullman} & \text{Spokane} & \text{Seattle} & \text{Returned To:} \end{array}$$
$$\begin{bmatrix} .30 & .15 & .05 \\ .30 & .70 & .05 \\ .40 & .15 & .90 \end{bmatrix} \begin{array}{l} \text{Airport} \\ \text{East} \\ \text{West} \end{array}$$

T 12. Budget® Rent a Car in Wichita, Kansas, has a fleet of about 500 cars, at three locations. A car rented at one location may be returned to any of the three locations. The various fractions of cars returned to the three locations are shown in the matrix below. Suppose that on Monday there are 295 cars at the airport (or rented from there), 55 cars at the east side office, and 150 cars at the west side office. What will be the approximate distribution of cars on Wednesday?

$$\begin{array}{cccl} & \text{Cars Rented From:} & & \\ \text{Airport} & \text{East} & \text{West} & \text{Returned To:} \end{array}$$
$$\begin{bmatrix} .97 & .05 & .10 \\ .00 & .90 & .05 \\ .03 & .05 & .85 \end{bmatrix} \begin{array}{l} \text{Airport} \\ \text{East} \\ \text{West} \end{array}$$

T 13. Let M and $\mathbf{x}_0$ be as in Example 3.

a. Compute the population vectors $\mathbf{x}_k$ for $k = 1, \ldots, 20$. Discuss what you find.

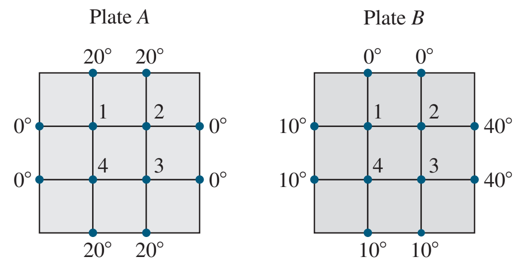
b. Repeat part (a) with an initial population of 350,000 in the city and 650,000 in the suburbs. What do you find?

T 14. Study how changes in boundary temperatures on a steel plate affect the temperatures at interior points on the plate.

a. Begin by estimating the temperatures T_1, T_2, T_3, T_4 at each of the sets of four points on the steel plate shown in the figure. In each case, the value of T_k is approximated by the average of the temperatures at the four closest points. See Exercises 43 and 44 in Section 1.1, where the values (in degrees) turn out to be (20, 27.5, 30, 22.5). How is this list of values related to your results for the points in set (a) and set (b)?

b. Without making any computations, guess the interior temperatures in (a) when the boundary temperatures are all multiplied by 3. Check your guess.

c. Finally, make a general conjecture about the correspondence from the list of eight boundary temperatures to the list of four interior temperatures.



(a) (b)

Solution to Practice Problem

$$A = \begin{bmatrix} 36 & 51 & 13 \\ 52 & 34 & 74 \\ 0 & 7 & 1.1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 33 \\ 45 \\ 3 \end{bmatrix}$$

CHAPTER 1 PROJECTS

Chapter 1 projects are available online.

A. *Interpolating Polynomials*: This project shows how to use a system of linear equations to fit a polynomial through a set of points.

B. *Splines*: This project also shows how to use a system of linear equations to fit a piecewise polynomial curve through a set of points.

C. *Network Flows*: The purpose of this project is to show how systems of linear equations may be used to model flow through a network.

D. *The Art of Linear Transformations*: In this project, it is illustrated how to graph a polygon and then use linear transformations to change its shape and create a design.

E. *Loop Currents*: The purpose of this project is to provide more and larger examples of loop currents.

F. *Diet*: The purpose of this project is to provide examples of vector equations that result from balancing nutrients in a diet.

CHAPTER 1 SUPPLEMENTARY EXERCISES

Mark each statement True or False **(T/F)**. Justify each answer. (If true, cite appropriate facts or theorems. If false, explain why or give a counterexample that shows why the statement is not true in every case.

1. **(T/F)** Every matrix is row equivalent to a unique matrix in echelon form.

2. **(T/F)** Any system of n linear equations in n variables has at most n solutions.

3. **(T/F)** If a system of linear equations has two different solutions, it must have infinitely many solutions.

4. **(T/F)** If a system of linear equations has no free variables, then it has a unique solution.

5. **(T/F)** If an augmented matrix $[\,A \quad \mathbf{b}\,]$ is transformed into $[\,C \quad \mathbf{d}\,]$ by elementary row operations, then the equations $A\mathbf{x} = \mathbf{b}$ and $C\mathbf{x} = \mathbf{d}$ have exactly the same solution sets.

6. **(T/F)** If a system $A\mathbf{x} = \mathbf{b}$ has more than one solution, then so does the system $A\mathbf{x} = \mathbf{0}$.

7. **(T/F)** If A is an $m \times n$ matrix and the equation $A\mathbf{x} = \mathbf{b}$ is consistent for some $\mathbf{b}$, then the columns of A span $\mathbb{R}^m$.

8. **(T/F)** If an augmented matrix $[\,A \quad \mathbf{b}\,]$ can be transformed by elementary row operations into reduced echelon form, then the equation $A\mathbf{x} = \mathbf{b}$ is consistent.

9. **(T/F)** If matrices A and B are row equivalent, they have the same reduced echelon form.

10. **(T/F)** The equation $A\mathbf{x} = \mathbf{0}$ has the trivial solution if and only if there are no free variables.

11. **(T/F)** If A is an $m \times n$ matrix and the equation $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b}$ in $\mathbb{R}^m$, then A has m pivot columns.

12. **(T/F)** If an $m \times n$ matrix A has a pivot position in every row, then the equation $A\mathbf{x} = \mathbf{b}$ has a unique solution for each $\mathbf{b}$ in $\mathbb{R}^m$.

13. **(T/F)** If an $n \times n$ matrix A has n pivot positions, then the reduced echelon form of A is the $n \times n$ identity matrix.

14. **(T/F)** If 3×3 matrices A and B each have three pivot positions, then A can be transformed into B by elementary row operations.

15. **(T/F)** If A is an $m \times n$ matrix, if the equation $A\mathbf{x} = \mathbf{b}$ has at least two different solutions, and if the equation $A\mathbf{x} = \mathbf{c}$ is consistent, then the equation $A\mathbf{x} = \mathbf{c}$ has many solutions.

16. **(T/F)** If A and B are row equivalent $m \times n$ matrices and if the columns of A span $\mathbb{R}^m$, then so do the columns of B.

17. **(T/F)** If none of the vectors in the set $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ in $\mathbb{R}^3$ is a multiple of one of the other vectors, then S is linearly independent.

18. **(T/F)** If $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly independent, then $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are not in $\mathbb{R}^2$.

19. **(T/F)** In some cases, it is possible for four vectors to span $\mathbb{R}^5$.

20. **(T/F)** If $\mathbf{u}$ and $\mathbf{v}$ are in $\mathbb{R}^m$, then $-\mathbf{u}$ is in Span$\{\mathbf{u}, \mathbf{v}\}$.

21. **(T/F)** If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are nonzero vectors in $\mathbb{R}^2$, then $\mathbf{w}$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$.

22. **(T/F)** If $\mathbf{w}$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, then $\mathbf{u}$ is a linear combination of $\mathbf{v}$ and $\mathbf{w}$.

23. **(T/F)** Suppose that $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$ are in $\mathbb{R}^5$, $\mathbf{v}_2$ is not a multiple of $\mathbf{v}_1$, and $\mathbf{v}_3$ is not a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. Then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent.

24. **(T/F)** A linear transformation is a function.

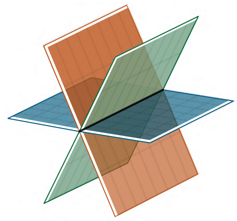
25. **(T/F)** If A is a 6×5 matrix, the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ cannot map $\mathbb{R}^5$ onto $\mathbb{R}^6$.

26. Let a and b represent real numbers. Describe the possible solution sets of the (linear) equation $ax = b$. [*Hint:* The number of solutions depends upon a and b.]

27. The solutions (x, y, z) of a single linear equation

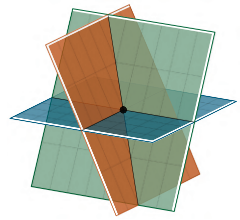
$$ax + by + cz = d$$

form a plane in $\mathbb{R}^3$ when a, b, and c are not all zero. Construct sets of three linear equations whose graphs (a) intersect in a single line, (b) intersect in a single point, and (c) have no points in common. Typical graphs are illustrated in the figure.



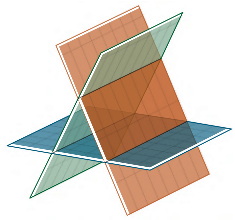
Three planes intersecting in a line

(a)



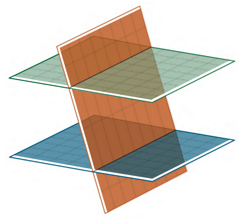
Three planes intersecting in a point

(b)



Three planes with no intersection

(c)



Three planes with no intersection

(c')

28. Suppose the coefficient matrix of a linear system of three equations in three variables has a pivot position in each column. Explain why the system has a unique solution.

29. Determine h and k such that the solution set of the system (i) is empty, (ii) contains a unique solution, and (iii) contains infinitely many solutions.

a. $\begin{aligned} x_1 + 3x_2 &= k \\ 4x_1 + hx_2 &= 8 \end{aligned}$ b. $\begin{aligned} -2x_1 + hx_2 &= 1 \\ 6x_1 + kx_2 &= -2 \end{aligned}$

30. Consider the problem of determining whether the following system of equations is consistent:

$$\begin{aligned} 4x_1 - 2x_2 + 7x_3 &= -5 \\ 8x_1 - 3x_2 + 10x_3 &= -3 \end{aligned}$$

a. Define appropriate vectors, and restate the problem in terms of linear combinations. Then solve that problem.

b. Define an appropriate matrix, and restate the problem using the phrase "columns of A."

c. Define an appropriate linear transformation T using the matrix in (b), and restate the problem in terms of T.

31. Consider the problem of determining whether the following system of equations is consistent for all b_1, b_2, b_3:

$$\begin{aligned} 2x_1 - 4x_2 - 2x_3 &= b_1 \\ -5x_1 + x_2 + x_3 &= b_2 \\ 7x_1 - 5x_2 - 3x_3 &= b_3 \end{aligned}$$

a. Define appropriate vectors, and restate the problem in terms of Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. Then solve that problem.

b. Define an appropriate matrix, and restate the problem using the phrase "columns of A."

c. Define an appropriate linear transformation T using the matrix in (b), and restate the problem in terms of T.

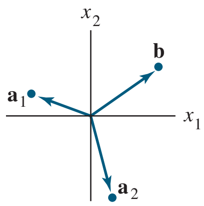
32. Describe the possible echelon forms of the matrix A. Use the notation of Example 1 in Section 1.2.

a. A is a 2×3 matrix whose columns span $\mathbb{R}^2$.

b. A is a 3×3 matrix whose columns span $\mathbb{R}^3$.

33. Write the vector $\begin{bmatrix} 5 \\ 6 \end{bmatrix}$ as the sum of two vectors, one on the line $\{(x, y) : y = 2x\}$ and one on the line $\{(x, y) : y = x/2\}$.

34. Let $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{b}$ be the vectors in $\mathbb{R}^2$ shown in the figure, and let $A = [\mathbf{a}_1 \quad \mathbf{a}_2]$. Does the equation $A\mathbf{x} = \mathbf{b}$ have a solution? If so, is the solution unique? Explain.



35. Construct a 2×3 matrix A, not in echelon form, such that the solution of $A\mathbf{x} = \mathbf{0}$ is a line in $\mathbb{R}^3$.

36. Construct a 2×3 matrix A, not in echelon form, such that the solution of $A\mathbf{x} = \mathbf{0}$ is a plane in $\mathbb{R}^3$.

37. Write the *reduced* echelon form of a 3×3 matrix A such that the first two columns of A are pivot columns and $A\begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

38. Determine the value(s) of a such that $\left\{\begin{bmatrix} 1 \\ a \end{bmatrix}, \begin{bmatrix} a+2 \\ a+6 \end{bmatrix}\right\}$ is linearly independent.

39. In (a) and (b), suppose the vectors are linearly independent. What can you say about the numbers $a, \dots, f$? Justify your answers. [*Hint:* Use a theorem for (b).]

a. $\begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} b \\ c \\ 0 \end{bmatrix}, \begin{bmatrix} d \\ e \\ f \end{bmatrix}$ b. $\begin{bmatrix} a \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} b \\ c \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} d \\ e \\ f \\ 1 \end{bmatrix}$

40. Use Theorem 7 in Section 1.7 to explain why the columns of the matrix A are linearly independent.

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 5 & 0 & 0 \\ 3 & 6 & 8 & 0 \\ 4 & 7 & 9 & 10 \end{bmatrix}$$

41. Explain why a set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ in $\mathbb{R}^5$ must be linearly independent when $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent and $\mathbf{v}_4$ is *not* in Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

42. Suppose $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a linearly independent set in $\mathbb{R}^n$. Show that $\{\mathbf{v}_1 + \mathbf{v}_2, \mathbf{v}_1 - \mathbf{v}_2\}$ is also linearly independent.

43. Suppose $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are distinct points on one line in $\mathbb{R}^3$. The line need not pass through the origin. Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent.

44. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and suppose $T(\mathbf{u}) = \mathbf{v}$. Show that $T(-\mathbf{u}) = -\mathbf{v}$.

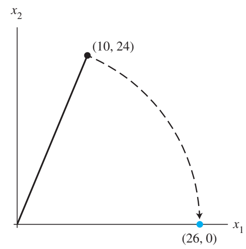
45. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation that reflects each vector through the plane $x_2 = 0$. That is, $T(x_1, x_2, x_3) = (x_1, -x_2, x_3)$. Find the standard matrix of T.

46. Let A be a 3×3 matrix with the property that the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^3$ onto $\mathbb{R}^3$. Explain why the transformation must be one-to-one.

47. A *Givens rotation* is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$ used in computer programs to create a zero entry in a vector (usually a column of a matrix). The standard matrix of a Givens rotation in $\mathbb{R}^2$ has the form

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}, \qquad a^2 + b^2 = 1$$

Find a and b such that $\begin{bmatrix} 10 \\ 24 \end{bmatrix}$ is rotated into $\begin{bmatrix} 26 \\ 0 \end{bmatrix}$.



A Givens rotation in $\mathbb{R}^2$.

48. The following equation describes a Givens rotation in $\mathbb{R}^3$. Find a and b.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & a & -b \\ 0 & b & a \end{bmatrix}\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 0 \end{bmatrix}, \qquad a^2 + b^2 = 1$$

49. A large apartment building is to be built using modular construction techniques. The arrangement of apartments on any particular floor is to be chosen from one of three basic floor plans. Plan A has 18 apartments on one floor, including 3 three-bedroom units, 7 two-bedroom units, and 8 one-bedroom units. Each floor of plan B includes 4 three-bedroom units, 4 two-bedroom units, and 8 one-bedroom units. Each floor of plan C includes 5 three-bedroom units, 3 two-bedroom units, and 9 one-bedroom units. Suppose the building contains a total of x_1 floors of plan A, x_2 floors of plan B, and x_3 floors of plan C.

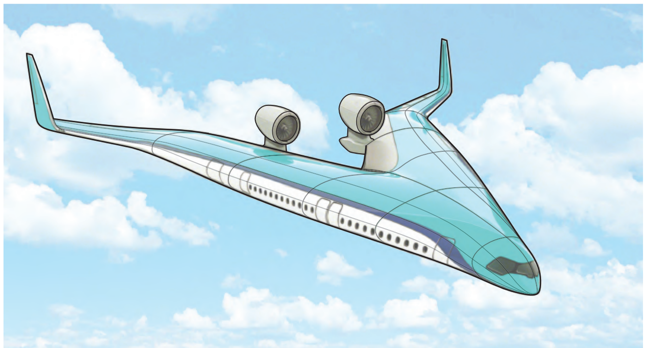
a. What interpretation can be given to the vector $x_1\begin{bmatrix}3\\7\\8\end{bmatrix}$?

b. Write a formal linear combination of vectors that expresses the total numbers of three-, two-, and one-bedroom apartments contained in the building.

T c. Is it possible to design the building with exactly 66 three-bedroom units, 74 two-bedroom units, and 136 one-bedroom units? If so, is there more than one way to do it? Explain your answer.

- *Matrix factorizations:* Even when written with partitioned matrices, the system of equations is complicated. To further simplify the computations, the CFD software at Boeing uses what is called an LU factorization of the coefficient matrix. Section 2.5 discusses LU and other useful matrix factorizations. Further details about factorizations appear at several points later in the text.

To analyze a solution of an airflow system, engineers want to visualize the airflow over the surface of the plane. They use computer graphics, and linear algebra provides the engine for the graphics. The wire-frame model of the plane's surface is stored as data in many matrices. Once the image has been rendered on a computer screen, engineers can change its scale, zoom in or out of small regions, and rotate the image to see parts that may be hidden from view. Each of these operations is accomplished by appropriate matrix multiplications. Section 2.7 explains the basic ideas.



TU-Delft and Air France-KLM are investigating a flying V aircraft design because of its potential for significantly better fuel economy.

Our ability to analyze and solve equations will be greatly enhanced when we can perform algebraic operations with matrices. Furthermore, the definitions and theorems in this chapter provide some basic tools for handling the many applications of linear algebra that involve two or more matrices. For $n \times n$ matrices, the Invertible Matrix Theorem in Section 2.3 ties together most of the concepts treated earlier in the text. Sections 2.4 and 2.5 examine partitioned matrices and matrix factorizations, which appear in most modern uses of linear algebra. Sections 2.6 and 2.7 describe two interesting applications of matrix algebra: to economics and to computer graphics. Sections 2.8 and 2.9 provide readers enough information about subspaces to move directly into Chapters 5, 6, and 7, without covering Chapter 4. You may want to omit these two sections if you plan to cover Chapter 4 before moving to Chapter 5.

2.1 Matrix Operations

If A is an $m \times n$ matrix—that is, a matrix with m rows and n columns—then the scalar entry in the ith row and jth column of A is denoted by a_{ij} and is called the (i, j)-entry of A. See Figure 1. For instance, the $(3, 2)$-entry is the number a_{32} in the third row, second column. Each column of A is a list of m real numbers, which identifies a vector in $\mathbb{R}^m$. Often, these columns are denoted by $\mathbf{a}_1, \ldots, \mathbf{a}_n$, and the matrix A is written as

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$$

Observe that the number a_{ij} is the ith entry (from the top) of the jth column vector $\mathbf{a}_j$.

The **diagonal entries** in an $m \times n$ matrix $A = [\, a_{ij} \,]$ are $a_{11}, a_{22}, a_{33}, \ldots,$ and they form the **main diagonal** of A. A **diagonal matrix** is a square $n \times n$ matrix whose nondiagonal entries are zero. An example is the $n \times n$ identity matrix, I_n. An $m \times n$ matrix whose entries are all zero is a **zero matrix** and is written as 0. The size of a zero matrix is usually clear from the context.

$$\begin{array}{ccccccc}
 & & & \text{Column} & & & \\
 & & & j & & & \\
\text{Row } i & \left[\begin{matrix} a_{11} \\ \vdots \\ a_{i1} \\ \vdots \\ a_{m1} \end{matrix}\right. & \begin{matrix} \cdots \\ \\ \cdots \\ \\ \cdots \end{matrix} & \begin{matrix} a_{1j} \\ \vdots \\ a_{ij} \\ \vdots \\ a_{mj} \end{matrix} & \begin{matrix} \cdots \\ \\ \cdots \\ \\ \cdots \end{matrix} & \left.\begin{matrix} a_{1n} \\ \vdots \\ a_{in} \\ \vdots \\ a_{mn} \end{matrix}\right] & = A \\
 & \uparrow & & \uparrow & & \uparrow & \\
 & \mathbf{a}_1 & & \mathbf{a}_j & & \mathbf{a}_n &
\end{array}$$

FIGURE 1 Matrix notation.

Sums and Scalar Multiples

The arithmetic for vectors described earlier has a natural extension to matrices. We say that two matrices are **equal** if they have the same size (i.e., the same number of rows and the same number of columns) and if their corresponding columns are equal, which amounts to saying that their corresponding entries are equal. If A and B are $m \times n$ matrices, then the **sum** $A + B$ is the $m \times n$ matrix whose columns are the sums of the corresponding columns in A and B. Since vector addition of the columns is done entrywise, each entry in $A + B$ is the sum of the corresponding entries in A and B. The sum $A + B$ is defined only when A and B are the same size.

EXAMPLE 1 Let

$$A = \begin{bmatrix} 4 & 0 & 5 \\ -1 & 3 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 5 & 7 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & -3 \\ 0 & 1 \end{bmatrix}$$

Then

$$A + B = \begin{bmatrix} 5 & 1 & 6 \\ 2 & 8 & 9 \end{bmatrix}$$

but $A + C$ is not defined because A and C have different sizes. ■

If r is a scalar and A is a matrix, then the **scalar multiple** rA is the matrix whose columns are r times the corresponding columns in A. As with vectors, $-A$ stands for $(-1)A$, and $A - B$ is the same as $A + (-1)B$.

EXAMPLE 2 If A and B are the matrices in Example 1, then

$$2B = 2\begin{bmatrix} 1 & 1 & 1 \\ 3 & 5 & 7 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 2 \\ 6 & 10 & 14 \end{bmatrix}$$

$$A - 2B = \begin{bmatrix} 4 & 0 & 5 \\ -1 & 3 & 2 \end{bmatrix} - \begin{bmatrix} 2 & 2 & 2 \\ 6 & 10 & 14 \end{bmatrix} = \begin{bmatrix} 2 & -2 & 3 \\ -7 & -7 & -12 \end{bmatrix}$$ ■

It was unnecessary in Example 2 to compute $A - 2B$ as $A + (-1)2B$ because the usual rules of algebra apply to sums and scalar multiples of matrices, as the following theorem shows.

THEOREM 1

Let A, B, and C be matrices of the same size, and let r and s be scalars.

a. $A + B = B + A$
b. $(A + B) + C = A + (B + C)$
c. $A + 0 = A$
d. $r(A + B) = rA + rB$
e. $(r + s)A = rA + sA$
f. $r(sA) = (rs)A$

Each equality in Theorem 1 is verified by showing that the matrix on the left side has the same size as the matrix on the right and that corresponding columns are equal. Size is no problem because A, B, and C are equal in size. The equality of columns follows immediately from analogous properties of vectors. For instance, if the jth columns of A, B, and C are $\mathbf{a}_j$, $\mathbf{b}_j$, and $\mathbf{c}_j$, respectively, then the jth columns of $(A+B)+C$ and $A+(B+C)$ are

$$(\mathbf{a}_j + \mathbf{b}_j) + \mathbf{c}_j \quad \text{and} \quad \mathbf{a}_j + (\mathbf{b}_j + \mathbf{c}_j)$$

respectively. Since these two vector sums are equal for each j, property (b) is verified.

Because of the associative property of addition, we can simply write $A+B+C$ for the sum, which can be computed either as $(A+B)+C$ or as $A+(B+C)$. The same applies to sums of four or more matrices.

Matrix Multiplication

When a matrix B multiplies a vector $\mathbf{x}$, it transforms $\mathbf{x}$ into the vector $B\mathbf{x}$. If this vector is then multiplied in turn by a matrix A, the resulting vector is $A(B\mathbf{x})$. See Figure 2.

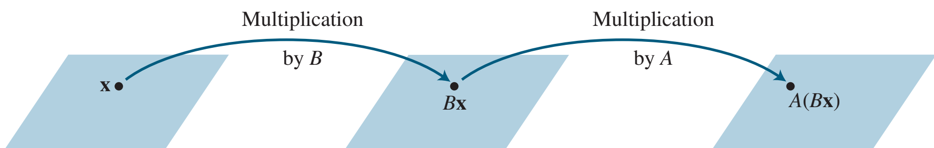


FIGURE 2 Multiplication by B and then A.

Thus $A(B\mathbf{x})$ is produced from $\mathbf{x}$ by a *composition* of mappings—the linear transformations studied in Section 1.8. Our goal is to represent this composite mapping as multiplication by a single matrix, denoted by AB, so that

$$A(B\mathbf{x}) = (AB)\mathbf{x} \tag{1}$$

See Figure 3.

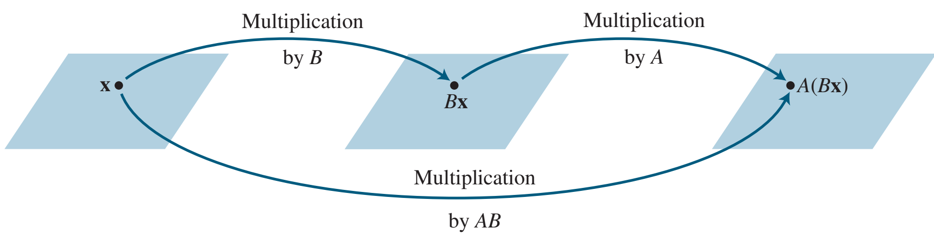


FIGURE 3 Multiplication by AB.

If A is $m \times n$, B is $n \times p$, and $\mathbf{x}$ is in $\mathbb{R}^p$, denote the columns of B by $\mathbf{b}_1, \dots, \mathbf{b}_p$ and the entries in $\mathbf{x}$ by $x_1, \dots, x_p$. Then

$$B\mathbf{x} = x_1\mathbf{b}_1 + \cdots + x_p\mathbf{b}_p$$

By the linearity of multiplication by A,

$$\begin{aligned} A(B\mathbf{x}) &= A(x_1\mathbf{b}_1) + \cdots + A(x_p\mathbf{b}_p) \\ &= x_1 A\mathbf{b}_1 + \cdots + x_p A\mathbf{b}_p \end{aligned}$$

The vector $A(B\mathbf{x})$ is a linear combination of the vectors $A\mathbf{b}_1, \dots, A\mathbf{b}_p$, using the entries in $\mathbf{x}$ as weights. In matrix notation, this linear combination is written as

$$A(B\mathbf{x}) = [\, A\mathbf{b}_1 \quad A\mathbf{b}_2 \quad \cdots \quad A\mathbf{b}_p \,]\mathbf{x}$$

Thus multiplication by $[\, A\mathbf{b}_1 \quad A\mathbf{b}_2 \quad \cdots \quad A\mathbf{b}_p \,]$ transforms $\mathbf{x}$ into $A(B\mathbf{x})$. We have found the matrix we sought!

DEFINITION

If A is an $m \times n$ matrix, and if B is an $n \times p$ matrix with columns $\mathbf{b}_1, \dots, \mathbf{b}_p$, then the product AB is the $m \times p$ matrix whose columns are $A\mathbf{b}_1, \dots, A\mathbf{b}_p$. That is,

$$AB = A\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_p \end{bmatrix} = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_p \end{bmatrix}$$

This definition makes equation (1) true for all $\mathbf{x}$ in $\mathbb{R}^p$. Equation (1) proves that the composite mapping in Figure 3 is a linear transformation and that its standard matrix is AB. *Multiplication of matrices corresponds to composition of linear transformations.*

EXAMPLE 3 Compute AB, where $A = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{bmatrix}$.

SOLUTION Write $B = [\, \mathbf{b}_1 \quad \mathbf{b}_2 \quad \mathbf{b}_3 \,]$, and compute:

$$A\mathbf{b}_1 = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 4 \\ 1 \end{bmatrix}, \quad A\mathbf{b}_2 = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 3 \\ -2 \end{bmatrix}, \quad A\mathbf{b}_3 = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 6 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} 11 \\ -1 \end{bmatrix} \qquad\qquad = \begin{bmatrix} 0 \\ 13 \end{bmatrix} \qquad\qquad = \begin{bmatrix} 21 \\ -9 \end{bmatrix}$$

Then

$$AB = A[\, \mathbf{b}_1 \quad \mathbf{b}_2 \quad \mathbf{b}_3 \,] = \begin{array}{c} \begin{bmatrix} 11 & 0 & 21 \\ -1 & 13 & -9 \end{bmatrix} \\ \begin{matrix} \uparrow & \uparrow & \uparrow \\ A\mathbf{b}_1 & A\mathbf{b}_2 & A\mathbf{b}_3 \end{matrix} \end{array}$$

■

Notice that since the first column of AB is $A\mathbf{b}_1$, this column is a linear combination of the columns of A using the entries in $\mathbf{b}_1$ as weights. A similar statement is true for each column of AB.

Each column of AB is a linear combination of the columns of A using weights from the corresponding column of B.

Obviously, the number of columns of A must match the number of rows in B in order for a linear combination such as $A\mathbf{b}_1$ to be defined. Also, the definition of AB shows that *AB has the same number of rows as A and the same number of columns as B.*

EXAMPLE 4 If A is a 3×5 matrix and B is a 5×2 matrix, what are the sizes of AB and BA, if they are defined?

SOLUTION Since A has 5 columns and B has 5 rows, the product AB is defined and is a 3×2 matrix:

$$\underset{3\times 5}{\overset{A}{\begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix}}}\ \underset{5\times 2}{\overset{B}{\begin{bmatrix} * & * \\ * & * \\ * & * \\ * & * \\ * & * \end{bmatrix}}} = \underset{3\times 2}{\overset{AB}{\begin{bmatrix} * & * \\ * & * \\ * & * \end{bmatrix}}}$$

Match

Size of AB

The product BA is *not* defined because the 2 columns of B do not match the 3 rows of A. ■

The definition of AB is important for theoretical work and applications, but the following rule provides a more efficient method for calculating the individual entries in AB when working small problems by hand.

ROW–COLUMN RULE FOR COMPUTING *AB*

If the product AB is defined, then the entry in row i and column j of AB is the sum of the products of corresponding entries from row i of A and column j of B. If $(AB)_{ij}$ denotes the (i, j)-entry in AB, and if A is an $m \times n$ matrix, then

$$(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

To verify this rule, let $B = [\,\mathbf{b}_1 \ \cdots \ \mathbf{b}_p\,]$. Column j of AB is $A\mathbf{b}_j$, and we can compute $A\mathbf{b}_j$ by the row–vector rule for computing $A\mathbf{x}$ from Section 1.4. The ith entry in $A\mathbf{b}_j$ is the sum of the products of corresponding entries from row i of A and the vector $\mathbf{b}_j$, which is precisely the computation described in the rule for computing the (i, j)-entry of AB.

EXAMPLE 5 Use the row–column rule to compute two of the entries in AB for the matrices in Example 3. An inspection of the numbers involved will make it clear how the two methods for calculating AB produce the same matrix.

SOLUTION To find the entry in row 1 and column 3 of AB, consider row 1 of A and column 3 of B. Multiply corresponding entries and add the results, as shown below:

$$AB = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{bmatrix} = \begin{bmatrix} \square & \square & 2(6)+3(3) \\ \square & \square & \square \end{bmatrix} = \begin{bmatrix} \square & \square & 21 \\ \square & \square & \square \end{bmatrix}$$

For the entry in row 2 and column 2 of AB, use row 2 of A and column 2 of B:

$$\begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{bmatrix} = \begin{bmatrix} \square & \square & 21 \\ \square & 1(3)+-5(-2) & \square \end{bmatrix} = \begin{bmatrix} \square & \square & 21 \\ \square & 13 & \square \end{bmatrix}$$

■

EXAMPLE 6 Find the entries in the second row of AB, where

$$A = \begin{bmatrix} 2 & -5 & 0 \\ -1 & 3 & -4 \\ 6 & -8 & -7 \\ -3 & 0 & 9 \end{bmatrix}, \qquad B = \begin{bmatrix} 4 & -6 \\ 7 & 1 \\ 3 & 2 \end{bmatrix}$$

SOLUTION By the row–column rule, the entries of the second row of AB come from row 2 of A (and the columns of B):

$$\begin{array}{r} \\ \rightarrow \end{array}\begin{bmatrix} 2 & -5 & 0 \\ -1 & 3 & -4 \\ 6 & -8 & -7 \\ -3 & 0 & 9 \end{bmatrix} \begin{array}{c} \begin{array}{cc} \downarrow & \downarrow \end{array} \\ \begin{bmatrix} 4 & -6 \\ 7 & 1 \\ 3 & 2 \end{bmatrix} \end{array}$$

$$= \begin{bmatrix} \square & \square \\ -4+21-12 & 6+3-8 \\ \square & \square \\ \square & \square \end{bmatrix} = \begin{bmatrix} \square & \square \\ 5 & 1 \\ \square & \square \\ \square & \square \end{bmatrix}$$

■

Notice that since Example 6 requested only the second row of AB, we could have written just the second row of A to the left of B and computed

$$\begin{bmatrix} -1 & 3 & -4 \end{bmatrix} \begin{bmatrix} 4 & -6 \\ 7 & 1 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 1 \end{bmatrix}$$

This observation about rows of AB is true in general and follows from the row–column rule. Let $\text{row}_i(A)$ denote the ith row of a matrix A. Then

$$\text{row}_i(AB) = \text{row}_i(A) \cdot B \tag{2}$$

Properties of Matrix Multiplication

The following theorem lists the standard properties of matrix multiplication. Recall that I_m represents the $m \times m$ identity matrix and $I_m\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^m$.

THEOREM 2

Let A be an $m \times n$ matrix, and let B and C have sizes for which the indicated sums and products are defined.

a. $A(BC) = (AB)C$ (associative law of multiplication)

b. $A(B + C) = AB + AC$ (left distributive law)

c. $(B + C)A = BA + CA$ (right distributive law)

d. $r(AB) = (rA)B = A(rB)$ for any scalar r

e. $I_mA = A = AI_n$ (identity for matrix multiplication)

PROOF Properties (b)–(e) are considered in the exercises. Property (a) follows from the fact that matrix multiplication corresponds to composition of linear transformations (which are functions), and it is known (or easy to check) that the composition of functions is associative. Here is another proof of (a) that rests on the "column definition" of the product of two matrices. Let

$$C = [\,\mathbf{c}_1 \quad \cdots \quad \mathbf{c}_p\,]$$

By the definition of matrix multiplication,

$$\begin{aligned} BC &= [\,B\mathbf{c}_1 \quad \cdots \quad B\mathbf{c}_p\,] \\ A(BC) &= [\,A(B\mathbf{c}_1) \quad \cdots \quad A(B\mathbf{c}_p)\,] \end{aligned}$$

Recall from equation (1) that the definition of AB makes $A(B\mathbf{x}) = (AB)\mathbf{x}$ for all $\mathbf{x}$, so

$$A(BC) = [\,(AB)\mathbf{c}_1 \quad \cdots \quad (AB)\mathbf{c}_p\,] = (AB)C$$ ■

The associative and distributive laws in Theorems 1 and 2 say essentially that pairs of parentheses in matrix expressions can be inserted and deleted in the same way as in the algebra of real numbers. In particular, we can write ABC for the product, which can be computed either as $A(BC)$ or as $(AB)C$.[1] Similarly, a product $ABCD$ of four matrices can be computed as $A(BCD)$ or $(ABC)D$ or $A(BC)D$, and so on. It does not matter how we group the matrices when computing the product, so long as the left-to-right order of the matrices is preserved.

The left-to-right order in products is critical because AB and BA are usually not the same. This is not surprising, because the columns of AB are linear combinations of the columns of A, whereas the columns of BA are constructed from the columns of B. The position of the factors in the product AB is emphasized by saying that A is *right-multiplied* by B or that B is *left-multiplied* by A. If $AB = BA$, we say that A and B **commute** with one another.

EXAMPLE 7 Let $A = \begin{bmatrix} 5 & 1 \\ 3 & -2 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix}$. Show that these matrices do not commute. That is, verify that $AB \neq BA$.

SOLUTION

$$\begin{aligned} AB &= \begin{bmatrix} 5 & 1 \\ 3 & -2 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 14 & 3 \\ -2 & -6 \end{bmatrix} \\ BA &= \begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix}\begin{bmatrix} 5 & 1 \\ 3 & -2 \end{bmatrix} = \begin{bmatrix} 10 & 2 \\ 29 & -2 \end{bmatrix} \end{aligned}$$ ■

Example 7 illustrates the first of the following list of important differences between matrix algebra and the ordinary algebra of real numbers. See Exercises 9–12 for examples of these situations.

***Warnings*:**

1. In general, $AB \neq BA$.
2. The cancellation laws do *not* hold for matrix multiplication. That is, if $AB = AC$, then it is *not* true in general that $B = C$. (See Exercise 10.)
3. If a product AB is the zero matrix, you *cannot* conclude in general that either $A = 0$ or $B = 0$. (See Exercise 12.)

---

[1] When B is square and C has fewer columns than A has rows, it is more efficient to compute $A(BC)$ than $(AB)C$.

Powers of a Matrix

If A is an $n \times n$ matrix and if k is a positive integer, then A^k denotes the product of k copies of A:

$$A^k = \underbrace{A \cdots A}_{k}$$

If A is nonzero and if $\mathbf{x}$ is in $\mathbb{R}^n$, then $A^k\mathbf{x}$ is the result of left-multiplying $\mathbf{x}$ by A repeatedly k times. If $k = 0$, then $A^0\mathbf{x}$ should be $\mathbf{x}$ itself. Thus A^0 is interpreted as the identity matrix. Matrix powers are useful in both theory and applications (Sections 2.6, 5.9, and later in the text).

The Transpose of a Matrix

Given an $m \times n$ matrix A, the **transpose** of A is the $n \times m$ matrix, denoted by A^T, whose columns are formed from the corresponding rows of A.

EXAMPLE 8 Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad B = \begin{bmatrix} -5 & 2 \\ 1 & -3 \\ 0 & 4 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -3 & 5 & -2 & 7 \end{bmatrix}$$

Then

$$A^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}, \quad B^T = \begin{bmatrix} -5 & 1 & 0 \\ 2 & -3 & 4 \end{bmatrix}, \quad C^T = \begin{bmatrix} 1 & -3 \\ 1 & 5 \\ 1 & -2 \\ 1 & 7 \end{bmatrix}$$

■

THEOREM 3

Let A and B denote matrices whose sizes are appropriate for the following sums and products.

a. $(A^T)^T = A$

b. $(A + B)^T = A^T + B^T$

c. For any scalar r, $(rA)^T = rA^T$

d. $(AB)^T = B^T A^T$

Proofs of (a)–(c) are straightforward and are omitted. For (d), see Exercise 41. Usually, $(AB)^T$ is not equal to $A^T B^T$, even when A and B have sizes such that the product $A^T B^T$ is defined.

The generalization of Theorem 3(d) to products of more than two factors can be stated in words as follows:

The transpose of a product of matrices equals the product of their transposes in the *reverse* order.

The exercises contain numerical examples that illustrate properties of transposes.

Artificial intelligence (AI) involves having a computer learn to recognize important information about anything that can be presented in a digitized format. One important area of AI is identifying whether the object in a picture matches a chosen object such as a number, fingerprint, or face.

In the next example, matrix transposition and matrix multiplication are used to tell whether or not a 2×2 block of colored squares matches the chosen checkerboard pattern in Figure 4.

EXAMPLE 9 In order to feed a 2×2 colored block into the computer, it first gets converted into a 4×1 vector by assigning a 1 to each block that is blue and a 0 to each block that is white. Then, the computer converts the block of numbers into a vector by placing the numbers in each column below the numbers in the column to its left.

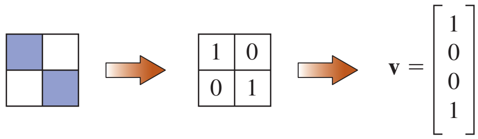


FIGURE 4

Let $M = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}$.

Notice that $\mathbf{v}^T M \mathbf{v} = \begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = 0$,

and $\mathbf{w}^T M \mathbf{w} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = 0$, where $\mathbf{w}$ is the vector generated by a 2×2 block of all white squares. It can be verified that for any other vector $\mathbf{x}$ generated from a 2×2 block of white and blue squares, if $\mathbf{x}$ is not $\mathbf{v}$ or $\mathbf{w}$, then the product $\mathbf{x}^T M \mathbf{x}$ is nonzero. Thus, if a computer checks the value of $\mathbf{x}^T M \mathbf{x}$ and finds it is nonzero, the computer knows that the pattern corresponding to $\mathbf{x}$ is not the checkerboard with a blue square in the top left corner.

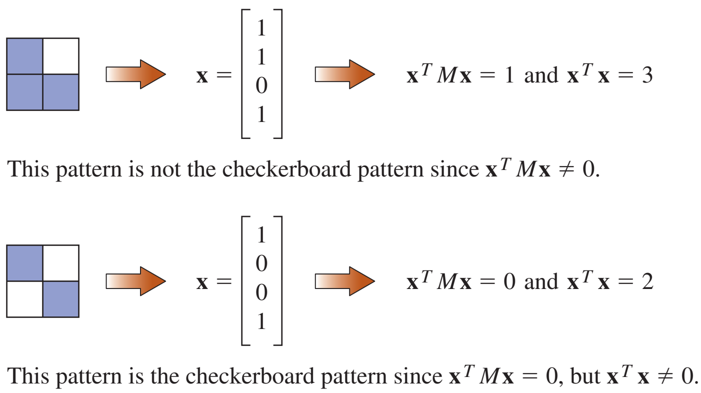


FIGURE 5

However, if the computer finds that $\mathbf{x}^T M\mathbf{x} = 0$, then $\mathbf{x}$ could be either $\mathbf{v}$ or $\mathbf{w}$. To distinguish between the two, the computer can calculate the product $\mathbf{x}^T\mathbf{x}$, for $\mathbf{x}^T\mathbf{x}$ is zero if and only if $\mathbf{x}$ is $\mathbf{w}$.[2] Thus, to conclude that $\mathbf{x}$ is equal to $\mathbf{v}$, the computer must have $\mathbf{x}^T M\mathbf{x} = 0$ and $\mathbf{x}^T\mathbf{x} \neq 0$. ■

Example 5 of Section 6.3 illustrates one way to choose a matrix M so that matrix multiplication and transposition can be used to identify a particular pattern of colored squares.

Another important aspect of AI starts even before the data is fed to the machine. In Section 1.9, it is illustrated how matrix multiplication can be used to move vectors around in space. In the next example, matrix multiplication is used to *scrub* data and prepare it for processing.

EXAMPLE 10 The dates of ground crew accidents for January and February of 2020 are listed in the columns of matrix T for Toronto Pearson Airport and matrix C for Chicago O'Hare Airport:

$$\text{Toronto: } T = \begin{bmatrix} 1 & 12 & 14 & 15 & 21 & 22 & 23 & 1 & 2 & 3 & 12 & 15 & 17 & 19 & 26 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \end{bmatrix}$$

$$\text{Chicago: } C = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 \\ 1 & 11 & 22 & 23 & 24 & 1 & 2 & 5 & 20 & 21 \end{bmatrix}$$

Clearly the data is listed differently in the two matrices. Canada and the United states have different traditions for whether the month or day comes first when writing a date. For matrix T, the day is listed in the first row and the month is listed in the second row. For matrix C, the month is listed in the first row and the day is listed in the second row. In order to use this data, the first and second rows need to be swapped in one of the matrices. Reviewing the effects of matrix multiplication in Table 1 of Section 1.9, notice that the matrix $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ switches the x_1 and x_2 coordinates of any vector $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ it is applied to and indeed

$$AT = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 1 & 12 & 14 & 15 & 21 & 22 & 23 & 1 & 2 & 3 & 12 & 15 & 17 & 19 & 26 \end{bmatrix}$$

has the data listed in the same order as it is listed in matrix C. The matrices AT and C can now be fed into the same machine. ■

In Exercises 51 and 52 you will be asked to *scrub* further data for this project.[3]

---

[2] To see why $\mathbf{x}^T\mathbf{x}$ is zero if and only if $\mathbf{x}$ is $\mathbf{w}$, let $\mathbf{x}^T = [x_1\ x_2\ x_3\ x_4]$. Then $\mathbf{x}^T\mathbf{x} = x_1^2 + x_2^2 + x_3^2 + x_4^2$ and this sum is zero if and only if the coordinates of $\mathbf{x}$ are all zero. That is, if and only if $\mathbf{x} = \mathbf{w}$.

[3] Although the data in this example and the corresponding exercises are fictitious, Data Analytics students at Washington State University identified scrubbing the data they received as an important first step in their actual analysis of ground crew accidents at three major airports in the United States.

Numerical Notes

1. The fastest way to obtain AB on a computer depends on the way in which the computer stores matrices in its memory. The standard high-performance algorithms, such as in LAPACK, calculate AB by columns, as in our definition of the product. (A version of LAPACK written in C++ calculates AB by rows.)
2. The definition of AB lends itself well to parallel processing on a computer. The columns of B are assigned individually or in groups to different processors, which independently and hence simultaneously compute the corresponding columns of AB.

Practice Problems

1. Since vectors in $\mathbb{R}^n$ may be regarded as $n \times 1$ matrices, the properties of transposes in Theorem 3 apply to vectors, too. Let

$$A = \begin{bmatrix} 1 & -3 \\ -2 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

Compute $(A\mathbf{x})^T$, $\mathbf{x}^T A^T$, $\mathbf{x}\mathbf{x}^T$, and $\mathbf{x}^T\mathbf{x}$. Is $A^T\mathbf{x}^T$ defined?

2. Let A be a 4×4 matrix and let $\mathbf{x}$ be a vector in $\mathbb{R}^4$. What is the fastest way to compute $A^2\mathbf{x}$? Count the multiplications.
3. Suppose A is an $m \times n$ matrix, all of whose rows are identical. Suppose B is an $n \times p$ matrix, all of whose columns are identical. What can be said about the entries in AB?

2.1 Exercises

In Exercises 1 and 2, compute each matrix sum or product if it is defined. If an expression is undefined, explain why. Let

$$A = \begin{bmatrix} 2 & 0 & -1 \\ 4 & -3 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 7 & -5 & 1 \\ 1 & -4 & -3 \end{bmatrix},$$
$$C = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 3 & 5 \\ -1 & 4 \end{bmatrix}, \quad E = \begin{bmatrix} -5 \\ 3 \end{bmatrix}$$

1. $-2A$, $B - 2A$, AC, CD
2. $A + 2B$, $3C - E$, CB, EB

In the rest of this exercise set and in those to follow, you should assume that each matrix expression is defined. That is, the sizes of the matrices (and vectors) involved "match" appropriately.

3. Let $A = \begin{bmatrix} 4 & -1 \\ 5 & -2 \end{bmatrix}$. Compute $3I_2 - A$ and $(3I_2)A$.
4. Compute $A - 5I_3$ and $(5I_3)A$, when

$$A = \begin{bmatrix} 9 & -1 & 3 \\ -8 & 7 & -3 \\ -4 & 1 & 8 \end{bmatrix}.$$

In Exercises 5 and 6, compute the product AB in two ways: (a) by the definition, where $A\mathbf{b}_1$ and $A\mathbf{b}_2$ are computed separately, and (b) by the row–column rule for computing AB.

5. $A = \begin{bmatrix} -1 & 2 \\ 5 & 4 \\ 2 & -3 \end{bmatrix}$, $B = \begin{bmatrix} 3 & -4 \\ -2 & 1 \end{bmatrix}$
6. $A = \begin{bmatrix} 4 & -2 \\ -3 & 0 \\ 3 & 5 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 3 \\ 4 & -1 \end{bmatrix}$
7. If a matrix A is 5×3 and the product AB is 5×7, what is the size of B?
8. How many rows does B have if BC is a 3×4 matrix?
9. Let $A = \begin{bmatrix} 3 & 4 \\ -2 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 5 & -6 \\ 3 & k \end{bmatrix}$. What value(s) of k, if any, will make $AB = BA$?
10. Let $A = \begin{bmatrix} 3 & -6 \\ -4 & 8 \end{bmatrix}$, $B = \begin{bmatrix} 8 & 6 \\ 5 & 7 \end{bmatrix}$, $C = \begin{bmatrix} 6 & -2 \\ 4 & 3 \end{bmatrix}$. Verify that $AB = AC$ and yet $B \neq C$.
11. Let $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 5 \end{bmatrix}$ and $D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$. Compute AD and DA. Explain how the columns or rows of A change when A is multiplied by D on the right or on the left. Find a 3×3 matrix B, not the identity matrix or the zero matrix, such that $AB = BA$.

12. Let $A = \begin{bmatrix} 2 & -8 \\ -1 & 4 \end{bmatrix}$. Construct a 2×2 matrix B such that AB is the zero matrix. Use two different nonzero columns for B.

13. Let $\mathbf{r}_1, \dots, \mathbf{r}_p$ be vectors in $\mathbb{R}^n$, and let Q be an $m \times n$ matrix. Write the matrix $[\, Q\mathbf{r}_1 \cdots Q\mathbf{r}_p \,]$ as a *product* of two matrices (neither of which is an identity matrix).

14. Let U be the 3×2 cost matrix described in Example 6 of Section 1.8. The first column of U lists the costs per dollar of output for manufacturing product B, and the second column lists the costs per dollar of output for product C. (The costs are categorized as materials, labor, and overhead.) Let $\mathbf{q}_1$ be a vector in $\mathbb{R}^2$ that lists the output (measured in dollars) of products B and C manufactured during the first quarter of the year, and let $\mathbf{q}_2, \mathbf{q}_3$, and $\mathbf{q}_4$ be the analogous vectors that list the amounts of products B and C manufactured in the second, third, and fourth quarters, respectively. Give an economic description of the data in the matrix UQ, where $Q = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \mathbf{q}_3 \quad \mathbf{q}_4]$.

Exercises 15–24 concern arbitrary matrices A, B, and C for which the indicated sums and products are defined. Mark each statement True or False **(T/F)**. Justify each answer.

15. **(T/F)** If A and B are 2×2 with columns $\mathbf{a}_1, \mathbf{a}_2$, and $\mathbf{b}_1, \mathbf{b}_2$, respectively, then $AB = [\mathbf{a}_1\mathbf{b}_1 \quad \mathbf{a}_2\mathbf{b}_2]$.

16. **(T/F)** If A and B are 3×3 and $B = [\mathbf{b}_1 \quad \mathbf{b}_2 \quad \mathbf{b}_3]$, then $AB = [A\mathbf{b}_1 + A\mathbf{b}_2 + A\mathbf{b}_3]$.

17. **(T/F)** Each column of AB is a linear combination of the columns of B using weights from the corresponding column of A.

18. **(T/F)** The second row of AB is the second row of A multiplied on the right by B.

19. **(T/F)** $AB + AC = A(B + C)$

20. **(T/F)** $A^T + B^T = (A + B)^T$

21. **(T/F)** $(AB)C = (AC)B$

22. **(T/F)** $(AB)^T = A^T B^T$

23. **(T/F)** The transpose of a product of matrices equals the product of their transposes in the same order.

24. **(T/F)** The transpose of a sum of matrices equals the sum of their transposes.

25. If $A = \begin{bmatrix} 1 & -3 \\ -3 & 8 \end{bmatrix}$ and $AB = \begin{bmatrix} -1 & 3 & -2 \\ 1 & -7 & 3 \end{bmatrix}$, determine the first and second columns of B.

26. Suppose the first two columns, $\mathbf{b}_1$ and $\mathbf{b}_2$, of B are equal. What can you say about the columns of AB (if AB is defined)? Why?

27. Suppose the third column of B is the sum of the first two columns. What can you say about the third column of AB? Why?

28. Suppose the second column of B is all zeros. What can you say about the second column of AB?

29. Suppose the last column of AB is all zeros, but B itself has no column of zeros. What can you say about the columns of A?

30. Show that if the columns of B are linearly dependent, then so are the columns of AB.

31. Suppose $CA = I_n$ (the $n \times n$ identity matrix). Show that the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Explain why A cannot have more columns than rows.

32. Suppose $AD = I_m$ (the $m \times m$ identity matrix). Show that for any $\mathbf{b}$ in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution. [*Hint:* Think about the equation $AD\mathbf{b} = \mathbf{b}$.] Explain why A cannot have more rows than columns.

33. Suppose A is an $m \times n$ matrix and there exist $n \times m$ matrices C and D such that $CA = I_n$ and $AD = I_m$. Prove that $m = n$ and $C = D$. [*Hint:* Think about the product CAD.]

34. Suppose A is a $3 \times n$ matrix whose columns span $\mathbb{R}^3$. Explain how to construct an $n \times 3$ matrix D such that $AD = I_3$.

In Exercises 35 and 36, view vectors in $\mathbb{R}^n$ as $n \times 1$ matrices. For $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, the matrix product $\mathbf{u}^T\mathbf{v}$ is a 1×1 matrix, called the **scalar product**, or **inner product**, of $\mathbf{u}$ and $\mathbf{v}$. It is usually written as a single real number without brackets. The matrix product $\mathbf{u}\mathbf{v}^T$ is an $n \times n$ matrix, called the **outer product** of $\mathbf{u}$ and $\mathbf{v}$. The products $\mathbf{u}^T\mathbf{v}$ and $\mathbf{u}\mathbf{v}^T$ will appear later in the text.

35. Let $\mathbf{u} = \begin{bmatrix} -2 \\ 3 \\ -4 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$. Compute $\mathbf{u}^T\mathbf{v}$, $\mathbf{v}^T\mathbf{u}$, $\mathbf{u}\mathbf{v}^T$, and $\mathbf{v}\mathbf{u}^T$.

36. If $\mathbf{u}$ and $\mathbf{v}$ are in $\mathbb{R}^n$, how are $\mathbf{u}^T\mathbf{v}$ and $\mathbf{v}^T\mathbf{u}$ related? How are $\mathbf{u}\mathbf{v}^T$ and $\mathbf{v}\mathbf{u}^T$ related?

37. Prove Theorem 2(b) and 2(c). Use the row–column rule. The (i, j)-entry in $A(B + C)$ can be written as

$$a_{i1}(b_{1j} + c_{1j}) + \cdots + a_{in}(b_{nj} + c_{nj}) \text{ or } \sum_{k=1}^{n} a_{ik}(b_{kj} + c_{kj})$$

38. Prove Theorem 2(d). [*Hint:* The (i, j)-entry in $(rA)B$ is $(ra_{i1})b_{1j} + \cdots + (ra_{in})b_{nj}$.]

39. Show that $I_m A = A$ when A is an $m \times n$ matrix. You can assume $I_m\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^m$.

40. Show that $AI_n = A$ when A is an $m \times n$ matrix. [*Hint:* Use the (column) definition of AI_n.]

41. Prove Theorem 3(d). [*Hint:* Consider the jth row of $(AB)^T$.]

42. Give a formula for $(AB\mathbf{x})^T$, where $\mathbf{x}$ is a vector and A and B are matrices of appropriate sizes.

T **43.** Use a web search engine such as Google to find documentation for your matrix program, and write the commands that will produce the following matrices (without keying in each entry of the matrix).

a. A 5×6 matrix of zeros

b. A 3×5 matrix of ones

c. The 6×6 identity matrix

d. A 5×5 diagonal matrix, with diagonal entries 3, 5, 7, 2, 4

A useful way to test new ideas in matrix algebra, or to make conjectures, is to make calculations with matrices selected at random. Checking a property for a few matrices does not prove that the property holds in general, but it makes the property more believable. Also, if the property is actually false, you may discover this when you make a few calculations.

T 44. Write the command(s) that will create a 6×4 matrix with random entries. In what range of numbers do the entries lie? Tell how to create a 3×3 matrix with random integer entries between -9 and 9. [*Hint:* If x is a random number such that $0 < x < 1$, then $-9.5 < 19(x - .5) < 9.5$.]

T 45. Construct a random 4×4 matrix A and test whether $(A + I)(A - I) = A^2 - I$. The best way to do this is to compute $(A + I)(A - I) - (A^2 - I)$ and verify that this difference is the zero matrix. Do this for three random matrices. Then test $(A + B)(A - B) = A^2 - B^2$ the same way for three pairs of random 4×4 matrices. Report your conclusions.

T 46. Use at least three pairs of random 4×4 matrices A and B to test the equalities $(A + B)^T = A^T + B^T$ and $(AB)^T = A^T B^T$. (See Exercise 45.) Report your conclusions. [*Note:* Most matrix programs use A' for A^T.]

T 47. Let

$$S = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Compute S^k for $k = 2, \dots, 6$.

T 48. Describe in words what happens when you compute A^5, A^{10}, A^{20}, and A^{30} for

$$A = \begin{bmatrix} 1/6 & 1/2 & 1/3 \\ 1/2 & 1/4 & 1/4 \\ 1/3 & 1/4 & 5/12 \end{bmatrix}$$

T 49. The matrix M can detect a particular 2×2 colored pattern like in Example 9. Create a nonzero 4×1 vector $\mathbf{x}$ by choosing each entry to be a zero or one. Test to see if $\mathbf{x}$ corresponds to the right pattern by calculating $\mathbf{x}^T M \mathbf{x}$. If $\mathbf{x}^T M \mathbf{x} = 0$, then $\mathbf{x}$ is the pattern identified by M. If $\mathbf{x}^T M \mathbf{x} \neq 0$, try a different nonzero vector of zeros and ones. You may want to be systematic in the way that you choose each $\mathbf{x}$ in order to avoid testing the same vector twice. You are using "guess and check" to determine which pattern of 2×2 colored squares the matrix M detects.

$$M = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

T 50. Repeat Exercise 49 with the matrix

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

T 51. Use the matrix $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ to switch the first and second rows of the matrix M containing dates of accidents at the Montreal Trudeau Airport.

Montreal:

$$M = \begin{bmatrix} 2 & 3 & 16 & 24 & 25 & 26 & 6 & 7 & 19 & 26 \\ 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 \end{bmatrix}$$

This data in matrix M has been scrubbed in matrix AM and can be fed into the same machine as the other data from Example 10.

T 52. Use the matrix $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ to remove the last row from the matrix N containing dates of accidents at the New York JFK Airport.

New York:

$$N = \begin{bmatrix} 1 & 1 & 1 & 1 & 2 & 2 & 2 \\ 1 & 12 & 21 & 22 & 3 & 20 & 21 \\ 2020 & 2020 & 2020 & 2020 & 2020 & 2020 & 2020 \end{bmatrix}$$

The data in matrix N has been scrubbed in matrix BN and can be fed into the same machine as the other data from Example 10.

Solutions to Practice Problems

1. $A\mathbf{x} = \begin{bmatrix} 1 & -3 \\ -2 & 4 \end{bmatrix}\begin{bmatrix} 5 \\ 3 \end{bmatrix} = \begin{bmatrix} -4 \\ 2 \end{bmatrix}$. So $(A\mathbf{x})^T = \begin{bmatrix} -4 & 2 \end{bmatrix}$. Also,

$$\mathbf{x}^T A^T = \begin{bmatrix} 5 & 3 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ -3 & 4 \end{bmatrix} = \begin{bmatrix} -4 & 2 \end{bmatrix}.$$

The quantities $(A\mathbf{x})^T$ and $\mathbf{x}^T A^T$ are equal, by Theorem 3(d). Next,

$$\mathbf{x}\mathbf{x}^T = \begin{bmatrix} 5 \\ 3 \end{bmatrix}\begin{bmatrix} 5 & 3 \end{bmatrix} = \begin{bmatrix} 25 & 15 \\ 15 & 9 \end{bmatrix}$$

$$\mathbf{x}^T\mathbf{x} = \begin{bmatrix} 5 & 3 \end{bmatrix}\begin{bmatrix} 5 \\ 3 \end{bmatrix} = [\,25 + 9\,] = 34$$

A 1×1 matrix such as $\mathbf{x}^T\mathbf{x}$ is usually written without the brackets. Finally, $A^T\mathbf{x}^T$ is not defined, because $\mathbf{x}^T$ does not have two rows to match the two columns of A^T.

2. The fastest way to compute $A^2\mathbf{x}$ is to compute $A(A\mathbf{x})$. The product $A\mathbf{x}$ requires 16 multiplications, 4 for each entry, and $A(A\mathbf{x})$ requires 16 more. In contrast, the product A^2 requires 64 multiplications, 4 for each of the 16 entries in A^2. After that, $A^2\mathbf{x}$ takes 16 more multiplications, for a total of 80.

3. First observe that by the definition of matrix multiplication,

$$AB = [A\mathbf{b}_1 \quad A\mathbf{b}_2 \quad \cdots \quad A\mathbf{b}_n] = [A\mathbf{b}_1 \quad A\mathbf{b}_1 \quad \cdots \quad A\mathbf{b}_1],$$

so the columns of AB are identical. Next, recall that $\text{row}_i(AB) = \text{row}_i(A) \cdot B$. Since all the rows of A are identical, all the rows of AB are identical. Putting this information about the rows and columns together, it follows that all the entries in AB are the same.

2.2 The Inverse of a Matrix

Matrix algebra provides tools for manipulating matrix equations and creating various useful formulas in ways similar to doing ordinary algebra with real numbers. This section investigates the matrix analogue of the reciprocal, or multiplicative inverse, of a nonzero number.

Recall that the multiplicative inverse of a number such as 5 is 1/5 or 5^{-1}. This inverse satisfies the equations

$$5^{-1}(5) = 1 \quad \text{and} \quad 5(5^{-1}) = 1$$

The matrix generalization requires *both* equations and avoids the slanted-line notation (for division) because matrix multiplication is not commutative. Furthermore, a full generalization is possible only if the matrices involved are square.[1]

An $n \times n$ matrix A is said to be **invertible** if there is an $n \times n$ matrix C such that

$$CA = I \quad \text{and} \quad AC = I$$

where $I = I_n$, the $n \times n$ identity matrix. In this case, C is an **inverse** of A. In fact, C is uniquely determined by A, because if B were another inverse of A, then $B = BI = B(AC) = (BA)C = IC = C$. This unique inverse is denoted by A^{-1}, so that

$$A^{-1}A = I \quad \text{and} \quad AA^{-1} = I$$

A matrix that is *not* invertible is sometimes called a **singular matrix**, and an invertible matrix is called a **nonsingular matrix**.

[1] One could say that an $m \times n$ matrix A is invertible if there exist $n \times m$ matrices C and D such that $CA = I_n$ and $AD = I_m$. However, these equations imply that A is square and $C = D$. Thus, A is invertible as defined above. See Exercises 31–33 in Section 2.1.

EXAMPLE 1 If $A = \begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix}$ and $C = \begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix}$, then

$$AC = \begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix}\begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and}$$

$$CA = \begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Thus $C = A^{-1}$. ■

Here is a simple formula for the inverse of a 2×2 matrix, along with a test to tell if the inverse exists.

THEOREM 4

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. If $ad - bc \neq 0$, then A is invertible and

$$A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

If $ad - bc = 0$, then A is not invertible.

The simple proof of Theorem 4 is outlined in Exercises 35 and 36. The quantity $ad - bc$ is called the **determinant** of A, and we write

$$\det A = ad - bc$$

Theorem 4 says that a 2×2 matrix A is invertible if and only if $\det A \neq 0$.

EXAMPLE 2 Find the inverse of $A = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}$.

SOLUTION Since $\det A = 3(6) - 4(5) = -2 \neq 0$, A is invertible, and

$$A^{-1} = \frac{1}{-2}\begin{bmatrix} 6 & -4 \\ -5 & 3 \end{bmatrix} = \begin{bmatrix} 6/(-2) & -4/(-2) \\ -5/(-2) & 3/(-2) \end{bmatrix} = \begin{bmatrix} -3 & 2 \\ 5/2 & -3/2 \end{bmatrix}$$ ■

Invertible matrices are indispensable in linear algebra—mainly for algebraic calculations and formula derivations, as in the next theorem. There are also occasions when an inverse matrix provides insight into a mathematical model of a real-life situation, as in Example 3.

THEOREM 5

If A is an invertible $n \times n$ matrix, then for each $\mathbf{b}$ in $\mathbb{R}^n$, the equation $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

PROOF Take any $\mathbf{b}$ in $\mathbb{R}^n$. A solution exists because if $A^{-1}\mathbf{b}$ is substituted for $\mathbf{x}$, then $A\mathbf{x} = A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = I\mathbf{b} = \mathbf{b}$. So $A^{-1}\mathbf{b}$ is a solution. To prove that the solution is unique, show that if $\mathbf{u}$ is any solution, then $\mathbf{u}$, in fact, must be $A^{-1}\mathbf{b}$. Indeed, if $A\mathbf{u} = \mathbf{b}$, we can multiply both sides by A^{-1} and obtain

$$A^{-1}A\mathbf{u} = A^{-1}\mathbf{b}, \quad I\mathbf{u} = A^{-1}\mathbf{b}, \quad \text{and} \quad \mathbf{u} = A^{-1}\mathbf{b}$$ ■

EXAMPLE 3 A horizontal elastic beam is supported at each end and is subjected to forces at points 1, 2, and 3, as shown in Figure 1. Let $\mathbf{f}$ in $\mathbb{R}^3$ list the forces at these points, and let $\mathbf{y}$ in $\mathbb{R}^3$ list the amounts of deflection (that is, movement) of the beam at the three points. Using Hooke's law from physics, it can be shown that

$$\mathbf{y} = D\mathbf{f}$$

where D is a *flexibility matrix*. Its inverse is called the *stiffness matrix*. Describe the physical significance of the columns of D and D^{-1}.

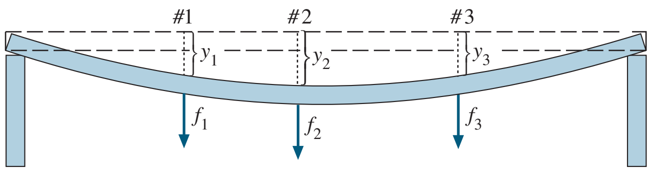


FIGURE 1 Deflection of an elastic beam.

SOLUTION Write $I_3 = [\,\mathbf{e}_1 \quad \mathbf{e}_2 \quad \mathbf{e}_3\,]$ and observe that

$$D = DI_3 = [\,D\mathbf{e}_1 \quad D\mathbf{e}_2 \quad D\mathbf{e}_3\,]$$

Interpret the vector $\mathbf{e}_1 = (1, 0, 0)$ as a unit force applied downward at point 1 on the beam (with zero force at the other two points). Then $D\mathbf{e}_1$, the first column of D, lists the beam deflections due to a unit force at point 1. Similar descriptions apply to the second and third columns of D.

To study the stiffness matrix D^{-1}, observe that the equation $\mathbf{f} = D^{-1}\mathbf{y}$ computes a force vector $\mathbf{f}$ when a deflection vector $\mathbf{y}$ is given. Write

$$D^{-1} = D^{-1}I_3 = [\,D^{-1}\mathbf{e}_1 \quad D^{-1}\mathbf{e}_2 \quad D^{-1}\mathbf{e}_3\,]$$

Now interpret $\mathbf{e}_1$ as a deflection vector. Then $D^{-1}\mathbf{e}_1$ lists the forces that create the deflection. That is, the first column of D^{-1} lists the forces that must be applied at the three points to produce a unit deflection at point 1 and zero deflections at the other points. Similarly, columns 2 and 3 of D^{-1} list the forces required to produce unit deflections at points 2 and 3, respectively. In each column, one or two of the forces must be negative (point upward) to produce a unit deflection at the desired point and zero deflections at the other two points. If the flexibility is measured, for example, in inches of deflection per pound of load, then the stiffness matrix entries are given in pounds of load per inch of deflection. ■

The formula in Theorem 5 is seldom used to solve an equation $A\mathbf{x} = \mathbf{b}$ numerically because row reduction of $[\,A \quad \mathbf{b}\,]$ is nearly always faster. (Row reduction is usually more accurate, too, when computations involve rounding off numbers.) One possible exception is the 2×2 case. In this case, mental computations to solve $A\mathbf{x} = \mathbf{b}$ are sometimes easier using the formula for A^{-1}, as in the next example.

EXAMPLE 4 Use the inverse of the matrix A in Example 2 to solve the system

$$\begin{aligned} 3x_1 + 4x_2 &= 3 \\ 5x_1 + 6x_2 &= 7 \end{aligned}$$

SOLUTION This system is equivalent to $A\mathbf{x} = \mathbf{b}$, so

$$\mathbf{x} = A^{-1}\mathbf{b} = \begin{bmatrix} -3 & 2 \\ 5/2 & -3/2 \end{bmatrix}\begin{bmatrix} 3 \\ 7 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$$ ■

The next theorem provides three useful facts about invertible matrices.

THEOREM 6

a. If A is an invertible matrix, then A^{-1} is invertible and

$$(A^{-1})^{-1} = A$$

b. If A and B are $n \times n$ invertible matrices, then so is AB, and the inverse of AB is the product of the inverses of A and B in the reverse order. That is,

$$(AB)^{-1} = B^{-1}A^{-1}$$

c. If A is an invertible matrix, then so is A^T, and the inverse of A^T is the transpose of A^{-1}. That is,

$$(A^T)^{-1} = (A^{-1})^T$$

PROOF To verify statement (a), find a matrix C such that

$$A^{-1}C = I \quad \text{and} \quad CA^{-1} = I$$

In fact, these equations are satisfied with A in place of C. Hence A^{-1} is invertible, and A is its inverse. Next, to prove statement (b), compute:

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$$

A similar calculation shows that $(B^{-1}A^{-1})(AB) = I$. For statement (c), use Theorem 3(d), read from right to left, $(A^{-1})^T A^T = (AA^{-1})^T = I^T = I$. Similarly, $A^T(A^{-1})^T = I^T = I$. Hence A^T is invertible, and its inverse is $(A^{-1})^T$. ■

Remark: Part (b) illustrates the important role that definitions play in proofs. The theorem claims that $B^{-1}A^{-1}$ is the inverse of AB. The proof establishes this by showing that $B^{-1}A^{-1}$ satisfies the definition of what it means to be the inverse of AB. Now, the inverse of AB is a matrix that when multiplied on the left (or right) by AB, the product is the identity matrix I. So the proof consists of showing that $B^{-1}A^{-1}$ has this property.

The following generalization of Theorem 6(b) is needed later.

The product of $n \times n$ invertible matrices is invertible, and the inverse is the product of their inverses in the reverse order.

There is an important connection between invertible matrices and row operations that leads to a method for computing inverses. As we shall see, an invertible matrix A is row equivalent to an identity matrix, and we can find A^{-1} by *watching the row reduction of A to I.*

Elementary Matrices

An **elementary matrix** is one that is obtained by performing a single elementary row operation on an identity matrix. The next example illustrates the three kinds of elementary matrices.

EXAMPLE 5 Let

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix},$$

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

Compute E_1A, E_2A, and E_3A, and describe how these products can be obtained by elementary row operations on A.

SOLUTION Verify that

$$E_1A = \begin{bmatrix} a & b & c \\ d & e & f \\ g-4a & h-4b & i-4c \end{bmatrix}, \quad E_2A = \begin{bmatrix} d & e & f \\ a & b & c \\ g & h & i \end{bmatrix},$$

$$E_3A = \begin{bmatrix} a & b & c \\ d & e & f \\ 5g & 5h & 5i \end{bmatrix}.$$

Addition of -4 times row 1 of A to row 3 produces E_1A. (This is a row replacement operation.) An interchange of rows 1 and 2 of A produces E_2A, and multiplication of row 3 of A by 5 produces E_3A. ■

Left-multiplication (that is, multiplication on the left) by E_1 in Example 5 has the same effect on any $3 \times n$ matrix. It adds -4 times row 1 to row 3. In particular, since $E_1 \cdot I = E_1$, we see that E_1 *itself* is produced by this same row operation on the identity. Thus Example 5 illustrates the following general fact about elementary matrices. See Exercises 37 and 38.

If an elementary row operation is performed on an $m \times n$ matrix A, the resulting matrix can be written as EA, where the $m \times m$ matrix E is created by performing the same row operation on I_m.

Since row operations are reversible, as shown in Section 1.1, elementary matrices are invertible, for if E is produced by a row operation on I, then there is another row operation of the same type that changes E back into I. Hence there is an elementary matrix F such that $FE = I$. Since E and F correspond to reverse operations, $EF = I$, too.

Each elementary matrix E is invertible. The inverse of E is the elementary matrix of the same type that transforms E back into I.

EXAMPLE 6 Find the inverse of $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}$.

SOLUTION To transform E_1 into I, add $+4$ times row 1 to row 3. The elementary matrix that does this is

$$E_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ +4 & 0 & 1 \end{bmatrix}$$

■

The following theorem provides the best way to "visualize" an invertible matrix, and the theorem leads immediately to a method for finding the inverse of a matrix.

THEOREM 7

An $n \times n$ matrix A is invertible if and only if A is row equivalent to I_n, and in this case, any sequence of elementary row operations that reduces A to I_n also transforms I_n into A^{-1}.

Remark: The comment on the proof of Theorem 11 in Chapter 1 noted that "P if and only if Q" is equivalent to two statements: (1) "If P then Q" and (2) "If Q then P." The second statement is called the *converse* of the first and explains the use of the word *conversely* in the second paragraph of this proof.

PROOF Suppose that A is invertible. Then, since the equation $A\mathbf{x} = \mathbf{b}$ has a solution for each $\mathbf{b}$ (Theorem 5), A has a pivot position in every row (Theorem 4 in Section 1.4). Because A is square, the n pivot positions must be on the diagonal, which implies that the reduced echelon form of A is I_n. That is, $A \sim I_n$.

Now suppose, conversely, that $A \sim I_n$. Then, since each step of the row reduction of A corresponds to left-multiplication by an elementary matrix, there exist elementary matrices $E_1, \dots, E_p$ such that

$$A \sim E_1 A \sim E_2(E_1 A) \sim \cdots \sim E_p(E_{p-1} \cdots E_1 A) = I_n$$

That is,

$$E_p \cdots E_1 A = I_n \tag{1}$$

Since the product $E_p \cdots E_1$ of invertible matrices is invertible, (1) leads to

$$\begin{aligned}(E_p \cdots E_1)^{-1}(E_p \cdots E_1)A &= (E_p \cdots E_1)^{-1} I_n \\ A &= (E_p \cdots E_1)^{-1}\end{aligned}$$

Thus A is invertible, as it is the inverse of an invertible matrix (Theorem 6). Also,

$$A^{-1} = [\,(E_p \cdots E_1)^{-1}\,]^{-1} = E_p \cdots E_1$$

Then $A^{-1} = E_p \cdots E_1 I_n$, which says that A^{-1} results from applying $E_1, \dots, E_p$ successively to I_n. This is the same sequence in (1) that reduced A to I_n. ■

An Algorithm for Finding A^{-1}

If we place A and I side by side to form an augmented matrix $[\,A \quad I\,]$, then row operations on this matrix produce identical operations on A and on I. By Theorem 7, either there are row operations that transform A to I_n and I_n to A^{-1} or else A is not invertible.

ALGORITHM FOR FINDING A^{-1}

Row reduce the augmented matrix $[\,A \quad I\,]$. If A is row equivalent to I, then $[\,A \quad I\,]$ is row equivalent to $[\,I \quad A^{-1}\,]$. Otherwise, A does not have an inverse.

EXAMPLE 7 Find the inverse of the matrix $A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix}$, if it exists.

SOLUTION

$$[\,A \quad I\,] = \begin{bmatrix} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & -3 & -4 & 0 & -4 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 3 & -4 & 1 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 3/2 & -2 & 1/2 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 0 & 0 & -9/2 & 7 & -3/2 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & 3/2 & -2 & 1/2 \end{bmatrix}$$

Theorem 7 shows, since $A \sim I$, that A is invertible, and

$$A^{-1} = \begin{bmatrix} -9/2 & 7 & -3/2 \\ -2 & 4 & -1 \\ 3/2 & -2 & 1/2 \end{bmatrix}$$

■

Reasonable Answers

Once you have found a candidate for the inverse of a matrix, you can check that your answer is correct by finding the product of A with A^{-1}. For the inverse found for matrix A in Example 7, notice

$$AA^{-1} = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix} \begin{bmatrix} -9/2 & 7 & -3/2 \\ -2 & 4 & -1 \\ 3/2 & -2 & 1/2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

confirming that answer is correct. It is not necessary to check that $A^{-1}A = I$ since A is invertible.

Another View of Matrix Inversion

Denote the columns of I_n by $\mathbf{e}_1, \dots, \mathbf{e}_n$. Then row reduction of $[\,A \quad I\,]$ to $[\,I \quad A^{-1}\,]$ can be viewed as the simultaneous solution of the n systems

$$A\mathbf{x} = \mathbf{e}_1, \quad A\mathbf{x} = \mathbf{e}_2, \quad \dots, \quad A\mathbf{x} = \mathbf{e}_n \tag{2}$$

where the "augmented columns" of these systems have all been placed next to A to form $[\,A \quad \mathbf{e}_1 \quad \mathbf{e}_2 \quad \cdots \quad \mathbf{e}_n\,] = [\,A \quad I\,]$. The equation $AA^{-1} = I$ and the definition of matrix multiplication show that the columns of A^{-1} are precisely the solutions of the systems in (2). This observation is useful because some applied problems may require finding only one or two columns of A^{-1}. In this case, only the corresponding systems in (2) need to be solved.

Numerical Note

In practical work, A^{-1} is seldom computed, unless the entries of A^{-1} are needed. Computing both A^{-1} and $A^{-1}\mathbf{b}$ takes about three times as many arithmetic operations as solving $A\mathbf{x} = \mathbf{b}$ by row reduction, and row reduction may be more accurate.

Practice Problems

1. Use determinants to determine which of the following matrices are invertible.

   a. $\begin{bmatrix} 3 & -9 \\ 2 & 6 \end{bmatrix}$ b. $\begin{bmatrix} 4 & -9 \\ 0 & 5 \end{bmatrix}$ c. $\begin{bmatrix} 6 & -9 \\ -4 & 6 \end{bmatrix}$

2. Find the inverse of the matrix $A = \begin{bmatrix} 1 & -2 & -1 \\ -1 & 5 & 6 \\ 5 & -4 & 5 \end{bmatrix}$, if it exists.

3. If A is an invertible matrix, prove that $5A$ is an invertible matrix.

2.2 Exercises

Find the inverses of the matrices in Exercises 1–4.

1. $\begin{bmatrix} 8 & 3 \\ 5 & 2 \end{bmatrix}$

2. $\begin{bmatrix} 5 & 4 \\ 9 & 7 \end{bmatrix}$

3. $\begin{bmatrix} 8 & 3 \\ -7 & -3 \end{bmatrix}$

4. $\begin{bmatrix} 3 & -2 \\ 7 & -4 \end{bmatrix}$

5. Verify that the inverse you found in Exercise 1 is correct.

6. Verify that the inverse you found in Exercise 2 is correct.

7. Use the inverse found in Exercise 1 to solve the system

$$\begin{aligned} 8x_1 + 3x_2 &= 2 \\ 5x_1 + 2x_2 &= -1 \end{aligned}$$

8. Use the inverse found in Exercise 2 to solve the system

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

9. Let $A = \begin{bmatrix} 1 & 2 \\ 5 & 12 \end{bmatrix}$, $\mathbf{b}_1 = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} 1 \\ -5 \end{bmatrix}$, $\mathbf{b}_3 = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$, and $\mathbf{b}_4 = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$.

   a. Find A^{-1}, and use it to solve the four equations $A\mathbf{x} = \mathbf{b}_1$, $A\mathbf{x} = \mathbf{b}_2$, $A\mathbf{x} = \mathbf{b}_3$, $A\mathbf{x} = \mathbf{b}_4$

   b. The four equations in part (a) can be solved by the same set of row operations, since the coefficient matrix is the same in each case. Solve the four equations in part (a) by row reducing the augmented matrix $\begin{bmatrix} A & \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \mathbf{b}_4 \end{bmatrix}$

10. Use matrix algebra to show that if A is invertible and D satisfies $AD = I$, then $D = A^{-1}$.

In Exercises 11–20, mark each statement True or False **(T/F)**. Justify each answer.

11. **(T/F)** In order for a matrix B to be the inverse of A, both equations $AB = I$ and $BA = I$ must be true.

12. **(T/F)** A product of invertible $n \times n$ matrices is invertible, and the inverse of the product is the product of their inverses in the same order.

13. **(T/F)** If A and B are $n \times n$ and invertible, then $A^{-1}B^{-1}$ is the inverse of AB.

14. **(T/F)** If A is invertible, then the inverse of A^{-1} is A itself.

15. **(T/F)** If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $ab - cd \neq 0$, then A is invertible.

16. **(T/F)** If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $ad = bc$, then A is not invertible.

17. **(T/F)** If A is an invertible $n \times n$ matrix, then the equation $A\mathbf{x} = \mathbf{b}$ is consistent for *each* $\mathbf{b}$ in $\mathbb{R}^n$.

18. **(T/F)** If A can be row reduced to the identity matrix, then A must be invertible.

19. **(T/F)** Each elementary matrix is invertible.

20. **(T/F)** If A is invertible, then the elementary row operations that reduce A to the identity I_n also reduce A^{-1} to I_n.

21. Let A be an invertible $n \times n$ matrix, and let B be an $n \times p$ matrix. Show that the equation $AX = B$ has a unique solution $A^{-1}B$.

22. Let A be an invertible $n \times n$ matrix, and let B be an $n \times p$ matrix. Explain why $A^{-1}B$ can be computed by row reduction:

$$\text{If } [A \;\; B] \sim \cdots \sim [I \;\; X], \text{ then } X = A^{-1}B.$$

If A is larger than 2×2, then row reduction of $[A \;\; B]$ is much faster than computing both A^{-1} and $A^{-1}B$.

23. Suppose $AB = AC$, where B and C are $n \times p$ matrices and A is invertible. Show that $B = C$. Is this true, in general, when A is not invertible?

24. Suppose $(B - C)D = 0$, where B and C are $m \times n$ matrices and D is invertible. Show that $B = C$.

25. Suppose A, B, and C are invertible $n \times n$ matrices. Show that ABC is also invertible by producing a matrix D such that $(ABC)\,D = I$ and $D\,(ABC) = I$.

26. Suppose A and B are $n \times n$, B is invertible, and AB is invertible. Show that A is invertible. [*Hint:* Let $C = AB$, and solve this equation for A.]

27. Solve the equation $AB = BC$ for A, assuming that A, B, and C are square and B is invertible.

28. Suppose P is invertible and $A = PBP^{-1}$. Solve for B in terms of A.

29. If A, B, and C are $n \times n$ invertible matrices, does the equation $C^{-1}(A + X)B^{-1} = I_n$ have a solution, X? If so, find it.

30. Suppose A, B, and X are $n \times n$ matrices with A, X, and $A - AX$ invertible, and suppose

$$(A - AX)^{-1} = X^{-1}B \tag{3}$$

a. Explain why B is invertible.

b. Solve (3) for X. If you need to invert a matrix, explain why that matrix is invertible.

31. Explain why the columns of an $n \times n$ matrix A are linearly independent when A is invertible.

32. Explain why the columns of an $n \times n$ matrix A span $\mathbb{R}^n$ when A is invertible. [*Hint:* Review Theorem 4 in Section 1.4.]

33. Suppose A is $n \times n$ and the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Explain why A has n pivot columns and A is row equivalent to I_n. By Theorem 7, this shows that A must be invertible. (This exercise and Exercise 34 will be cited in Section 2.3.)

34. Suppose A is $n \times n$ and the equation $A\mathbf{x} = \mathbf{b}$ has a solution for each $\mathbf{b}$ in $\mathbb{R}^n$. Explain why A must be invertible. [*Hint:* Is A row equivalent to I_n?]

Exercises 35 and 36 prove Theorem 4 for $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

35. Show that if $ad - bc = 0$, then the equation $A\mathbf{x} = \mathbf{0}$ has more than one solution. Why does this imply that A is not invertible? [*Hint:* First, consider $a = b = 0$. Then, if a and b are not both zero, consider the vector $\mathbf{x} = \begin{bmatrix} -b \\ a \end{bmatrix}$.]

36. Show that if $ad - bc \neq 0$, the formula for A^{-1} works.

Exercises 37 and 38 prove special cases of the facts about elementary matrices stated in the box following Example 5. Here A is a 3×3 matrix and $I = I_3$. (A general proof would require slightly more notation.)

37. a. Use equation (1) from Section 2.1 to show that $\text{row}_i(A) = \text{row}_i(I) \cdot A$, for $i = 1, 2, 3$.

b. Show that if rows 1 and 2 of A are interchanged, then the result may be written as EA, where E is an elementary matrix formed by interchanging rows 1 and 2 of I.

c. Show that if row 3 of A is multiplied by 5, then the result may be written as EA, where E is formed by multiplying row 3 of I by 5.

38. Show that if row 3 of A is replaced by $\text{row}_3(A) - 4\text{row}_1(A)$, the result is EA, where E is formed from I by replacing $\text{row}_3(I)$ by $\text{row}_3(I) - 4\text{row}_1(I)$.

Find the inverses of the matrices in Exercises 39–42, if they exist. Use the algorithm introduced in this section.

39. $\begin{bmatrix} 1 & 2 \\ 4 & 7 \end{bmatrix}$

40. $\begin{bmatrix} 9 & 7 \\ 8 & 6 \end{bmatrix}$

41. $\begin{bmatrix} 1 & 0 & -2 \\ -3 & 1 & 4 \\ 2 & -3 & 4 \end{bmatrix}$

42. $\begin{bmatrix} 1 & -2 & 1 \\ 4 & -7 & 3 \\ -2 & 6 & -4 \end{bmatrix}$

43. Use the algorithm from this section to find the inverses of

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

Let A be the corresponding $n \times n$ matrix, and let B be its inverse. Guess the form of B, and then prove that $AB = I$ and $BA = I$.

44. Repeat the strategy of Exercise 43 to guess the inverse of $A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 2 & 0 & & 0 \\ 1 & 2 & 3 & & 0 \\ \vdots & & & \ddots & \vdots \\ 1 & 2 & 3 & \cdots & n \end{bmatrix}$. Prove that your guess is correct.

45. Let $A = \begin{bmatrix} -2 & -7 & -9 \\ 2 & 5 & 6 \\ 1 & 3 & 4 \end{bmatrix}$. Find the third column of A^{-1} without computing the other columns.

T **46.** Let $A = \begin{bmatrix} -25 & -9 & -27 \\ 546 & 180 & 537 \\ 154 & 50 & 149 \end{bmatrix}$. Find the second and third columns of A^{-1} without computing the first column.

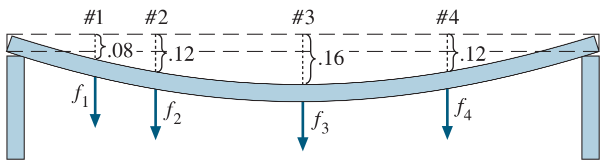
47. Let $A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \end{bmatrix}$. Construct a 2×3 matrix C (by trial and error) using only 1, -1, and 0 as entries, such that $CA = I_2$. Compute AC and note that $AC \neq I_3$.

48. Let $A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}$. Construct a 4×2 matrix D using only 1 and 0 as entries, such that $AD = I_2$. Is it possible that $CA = I_4$ for some 4×2 matrix C? Why or why not?

49. Let $D = \begin{bmatrix} .005 & .002 & .001 \\ .002 & .004 & .002 \\ .001 & .002 & .005 \end{bmatrix}$ be a flexibility matrix, with flexibility measured in inches per pound. Suppose that forces of 30, 50, and 20 lb are applied at points 1, 2, and 3, respectively, in Figure 1 of Example 3. Find the corresponding deflections.

50. Compute the stiffness matrix D^{-1} for D in Exercise 49. List the forces needed to produce a deflection of .04 in. at point 3, with zero deflections at the other points.

51. Let $D = \begin{bmatrix} .0040 & .0030 & .0010 & .0005 \\ .0030 & .0050 & .0030 & .0010 \\ .0010 & .0030 & .0050 & .0030 \\ .0005 & .0010 & .0030 & .0040 \end{bmatrix}$ be a flexibility matrix for an elastic beam with four points at which force is applied. Units are centimeters per newton of force. Measurements at the four points show deflections of .08, .12, .16, and .12 cm. Determine the forces at the four points.



Deflection of elastic beam in Exercises 51 and 52.

52. With D as in Exercise 51, determine the forces that produce a deflection of .24 cm at the second point on the beam, with zero deflections at the other three points. How is the answer related to the entries in D^{-1}? [*Hint:* First answer the question when the deflection is 1 cm at the second point.]

Solutions to Practice Problems

1. a. $\det \begin{bmatrix} 3 & -9 \\ 2 & 6 \end{bmatrix} = 3 \cdot 6 - (-9) \cdot 2 = 18 + 18 = 36$. The determinant is nonzero, so the matrix is invertible.

b. $\det \begin{bmatrix} 4 & -9 \\ 0 & 5 \end{bmatrix} = 4 \cdot 5 - (-9) \cdot 0 = 20 \neq 0$. The matrix is invertible.

c. $\det \begin{bmatrix} 6 & -9 \\ -4 & 6 \end{bmatrix} = 6 \cdot 6 - (-9)(-4) = 36 - 36 = 0$. The matrix is not invertible.

2.
$$[\,A \quad I\,] \sim \begin{bmatrix} 1 & -2 & -1 & 1 & 0 & 0 \\ -1 & 5 & 6 & 0 & 1 & 0 \\ 5 & -4 & 5 & 0 & 0 & 1 \end{bmatrix}$$
$$\sim \begin{bmatrix} 1 & -2 & -1 & 1 & 0 & 0 \\ 0 & 3 & 5 & 1 & 1 & 0 \\ 0 & 6 & 10 & -5 & 0 & 1 \end{bmatrix}$$
$$\sim \begin{bmatrix} 1 & -2 & -1 & 1 & 0 & 0 \\ 0 & 3 & 5 & 1 & 1 & 0 \\ 0 & 0 & 0 & -7 & -2 & 1 \end{bmatrix}$$

So $[\,A \quad I\,]$ is row equivalent to a matrix of the form $[\,B \quad D\,]$, where B is square and has a row of zeros. Further row operations will not transform B into I, so we stop. A does not have an inverse.

3. Since A is an invertible matrix, there exists a matrix C such that $AC = I = CA$. The goal is to find a matrix D so that $(5A)D = I = D(5A)$. Set $D = 1/5\,C$. Applying Theorem 2 from Section 2.1 establishes that $(5A)(1/5\,C) = (5)(1/5)(AC) = 1\,I = I$, and $(1/5\,C)(5A) = (1/5)(5)(CA) = 1\,I = I$. Thus $1/5\,C$ is indeed the inverse of A, proving that A is invertible.

2.3 Characterizations of Invertible Matrices

This section provides a review of most of the concepts introduced in Chapter 1, in relation to systems of n linear equations in n unknowns and to *square* matrices. The main result is Theorem 8.

THEOREM 8 **The Invertible Matrix Theorem**

Let A be a square $n \times n$ matrix. Then the following statements are equivalent. That is, for a given A, the statements are either all true or all false.

a. A is an invertible matrix.
b. A is row equivalent to the $n \times n$ identity matrix.
c. A has n pivot positions.
d. The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
e. The columns of A form a linearly independent set.
f. The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.
g. The equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$.
h. The columns of A span $\mathbb{R}^n$.
i. The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.
j. There is an $n \times n$ matrix C such that $CA = I$.
k. There is an $n \times n$ matrix D such that $AD = I$.
l. A^T is an invertible matrix.

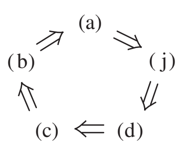
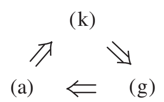


FIGURE 1

First, we need some notation. If the truth of statement (a) always implies that statement (j) is true, we say that (a) *implies* (j) and write (a) $\Rightarrow$ (j). The proof will establish the "circle" of implications shown in Figure 1. If any one of these five statements is true, then so are the others. Finally, the proof will link the remaining statements of the theorem to the statements in this circle.

PROOF If statement (a) is true, then A^{-1} works for C in (j), so (a) $\Rightarrow$ (j). Next, (j) $\Rightarrow$ (d) by Exercise 31 in Section 2.1. (Turn back and read the exercise.) Also, (d) $\Rightarrow$ (c) by Exercise 33 in Section 2.2. If A is square and has n pivot positions, then the pivots must lie on the main diagonal, in which case the reduced echelon form of A is I_n. Thus (c) $\Rightarrow$ (b). Also, (b) $\Rightarrow$ (a) by Theorem 7 in Section 2.2. This completes the circle in Figure 1.



Next, (a) $\Rightarrow$ (k) because A^{-1} works for D. Also, (k) $\Rightarrow$ (g) by Exercise 32 in Section 2.1, and (g) $\Rightarrow$ (a) by Exercise 34 in Section 2.2. So (k) and (g) are linked to the circle. Further, (g), (h), and (i) are equivalent for any matrix, by Theorem 4 in Section 1.4 and Theorem 12(a) in Section 1.9. Thus, (h) and (i) are linked through (g) to the circle.

$$\text{(g)} \Leftrightarrow \text{(h)} \Leftrightarrow \text{(i)}$$

$$\text{(d)} \Leftrightarrow \text{(e)} \Leftrightarrow \text{(f)}$$

$$\text{(a)} \Leftrightarrow \text{(l)}$$

Since (d) is linked to the circle, so are (e) and (f), because (d), (e), and (f) are all equivalent for *any* matrix A. (See Section 1.7 and Theorem 12(b) in Section 1.9.) Finally, (a) $\Rightarrow$ (l) by Theorem 6(c) in Section 2.2, and (l) $\Rightarrow$ (a) by the same theorem with A and A^T interchanged. This completes the proof. ■

Because of Theorem 5 in Section 2.2, statement (g) in Theorem 8 could also be written as "The equation $A\mathbf{x} = \mathbf{b}$ has a *unique* solution for each $\mathbf{b}$ in $\mathbb{R}^n$." This statement certainly implies (b) and hence implies that A is invertible.

The next fact follows from Theorem 8 and Exercise 10 in Section 2.2.

Let A and B be square matrices. If $AB = I$, then A and B are both invertible, with $B = A^{-1}$ and $A = B^{-1}$.

The Invertible Matrix Theorem divides the set of all $n \times n$ matrices into two disjoint classes: the invertible (nonsingular) matrices, and the noninvertible (singular) matrices. Each statement in the theorem describes a property of every $n \times n$ invertible matrix. The *negation* of a statement in the theorem describes a property of every $n \times n$ singular matrix. For instance, an $n \times n$ singular matrix is *not* row equivalent to I_n, does *not* have n pivot positions, and has linearly *dependent* columns. Negations of other statements are considered in the exercises.

EXAMPLE 1 Use the Invertible Matrix Theorem to decide if A is invertible:

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 3 & 1 & -2 \\ -5 & -1 & 9 \end{bmatrix}$$

SOLUTION

$$A \sim \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 4 \\ 0 & -1 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 4 \\ 0 & 0 & 3 \end{bmatrix}$$

So A has three pivot positions and hence is invertible, by the Invertible Matrix Theorem, statement (c). ■

STUDY GUIDE offers an expanded table for the Invertible Matrix Theorem.

The power of the Invertible Matrix Theorem lies in the connections it provides among so many important concepts, such as linear independence of columns of a matrix A and the existence of solutions to equations of the form $A\mathbf{x} = \mathbf{b}$. It should be emphasized, however, that the Invertible Matrix Theorem *applies only to square matrices*. For example, if the columns of a 4×3 matrix are linearly independent, we cannot use the Invertible Matrix Theorem to conclude anything about the existence or nonexistence of solutions to equations of the form $A\mathbf{x} = \mathbf{b}$.

Invertible Linear Transformations

Recall from Section 2.1 that matrix multiplication corresponds to composition of linear transformations. When a matrix A is invertible, the equation $A^{-1}A\mathbf{x} = \mathbf{x}$ can be viewed as a statement about linear transformations. See Figure 2.

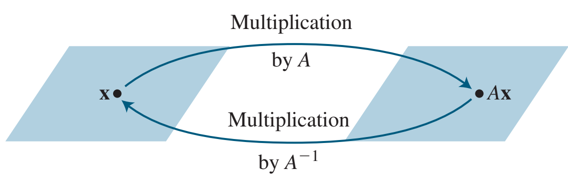


FIGURE 2 A^{-1} transforms $A\mathbf{x}$ back to $\mathbf{x}$.

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is said to be **invertible** if there exists a function $S : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$S(T(\mathbf{x})) = \mathbf{x} \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^n \tag{1}$$

$$T(S(\mathbf{x})) = \mathbf{x} \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^n \tag{2}$$

The next theorem shows that if such an S exists, it is unique and must be a linear transformation. We call S the **inverse** of T and write it as T^{-1}.

THEOREM 9

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation and let A be the standard matrix for T. Then T is invertible if and only if A is an invertible matrix. In that case, the linear transformation S given by $S(\mathbf{x}) = A^{-1}\mathbf{x}$ is the unique function satisfying equations (1) and (2).

Remark: See the comment on the proof of Theorem 7.

PROOF Suppose that T is invertible. Then (2) shows that T is onto $\mathbb{R}^n$, for if $\mathbf{b}$ is in $\mathbb{R}^n$ and $\mathbf{x} = S(\mathbf{b})$, then $T(\mathbf{x}) = T(S(\mathbf{b})) = \mathbf{b}$, so each $\mathbf{b}$ is in the range of T. Thus A is invertible, by the Invertible Matrix Theorem, statement (i).

Conversely, suppose that A is invertible, and let $S(\mathbf{x}) = A^{-1}\mathbf{x}$. Then, S is a linear transformation, and S obviously satisfies (1) and (2). For instance,

$$S(T(\mathbf{x})) = S(A\mathbf{x}) = A^{-1}(A\mathbf{x}) = \mathbf{x}$$

Thus T is invertible. The proof that S is unique is outlined in Exercise 47. ■

EXAMPLE 2 What can you say about a one-to-one linear transformation T from $\mathbb{R}^n$ into $\mathbb{R}^n$?

SOLUTION The columns of the standard matrix A of T are linearly independent (by Theorem 12 in Section 1.9). So A is invertible, by the Invertible Matrix Theorem, and T maps $\mathbb{R}^n$ *onto* $\mathbb{R}^n$. Also, T is invertible, by Theorem 9. ■

Numerical Notes

In practical work, you might occasionally encounter a “nearly singular” or **ill-conditioned** matrix—an invertible matrix that can become singular if some of its entries are changed ever so slightly. In this case, row reduction may produce fewer than n pivot positions, as a result of roundoff error. Also, roundoff error can sometimes make a singular matrix appear to be invertible.

Some matrix programs will compute a **condition number** for a square matrix. The larger the condition number, the closer the matrix is to being singular. The condition number of the identity matrix is 1. A singular matrix has an infinite condition number. In extreme cases, a matrix program may not be able to distinguish between a singular matrix and an ill-conditioned matrix.

Exercises 49–53 show that matrix computations can produce substantial error when a condition number is large.

Practice Problems

1. Determine if $A = \begin{bmatrix} 2 & 3 & 4 \\ 2 & 3 & 4 \\ 2 & 3 & 4 \end{bmatrix}$ is invertible.
2. Suppose that for a certain $n \times n$ matrix A, statement (g) of the Invertible Matrix Theorem is *not* true. What can you say about equations of the form $A\mathbf{x} = \mathbf{b}$?
3. Suppose that A and B are $n \times n$ matrices and the equation $AB\mathbf{x} = \mathbf{0}$ has a nontrivial solution. What can you say about the matrix AB?

2.3 Exercises

Unless otherwise specified, assume that all matrices in these exercises are $n \times n$. Determine which of the matrices in Exercises 1–10 are invertible. Use as few calculations as possible. Justify your answers.

1. $\begin{bmatrix} 5 & 7 \\ -3 & -6 \end{bmatrix}$

2. $\begin{bmatrix} -4 & 6 \\ 6 & -9 \end{bmatrix}$

3. $\begin{bmatrix} 5 & 0 & 0 \\ -3 & -7 & 0 \\ 8 & 5 & -1 \end{bmatrix}$

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

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

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

7. $\begin{bmatrix} -1 & 0 & 2 & 1 \\ -5 & -3 & 9 & 3 \\ 3 & 0 & 1 & -3 \\ 0 & 3 & 1 & 2 \end{bmatrix}$

8. $\begin{bmatrix} 1 & 3 & 7 & 4 \\ 0 & 5 & 9 & 6 \\ 0 & 0 & 2 & 8 \\ 0 & 0 & 0 & 10 \end{bmatrix}$

T 9. $\begin{bmatrix} 4 & 0 & -7 & -7 \\ -6 & 1 & 11 & 9 \\ 7 & -5 & 10 & 19 \\ -1 & 2 & 3 & -1 \end{bmatrix}$

T 10. $\begin{bmatrix} 5 & 3 & 1 & 7 & 9 \\ 6 & 4 & 2 & 8 & -8 \\ 7 & 5 & 3 & 10 & 9 \\ 9 & 6 & 4 & -9 & -5 \\ 8 & 5 & 2 & 11 & 4 \end{bmatrix}$

In Exercises 11–20, the matrices are all $n \times n$. Each part of the exercises is an *implication* of the form "If 'statement 1', then 'statement 2'." Mark an implication as True if the truth of "statement 2" *always* follows whenever "statement 1" happens to be true. An implication is False if there is an instance in which "statement 2" is false but "statement 1" is true. Justify each answer.

11. **(T/F)** If the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, then A is row equivalent to the $n \times n$ identity matrix.

12. **(T/F)** If there is an $n \times n$ matrix D such that $AD = I$, then there is also an $n \times n$ matrix C such that $CA = I$.

13. **(T/F)** If the columns of A span $\mathbb{R}^n$, then the columns are linearly independent.

14. **(T/F)** If the columns of A are linearly independent, then the columns of A span $\mathbb{R}^n$.

15. **(T/F)** If A is an $n \times n$ matrix, then the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$.

16. **(T/F)** If the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$, then the solution is unique for each $\mathbf{b}$.

17. **(T/F)** If the equation $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution, then A has fewer than n pivot positions.

18. **(T/F)** If the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ into $\mathbb{R}^n$, then A has n pivot positions.

19. **(T/F)** If A^T is not invertible, then A is not invertible.

20. **(T/F)** If there is a $\mathbf{b}$ in $\mathbb{R}^n$ such that the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent, then the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is not one-to-one.

21. An $m \times n$ **upper triangular matrix** is one whose entries *below* the main diagonal are 0's (as in Exercise 8). When is a square upper triangular matrix invertible? Justify your answer.

22. An $m \times n$ **lower triangular matrix** is one whose entries *above* the main diagonal are 0's (as in Exercise 3). When is a square lower triangular matrix invertible? Justify your answer.

23. Can a square matrix with two identical columns be invertible? Why or why not?

24. Is it possible for a 5×5 matrix to be invertible when its columns do not span $\mathbb{R}^5$? Why or why not?

25. If A is invertible, then the columns of A^{-1} are linearly independent. Explain why.

26. If C is 6×6 and the equation $C\mathbf{x} = \mathbf{v}$ is consistent for every $\mathbf{v}$ in $\mathbb{R}^6$, is it possible that for some $\mathbf{v}$, the equation $C\mathbf{x} = \mathbf{v}$ has more than one solution? Why or why not?

27. If the columns of a 7×7 matrix D are linearly independent, what can you say about solutions of $D\mathbf{x} = \mathbf{b}$? Why?

28. If $n \times n$ matrices E and F have the property that $EF = I$, then E and F commute. Explain why.

29. If the equation $G\mathbf{x} = \mathbf{y}$ has more than one solution for some $\mathbf{y}$ in $\mathbb{R}^n$, can the columns of G span $\mathbb{R}^n$? Why or why not?

30. If the equation $H\mathbf{x} = \mathbf{c}$ is inconsistent for some $\mathbf{c}$ in $\mathbb{R}^n$, what can you say about the equation $H\mathbf{x} = \mathbf{0}$? Why?

31. If an $n \times n$ matrix K cannot be row reduced to I_n, what can you say about the columns of K? Why?

32. If L is $n \times n$ and the equation $L\mathbf{x} = \mathbf{0}$ has the trivial solution, do the columns of L span $\mathbb{R}^n$? Why?

33. Verify the boxed statement preceding Example 1.

34. Explain why the columns of A^2 span $\mathbb{R}^n$ whenever the columns of A are linearly independent.

35. Show that if AB is invertible, so is A. You cannot use Theorem 6(b), because you cannot *assume* that A and B are invertible. [*Hint:* There is a matrix W such that $ABW = I$. Why?]

36. Show that if AB is invertible, so is B.

37. If A is an $n \times n$ matrix and the equation $A\mathbf{x} = \mathbf{b}$ has more than one solution for some $\mathbf{b}$, then the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is

not one-to-one. What else can you say about this transformation? Justify your answer.

38. If A is an $n \times n$ matrix and the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one, what else can you say about this transformation? Justify your answer.

39. Suppose A is an $n \times n$ matrix with the property that the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$. Without using Theorems 5 or 8, explain why each equation $A\mathbf{x} = \mathbf{b}$ has in fact exactly one solution.

40. Suppose A is an $n \times n$ matrix with the property that the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Without using the Invertible Matrix Theorem, explain directly why the equation $A\mathbf{x} = \mathbf{b}$ must have a solution for each $\mathbf{b}$ in $\mathbb{R}^n$.

In Exercises 41 and 42, T is a linear transformation from $\mathbb{R}^2$ into $\mathbb{R}^2$. Show that T is invertible and find a formula for T^{-1}.

41. $T(x_1, x_2) = (-9x_1 + 7x_2, 4x_1 - 3x_2)$

42. $T(x_1, x_2) = (6x_1 - 8x_2, -5x_1 + 7x_2)$

43. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear transformation. Explain why T is both one-to-one and onto $\mathbb{R}^n$. Use equations (1) and (2). Then give a second explanation using one or more theorems.

44. Let T be a linear transformation that maps $\mathbb{R}^n$ onto $\mathbb{R}^n$. Show that T^{-1} exists and maps $\mathbb{R}^n$ onto $\mathbb{R}^n$. Is T^{-1} also one-to-one?

45. Suppose T and U are linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^n$ such that $T(U\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^n$. Is it true that $U(T\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^n$? Why or why not?

46. Suppose a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ has the property that $T(\mathbf{u}) = T(\mathbf{v})$ for some pair of distinct vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$. Can T map $\mathbb{R}^n$ onto $\mathbb{R}^n$? Why or why not?

47. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear transformation, and let S and U be functions from $\mathbb{R}^n$ into $\mathbb{R}^n$ such that $S(T(\mathbf{x})) = \mathbf{x}$ and $U(T(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^n$. Show that $U(\mathbf{v}) = S(\mathbf{v})$ for all $\mathbf{v}$ in $\mathbb{R}^n$. This will show that T has a unique inverse, as asserted in Theorem 9. [*Hint:* Given any $\mathbf{v}$ in $\mathbb{R}^n$, we can write $\mathbf{v} = T(\mathbf{x})$ for some $\mathbf{x}$. Why? Compute $S(\mathbf{v})$ and $U(\mathbf{v})$.]

48. Suppose T and S satisfy the invertibility equations (1) and (2), where T is a linear transformation. Show directly that S is a linear transformation. [*Hint:* Given $\mathbf{u}$, $\mathbf{v}$ in $\mathbb{R}^n$, let $\mathbf{x} = S(\mathbf{u})$, $\mathbf{y} = S(\mathbf{v})$. Then $T(\mathbf{x}) = \mathbf{u}$, $T(\mathbf{y}) = \mathbf{v}$. Why? Apply S to both sides of the equation $T(\mathbf{x}) + T(\mathbf{y}) = T(\mathbf{x} + \mathbf{y})$. Also, consider $T(c\mathbf{x}) = cT(\mathbf{x})$.]

T 49. Suppose an experiment leads to the following system of equations:

$$\begin{aligned} 4.5x_1 + 3.1x_2 &= 19.249 \\ 1.6x_1 + 1.1x_2 &= 6.843 \end{aligned} \tag{3}$$

a. Solve system (3), and then solve system (4), below, in which the data on the right have been rounded to two decimal places. In each case, find the *exact* solution.

$$\begin{aligned} 4.5x_1 + 3.1x_2 &= 19.25 \\ 1.6x_1 + 1.1x_2 &= 6.84 \end{aligned} \tag{4}$$

b. The entries in (4) differ from those in (3) by less than .05%. Find the percentage error when using the solution of (4) as an approximation for the solution of (3).

c. Use your matrix program to produce the condition number of the coefficient matrix in (3).

Exercises 50–52 show how to use the condition number of a matrix A to estimate the accuracy of a computed solution of $A\mathbf{x} = \mathbf{b}$. If the entries of A and $\mathbf{b}$ are accurate to about r significant digits and if the condition number of A is approximately 10^k (with k a positive integer), then the computed solution of $A\mathbf{x} = \mathbf{b}$ should usually be accurate to at least $r - k$ significant digits.

T 50. Find the condition number of the matrix A in Exercise 9. Construct a random vector $\mathbf{x}$ in $\mathbb{R}^4$ and compute $\mathbf{b} = A\mathbf{x}$. Then use your matrix program to compute the solution $\mathbf{x}_1$ of $A\mathbf{x} = \mathbf{b}$. To how many digits do $\mathbf{x}$ and $\mathbf{x}_1$ agree? Find out the number of digits your matrix program stores accurately, and report how many digits of accuracy are lost when $\mathbf{x}_1$ is used in place of the exact solution $\mathbf{x}$.

T 51. Repeat Exercise 50 for the matrix in Exercise 10.

T 52. Solve an equation $A\mathbf{x} = \mathbf{b}$ for a suitable $\mathbf{b}$ to find the last column of the inverse of the *fifth-order Hilbert matrix*

$$A = \begin{bmatrix} 1 & 1/2 & 1/3 & 1/4 & 1/5 \\ 1/2 & 1/3 & 1/4 & 1/5 & 1/6 \\ 1/3 & 1/4 & 1/5 & 1/6 & 1/7 \\ 1/4 & 1/5 & 1/6 & 1/7 & 1/8 \\ 1/5 & 1/6 & 1/7 & 1/8 & 1/9 \end{bmatrix}$$

How many digits in each entry of $\mathbf{x}$ do you expect to be correct? Explain. [*Note:* The exact solution is (630, −12600, 56700, −88200, 44100).]

T 53. Some matrix programs, such as MATLAB, have a command to create Hilbert matrices of various sizes. If possible, use an inverse command to compute the inverse of a twelfth-order or larger Hilbert matrix, A. Compute AA^{-1}. Report what you find.

STUDY GUIDE offers additional resources for reviewing and reflecting on what you have learned.

Solutions to Practice Problems

1. The columns of A are obviously linearly dependent because columns 2 and 3 are multiples of column 1. Hence, A cannot be invertible (by the Invertible Matrix Theorem).

Solutions to Practice Problems (Continued)

2. If statement (g) is *not* true, then the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent for at least one $\mathbf{b}$ in $\mathbb{R}^n$.
3. Apply the Invertible Matrix Theorem to the matrix AB in place of A. Then statement (d) becomes: $AB\mathbf{x} = \mathbf{0}$ has only the trivial solution. This is not true. So AB is not invertible.

2.4 Partitioned Matrices

A key feature of our work with matrices has been the ability to regard a matrix A as a list of column vectors rather than just a rectangular array of numbers. This point of view has been so useful that we wish to consider other **partitions** of A, indicated by horizontal and vertical dividing rules, as in Example 1 below. Partitioned matrices appear in most modern applications of linear algebra because the notation highlights essential structures in matrix analysis, as in the chapter introductory example on aircraft design. This section provides an opportunity to review matrix algebra and use the Invertible Matrix Theorem.

EXAMPLE 1 The matrix

$$A = \left[\begin{array}{rrr|rr|r} 3 & 0 & -1 & 5 & 9 & -2 \\ -5 & 2 & 4 & 0 & -3 & 1 \\ \hline -8 & -6 & 3 & 1 & 7 & -4 \end{array}\right]$$

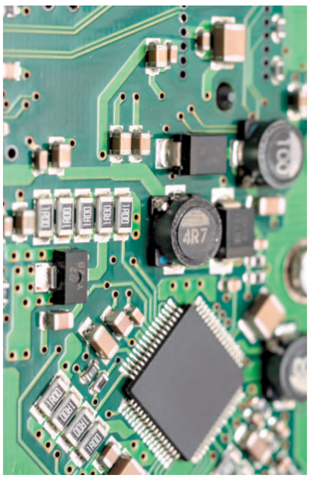
can also be written as the 2×3 **partitioned** (or **block**) **matrix**

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}$$

whose entries are the *blocks* (or *submatrices*)

$$A_{11} = \begin{bmatrix} 3 & 0 & -1 \\ -5 & 2 & 4 \end{bmatrix}, \quad A_{12} = \begin{bmatrix} 5 & 9 \\ 0 & -3 \end{bmatrix}, \quad A_{13} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

$$A_{21} = \begin{bmatrix} -8 & -6 & 3 \end{bmatrix}, \quad A_{22} = \begin{bmatrix} 1 & 7 \end{bmatrix}, \quad A_{23} = \begin{bmatrix} -4 \end{bmatrix}$$ ■



EXAMPLE 2 When a matrix A appears in a mathematical model of a physical system such as an electrical network, a transportation system, or a large corporation, it may be natural to regard A as a partitioned matrix. For instance, if a microcomputer circuit board consists mainly of three VLSI (very large-scale integrated) microchips, then the matrix for the circuit board might have the general form

$$A = \left[\begin{array}{c|c|c} A_{11} & A_{12} & A_{13} \\ \hline A_{21} & A_{22} & A_{23} \\ \hline A_{31} & A_{32} & A_{33} \end{array}\right]$$

The submatrices on the "diagonal" of A—namely A_{11}, A_{22}, and A_{33}—concern the three VLSI chips, while the other submatrices depend on the interconnections among those microchips. ■

Addition and Scalar Multiplication

If matrices A and B are the same size and are partitioned in exactly the same way, then it is natural to make the same partition of the ordinary matrix sum $A + B$. In this

27. Without using row reduction, find the inverse of

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

T 28. For block operations, it may be necessary to access or enter submatrices of a large matrix. Describe the functions or commands of your matrix program that accomplish the following tasks. Suppose A is a 20×30 matrix.

a. Display the submatrix of A from rows 15 to 20 and columns 5 to 10.

b. Insert a 5×10 matrix B into A, beginning at row 10 and column 20.

c. Create a 50×50 matrix of the form $B = \begin{bmatrix} A & 0 \\ 0 & A^T \end{bmatrix}$.

[*Note:* It may not be necessary to specify the zero blocks in B.]

T 29. Suppose memory or size restrictions prevent your matrix program from working with matrices having more than 32 rows and 32 columns, and suppose some project involves 50×50 matrices A and B. Describe the commands or operations of your matrix program that accomplish the following tasks.

a. Compute $A + B$.

b. Compute AB.

c. Solve $A\mathbf{x} = \mathbf{b}$ for some vector $\mathbf{b}$ in $\mathbb{R}^{50}$, assuming that A can be partitioned into a 2×2 block matrix $[A_{ij}]$, with A_{11} an invertible 20×20 matrix, A_{22} an invertible 30×30 matrix, and A_{12} a zero matrix. [*Hint:* Describe appropriate smaller systems to solve, without using any matrix inverses.]

Solutions to Practice Problems

1. If $\begin{bmatrix} I & 0 \\ A & I \end{bmatrix}$ is invertible, its inverse has the form $\begin{bmatrix} W & X \\ Y & Z \end{bmatrix}$. Verify that

$$\begin{bmatrix} I & 0 \\ A & I \end{bmatrix}\begin{bmatrix} W & X \\ Y & Z \end{bmatrix} = \begin{bmatrix} W & X \\ AW + Y & AX + Z \end{bmatrix}$$

So W, X, Y, and Z must satisfy $W = I$, $X = 0$, $AW + Y = 0$, and $AX + Z = I$. It follows that $Y = -A$ and $Z = I$. Hence

$$\begin{bmatrix} I & 0 \\ A & I \end{bmatrix}\begin{bmatrix} I & 0 \\ -A & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

The product in the reverse order is also the identity, so the block matrix is invertible, and its inverse is $\begin{bmatrix} I & 0 \\ -A & I \end{bmatrix}$. (You could also appeal to the Invertible Matrix Theorem.)

2. $X^T X = \begin{bmatrix} X_1^T \\ X_2^T \end{bmatrix}\begin{bmatrix} X_1 & X_2 \end{bmatrix} = \begin{bmatrix} X_1^T X_1 & X_1^T X_2 \\ X_2^T X_1 & X_2^T X_2 \end{bmatrix}$. The partitions of X^T and X are automatically conformable for block multiplication because the columns of X^T are the rows of X. This partition of $X^T X$ is used in several computer algorithms for matrix computations.

2.5 Matrix Factorizations

A *factorization* of a matrix A is an equation that expresses A as a product of two or more matrices. Whereas matrix multiplication involves a *synthesis* of data (combining the effects of two or more linear transformations into a single matrix), matrix factorization is an *analysis* of data. In the language of computer science, the expression of A as a product amounts to a *preprocessing* of the data in A, organizing that data into two or more parts whose structures are more useful in some way, perhaps more accessible for computation.

Matrix factorizations and, later, factorizations of linear transformations will appear at a number of key points throughout the text. This section focuses on a factorization that lies at the heart of several important computer programs widely used in applications, such as the airflow problem described in the chapter introduction. Several other factorizations, to be studied later, are introduced in the exercises.

The LU Factorization

The LU factorization, described below, is motivated by the fairly common industrial and business problem of solving a sequence of equations, all with the same coefficient matrix:

$$A\mathbf{x} = \mathbf{b}_1, \quad A\mathbf{x} = \mathbf{b}_2, \quad \ldots, \quad A\mathbf{x} = \mathbf{b}_p \tag{1}$$

See Exercise 32, for example. Also see Section 5.8, where the inverse power method is used to estimate eigenvalues of a matrix by solving equations like those in sequence (1), one at a time.

When A is invertible, one could compute A^{-1} and then compute $A^{-1}\mathbf{b}_1$, $A^{-1}\mathbf{b}_2$, and so on. However, it is more efficient to solve the first equation in sequence (1) by row reduction and obtain an LU factorization of A at the same time. Thereafter, the remaining equations in sequence (1) are solved with the LU factorization.

At first, assume that A is an $m \times n$ matrix that can be row reduced to echelon form, *without row interchanges*. (Later, we will treat the general case.) Then A can be written in the form $A = LU$, where L is an $m \times m$ lower triangular matrix with 1's on the diagonal and U is an $m \times n$ echelon form of A. For instance, see Figure 1. Such a factorization is called an **LU factorization** of A. The matrix L is invertible and is called a *unit* lower triangular matrix.

$$A = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ * & 1 & 0 & 0 \\ * & * & 1 & 0 \\ * & * & * & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} \blacksquare & * & * & * & * \\ 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & \blacksquare & * \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{U}$$

FIGURE 1 An LU factorization.

Before studying how to construct L and U, we should look at why they are so useful. When $A = LU$, the equation $A\mathbf{x} = \mathbf{b}$ can be written as $L(U\mathbf{x}) = \mathbf{b}$. Writing $\mathbf{y}$ for $U\mathbf{x}$, we can find $\mathbf{x}$ by solving the *pair* of equations

$$\boxed{\begin{aligned} L\mathbf{y} &= \mathbf{b} \\ U\mathbf{x} &= \mathbf{y} \end{aligned}} \tag{2}$$

First solve $L\mathbf{y} = \mathbf{b}$ for $\mathbf{y}$, and then solve $U\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$. See Figure 2. Each equation is easy to solve because L and U are triangular.

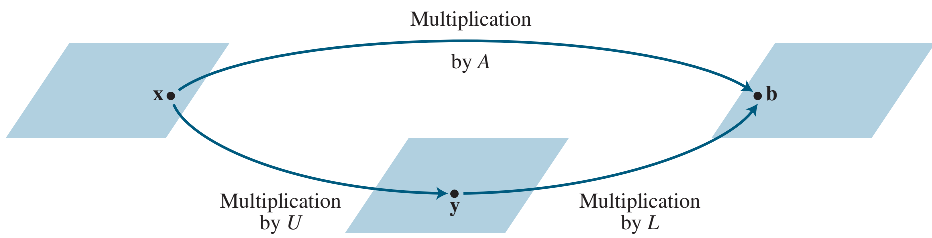


FIGURE 2 Factorization of the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

EXAMPLE 1 It can be verified that

$$A = \begin{bmatrix} 3 & -7 & -2 & 2 \\ -3 & 5 & 1 & 0 \\ 6 & -4 & 0 & -5 \\ -9 & 5 & -5 & 12 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 2 & -5 & 1 & 0 \\ -3 & 8 & 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & -7 & -2 & 2 \\ 0 & -2 & -1 & 2 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix} = LU$$

Use this LU factorization of A to solve $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b} = \begin{bmatrix} -9 \\ 5 \\ 7 \\ 11 \end{bmatrix}$.

SOLUTION The solution of $L\mathbf{y} = \mathbf{b}$ needs only 6 multiplications and 6 additions, because the arithmetic takes place only in column 5. (The zeros below each pivot in L are created automatically by the choice of row operations.)

$$\begin{bmatrix} L & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & -9 \\ -1 & 1 & 0 & 0 & 5 \\ 2 & -5 & 1 & 0 & 7 \\ -3 & 8 & 3 & 1 & 11 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 & -9 \\ 0 & 1 & 0 & 0 & -4 \\ 0 & 0 & 1 & 0 & 5 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} I & \mathbf{y} \end{bmatrix}$$

Then, for $U\mathbf{x} = \mathbf{y}$, the "backward" phase of row reduction requires 4 divisions, 6 multiplications, and 6 additions. (For instance, creating the zeros in column 4 of $\begin{bmatrix} U & \mathbf{y} \end{bmatrix}$ requires 1 division in row 4 and 3 multiplication–addition pairs to add multiples of row 4 to the rows above.)

$$\begin{bmatrix} U & \mathbf{y} \end{bmatrix} = \begin{bmatrix} 3 & -7 & -2 & 2 & -9 \\ 0 & -2 & -1 & 2 & -4 \\ 0 & 0 & -1 & 1 & 5 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 & -6 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 3 \\ 4 \\ -6 \\ -1 \end{bmatrix}$$

To find $\mathbf{x}$ requires 28 arithmetic operations, or "flops" (floating point operations), excluding the cost of finding L and U. In contrast, row reduction of $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$ to $\begin{bmatrix} I & \mathbf{x} \end{bmatrix}$ takes 62 operations. ■

The computational efficiency of the LU factorization depends on knowing L and U. The next algorithm shows that the row reduction of A to an echelon form U amounts to an LU factorization because it produces L with essentially no extra work. After the first row reduction, L and U are available for solving additional equations whose coefficient matrix is A.

An LU Factorization Algorithm

Suppose A can be reduced to an echelon form U using only row replacements that add a multiple of one row to another row *below it*. In this case, there exist unit lower triangular elementary matrices $E_1, \ldots, E_p$ such that

$$E_p \cdots E_1 A = U \tag{3}$$

Then

$$A = (E_p \cdots E_1)^{-1} U = LU$$

where

$$L = (E_p \cdots E_1)^{-1} \tag{4}$$

It can be shown that products and inverses of unit lower triangular matrices are also unit lower triangular. (For instance, see Exercise 19.) Thus L is unit lower triangular.

Note that the row operations in equation (3), which reduce A to U, also reduce the L in equation (4) to I, because $E_p \cdots E_1 L = (E_p \cdots E_1)(E_p \cdots E_1)^{-1} = I$. This observation is the key to *constructing* L.

ALGORITHM FOR AN LU FACTORIZATION

1. Reduce A to an echelon form U by a sequence of row replacement operations, if possible.
2. Place entries in L such that the *same sequence of row operations* reduces L to I.

Step 1 is not always possible, but when it is, the argument above shows that an LU factorization exists. Example 2 will show how to implement step 2. By construction, L will satisfy

$$(E_p \cdots E_1)L = I$$

using the same $E_1, \dots, E_p$ as in equation (3). Thus L will be invertible, by the Invertible Matrix Theorem, with $(E_p \cdots E_1) = L^{-1}$. From (3), $L^{-1}A = U$, and $A = LU$. So step 2 will produce an acceptable L.

EXAMPLE 2 Find an LU factorization of

$$A = \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ -4 & -5 & 3 & -8 & 1 \\ 2 & -5 & -4 & 1 & 8 \\ -6 & 0 & 7 & -3 & 1 \end{bmatrix}$$

SOLUTION Since A has four rows, L should be 4×4. The first column of L is the first column of A divided by the top pivot entry:

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 1 & & 1 & 0 \\ -3 & & & 1 \end{bmatrix}$$

Compare the first columns of A and L. *The row operations that create zeros in the first column of A will also create zeros in the first column of L.* To make this same correspondence of row operations on A hold for the rest of L, watch a row reduction of A to an echelon form U. That is, *highlight the entries* in each matrix that are used to determine the sequence of row operations that transform A into U. [See the highlighted entries in equation (5).]

$$A = \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ -4 & -5 & 3 & -8 & 1 \\ 2 & -5 & -4 & 1 & 8 \\ -6 & 0 & 7 & -3 & 1 \end{bmatrix} \sim \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & 3 & 1 & 2 & -3 \\ 0 & -9 & -3 & -4 & 10 \\ 0 & 12 & 4 & 12 & -5 \end{bmatrix} = A_1 \tag{5}$$

$$\sim A_2 = \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & 3 & 1 & 2 & -3 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 4 & 7 \end{bmatrix} \sim \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & 3 & 1 & 2 & -3 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix} = U$$

These highlighted entries determine the row reduction of A to U. At each pivot column, divide the highlighted entries by the pivot and place the result into L:

$$\begin{bmatrix} 2 \\ -4 \\ 2 \\ -6 \end{bmatrix} \begin{bmatrix} 3 \\ -9 \\ 12 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} \begin{bmatrix} 5 \end{bmatrix}$$

$$\div 2 \quad \div 3 \quad \div 2 \quad \div 5$$

$$\downarrow \quad \downarrow \quad \downarrow \quad \downarrow$$

$$\begin{bmatrix} 1 & & & \\ -2 & 1 & & \\ 1 & -3 & 1 & \\ -3 & 4 & 2 & 1 \end{bmatrix}, \quad \text{and} \quad L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 1 & -3 & 1 & 0 \\ -3 & 4 & 2 & 1 \end{bmatrix}$$

An easy calculation verifies that this L and U satisfy $LU = A$. ■

In practical work, row interchanges are nearly always needed, because partial pivoting is used for high accuracy. (Recall that this procedure selects, among the possible choices for a pivot, an entry in the column having the largest absolute value.) To handle row interchanges, the LU factorization above can be modified easily to produce an L that is *permuted lower triangular*, in the sense that a rearrangement (called a permutation) of the rows of L can make L (unit) lower triangular. The resulting *permuted LU factorization* solves $A\mathbf{x} = \mathbf{b}$ in the same way as before, except that the reduction of $[\,L \quad \mathbf{b}\,]$ to $[\,I \quad \mathbf{y}\,]$ follows the order of the pivots in L from left to right, starting with the pivot in the first column. A reference to an "LU factorization" usually includes the possibility that L might be permuted lower triangular. For details, see the *Study Guide*.

STUDY GUIDE offers information about permuted LU factorizations.

Numerical Notes

The following operation counts apply to an $n \times n$ dense matrix A (with most entries nonzero) for n moderately large, say, $n \geq 30$.[1]

1. Computing an LU factorization of A takes about $2n^3/3$ flops (about the same as row reducing $[\,A \quad \mathbf{b}\,]$), whereas finding A^{-1} requires about $2n^3$ flops.
2. Solving $L\mathbf{y} = \mathbf{b}$ and $U\mathbf{x} = \mathbf{y}$ requires about $2n^2$ flops, because any $n \times n$ triangular system can be solved in about n^2 flops.
3. Multiplication of $\mathbf{b}$ by A^{-1} also requires about $2n^2$ flops, but the result may not be as accurate as that obtained from L and U (because of roundoff error when computing both A^{-1} and $A^{-1}\mathbf{b}$).
4. If A is sparse (with mostly zero entries), then L and U may be sparse, too, whereas A^{-1} is likely to be dense. In this case, a solution of $A\mathbf{x} = \mathbf{b}$ with an LU factorization is *much* faster than using A^{-1}. See Exercise 31.

A Matrix Factorization in Electrical Engineering

Matrix factorization is intimately related to the problem of constructing an electrical network with specified properties. The following discussion gives just a glimpse of the connection between factorization and circuit design.

[1] See Section 3.8 in *Applied Linear Algebra*, 3rd ed., by Ben Noble and James W. Daniel (Englewood Cliffs, NJ: Prentice-Hall, 1988). Recall that for our purposes, a *flop* is $+$, $-$, $\times$, or $\div$.

Suppose the box in Figure 3 represents some sort of electric circuit, with an input and output. Record the input voltage and current by $\begin{bmatrix} v_1 \\ i_1 \end{bmatrix}$ (with voltage v in volts and current i in amps), and record the output voltage and current by $\begin{bmatrix} v_2 \\ i_2 \end{bmatrix}$. Frequently, the transformation $\begin{bmatrix} v_1 \\ i_1 \end{bmatrix} \mapsto \begin{bmatrix} v_2 \\ i_2 \end{bmatrix}$ is linear. That is, there is a matrix A, called the *transfer matrix*, such that

$$\begin{bmatrix} v_2 \\ i_2 \end{bmatrix} = A \begin{bmatrix} v_1 \\ i_1 \end{bmatrix}$$

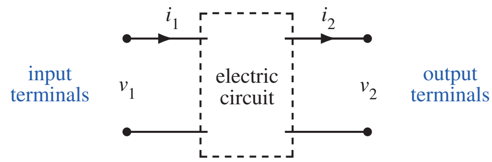


FIGURE 3 A circuit with input and output terminals.

Figure 4 shows a *ladder network*, where two circuits (there could be more) are connected in series, so that the output of one circuit becomes the input of the next circuit. The left circuit in Figure 4 is called a *series circuit*, with resistance R_1 (in ohms).

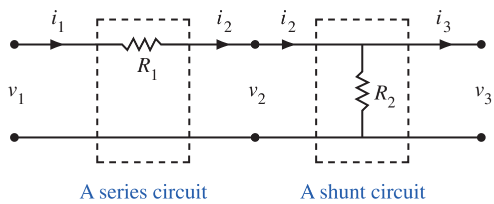


FIGURE 4 A ladder network.

The right circuit in Figure 4 is a *shunt circuit*, with resistance R_2. Using Ohm's law and Kirchhoff's laws, one can show that the transfer matrices of the series and shunt circuits, respectively, are

$$\underset{\text{Transfer matrix of series circuit}}{\begin{bmatrix} 1 & -R_1 \\ 0 & 1 \end{bmatrix}} \quad \text{and} \quad \underset{\text{Transfer matrix of shunt circuit}}{\begin{bmatrix} 1 & 0 \\ -1/R_2 & 1 \end{bmatrix}}$$

EXAMPLE 3

a. Compute the transfer matrix of the ladder network in Figure 4.

b. Design a ladder network whose transfer matrix is $\begin{bmatrix} 1 & -8 \\ -.5 & 5 \end{bmatrix}$.

SOLUTION

a. Let A_1 and A_2 be the transfer matrices of the series and shunt circuits, respectively. Then an input vector $\mathbf{x}$ is transformed first into $A_1\mathbf{x}$ and then into $A_2(A_1\mathbf{x})$. The series connection of the circuits corresponds to composition of linear transformations, and the transfer matrix of the ladder network is (note the order)

$$A_2A_1 = \begin{bmatrix} 1 & 0 \\ -1/R_2 & 1 \end{bmatrix} \begin{bmatrix} 1 & -R_1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -R_1 \\ -1/R_2 & 1 + R_1/R_2 \end{bmatrix} \tag{6}$$

b. To factor the matrix $\begin{bmatrix} 1 & -8 \\ -.5 & 5 \end{bmatrix}$ into the product of transfer matrices, as in equation (6), look for R_1 and R_2 in Figure 4 to satisfy

$$\begin{bmatrix} 1 & -R_1 \\ -1/R_2 & 1 + R_1/R_2 \end{bmatrix} = \begin{bmatrix} 1 & -8 \\ -.5 & 5 \end{bmatrix}$$

From the $(1, 2)$-entries, $R_1 = 8$ ohms, and from the $(2, 1)$-entries, $1/R_2 = .5$ ohm and $R_2 = 1/.5 = 2$ ohms. With these values, the network in Figure 4 has the desired transfer matrix. ■

A network transfer matrix summarizes the input–output behavior (the design specifications) of the network without reference to the interior circuits. To physically build a network with specified properties, an engineer first determines if such a network can be constructed (or *realized*). Then the engineer tries to factor the transfer matrix into matrices corresponding to smaller circuits that perhaps are already manufactured and ready for assembly. In the common case of alternating current, the entries in the transfer matrix are usually rational complex-valued functions. (See Exercises 21 and 22 in Section 2.4.) A standard problem is to find a *minimal realization* that uses the smallest number of electrical components.

Practice Problem

Find an LU factorization of $A = \begin{bmatrix} 2 & -4 & -2 & 3 \\ 6 & -9 & -5 & 8 \\ 2 & -7 & -3 & 9 \\ 4 & -2 & -2 & -1 \\ -6 & 3 & 3 & 4 \end{bmatrix}$. [*Note:* It will turn out that A has only three pivot columns, so the method of Example 2 will produce only the first three columns of L. The remaining two columns of L come from I_5.]

2.5 Exercises

In Exercises 1–6, solve the equation $A\mathbf{x} = \mathbf{b}$ by using the LU factorization given for A. In Exercises 1 and 2, also solve $A\mathbf{x} = \mathbf{b}$ by ordinary row reduction.

1. $A = \begin{bmatrix} 3 & -7 & -2 \\ -3 & 5 & 1 \\ 6 & -4 & 0 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -7 \\ 5 \\ 2 \end{bmatrix}$

$A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & -5 & 1 \end{bmatrix} \begin{bmatrix} 3 & -7 & -2 \\ 0 & -2 & -1 \\ 0 & 0 & -1 \end{bmatrix}$

2. $A = \begin{bmatrix} 4 & 3 & -5 \\ -4 & -5 & 7 \\ 8 & 6 & -8 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ -4 \\ 6 \end{bmatrix}$

$A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & 3 & -5 \\ 0 & -2 & 2 \\ 0 & 0 & 2 \end{bmatrix}$

3. $A = \begin{bmatrix} 2 & -1 & 2 \\ -6 & 0 & -2 \\ 8 & -1 & 5 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}$

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

4. $A = \begin{bmatrix} 2 & -2 & 4 \\ 1 & -3 & 1 \\ 3 & 7 & 5 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 0 \\ -5 \\ 7 \end{bmatrix}$

$A = \begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ 3/2 & -5 & 1 \end{bmatrix} \begin{bmatrix} 2 & -2 & 4 \\ 0 & -2 & -1 \\ 0 & 0 & -6 \end{bmatrix}$

5. $A = \begin{bmatrix} 1 & -2 & -4 & -3 \\ 2 & -7 & -7 & -6 \\ -1 & 2 & 6 & 4 \\ -4 & -1 & 9 & 8 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ 7 \\ 0 \\ 3 \end{bmatrix}$

$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -4 & 3 & -5 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & -4 & -3 \\ 0 & -3 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

6. $A = \begin{bmatrix} 1 & 3 & 4 & 0 \\ -3 & -6 & -7 & 2 \\ 3 & 3 & 0 & -4 \\ -5 & -3 & 2 & 9 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ -2 \\ -1 \\ 2 \end{bmatrix}$

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ 3 & -2 & 1 & 0 \\ -5 & 4 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 4 & 0 \\ 0 & 3 & 5 & 2 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Find an LU factorization of the matrices in Exercises 7–16 (with L unit lower triangular). Note that MATLAB will usually produce a permuted LU factorization because it uses partial pivoting for numerical accuracy.

7. $\begin{bmatrix} 2 & 5 \\ -3 & -4 \end{bmatrix}$

8. $\begin{bmatrix} 6 & 9 \\ 4 & 5 \end{bmatrix}$

9. $\begin{bmatrix} 3 & -1 & 2 \\ -3 & -2 & 10 \\ 9 & -5 & 6 \end{bmatrix}$

10. $\begin{bmatrix} -5 & 3 & 4 \\ 10 & -8 & -9 \\ 15 & 1 & 2 \end{bmatrix}$

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

12. $\begin{bmatrix} 2 & -4 & 2 \\ 1 & 5 & -4 \\ -6 & -2 & 4 \end{bmatrix}$

13. $\begin{bmatrix} 1 & 3 & -5 & -3 \\ -1 & -5 & 8 & 4 \\ 4 & 2 & -5 & -7 \\ -2 & -4 & 7 & 5 \end{bmatrix}$

14. $\begin{bmatrix} 1 & 4 & -1 & 5 \\ 3 & 7 & -2 & 9 \\ -2 & -3 & 1 & -4 \\ -1 & 6 & -1 & 7 \end{bmatrix}$

15. $\begin{bmatrix} 2 & -4 & 4 & -2 \\ 6 & -9 & 7 & -3 \\ -1 & -4 & 8 & 0 \end{bmatrix}$

16. $\begin{bmatrix} 2 & -6 & 6 \\ -4 & 5 & -7 \\ 3 & 5 & -1 \\ -6 & 4 & -8 \\ 8 & -3 & 9 \end{bmatrix}$

17. When A is invertible, MATLAB finds A^{-1} by factoring $A = LU$ (where L may be permuted lower triangular), inverting L and U, and then computing $U^{-1}L^{-1}$. Use this method to compute the inverse of A in Exercise 2. (Apply the algorithm of Section 2.2 to L and to U.)

18. Find A^{-1} as in Exercise 17, using A from Exercise 3.

19. Let A be a lower triangular $n \times n$ matrix with nonzero entries on the diagonal. Show that A is invertible and A^{-1} is lower triangular. [*Hint:* Explain why A can be changed into I using only row replacements and scaling. (Where are the pivots?) Also, explain why the row operations that reduce A to I change I into a lower triangular matrix.]

20. Let $A = LU$ be an LU factorization. Explain why A can be row reduced to U using only replacement operations. (This fact is the converse of what was proved in the text.)

21. Suppose $A = BC$, where B is invertible. Show that any sequence of row operations that reduces B to I also reduces A to C. The converse is not true, since the zero matrix may be factored as $0 = B(0)$.

Exercises 22–26 provide a glimpse of some widely used matrix factorizations, some of which are discussed later in the text.

22. (*Reduced LU Factorization*) With A as in the Practice Problem, find a 5×3 matrix B and a 3×4 matrix C such that $A = BC$. Generalize this idea to the case where A is $m \times n$, $A = LU$, and U has only three nonzero rows.

23. (*Rank Factorization*) Suppose an $m \times n$ matrix A admits a factorization $A = CD$ where C is $m \times 4$ and D is $4 \times n$.

a. Show that A is the sum of four outer products. (See Section 2.4.)

b. Let $m = 400$ and $n = 100$. Explain why a computer programmer might prefer to store the data from A in the form of two matrices C and D.

24. (*QR Factorization*) Suppose $A = QR$, where Q and R are $n \times n$, R is invertible and upper triangular, and Q has the property that $Q^TQ = I$. Show that for each $\mathbf{b}$ in $\mathbb{R}^n$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique solution. What computations with Q and R will produce the solution?

25. (*Singular Value Decomposition*) Suppose $A = UDV^T$, where U and V are $n \times n$ matrices with the property that $U^TU = I$ and $V^TV = I$, and where D is a diagonal matrix with positive numbers $\sigma_1, \ldots, \sigma_n$ on the diagonal. Show that A is invertible, and find a formula for A^{-1}.

26. (*Spectral Factorization*) Suppose a 3×3 matrix A admits a factorization as $A = PDP^{-1}$, where P is some invertible 3×3 matrix and D is the diagonal matrix

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}$$

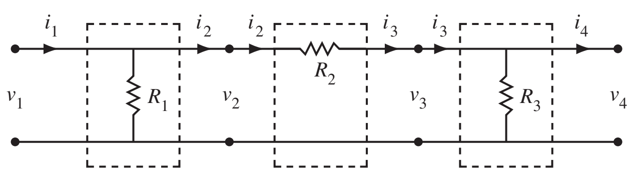
Show that this factorization is useful when computing high powers of A. Find fairly simple formulas for A^2, A^3, and A^k (k a positive integer), using P and the entries in D.

27. Design two different ladder networks that each output 9 volts and 4 amps when the input is 12 volts and 6 amps.

28. Show that if three shunt circuits (with resistances R_1, R_2, R_3) are connected in series, the resulting network has the same transfer matrix as a single shunt circuit. Find a formula for the resistance in that circuit.

29. a. Compute the transfer matrix of the network in the figure.

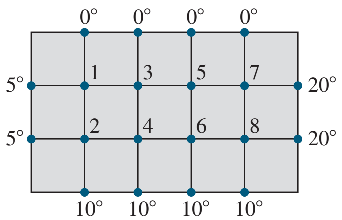
b. Let $A = \begin{bmatrix} 4/3 & -12 \\ -1/4 & 3 \end{bmatrix}$. Design a ladder network whose transfer matrix is A by finding a suitable matrix factorization of A.



30. Find a different factorization of the A in Exercise 29, and thereby design a different ladder network whose transfer matrix is A.

T 31. The solution to the steady-state heat flow problem for the plate in the figure is approximated by the solution to the equation $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b} = (5, 15, 0, 10, 0, 10, 20, 30)$ and

$$A = \begin{bmatrix} 4 & -1 & -1 & & & & & \\ -1 & 4 & 0 & -1 & & & & \\ -1 & 0 & 4 & -1 & -1 & & & \\ & -1 & -1 & 4 & 0 & -1 & & \\ & & -1 & 0 & 4 & -1 & -1 & \\ & & & -1 & -1 & 4 & 0 & -1 \\ & & & & -1 & 0 & 4 & -1 \\ & & & & & -1 & -1 & 4 \end{bmatrix}$$



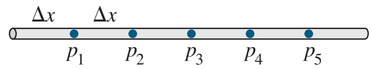
(Refer to Exercise 43 of Section 1.1.) The missing entries in A are zeros. The nonzero entries of A lie within a band along the main diagonal. Such *band matrices* occur in a variety of applications and often are extremely large (with thousands of rows and columns but relatively narrow bands).

a. Use the method of Example 2 to construct an LU factorization of A, and note that both factors are band matrices (with two nonzero diagonals below or above the main diagonal). Compute $LU - A$ to check your work.

b. Use the LU factorization to solve $A\mathbf{x} = \mathbf{b}$.

c. Obtain A^{-1} and note that A^{-1} is a dense matrix with no band structure. When A is large, L and U can be stored in much less space than A^{-1}. This fact is another reason for preferring the LU factorization of A to A^{-1} itself.

T 32. The band matrix A shown below can be used to estimate the unsteady conduction of heat in a rod when the temperatures at points $p_1, \ldots, p_5$ on the rod change with time.[2]



The constant C in the matrix depends on the physical nature of the rod, the distance Δx between the points on the rod, and the length of time Δt between successive temperature measurements. Suppose that for $k = 0, 1, 2, \ldots$, a vector $\mathbf{t}_k$ in $\mathbb{R}^5$ lists the temperatures at time $k\Delta t$. If the two ends of the rod are maintained at $0°$, then the temperature vectors satisfy the equation $A\mathbf{t}_{k+1} = \mathbf{t}_k$ $(k = 0, 1, \ldots)$, where

$$A = \begin{bmatrix} (1+2C) & -C & & & \\ -C & (1+2C) & -C & & \\ & -C & (1+2C) & -C & \\ & & -C & (1+2C) & -C \\ & & & -C & (1+2C) \end{bmatrix}$$

a. Find the LU factorization of A when $C = 1$. A matrix such as A with three nonzero diagonals is called a *tridiagonal matrix*. The L and U factors are *bidiagonal matrices*.

b. Suppose $C = 1$ and $\mathbf{t}_0 = (10, 12, 12, 12, 10)$. Use the LU factorization of A to find the temperature distributions $\mathbf{t}_1$, $\mathbf{t}_2$, $\mathbf{t}_3$, and $\mathbf{t}_4$.

[2] See Biswa N. Datta, *Numerical Linear Algebra and Applications* (Pacific Grove, CA: Brooks/Cole, 1994), pp. 200–201.

Solution to Practice Problem

$$A = \begin{bmatrix} 2 & -4 & -2 & 3 \\ 6 & -9 & -5 & 8 \\ 2 & -7 & -3 & 9 \\ 4 & -2 & -2 & -1 \\ -6 & 3 & 3 & 4 \end{bmatrix} \sim \begin{bmatrix} 2 & -4 & -2 & 3 \\ 0 & 3 & 1 & -1 \\ 0 & -3 & -1 & 6 \\ 0 & 6 & 2 & -7 \\ 0 & -9 & -3 & 13 \end{bmatrix}$$

$$\sim \begin{bmatrix} 2 & -4 & -2 & 3 \\ 0 & 3 & 1 & -1 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & -5 \\ 0 & 0 & 0 & 10 \end{bmatrix} \sim \begin{bmatrix} 2 & -4 & -2 & 3 \\ 0 & 3 & 1 & -1 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = U$$

Divide the entries in each highlighted column by the pivot at the top. The resulting columns form the first three columns in the lower half of L. This suffices to make row reduction of L to I correspond to reduction of A to U. Use the last two columns of I_5

to make L unit lower triangular.

$$\begin{array}{ccc} \begin{bmatrix} 2 \\ 6 \\ 2 \\ 4 \\ -6 \end{bmatrix} & \begin{bmatrix} 3 \\ -3 \\ 6 \\ -9 \end{bmatrix} & \begin{bmatrix} 5 \\ -5 \\ 10 \end{bmatrix} \\ \div 2 & \div 3 & \div 5 \\ \downarrow & \downarrow & \downarrow \end{array}$$

$$\begin{bmatrix} 1 & & & & \\ 3 & 1 & & & \\ 1 & -1 & 1 & \cdots & \\ 2 & 2 & -1 & & \\ -3 & -3 & 2 & & \end{bmatrix}, \quad L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 \\ 2 & 2 & -1 & 1 & 0 \\ -3 & -3 & 2 & 0 & 1 \end{bmatrix}$$

2.6 The Leontief Input–Output Model

Linear algebra played an essential role in the Nobel prize–winning work of Wassily Leontief, as mentioned at the beginning of Chapter 1. The economic model described in this section is the basis for more elaborate models used in many parts of the world.

Suppose a nation's economy is divided into n sectors that produce goods or services, and let $\mathbf{x}$ be a **production vector** in $\mathbb{R}^n$ that lists the output of each sector for one year. Also, suppose another part of the economy (called the *open sector*) does not produce goods or services but only consumes them, and let $\mathbf{d}$ be a **final demand vector** (or **bill of final demands**) that lists the values of the goods and services demanded from the various sectors by the nonproductive part of the economy. The vector $\mathbf{d}$ can represent consumer demand, government consumption, surplus production, exports, or other external demands.

As the various sectors produce goods to meet consumer demand, the producers themselves create additional **intermediate demand** for goods they need as inputs for their own production. The interrelations between the sectors are very complex, and the connection between the final demand and the production is unclear. Leontief asked if there is a production level $\mathbf{x}$ such that the amounts produced (or "supplied") will exactly balance the total demand for that production, so that

$$\left\{\begin{matrix} \text{amount} \\ \text{produced} \\ \mathbf{x} \end{matrix}\right\} = \left\{\begin{matrix} \text{intermediate} \\ \text{demand} \end{matrix}\right\} + \left\{\begin{matrix} \text{final} \\ \text{demand} \\ \mathbf{d} \end{matrix}\right\} \tag{1}$$

The basic assumption of Leontief's input–output model is that for each sector, there is a **unit consumption vector** in $\mathbb{R}^n$ that lists the inputs needed *per unit of output* of the sector. All input and output units are measured in millions of dollars, rather than in quantities such as tons or bushels. (Prices of goods and services are held constant.)

As a simple example, suppose the economy consists of three sectors—manufacturing, agriculture, and services—with unit consumption vectors $\mathbf{c}_1$, $\mathbf{c}_2$, and $\mathbf{c}_3$, as shown in the table that follows.

If $D = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 \end{bmatrix}$, then the determinant of $D^T D = 256$, and hence this is a better design. Notice that the first weighing of this design is the same as the previous one, but then the remaining weighings each have two objects in each pan.

Calculating determinants of matrices and understanding their properties is the theme of this chapter. As you learn more about determinants, you may also come up with strategies for good and bad choices for a weighing design.

Another important use of the determinant is to calculate the area of a parallelogram or the volume of a parallelepiped. In Section 1.9, we saw that matrix multiplication can be used to change the shape of a box or other object. The determinant of the matrix used determines how much the area changes when it is multiplied by a matrix, just as a fish story can transform the size of the fish caught.

Indeed, the determinant has so many uses that a summary of the applications known in the early 1900s filled a four-volume treatise by Thomas Muir. With changes in emphasis and the greatly increased sizes of the matrices used in modem applications, many uses that were important then are no longer critical today. Nevertheless, the determinant still plays many important theoretical and practical roles.

Beyond introducing the determinant in Section 3.1, this chapter presents two important ideas. Section 3.2 derives an invertibility criterion for a square matrix that plays a pivotal role in Chapter 5. Section 3.3 shows how the determinant measures the amount by which a linear transformation changes the area of a figure. When applied locally, this technique answers the question of a map's expansion rate near the poles. This idea plays a critical role in multivariable calculus in the form of the Jacobian.

3.1 Introduction to Determinants

Recall from Section 2.2 that a 2×2 matrix is invertible if and only if its determinant is nonzero. To extend this useful fact to larger matrices, we need a definition for the determinant of an $n \times n$ matrix. We can discover the definition for the 3×3 case by watching what happens when an invertible 3×3 matrix A is row reduced.

Consider $A = [a_{ij}]$ with $a_{11} \neq 0$. If we multiply the second and third rows of A by a_{11} and then subtract appropriate multiples of the first row from the other two rows, we find that A is row equivalent to the following two matrices:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{11}a_{21} & a_{11}a_{22} & a_{11}a_{23} \\ a_{11}a_{31} & a_{11}a_{32} & a_{11}a_{33} \end{bmatrix} \sim \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{11}a_{22} - a_{12}a_{21} & a_{11}a_{23} - a_{13}a_{21} \\ 0 & a_{11}a_{32} - a_{12}a_{31} & a_{11}a_{33} - a_{13}a_{31} \end{bmatrix} \quad (1)$$

Since A is invertible, either the $(2,2)$-entry or the $(3,2)$-entry on the right in (1) is nonzero. Let us suppose that the $(2,2)$-entry is nonzero. (Otherwise, we can make a row interchange before proceeding.) Multiply row 3 by $a_{11}a_{22} - a_{12}a_{21}$, and then to the new row 3 add $-(a_{11}a_{32} - a_{12}a_{31})$ times row 2. This will show that

$$A \sim \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{11}a_{22} - a_{12}a_{21} & a_{11}a_{23} - a_{13}a_{21} \\ 0 & 0 & a_{11}\Delta \end{bmatrix}$$

where

$$\Delta = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \quad (2)$$

Since A is invertible, Δ must be nonzero. The converse is true, too, as we will see in Section 3.2. We call Δ in (2) the **determinant** of the 3×3 matrix A.

Recall that the determinant of a 2×2 matrix, $A = [a_{ij}]$, is the number

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

For a 1×1 matrix—say, $A = [a_{11}]$—we define $\det A = a_{11}$. To generalize the definition of the determinant to larger matrices, we'll use 2×2 determinants to rewrite the 3×3 determinant Δ described above. Since the terms in Δ can be grouped as $(a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32}) - (a_{12}a_{21}a_{33} - a_{12}a_{23}a_{31}) + (a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31})$,

$$\Delta = a_{11} \det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{12} \det \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} + a_{13} \det \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

For brevity, write

$$\Delta = a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13} \quad (3)$$

where A_{11}, A_{12}, and A_{13} are obtained from A by deleting the first row and one of the three columns. For any square matrix A, let A_{ij} denote the submatrix formed by deleting the ith row and jth column of A. For instance, if

$$A = \begin{bmatrix} 1 & -2 & 5 & 0 \\ 2 & 0 & 4 & -1 \\ 3 & 1 & 0 & 7 \\ 0 & 4 & -2 & 0 \end{bmatrix}$$

then A_{32} is obtained by crossing out row 3 and column 2,

$$\begin{bmatrix} 1 & -2 & 5 & 0 \\ 2 & 0 & 4 & -1 \\ 3 & 1 & 0 & 7 \\ 0 & 4 & -2 & 0 \end{bmatrix}$$

so that

$$A_{32} = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}$$

We can now give a *recursive* definition of a determinant. When $n = 3$, $\det A$ is defined using determinants of the 2×2 submatrices A_{1j}, as in (3) above. When $n = 4$, $\det A$ uses determinants of the 3×3 submatrices A_{1j}. In general, an $n \times n$ determinant is defined by determinants of $(n-1) \times (n-1)$ submatrices.

DEFINITION

For $n \geq 2$, the **determinant** of an $n \times n$ matrix $A = [a_{ij}]$ is the sum of n terms of the form $\pm a_{1j} \det A_{1j}$, with plus and minus signs alternating, where the entries $a_{11}, a_{12}, \ldots, a_{1n}$ are from the first row of A. In symbols,

$$\begin{aligned} \det A &= a_{11} \det A_{11} - a_{12} \det A_{12} + \cdots + (-1)^{1+n} a_{1n} \det A_{1n} \\ &= \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{1j} \end{aligned}$$

EXAMPLE 1 Compute the determinant of

$$A = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}$$

SOLUTION Compute $\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13}$:

$$\begin{aligned} \det A &= 1 \det \begin{bmatrix} 4 & -1 \\ -2 & 0 \end{bmatrix} - 5 \det \begin{bmatrix} 2 & -1 \\ 0 & 0 \end{bmatrix} + 0 \det \begin{bmatrix} 2 & 4 \\ 0 & -2 \end{bmatrix} \\ &= 1(0-2) - 5(0-0) + 0(-4-0) = -2 \end{aligned}$$

■

Another common notation for the determinant of a matrix uses a pair of vertical lines in place of brackets. Thus the calculation in Example 1 can be written as

$$\det A = 1 \begin{vmatrix} 4 & -1 \\ -2 & 0 \end{vmatrix} - 5 \begin{vmatrix} 2 & -1 \\ 0 & 0 \end{vmatrix} + 0 \begin{vmatrix} 2 & 4 \\ 0 & -2 \end{vmatrix} = \cdots = -2$$

To state the next theorem, it is convenient to write the definition of $\det A$ in a slightly different form. Given $A = [a_{ij}]$, the **(i, j)-cofactor** of A is the number C_{ij} given by

$$C_{ij} = (-1)^{i+j} \det A_{ij} \tag{4}$$

Then

$$\det A = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}$$

This formula is called a **cofactor expansion across the first row** of A. We omit the proof of the following fundamental theorem to avoid a lengthy digression.

THEOREM 1

The determinant of an $n \times n$ matrix A can be computed by a cofactor expansion across any row or down any column. The expansion across the ith row using the cofactors in (4) is

$$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$$

The cofactor expansion down the jth column is

$$\det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}$$

The plus or minus sign in the (i, j)-cofactor depends on the position of a_{ij} in the matrix, regardless of the sign of a_{ij} itself. The factor $(-1)^{i+j}$ determines the following checkerboard pattern of signs:

$$\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \\ + & - & + & \\ \vdots & & & \ddots \end{bmatrix}$$

EXAMPLE 2 Use a cofactor expansion across the third row to compute $\det A$, where

$$A = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}$$

SOLUTION Compute

$$\begin{aligned}
\det A &= a_{31}C_{31} + a_{32}C_{32} + a_{33}C_{33} \\
&= (-1)^{3+1}a_{31}\det A_{31} + (-1)^{3+2}a_{32}\det A_{32} + (-1)^{3+3}a_{33}\det A_{33} \\
&= 0\begin{vmatrix} 5 & 0 \\ 4 & -1 \end{vmatrix} - (-2)\begin{vmatrix} 1 & 0 \\ 2 & -1 \end{vmatrix} + 0\begin{vmatrix} 1 & 5 \\ 2 & 4 \end{vmatrix} \\
&= 0 + 2(-1) + 0 = -2
\end{aligned}$$

■

Theorem 1 is helpful for computing the determinant of a matrix that contains many zeros. For example, if a row is mostly zeros, then the cofactor expansion across that row has many terms that are zero, and the cofactors in those terms need not be calculated. The same approach works with a column that contains many zeros.

EXAMPLE 3 Compute $\det A$, where

$$A = \begin{bmatrix} 3 & -7 & 8 & 9 & -6 \\ 0 & 2 & -5 & 7 & 3 \\ 0 & 0 & 1 & 5 & 0 \\ 0 & 0 & 2 & 4 & -1 \\ 0 & 0 & 0 & -2 & 0 \end{bmatrix}$$

SOLUTION The cofactor expansion down the first column of A has all terms equal to zero except the first. Thus

$$\det A = 3\begin{vmatrix} 2 & -5 & 7 & 3 \\ 0 & 1 & 5 & 0 \\ 0 & 2 & 4 & -1 \\ 0 & 0 & -2 & 0 \end{vmatrix} + 0\,C_{21} + 0\,C_{31} + 0\,C_{41} + 0\,C_{51}$$

Henceforth we will omit the zero terms in the cofactor expansion. Next, expand this 4×4 determinant down the first column to take advantage of the zeros there. We have

$$\det A = 3(2)\begin{vmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{vmatrix}$$

This 3×3 determinant was computed in Example 1 and found to equal -2. Hence $\det A = 3(2)(-2) = -12$. ■

The matrix in Example 3 was nearly triangular. The method in that example is easily adapted to prove the following theorem.

THEOREM 2

If A is a triangular matrix, then $\det A$ is the product of the entries on the main diagonal of A.

The strategy in Example 3 of looking for zeros works extremely well when an entire row or column consists of zeros. In such a case, the cofactor expansion along such a row or column is a sum of zeros! So the determinant is zero. Unfortunately, most cofactor expansions are not so quickly evaluated.

Reasonable Answers

How big can a determinant be? Let A be an $n \times n$ matrix. Notice that taking the determinant of A consists of adding and subtracting terms with n products each. If p is the product of the n largest elements in absolute value (the same number may be repeated if it occurs more than once as a matrix entry), then the determinant must be between $-np$ and np. For example, consider $A = \begin{bmatrix} 6 & 5 \\ -7 & 9 \end{bmatrix}$ and $B = \begin{bmatrix} 7 & 6 \\ 7 & -9 \end{bmatrix}$. The largest number in absolute value of each matrix is 9, and the second largest number is 7. In these two cases, $p = 7(9) = 63$ and $np = 126$. The determinant of each of these matrices should be a number between -126 and 126. Notice that $\det A = 6(9) - 5(-7) = 54 + 35 = 89$, $\det B = 7(-9) - 6(7) = -63 - 42 = -105$, illustrating that because the products are added and subtracted, any number in the range between -126 and 126 could turn out to be the determinant.

Next, consider $C = \begin{bmatrix} 7 & 9 \\ 7 & 9 \end{bmatrix}$ and $D = \begin{bmatrix} -9 & 9 \\ 9 & 9 \end{bmatrix}$. In matrices C and D, the number 9 appears twice and so should be selected twice. In this case, $p = 9(9) = 81$ and $np = 162$, so the determinants of C and D should be numbers between -162 and 162. Indeed, $\det C = (7)(9) - (7)(9) = 0$ and $\det D = (-9)(9) - (9)(9) = -162$. Notice that it is important to choose 9 twice as the two largest numbers in matrix D in order to get the correct bounds for the determinant of D.

Numerical Note

By today's standards, a 25×25 matrix is small. Yet it would be impossible to calculate a 25×25 determinant by cofactor expansion. In general, a cofactor expansion requires more than $n!$ multiplications, and $25!$ is approximately 1.55×10^{25}.

If a computer performs one trillion multiplications per second, it would have to run for almost 500,000 years to compute a 25×25 determinant by this method. Fortunately, there are faster methods, as we'll soon discover.

Exercises 19–38 explore important properties of determinants, mostly for the 2×2 case. The results from Exercises 33–36 will be used in the next section to derive the analogous properties for $n \times n$ matrices.

Practice Problem

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

3.1 Exercises

Compute the determinants in Exercises 1–8 using a cofactor expansion across the first row. In Exercises 1–4, also compute the determinant by a cofactor expansion down the second column.

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

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

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

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

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

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

7. $\begin{vmatrix} 4 & 3 & 0 \\ 6 & 5 & 2 \\ 9 & 7 & 3 \end{vmatrix}$

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

Compute the determinants in Exercises 9–14 by cofactor expansions. At each step, choose a row or column that involves the least amount of computation.

9. $\begin{vmatrix} 7 & 6 & 8 & 4 \\ 0 & 0 & 0 & 6 \\ 8 & 7 & 9 & 3 \\ 0 & 4 & 0 & 5 \end{vmatrix}$

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

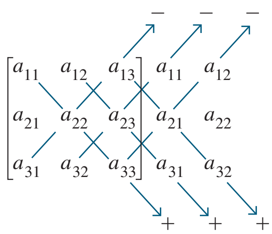
11. $\begin{vmatrix} 2 & -3 & 4 & 5 \\ 0 & 5 & 3 & -1 \\ 0 & 0 & -2 & 7 \\ 0 & 0 & 0 & 4 \end{vmatrix}$

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

13. $\begin{vmatrix} 4 & 0 & -7 & 3 & -5 \\ 0 & 0 & 2 & 0 & 0 \\ 7 & 3 & -6 & 4 & -8 \\ 5 & 0 & 5 & 2 & -3 \\ 0 & 0 & 9 & -1 & 2 \end{vmatrix}$

14. $\begin{vmatrix} 6 & 0 & 2 & 4 & 0 \\ 9 & 0 & -4 & 1 & 0 \\ 8 & -5 & 6 & 7 & 1 \\ 2 & 0 & 0 & 0 & 0 \\ 4 & 2 & 3 & 2 & 0 \end{vmatrix}$

The expansion of a 3×3 determinant can be remembered by the following device. Write a second copy of the first two columns to the right of the matrix, and compute the determinant by multiplying entries on six diagonals:



Add the downward diagonal products and subtract the upward products. Use this method to compute the determinants in Exercises 15–18. ***Warning:*** *This trick does not generalize in any reasonable way to* 4×4 *or larger matrices.*

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

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

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

18. $\begin{vmatrix} 1 & 4 & 5 \\ 3 & 4 & 3 \\ 3 & 3 & 4 \end{vmatrix}$

In Exercises 19–24, explore the effect of an elementary row operation on the determinant of a matrix. In each case, state the row operation and describe how it affects the determinant.

19. $\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \begin{bmatrix} c & d \\ a & b \end{bmatrix}$

20. $\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \begin{bmatrix} a & b \\ kc & kd \end{bmatrix}$

21. $\begin{bmatrix} 6 & 5 \\ 3 & 4 \end{bmatrix}, \begin{bmatrix} 6 & 5 \\ 3+6k & 4+5k \end{bmatrix}$

22. $\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \begin{bmatrix} a+kc & b+kd \\ c & d \end{bmatrix}$

23. $\begin{bmatrix} 1 & -2 & 3 \\ 2 & 3 & -4 \\ 3 & -4 & 5 \end{bmatrix}, \begin{bmatrix} k & -2k & 3k \\ 2 & 3 & -4 \\ 3 & -4 & 5 \end{bmatrix}$

24. $\begin{bmatrix} a & b & c \\ 1 & 4 & 5 \\ 2 & 3 & 6 \end{bmatrix}, \begin{bmatrix} 2 & 3 & 6 \\ 1 & 4 & 5 \\ a & b & c \end{bmatrix}$

Compute the determinants of the elementary matrices given in Exercises 25–30. (See Section 2.2, Examples 5 and 6.)

25. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & k & 1 \end{bmatrix}$

26. $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

27. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ k & 0 & 1 \end{bmatrix}$

28. $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$

29. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{bmatrix}$

30. $\begin{bmatrix} k & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Use Exercises 25–30 to answer the questions in Exercises 31 and 32. Give reasons for your answers.

31. What is the determinant of an elementary row replacement matrix?

32. What is the determinant of an elementary scaling matrix with k on the diagonal?

In Exercises 33–36, verify that $\det EA = (\det E)(\det A)$, where E is the elementary matrix shown and $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

33. $\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$ **34.** $\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$

35. $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ **36.** $\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$

37. Let $A = \begin{bmatrix} 6 & 5 \\ 3 & 4 \end{bmatrix}$. Write $2A$. Is $\det 2A = 2 \det A$?

38. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and let k be a scalar. Find a formula that relates $\det kA$ to k and $\det A$.

In Exercises 39 through 42, A is an $n \times n$ matrix. Mark each statement True or False **(T/F)**. Justify each answer.

39. **(T/F)** An $n \times n$ determinant is defined by determinants of $(n-1) \times (n-1)$ submatrices.

40. **(T/F)** The (i, j)-cofactor of a matrix A is the matrix A_{ij} obtained by deleting from A its ith row and jth column.

41. **(T/F)** The cofactor expansion of $\det A$ down a column is equal to the cofactor expansion along a row.

42. **(T/F)** The determinant of a triangular matrix is the sum of the entries on the main diagonal.

43. Let $\mathbf{u} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Compute the area of the parallelogram determined by $\mathbf{u}$, $\mathbf{v}$, $\mathbf{u} + \mathbf{v}$, and $\mathbf{0}$, and compute the determinant of $[\,\mathbf{u} \;\; \mathbf{v}\,]$. How do they compare? Replace the first entry of $\mathbf{v}$ by an arbitrary number x, and repeat the problem. Draw a picture and explain what you find.

44. Let $\mathbf{u} = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} c \\ 0 \end{bmatrix}$, where a, b, and c are positive (for simplicity). Compute the area of the parallelogram determined by $\mathbf{u}$, $\mathbf{v}$, $\mathbf{u} + \mathbf{v}$, and $\mathbf{0}$, and compute the determinants of the matrices $[\,\mathbf{u} \;\; \mathbf{v}\,]$ and $[\,\mathbf{v} \;\; \mathbf{u}\,]$. Draw a picture and explain what you find.

45. Let A be a 2×2 matrix all of whose entries are numbers that are greater than or equal to -10 and less than or equal to 10. Decide if each of the following is a reasonable answer for $\det A$.

a. 0

b. 202

c. -110

d. 555

46. Let A be a 3×3 matrix all of whose entries are numbers that are greater than or equal to -5 and less than or equal to 5. Decide if each of the following is a reasonable answer for $\det A$.

a. 300

b. -220

c. 1000

d. 10

T 47. Construct a random 4×4 matrix A with integer entries between -9 and 9. How is $\det A^{-1}$ related to $\det A$? Experiment with random $n \times n$ integer matrices for $n = 4$, 5, and 6, and make a conjecture. *Note:* In the unlikely event that you encounter a matrix with a zero determinant, reduce it to echelon form and discuss what you find.

T 48. Is it true that $\det AB = (\det A)(\det B)$? To find out, generate random 5×5 matrices A and B, and compute $\det AB - (\det A \det B)$. Repeat the calculations for three other pairs of $n \times n$ matrices, for various values of n. Report your results.

T 49. Is it true that $\det(A + B) = \det A + \det B$? Experiment with four pairs of random matrices as in Exercise 48, and make a conjecture.

T 50. Construct a random 4×4 matrix A with integer entries between -9 and 9, and compare $\det A$ with $\det A^T$, $\det(-A)$, $\det(2A)$, and $\det(10A)$. Repeat with two other random 4×4 integer matrices, and make conjectures about how these determinants are related. (Refer to Exercise 44 in Section 2.1.) Then check your conjectures with several random 5×5 and 6×6 integer matrices. Modify your conjectures, if necessary, and report your results.

T 51. Recall from the introductory section that the larger the determinant of $D^T D$, where D is the design matrix, the better will be the accuracy of the calculated weights for small light objects. Which of the following matrices corresponds to the best design for four weighings of four objects? Describe which of the objects s_1, s_2, s_3, and s_4 you would put in the left and right pans for each weighing corresponding to the best design matrix.

a. $D = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 \end{bmatrix}$

b. $D = \begin{bmatrix} -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix}$

c. $D = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$

T 52. Repeat Exercise 51 for the case of five weighings of four objects and the following design matrices.

a. $D = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}$

b. $D = \begin{bmatrix} -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix}$

c. $D = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & -1 & -1 & -1 \end{bmatrix}$

Solution to Practice Problem

Take advantage of the zeros. Begin with a cofactor expansion down the third column to obtain a 3×3 matrix, which may be evaluated by an expansion down its first column.

$$\begin{vmatrix} 5 & -7 & 2 & 2 \\ 0 & 3 & 0 & -4 \\ -5 & -8 & 0 & 3 \\ 0 & 5 & 0 & -6 \end{vmatrix} = (-1)^{1+3}(2)\begin{vmatrix} 0 & 3 & -4 \\ -5 & -8 & 3 \\ 0 & 5 & -6 \end{vmatrix}$$

$$= 2\,(-1)^{2+1}(-5)\begin{vmatrix} 3 & -4 \\ 5 & -6 \end{vmatrix} = 20$$

The $(-1)^{2+1}$ in the next-to-last calculation came from the (2, 1)-position of the -5 in the 3×3 determinant.

3.2 Properties of Determinants

The secret of determinants lies in how they change when row operations are performed. The following theorem generalizes the results of Exercises 19–24 in Section 3.1. The proof is at the end of this section.

THEOREM 3 **Row Operations**

Let A be a square matrix.

a. If a multiple of one row of A is added to another row to produce a matrix B, then $\det B = \det A$.

b. If two rows of A are interchanged to produce B, then $\det B = -\det A$.

c. If one row of A is multiplied by k to produce B, then $\det B = k \det A$.

The following examples show how to use Theorem 3 to find determinants efficiently.

EXAMPLE 1 Compute $\det A$, where $A = \begin{bmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{bmatrix}$.

SOLUTION The strategy is to reduce A to echelon form and then to use the fact that the determinant of a triangular matrix is the product of the diagonal entries. The first two row replacements in column 1 do not change the determinant:

$$\det A = \begin{vmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{vmatrix} = \begin{vmatrix} 1 & -4 & 2 \\ 0 & 0 & -5 \\ -1 & 7 & 0 \end{vmatrix} = \begin{vmatrix} 1 & -4 & 2 \\ 0 & 0 & -5 \\ 0 & 3 & 2 \end{vmatrix}$$

An interchange of rows 2 and 3 reverses the sign of the determinant, so

$$\det A = -\begin{vmatrix} 1 & -4 & 2 \\ 0 & 3 & 2 \\ 0 & 0 & -5 \end{vmatrix} = -(1)(3)(-5) = 15$$

A common use of Theorem 3(c) in hand calculations is to *factor out a common multiple of one row* of a matrix. For instance,

$$\begin{vmatrix} * & * & * \\ 5k & -2k & 3k \\ * & * & * \end{vmatrix} = k\begin{vmatrix} * & * & * \\ 5 & -2 & 3 \\ * & * & * \end{vmatrix}$$

where the starred entries are unchanged. We use this step in the next example.

EXAMPLE 2 Compute $\det A$, where $A = \begin{bmatrix} 2 & -8 & 6 & 8 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{bmatrix}$.

SOLUTION To simplify the arithmetic, we want a 1 in the upper-left corner. We could interchange rows 1 and 4. Instead, we factor out 2 from the top row, and then proceed with row replacements in the first column:

$$\det A = 2\begin{vmatrix} 1 & -4 & 3 & 4 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{vmatrix} = 2\begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & 3 & -4 & -2 \\ 0 & -12 & 10 & 10 \\ 0 & 0 & -3 & 2 \end{vmatrix}$$

Next, we could factor out another 2 from row 3 or use the 3 in the second column as a pivot. We choose the latter operation, adding 4 times row 2 to row 3:

$$\det A = 2\begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & 3 & -4 & -2 \\ 0 & 0 & -6 & 2 \\ 0 & 0 & -3 & 2 \end{vmatrix}$$

Finally, adding $-1/2$ times row 3 to row 4, and computing the "triangular" determinant, we find that

$$\det A = 2\begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & 3 & -4 & -2 \\ 0 & 0 & -6 & 2 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 2\,(1)(3)(-6)(1) = -36$$

$$U = \begin{bmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & \blacksquare & * \\ 0 & 0 & 0 & \blacksquare \end{bmatrix}$$
$\det U \neq 0$

$$U = \begin{bmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & 0 & \blacksquare \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$\det U = 0$

FIGURE 1
Typical echelon forms of square matrices.

Suppose a square matrix A has been reduced to an echelon form U by row replacements and row interchanges. (This is always possible. See the row reduction algorithm in Section 1.2.) If there are r interchanges, then Theorem 3 shows that

$$\det A = (-1)^r \det U$$

Since U is in echelon form, it is triangular, and so $\det U$ is the product of the diagonal entries $u_{11}, \dots, u_{nn}$. If A is invertible, the entries u_{ii} are all pivots (because $A \sim I_n$ and the u_{ii} have not been scaled to 1's). Otherwise, at least u_{nn} is zero, and the product $u_{11} \cdots u_{nn}$ is zero. See Figure 1. Thus

$$\det A = \begin{cases} (-1)^r \begin{pmatrix} \text{product of} \\ \text{pivots in } U \end{pmatrix} & \text{when } A \text{ is invertible} \\ 0 & \text{when } A \text{ is not invertible} \end{cases} \tag{1}$$

It is interesting to note that although the echelon form U described above is not unique (because it is not completely row reduced), and the pivots are not unique, the *product* of the pivots *is* unique, except for a possible minus sign.

Formula (1) not only gives a concrete interpretation of $\det A$ but also proves the main theorem of this section:

THEOREM 4 A square matrix A is invertible if and only if $\det A \neq 0$.

Theorem 4 adds the statement "$\det A \neq 0$" to the Invertible Matrix Theorem. A useful corollary is that $\det A = 0$ when the columns of A are linearly dependent. Also, $\det A = 0$ when the *rows* of A are linearly dependent. (Rows of A are columns of A^T, and linearly dependent columns of A^T make A^T singular. When A^T is singular, so is A, by the Invertible Matrix Theorem.) In practice, linear dependence is obvious when two columns or two rows are the same or a column or a row is zero.

EXAMPLE 3 Compute $\det A$, where $A = \begin{bmatrix} 3 & -1 & 2 & -5 \\ 0 & 5 & -3 & -6 \\ -6 & 7 & -7 & 4 \\ -5 & -8 & 0 & 9 \end{bmatrix}$.

SOLUTION Add 2 times row 1 to row 3 to obtain

$$\det A = \det \begin{bmatrix} 3 & -1 & 2 & -5 \\ 0 & 5 & -3 & -6 \\ 0 & 5 & -3 & -6 \\ -5 & -8 & 0 & 9 \end{bmatrix} = 0$$

because the second and third rows of the second matrix are equal. ■

Numerical Notes

1. Most computer programs that compute $\det A$ for a general matrix A use the method of formula (1) above.
2. It can be shown that evaluation of an $n \times n$ determinant using row operations requires about $2n^3/3$ arithmetic operations. Any modern microcomputer can calculate a 25×25 determinant in a fraction of a second, since only about 10,000 operations are required.

Computers can also handle large "sparse" matrices, with special routines that take advantage of the presence of many zeros. Of course, zero entries can speed hand computations, too. The calculations in the next example combine the power of row operations with the strategy from Section 3.1 of using zero entries in cofactor expansions.

EXAMPLE 4 Compute $\det A$, where $A = \begin{bmatrix} 0 & 1 & 2 & -1 \\ 2 & 5 & -7 & 3 \\ 0 & 3 & 6 & 2 \\ -2 & -5 & 4 & -2 \end{bmatrix}$.

SOLUTION A good way to begin is to use the 2 in column 1 as a pivot, eliminating the -2 below it. Then use a cofactor expansion to reduce the size of the determinant, followed by another row replacement operation. Thus

$$\det A = \begin{vmatrix} 0 & 1 & 2 & -1 \\ 2 & 5 & -7 & 3 \\ 0 & 3 & 6 & 2 \\ 0 & 0 & -3 & 1 \end{vmatrix} = -2\begin{vmatrix} 1 & 2 & -1 \\ 3 & 6 & 2 \\ 0 & -3 & 1 \end{vmatrix} = -2\begin{vmatrix} 1 & 2 & -1 \\ 0 & 0 & 5 \\ 0 & -3 & 1 \end{vmatrix}$$

An interchange of rows 2 and 3 would produce a "triangular determinant." Another approach is to make a cofactor expansion down the first column:

$$\det A = (-2)(1)\begin{vmatrix} 0 & 5 \\ -3 & 1 \end{vmatrix} = -2\,(15) = -30$$

■

Column Operations

We can perform operations on the columns of a matrix in a way that is analogous to the row operations we have considered. The next theorem shows that column operations have the same effects on determinants as row operations.

Remark: The Principle of Mathematical Induction says the following: Let $P(n)$ be a statement that is either true or false for each natural number n. Then $P(n)$ is true for all $n \geq 1$ provided that $P(1)$ is true, and for each natural number k, if $P(k)$ is true, then $P(k+1)$ is true. The Principle of Mathematical Induction is used to prove the next theorem.

THEOREM 5

If A is an $n \times n$ matrix, then $\det A^T = \det A$.

PROOF The theorem is obvious for $n = 1$. Suppose the theorem is true for $k \times k$ determinants and let $n = k + 1$. Then the cofactor of a_{1j} in A equals the cofactor of a_{j1} in A^T, because the cofactors involve $k \times k$ determinants. Hence the cofactor expansion of $\det A$ along the first *row* equals the cofactor expansion of $\det A^T$ down the first *column*. That is, A and A^T have equal determinants. The theorem is true for $n = 1$, and the truth of the theorem for one value of n implies its truth for the next value of n. By the Principle of Mathematical Induction, the theorem is true for all $n \geq 1$. ■

Because of Theorem 5, each statement in Theorem 3 is true when the word *row* is replaced everywhere by *column*. To verify this property, one merely applies the original Theorem 3 to A^T. A row operation on A^T amounts to a column operation on A.

Column operations are useful for both theoretical purposes and hand computations. However, for simplicity we'll perform only row operations in numerical calculations.

Determinants and Matrix Products

The proof of the following useful theorem is at the end of the section. Applications are in the exercises.

THEOREM 6 **Multiplicative Property**

If A and B are $n \times n$ matrices, then $\det AB = (\det A)(\det B)$.

EXAMPLE 5 Verify Theorem 6 for $A = \begin{bmatrix} 6 & 1 \\ 3 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}$.

SOLUTION

$$AB = \begin{bmatrix} 6 & 1 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 25 & 20 \\ 14 & 13 \end{bmatrix}$$

and

$$\det AB = 25(13) - 20(14) = 325 - 280 = 45$$

Since $\det A = 9$ and $\det B = 5$,

$$(\det A)(\det B) = 9(5) = 45 = \det AB$$

■

Warning: A common misconception is that Theorem 6 has an analogue for *sums* of matrices. However, $\det(A + B)$ is *not* equal to $\det A + \det B$, in general.

A Linearity Property of the Determinant Function

For an $n \times n$ matrix A, we can consider $\det A$ as a function of the n column vectors in A. We will show that if all columns except one are held fixed, then $\det A$ is a *linear function* of that one (vector) variable.

Suppose that the jth column of A is allowed to vary, and write

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{j-1} & \mathbf{x} & \mathbf{a}_{j+1} & \cdots & \mathbf{a}_n \end{bmatrix}$$

Define a transformation T from $\mathbb{R}^n$ to $\mathbb{R}$ by

$$T(\mathbf{x}) = \det \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{j-1} & \mathbf{x} & \mathbf{a}_{j+1} & \cdots & \mathbf{a}_n \end{bmatrix}$$

Then,

$$T(c\mathbf{x}) = cT(\mathbf{x}) \quad \text{for all scalars } c \text{ and all } \mathbf{x} \text{ in } \mathbb{R}^n \tag{2}$$

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) \quad \text{for all } \mathbf{u}, \mathbf{v} \text{ in } \mathbb{R}^n \tag{3}$$

Property (2) is Theorem 3(c) applied to the columns of A. A proof of property (3) follows from a cofactor expansion of $\det A$ down the jth column. (See Exercise 49.) This (multi-) linearity property of the determinant turns out to have many useful consequences that are studied in more advanced courses.

Proofs of Theorems 3 and 6

It is convenient to prove Theorem 3 when it is stated in terms of the elementary matrices discussed in Section 2.2. We call an elementary matrix E a *row replacement* (*matrix*) if E is obtained from the identity I by adding a multiple of one row to another row; E is an *interchange* if E is obtained by interchanging two rows of I; and E is *a scale by* r if E is obtained by multiplying a row of I by a nonzero scalar r. With this terminology, Theorem 3 can be reformulated as follows:

If A is an $n \times n$ matrix and E is an $n \times n$ elementary matrix, then

$$\det EA = (\det E)(\det A)$$

where

$$\det E = \begin{cases} 1 & \text{if } E \text{ is a row replacement} \\ -1 & \text{if } E \text{ is an interchange} \\ r & \text{if } E \text{ is a scale by } r \end{cases}$$

PROOF OF THEOREM 3 The proof is by induction on the size of A. The case of a 2×2 matrix was verified in Exercises 33–36 of Section 3.1. Suppose the theorem has been verified for determinants of $k \times k$ matrices with $k \geq 2$, let $n = k + 1$, and let A be $n \times n$. The action of E on A involves either two rows or only one row. So we can expand $\det EA$ across a row that is unchanged by the action of E, say, row i. Let A_{ij} (respectively, B_{ij}) be the matrix obtained by deleting row i and column j from A (respectively, EA). Then the rows of B_{ij} are obtained from the rows of A_{ij} by the same type of elementary row operation that E performs on A. Since these submatrices are only $k \times k$, the induction assumption implies that

$$\det B_{ij} = \alpha \det A_{ij}$$

where $\alpha = 1, -1$, or r, depending on the nature of E. The cofactor expansion across row i is

$$\begin{aligned} \det EA &= a_{i1}(-1)^{i+1} \det B_{i1} + \cdots + a_{in}(-1)^{i+n} \det B_{in} \\ &= \alpha a_{i1}(-1)^{i+1} \det A_{i1} + \cdots + \alpha a_{in}(-1)^{i+n} \det A_{in} \\ &= \alpha \det A \end{aligned}$$

In particular, taking $A = I_n$, we see that $\det E = 1, -1$, or r, depending on the nature of E. Thus the theorem is true for $n = 2$, and the truth of the theorem for one value of n implies its truth for the next value of n. By the principle of induction, the theorem must be true for $n \geq 2$. The theorem is trivially true for $n = 1$. ■

PROOF OF THEOREM 6 If A is not invertible, then neither is AB, by Exercise 35 in Section 2.3. In this case, $\det AB = (\det A)(\det B)$, because both sides are zero, by Theorem 4. If A is invertible, then A and the identity matrix I_n are row equivalent by the Invertible Matrix Theorem. So there exist elementary matrices $E_1, \ldots, E_p$ such that

$$A = E_p E_{p-1} \cdots E_1 I_n = E_p E_{p-1} \cdots E_1$$

For brevity, write $|A|$ for det A. Then repeated application of Theorem 3, as rephrased above, shows that

$$\begin{aligned}|AB| &= |E_p \cdots E_1 B| = |E_p||E_{p-1} \cdots E_1 B| = \cdots \\ &= |E_p| \cdots |E_1||B| = \cdots = |E_p \cdots E_1||B| \\ &= |A||B|\end{aligned}$$

■

Practice Problems

1. Compute $\begin{vmatrix} 1 & -3 & 1 & -2 \\ 2 & -5 & -1 & -2 \\ 0 & -4 & 5 & 1 \\ -3 & 10 & -6 & 8 \end{vmatrix}$ in as few steps as possible.
2. Use a determinant to decide if $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are linearly independent, when

$$\mathbf{v}_1 = \begin{bmatrix} 5 \\ -7 \\ 9 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -3 \\ 3 \\ -5 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2 \\ -7 \\ 5 \end{bmatrix}$$

3. Let A be an $n \times n$ matrix such that $A^2 = I$. Show that det $A = \pm 1$.

3.2 Exercises

Each equation in Exercises 1–4 illustrates a property of determinants. State the property.

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

2. $\begin{vmatrix} 3 & -6 & 9 \\ 3 & 5 & -5 \\ 1 & 3 & 3 \end{vmatrix} = 3\begin{vmatrix} 1 & -2 & 3 \\ 3 & 5 & -5 \\ 1 & 3 & 3 \end{vmatrix}$

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

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

Find the determinants in Exercises 5–10 by row reduction to echelon form.

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

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

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

8. $\begin{vmatrix} 1 & 2 & -3 & 4 \\ 0 & 1 & 5 & 6 \\ -4 & -9 & 7 & -14 \\ 2 & 5 & 0 & 7 \end{vmatrix}$

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

10. $\begin{vmatrix} 1 & 3 & -1 & 0 & -2 \\ 0 & 1 & -2 & -1 & -3 \\ -2 & -6 & 2 & 3 & 10 \\ 1 & 5 & -6 & 2 & -3 \\ 0 & 2 & -4 & 5 & 9 \end{vmatrix}$

Combine the methods of row reduction and cofactor expansion to compute the determinants in Exercises 11–14.

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

12. $\begin{vmatrix} -2 & 6 & 0 & 9 \\ 3 & 4 & 8 & 2 \\ 4 & 3 & 0 & 1 \\ 3 & 1 & 2 & -1 \end{vmatrix}$

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

14. $\begin{vmatrix} 4 & 3 & 2 & 1 \\ 5 & 4 & -3 & 0 \\ 9 & -8 & -7 & 0 \\ 4 & 6 & 2 & 1 \end{vmatrix}$

Find the determinants in Exercises 15–20, where

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = 7.$$

15. $\begin{vmatrix} a & b & c \\ d & e & f \\ 3g & 3h & 3i \end{vmatrix}$

16. $\begin{vmatrix} a & b & c \\ d+3g & e+3h & f+3i \\ g & h & i \end{vmatrix}$

17. $\begin{vmatrix} a+d & b+e & c+f \\ d & e & f \\ g & h & i \end{vmatrix}$

18. $\begin{vmatrix} a & b & c \\ 8d & 8e & 8f \\ g & h & i \end{vmatrix}$

19. $\begin{vmatrix} a & b & c \\ 2d+a & 2e+b & 2f+c \\ g & h & i \end{vmatrix}$

20. $\begin{vmatrix} g & h & i \\ a & b & c \\ d & e & f \end{vmatrix}$

In Exercises 21–23, use determinants to find out if the matrix is invertible.

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

22. $\begin{bmatrix} 4 & 5 & 0 \\ 3 & 2 & 1 \\ 1 & -4 & 3 \end{bmatrix}$

23. $\begin{bmatrix} 3 & 0 & 0 & 2 \\ 6 & 8 & 9 & 0 \\ 4 & 5 & 6 & 0 \\ 0 & -8 & -9 & 4 \end{bmatrix}$

In Exercises 24–26, use determinants to decide if the set of vectors is linearly independent.

24. $\begin{bmatrix} 4 \\ 6 \\ 2 \end{bmatrix}, \begin{bmatrix} -6 \\ 0 \\ 6 \end{bmatrix}, \begin{bmatrix} -3 \\ -5 \\ -2 \end{bmatrix}$

25. $\begin{bmatrix} 7 \\ -4 \\ -6 \end{bmatrix}, \begin{bmatrix} -8 \\ 5 \\ 7 \end{bmatrix}, \begin{bmatrix} 7 \\ 0 \\ -5 \end{bmatrix}$

26. $\begin{bmatrix} 3 \\ 5 \\ -6 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ -6 \\ 0 \\ 7 \end{bmatrix}, \begin{bmatrix} -2 \\ -1 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ -2 \end{bmatrix}$

In Exercises 27–34, A and B are $n \times n$ matrices. Mark each statement True or False **(T/F)**. Justify each answer.

27. **(T/F)** A row replacement operation does not affect the determinant of a matrix.

28. **(T/F)** If $\det A$ is zero, then two rows or two columns are the same, or a row or a column is zero.

29. **(T/F)** If the columns of A are linearly dependent, then $\det A = 0$.

30. **(T/F)** The determinant of A is the product of the diagonal entries in A.

31. **(T/F)** If three row interchanges are made in succession, then the new determinant equals the old determinant.

32. **(T/F)** The determinant of A is the product of the pivots in any echelon form U of A, multiplied by $(-1)^r$, where r is the number of row interchanges made during row reduction from A to U.

33. **(T/F)** $\det(A+B) = \det A + \det B$.

34. **(T/F)** $\det A^{-1} = (-1)\det A$.

35. Compute $\det B^4$, where $B = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix}$.

36. Use Theorem 3 (but not Theorem 4) to show that if two rows of a square matrix A are equal, then $\det A = 0$. The same is true for two columns. Why?

In Exercises 37–42, mention an appropriate theorem in your explanation.

37. Show that if A is invertible, then $\det A^{-1} = \dfrac{1}{\det A}$.

38. Suppose that A is a square matrix such that $\det A^3 = 0$. Explain why A cannot be invertible.

39. Let A and B be square matrices. Show that even though AB and BA may not be equal, it is always true that $\det AB = \det BA$.

40. Let A and P be square matrices, with P invertible. Show that $\det(PAP^{-1}) = \det A$.

41. Let U be a square matrix such that $U^TU = I$. Show that $\det U = \pm 1$.

42. Find a formula for $\det(rA)$ when A is an $n \times n$ matrix.

Verify that $\det AB = (\det A)(\det B)$ for the matrices in Exercises 43 and 44. (Do not use Theorem 6.)

43. $A = \begin{bmatrix} 3 & 0 \\ 6 & 1 \end{bmatrix}, B = \begin{bmatrix} 2 & 0 \\ 5 & 4 \end{bmatrix}$

44. $A = \begin{bmatrix} 2 & 3 \\ -3 & -1 \end{bmatrix}, B = \begin{bmatrix} 2 & 4 \\ -3 & -6 \end{bmatrix}$

45. Let A and B be 3×3 matrices, with $\det A = -2$ and $\det B = 3$. Use properties of determinants (in the text and in the preceding exercises) to compute:
 a. $\det AB$ b. $\det 5A$ c. $\det B^T$
 d. $\det A^{-1}$ e. $\det A^3$

46. Let A and B be 4×4 matrices, with $\det A = 4$ and $\det B = -5$. Compute:
 a. $\det AB$ b. $\det 3A$ c. $\det B^4$
 d. $\det BA\,B^T$ e. $\det AB\,A^{-1}$

47. Verify that $\det A = \det B + \det C$, where

$$A = \begin{bmatrix} a+e & b+f \\ c & d \end{bmatrix},\ B = \begin{bmatrix} a & b \\ c & d \end{bmatrix},\ C = \begin{bmatrix} e & f \\ c & d \end{bmatrix}$$

48. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Show that $\det(A+B) = \det A + \det B$ if and only if $a + d = 0$.

49. Verify that $\det A = \det B + \det C$, where

$$A = \begin{bmatrix} a_{11} & a_{12} & u_1 + v_1 \\ a_{21} & a_{22} & u_2 + v_2 \\ a_{31} & a_{32} & u_3 + v_3 \end{bmatrix},$$

$$B = \begin{bmatrix} a_{11} & a_{12} & u_1 \\ a_{21} & a_{22} & u_2 \\ a_{31} & a_{32} & u_3 \end{bmatrix},\ C = \begin{bmatrix} a_{11} & a_{12} & v_1 \\ a_{21} & a_{22} & v_2 \\ a_{31} & a_{32} & v_3 \end{bmatrix}$$

Note, however, that A is *not* the same as $B + C$.

50. Right-multiplication by an elementary matrix E affects the *columns* of A in the same way that left-multiplication affects the *rows*. Use Theorems 5 and 3 and the obvious fact that E^T is another elementary matrix to show that

$$\det AE = (\det E)(\det A)$$

Do not use Theorem 6.

51. Suppose A is an $n \times n$ matrix and a computer suggests that $\det A = 5$ and $\det\left(A^{-1}\right) = 1$. Should you trust these answers? Why or why not?

52. Suppose A and B are $n \times n$ matrices and a computer suggests that $\det A = 5$, $\det B = 2$ and $\det AB = 7$. Should you trust these answers? Why or why not?

T **53.** Compute $\det A^T A$ and $\det AA^T$ for several random 4×5 matrices and several random 5×6 matrices. What can you say about $A^T A$ and AA^T when A has more columns than rows?

T **54.** If $\det A$ is close to zero, is the matrix A nearly singular? Experiment with the nearly singular 4×4 matrix

$$A = \begin{bmatrix} 4 & 0 & -7 & -7 \\ -6 & 1 & 11 & 9 \\ 7 & -5 & 10 & 19 \\ -1 & 2 & 3 & -1 \end{bmatrix}$$

Compute the determinants of A, $10A$, and $0.1A$. In contrast, compute the condition numbers of these matrices. Repeat these calculations when A is the 4×4 identity matrix. Discuss your results.

Solutions to Practice Problems

1. Perform row replacements to create zeros in the first column, and then create a row of zeros.

$$\begin{vmatrix} 1 & -3 & 1 & -2 \\ 2 & -5 & -1 & -2 \\ 0 & -4 & 5 & 1 \\ -3 & 10 & -6 & 8 \end{vmatrix} = \begin{vmatrix} 1 & -3 & 1 & -2 \\ 0 & 1 & -3 & 2 \\ 0 & -4 & 5 & 1 \\ 0 & 1 & -3 & 2 \end{vmatrix} = \begin{vmatrix} 1 & -3 & 1 & -2 \\ 0 & 1 & -3 & 2 \\ 0 & -4 & 5 & 1 \\ 0 & 0 & 0 & 0 \end{vmatrix} = 0$$

2. $\det[\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3\,] = \begin{vmatrix} 5 & -3 & 2 \\ -7 & 3 & -7 \\ 9 & -5 & 5 \end{vmatrix} = \begin{vmatrix} 5 & -3 & 2 \\ -2 & 0 & -5 \\ 9 & -5 & 5 \end{vmatrix}$ Row 1 added to row 2

$= -(-3)\begin{vmatrix} -2 & -5 \\ 9 & 5 \end{vmatrix} - (-5)\begin{vmatrix} 5 & 2 \\ -2 & -5 \end{vmatrix}$ Cofactors of column 2

$= 3\,(35) + 5\,(-21) = 0$

By Theorem 4, the matrix $[\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3\,]$ is not invertible. The columns are linearly dependent, by the Invertible Matrix Theorem.

3. Recall that $\det I = 1$. By Theorem 6, $\det(AA) = (\det A)(\det A)$. Putting these two observations together results in

$$1 = \det I = \det A^2 = \det(AA) = (\det A)(\det A) = (\det A)^2$$

Taking the square root of both sides establishes that $\det A = \pm 1$.

3.3 Cramer's Rule, Volume, and Linear Transformations

This section applies the theory of the preceding sections to obtain important theoretical formulas and a geometric interpretation of the determinant.

Cramer's Rule

Cramer's rule is needed in a variety of theoretical calculations. For instance, it can be used to study how the solution of $A\mathbf{x} = \mathbf{b}$ is affected by changes in the entries of $\mathbf{b}$. However, the formula is inefficient for hand calculations, except for 2×2 or perhaps 3×3 matrices.

For any $n \times n$ matrix A and any $\mathbf{b}$ in $\mathbb{R}^n$, let $A_i(\mathbf{b})$ be the matrix obtained from A by replacing column i by the vector $\mathbf{b}$.

$$A_i(\mathbf{b}) = [\mathbf{a}_1 \quad \cdots \quad \underset{\substack{\uparrow \\ \text{col } i}}{\mathbf{b}} \quad \cdots \quad \mathbf{a}_n]$$

THEOREM 7

Cramer's Rule

Let A be an invertible $n \times n$ matrix. For any $\mathbf{b}$ in $\mathbb{R}^n$, the unique solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ has entries given by

$$x_i = \frac{\det A_i(\mathbf{b})}{\det A}, \qquad i = 1, 2, \ldots, n \tag{1}$$

PROOF Denote the columns of A by $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and the columns of the $n \times n$ identity matrix I by $\mathbf{e}_1, \ldots, \mathbf{e}_n$. If $A\mathbf{x} = \mathbf{b}$, the definition of matrix multiplication shows that

$$\begin{aligned} A\,(I_i(\mathbf{x})) &= A\begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{x} & \cdots & \mathbf{e}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{e}_1 & \cdots & A\mathbf{x} & \cdots & A\mathbf{e}_n \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{b} & \cdots & \mathbf{a}_n \end{bmatrix} = A_i(\mathbf{b}) \end{aligned}$$

By the multiplicative property of determinants,

$$(\det A)(\det I_i(\mathbf{x})) = \det A_i(\mathbf{b})$$

The second determinant on the left is simply x_i. (Make a cofactor expansion along the ith row.) Hence $(\det A)\,x_i = \det A_i(\mathbf{b})$. This proves (1) because A is invertible and $\det A \neq 0$. ■

EXAMPLE 1 Use Cramer's rule to solve the system

$$\begin{aligned} 3x_1 - 2x_2 &= 6 \\ -5x_1 + 4x_2 &= 8 \end{aligned}$$

SOLUTION View the system as $A\mathbf{x} = \mathbf{b}$. Using the notation introduced above,

$$A = \begin{bmatrix} 3 & -2 \\ -5 & 4 \end{bmatrix}, \qquad A_1(\mathbf{b}) = \begin{bmatrix} 6 & -2 \\ 8 & 4 \end{bmatrix}, \qquad A_2(\mathbf{b}) = \begin{bmatrix} 3 & 6 \\ -5 & 8 \end{bmatrix}$$

Since $\det A = 2$, the system has a unique solution. By Cramer's rule,

$$x_1 = \frac{\det A_1(\mathbf{b})}{\det A} = \frac{24 + 16}{2} = 20$$
$$x_2 = \frac{\det A_2(\mathbf{b})}{\det A} = \frac{24 + 30}{2} = 27 \qquad \blacksquare$$

Application to Engineering

A number of important engineering problems, particularly in electrical engineering and control theory, can be analyzed by *Laplace transforms*. This approach converts an appropriate system of linear differential equations into a system of linear algebraic equations whose coefficients involve a parameter. The next example illustrates the type of algebraic system that may arise.

EXAMPLE 2 Consider the following system in which s is an unspecified parameter. Determine the values of s for which the system has a unique solution, and use Cramer's rule to describe the solution.

$$\begin{aligned} 3sx_1 - 2x_2 &= 4 \\ -6x_1 + sx_2 &= 1 \end{aligned}$$

SOLUTION View the system as $A\mathbf{x} = \mathbf{b}$. Then

$$A = \begin{bmatrix} 3s & -2 \\ -6 & s \end{bmatrix}, \quad A_1(\mathbf{b}) = \begin{bmatrix} 4 & -2 \\ 1 & s \end{bmatrix}, \quad A_2(\mathbf{b}) = \begin{bmatrix} 3s & 4 \\ -6 & 1 \end{bmatrix}$$

Since

$$\det A = 3s^2 - 12 = 3(s+2)(s-2)$$

the system has a unique solution precisely when $s \neq \pm 2$. For such an s, the solution is (x_1, x_2), where

$$x_1 = \frac{\det A_1(\mathbf{b})}{\det A} = \frac{4s + 2}{3(s+2)(s-2)}$$
$$x_2 = \frac{\det A_2(\mathbf{b})}{\det A} = \frac{3s + 24}{3(s+2)(s-2)} = \frac{s + 8}{(s+2)(s-2)} \qquad \blacksquare$$

A Formula for A^{-1}

Cramer's rule leads easily to a general formula for the inverse of an $n \times n$ matrix A. The jth column of A^{-1} is a vector $\mathbf{x}$ that satisfies

$$A\mathbf{x} = \mathbf{e}_j$$

where $\mathbf{e}_j$ is the jth column of the identity matrix, and the ith entry of $\mathbf{x}$ is the (i, j)-entry of A^{-1}. By Cramer's rule,

$$\left\{(i, j)\text{-entry of } A^{-1}\right\} = x_i = \frac{\det A_i(\mathbf{e}_j)}{\det A} \tag{2}$$

Recall that A_{ji} denotes the submatrix of A formed by deleting row j and column i. A cofactor expansion down column i of $A_i(\mathbf{e}_j)$ shows that

$$\det A_i(\mathbf{e}_j) = (-1)^{i+j} \det A_{ji} = C_{ji} \tag{3}$$

where C_{ji} is a cofactor of A. By (2), the (i, j)-entry of A^{-1} is the cofactor C_{ji} divided by $\det A$. [Note that the subscripts on C_{ji} are the reverse of (i, j).] Thus

$$A^{-1} = \frac{1}{\det A}\begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix} \tag{4}$$

The matrix of cofactors on the right side of (4) is called the **adjugate** (or **classical adjoint**) of A, denoted by adj A. (The term *adjoint* also has another meaning in advanced texts on linear transformations.) The next theorem simply restates (4).

THEOREM 8

An Inverse Formula

Let A be an invertible $n \times n$ matrix. Then

$$A^{-1} = \frac{1}{\det A}\operatorname{adj} A$$

EXAMPLE 3 Find the inverse of the matrix $A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & -1 & 1 \\ 1 & 4 & -2 \end{bmatrix}$.

SOLUTION The nine cofactors are

$$C_{11} = +\begin{vmatrix} -1 & 1 \\ 4 & -2 \end{vmatrix} = -2, \quad C_{12} = -\begin{vmatrix} 1 & 1 \\ 1 & -2 \end{vmatrix} = 3, \quad C_{13} = +\begin{vmatrix} 1 & -1 \\ 1 & 4 \end{vmatrix} = 5$$

$$C_{21} = -\begin{vmatrix} 1 & 3 \\ 4 & -2 \end{vmatrix} = 14, \quad C_{22} = +\begin{vmatrix} 2 & 3 \\ 1 & -2 \end{vmatrix} = -7, \quad C_{23} = -\begin{vmatrix} 2 & 1 \\ 1 & 4 \end{vmatrix} = -7$$

$$C_{31} = +\begin{vmatrix} 1 & 3 \\ -1 & 1 \end{vmatrix} = 4, \quad C_{32} = -\begin{vmatrix} 2 & 3 \\ 1 & 1 \end{vmatrix} = 1, \quad C_{33} = +\begin{vmatrix} 2 & 1 \\ 1 & -1 \end{vmatrix} = -3$$

The adjugate matrix is the *transpose* of the matrix of cofactors. [For instance, C_{12} goes in the (2, 1) position.] Thus

$$\operatorname{adj} A = \begin{bmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{bmatrix}$$

We could compute $\det A$ directly, but the following computation provides a check on the calculations for adj A *and* produces $\det A$:

$$(\operatorname{adj} A)\,A = \begin{bmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{bmatrix}\begin{bmatrix} 2 & 1 & 3 \\ 1 & -1 & 1 \\ 1 & 4 & -2 \end{bmatrix} = \begin{bmatrix} 14 & 0 & 0 \\ 0 & 14 & 0 \\ 0 & 0 & 14 \end{bmatrix} = 14I$$

Since $(\operatorname{adj} A)A = 14I$, Theorem 8 shows that $\det A = 14$ and

$$A^{-1} = \frac{1}{14}\begin{bmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{bmatrix} = \begin{bmatrix} -1/7 & 1 & 2/7 \\ 3/14 & -1/2 & 1/14 \\ 5/14 & -1/2 & -3/14 \end{bmatrix}$$

■

Numerical Notes

Theorem 8 is useful mainly for theoretical calculations. The formula for A^{-1} permits one to deduce properties of the inverse without actually calculating it. Except for special cases, the algorithm in Section 2.2 gives a much better way to compute A^{-1}, if the inverse is really needed.

Cramer's rule is also a theoretical tool. It can be used to study how sensitive the solution of $A\mathbf{x} = \mathbf{b}$ is to changes in an entry in $\mathbf{b}$ or in A (perhaps due to experimental error when acquiring the entries for $\mathbf{b}$ or A). When A is a 3×3 matrix with *complex* entries, Cramer's rule is sometimes selected for hand computation because row reduction of $[\, A \quad \mathbf{b} \,]$ with complex arithmetic can be messy, and the determinants are fairly easy to compute. For a larger $n \times n$ matrix (real or complex), Cramer's rule is hopelessly inefficient. Computing just *one* determinant takes about as much work as solving $A\mathbf{x} = \mathbf{b}$ by row reduction.

Determinants as Area or Volume

In the next application, we verify the geometric interpretation of determinants described in the chapter introduction. Although a general discussion of length and distance in $\mathbb{R}^n$ will not be given until Chapter 6, we assume here that the usual Euclidean concepts of length, area, and volume are already understood for $\mathbb{R}^2$ and $\mathbb{R}^3$.

THEOREM 9

If A is a 2×2 matrix, the area of the parallelogram determined by the columns of A is $|\det A|$. If A is a 3×3 matrix, the volume of the parallelepiped determined by the columns of A is $|\det A|$.

STUDY GUIDE provides a geometric proof of the determinant as area.

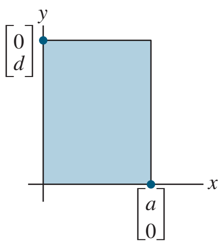


FIGURE 1
Area $= |ad|$.

PROOF The theorem is obviously true for any 2×2 diagonal matrix:

$$\left| \det \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} \right| = |ad| = \left\{ \begin{matrix} \text{area of} \\ \text{rectangle} \end{matrix} \right\}$$

See Figure 1. It will suffice to show that any 2×2 matrix $A = [\, \mathbf{a}_1 \quad \mathbf{a}_2 \,]$ can be transformed into a diagonal matrix in a way that changes neither the area of the associated parallelogram nor $|\det A|$. From Section 3.2, we know that the absolute value of the determinant is unchanged when two columns are interchanged or a multiple of one column is added to another. And it is easy to see that such operations suffice to transform A into a diagonal matrix. Column interchanges do not change the parallelogram at all. So it suffices to prove the following simple geometric observation that applies to vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$:

Let $\mathbf{a}_1$ and $\mathbf{a}_2$ be nonzero vectors. Then for any scalar c, the area of the parallelogram determined by $\mathbf{a}_1$ and $\mathbf{a}_2$ equals the area of the parallelogram determined by $\mathbf{a}_1$ and $\mathbf{a}_2 + c\mathbf{a}_1$.

To prove this statement, we may assume that $\mathbf{a}_2$ is not a multiple of $\mathbf{a}_1$, for otherwise the two parallelograms would be degenerate and have zero area. If L is the line through $\mathbf{0}$ and $\mathbf{a}_1$, then $\mathbf{a}_2 + L$ is the line through $\mathbf{a}_2$ parallel to L, and $\mathbf{a}_2 + c\mathbf{a}_1$ is on this line. See Figure 2. The points $\mathbf{a}_2$ and $\mathbf{a}_2 + c\mathbf{a}_1$ have the same perpendicular distance to L. Hence the two parallelograms in Figure 2 have the same area, since they share the base from $\mathbf{0}$ to $\mathbf{a}_1$. This completes the proof for $\mathbb{R}^2$.

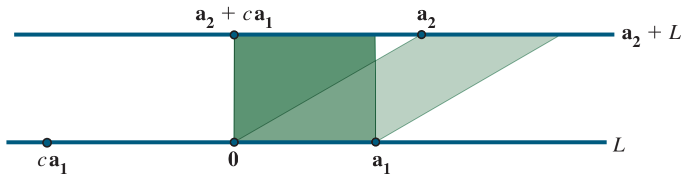


FIGURE 2 Two parallelograms of equal area.

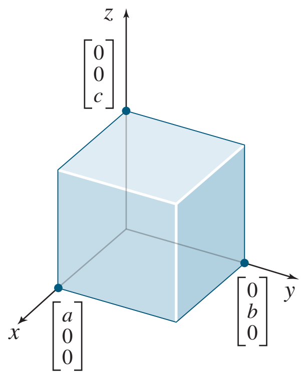


FIGURE 3
Volume $= |abc|$.

The proof for $\mathbb{R}^3$ is similar. The theorem is obviously true for a 3×3 diagonal matrix. See Figure 3. And any 3×3 matrix A can be transformed into a diagonal matrix using column operations that do not change $|\det A|$. (Think about doing row operations on A^T.) So it suffices to show that these operations do not affect the volume of the parallelepiped determined by the columns of A.

A parallelepiped is shown in Figure 4 as a shaded box with two sloping sides. Its volume is the area of the base in the plane Span $\{\mathbf{a}_1, \mathbf{a}_3\}$ times the altitude of $\mathbf{a}_2$ above Span $\{\mathbf{a}_1, \mathbf{a}_3\}$. Any vector $\mathbf{a}_2 + c\mathbf{a}_1$ has the same altitude because $\mathbf{a}_2 + c\mathbf{a}_1$ lies in the plane $\mathbf{a}_2 +$ Span $\{\mathbf{a}_1, \mathbf{a}_3\}$, which is parallel to Span $\{\mathbf{a}_1, \mathbf{a}_3\}$. Hence the volume of the parallelepiped is unchanged when $[\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3\,]$ is changed to $[\,\mathbf{a}_1 \quad \mathbf{a}_2 + c\mathbf{a}_1 \quad \mathbf{a}_3\,]$. Thus a column replacement operation does not affect the volume of the parallelepiped. Since column interchanges have no effect on the volume, the proof is complete. ■

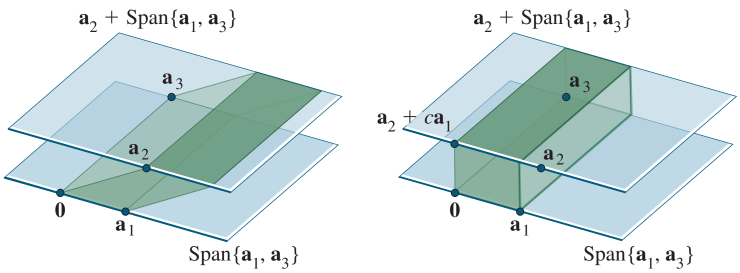


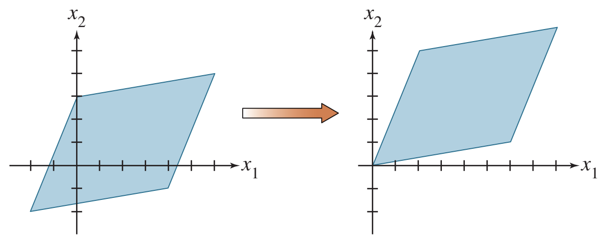
FIGURE 4 Two parallelepipeds of equal volume.

EXAMPLE 4 Calculate the area of the parallelogram determined by the points $(-2, -2)$, $(0, 3)$, $(4, -1)$, and $(6, 4)$. See Figure 5(a).

SOLUTION First translate the parallelogram to one having the origin as a vertex. For example, subtract the vertex $(-2, -2)$ from each of the four vertices. The new parallelogram has the same area, and its vertices are $(0, 0)$, $(2, 5)$, $(6, 1)$, and $(8, 6)$. See Figure 5(b). This parallelogram is determined by the columns of

$$A = \begin{bmatrix} 2 & 6 \\ 5 & 1 \end{bmatrix}$$

Since $|\det A| = |-28|$, the area of the parallelogram is 28. ■



(a) (b)

FIGURE 5 Translating a parallelogram does not change its area.

Linear Transformations

Determinants can be used to describe an important geometric property of linear transformations in the plane and in $\mathbb{R}^3$. If T is a linear transformation and S is a set in the domain of T, let $T(S)$ denote the set of images of points in S. We are interested in how the area (or volume) of $T(S)$ compares with the area (or volume) of the original set S. For convenience, when S is a region bounded by a parallelogram, we also refer to S as a parallelogram.

THEOREM 10

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation determined by a 2×2 matrix A. If S is a parallelogram in $\mathbb{R}^2$, then

$$\{\text{area of } T(S)\} = |\det A| \cdot \{\text{area of } S\} \tag{5}$$

If T is determined by a 3×3 matrix A, and if S is a parallelepiped in $\mathbb{R}^3$, then

$$\{\text{volume of } T(S)\} = |\det A| \cdot \{\text{volume of } S\} \tag{6}$$

PROOF Consider the 2×2 case, with $A = [\,\mathbf{a}_1 \quad \mathbf{a}_2\,]$. A parallelogram at the origin in $\mathbb{R}^2$ determined by vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ has the form

$$S = \{s_1\mathbf{b}_1 + s_2\mathbf{b}_2 : 0 \le s_1 \le 1,\ 0 \le s_2 \le 1\}$$

The image of S under T consists of points of the form

$$\begin{aligned} T(s_1\mathbf{b}_1 + s_2\mathbf{b}_2) &= s_1T(\mathbf{b}_1) + s_2T(\mathbf{b}_2) \\ &= s_1A\mathbf{b}_1 + s_2A\mathbf{b}_2 \end{aligned}$$

where $0 \le s_1 \le 1$, $0 \le s_2 \le 1$. It follows that $T(S)$ is the parallelogram determined by the columns of the matrix $[\,A\mathbf{b}_1 \quad A\mathbf{b}_2\,]$. This matrix can be written as AB, where $B = [\,\mathbf{b}_1 \quad \mathbf{b}_2\,]$. By Theorem 9 and the product theorem for determinants,

$$\begin{aligned} \{\text{area of } T(S)\} &= |\det AB| = |\det A|\,|\det B| \\ &= |\det A| \cdot \{\text{area of } S\} \end{aligned} \tag{7}$$

An arbitrary parallelogram has the form $\mathbf{p} + S$, where $\mathbf{p}$ is a vector and S is a parallelogram at the origin, as seen previously. It is easy to see that T transforms $\mathbf{p} + S$

into $T(\mathbf{p}) + T(S)$. (See Exercise 26.) Since translation does not affect the area of a set,

$$\begin{aligned} \{\text{area of } T(\mathbf{p} + S)\} &= \{\text{area of } T(\mathbf{p}) + T(S)\} \\ &= \{\text{area of } T(S)\} && \text{Translation} \\ &= |\det A| \cdot \{\text{area of } S\} && \text{By equation (7)} \\ &= |\det A| \cdot \{\text{area of } (\mathbf{p} + S)\} && \text{Translation} \end{aligned}$$

This shows that (5) holds for all parallelograms in $\mathbb{R}^2$. The proof of (6) for the 3×3 case is analogous. ∎

When we attempt to generalize Theorem 10 to a region in $\mathbb{R}^2$ or $\mathbb{R}^3$ that is not bounded by straight lines or planes, we must face the problem of how to define and compute its area or volume. This is a question studied in calculus, and we shall only outline the basic idea for $\mathbb{R}^2$. If R is a planar region that has a finite area, then R can be approximated by a grid of small squares that lie inside R. By making the squares sufficiently small, the area of R may be approximated as closely as desired by the sum of the areas of the small squares. See Figure 6.

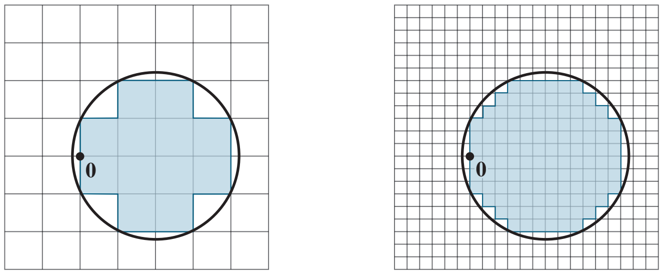


FIGURE 6 Approximating a planar region by a union of squares. The approximation improves as the grid becomes finer.

If T is a linear transformation associated with a 2×2 matrix A, then the image of a planar region R under T is approximated by the images of the small squares inside R. The proof of Theorem 10 shows that each such image is a parallelogram whose area is $|\det A|$ times the area of the square. If R' is the union of the squares inside R, then the area of $T(R')$ is $|\det A|$ times the area of R'. See Figure 7. Also, the area of $T(R')$ is close to the area of $T(R)$. An argument involving a limiting process may be given to justify the following generalization of Theorem 10.

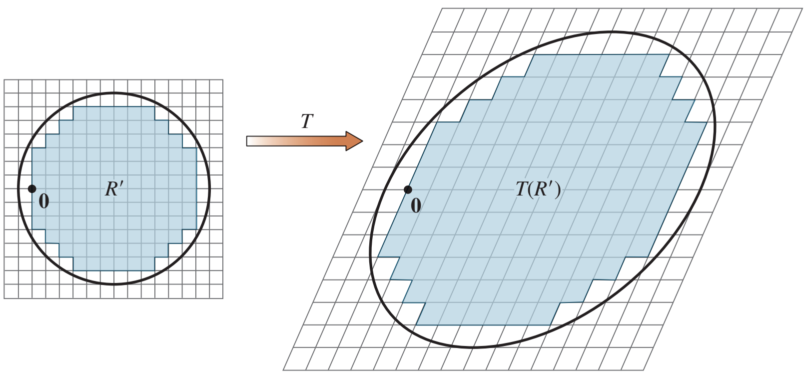
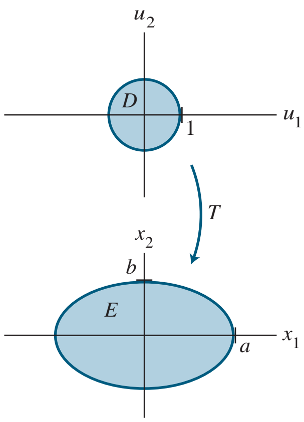


FIGURE 7 Approximating $T(R)$ by a union of parallelograms.

The conclusions of Theorem 10 hold whenever S is a region in $\mathbb{R}^2$ with finite area or a region in $\mathbb{R}^3$ with finite volume.

EXAMPLE 5 Let a and b be positive numbers. Find the area of the region E bounded by the ellipse whose equation is

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$$



SOLUTION We claim that E is the image of the unit disk D under the linear transformation T determined by the matrix $A = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$, because if $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, and $\mathbf{x} = A\mathbf{u}$, then

$$u_1 = \frac{x_1}{a} \quad \text{and} \quad u_2 = \frac{x_2}{b}$$

It follows that $\mathbf{u}$ is in the unit disk, with $u_1^2 + u_2^2 \le 1$, if and only if $\mathbf{x}$ is in E, with $(x_1/a)^2 + (x_2/b)^2 \le 1$. By the generalization of Theorem 10,

$$\begin{aligned} \{\text{area of ellipse}\} &= \{\text{area of } T(D)\} \\ &= |\det A| \cdot \{\text{area of } D\} \\ &= ab\pi(1)^2 = \pi ab \end{aligned}$$

■

Practice Problem

Let S be the parallelogram determined by the vectors $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$, and let $A = \begin{bmatrix} 1 & -.1 \\ 0 & 2 \end{bmatrix}$. Compute the area of the image of S under the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

3.3 Exercises

Use Cramer's rule to compute the solutions of the systems in Exercises 1–6.

1. $\begin{aligned} 5x_1 + 7x_2 &= 3 \\ 2x_1 + 4x_2 &= 1 \end{aligned}$

2. $\begin{aligned} 6x_1 + x_2 &= 3 \\ 5x_1 + 2x_2 &= 4 \end{aligned}$

3. $\begin{aligned} 3x_1 - 2x_2 &= 3 \\ -4x_1 + 6x_2 &= -5 \end{aligned}$

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

5. $\begin{aligned} x_1 + x_2 &= 2 \\ -5x_1 + 4x_3 &= 0 \\ x_2 - x_3 &= -1 \end{aligned}$

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

In Exercises 7–10, determine the values of the parameter s for which the system has a unique solution, and describe the solution.

7. $\begin{aligned} 2sx_1 + 5x_2 &= 8 \\ 6x_1 + 3sx_2 &= 4 \end{aligned}$

8. $\begin{aligned} 3sx_1 + 5x_2 &= 3 \\ 12x_1 + 5sx_2 &= 2 \end{aligned}$

9. $\begin{aligned} sx_1 + 2sx_2 &= -1 \\ 3x_1 + 6sx_2 &= 4 \end{aligned}$

10. $\begin{aligned} sx_1 - 2x_2 &= 1 \\ 4sx_1 + 4sx_2 &= 2 \end{aligned}$

In Exercises 11–16, compute the adjugate of the given matrix, and then use Theorem 8 to give the inverse of the matrix.

11. $\begin{bmatrix} 0 & -2 & -1 \\ 5 & 0 & 0 \\ -1 & 1 & 1 \end{bmatrix}$

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

13. $\begin{bmatrix} 3 & 5 & 4 \\ 1 & 0 & 1 \\ 2 & 1 & 1 \end{bmatrix}$

14. $\begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & 1 \\ 3 & 0 & 6 \end{bmatrix}$

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

16. $\begin{bmatrix} 1 & 2 & 4 \\ 0 & -3 & 1 \\ 0 & 0 & -2 \end{bmatrix}$

17. Show that if A is 2×2, then Theorem 8 gives the same formula for A^{-1} as that given by Theorem 4 in Section 2.2.

18. Suppose that all the entries in A are integers and $\det A = 1$. Explain why all the entries in A^{-1} are integers.

In Exercises 19–22, find the area of the parallelogram whose vertices are listed.

19. $(0,0), (5,2), (6,4), (11,6)$

20. $(0,0), (-3,7), (8,-9), (5,-2)$

21. $(-6,0), (0,5), (4,5), (-2,0)$

22. $(0,-2), (5,-2), (-3,1), (2,1)$

23. Find the volume of the parallelepiped with one vertex at the origin and adjacent vertices at $(1,0,-6)$, $(1,3,5)$, and $(6,7,0)$.

24. Find the volume of the parallelepiped with one vertex at the origin and adjacent vertices at $(1,5,0)$, $(-3,0,3)$, and $(-1,4,-1)$.

25. Use the concept of volume to explain why the determinant of a 3×3 matrix A is zero if and only if A is not invertible. Do not appeal to Theorem 4 in Section 3.2. [*Hint:* Think about the columns of A.]

26. Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation, and let $\mathbf{p}$ be a vector and S a set in $\mathbb{R}^m$. Show that the image of $\mathbf{p} + S$ under T is the translated set $T(\mathbf{p}) + T(S)$ in $\mathbb{R}^n$.

27. Let S be the parallelogram determined by the vectors $\mathbf{b}_1 = \begin{bmatrix} -3 \\ 5 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} -3 \\ 8 \end{bmatrix}$, and let $A = \begin{bmatrix} 3 & -4 \\ -4 & 6 \end{bmatrix}$. Compute the area of the image of S under the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

28. Repeat Exercise 27 with $\mathbf{b}_1 = \begin{bmatrix} -3 \\ 5 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} 0 \\ -3 \end{bmatrix}$, and $A = \begin{bmatrix} 3 & 4 \\ -2 & -2 \end{bmatrix}$.

29. Find a formula for the area of the triangle whose vertices are $\mathbf{0}$, $\mathbf{v}_1$, and $\mathbf{v}_2$ in $\mathbb{R}^2$.

30. Let R be the triangle with vertices at (x_1, y_1), (x_2, y_2), and (x_3, y_3). Show that

$$\{\text{area of triangle}\} = \frac{1}{2} \det \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix}$$

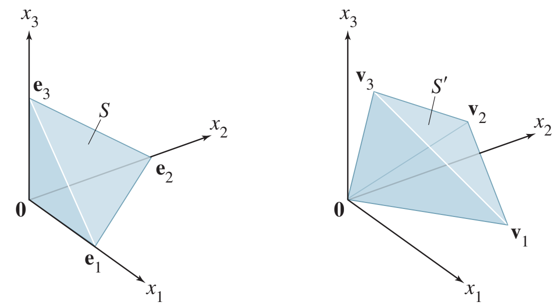
[*Hint:* Translate R to the origin by subtracting one of the vertices, and use Exercise 29.]

31. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation determined by the matrix $A = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$, where a, b, and c are positive numbers. Let S be the unit ball, whose bounding surface has the equation $x_1^2 + x_2^2 + x_3^2 = 1$.

a. Show that $T(S)$ is bounded by the ellipsoid with the equation $\dfrac{x_1^2}{a^2} + \dfrac{x_2^2}{b^2} + \dfrac{x_3^2}{c^2} = 1$.

b. Use the fact that the volume of the unit ball is $4\pi/3$ to determine the volume of the region bounded by the ellipsoid in part (a).

32. Let S be the tetrahedron in $\mathbb{R}^3$ with vertices at the vectors $\mathbf{0}$, $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$, and let S' be the tetrahedron with vertices at vectors $\mathbf{0}$, $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. See the figure.



a. Describe a linear transformation that maps S onto S'.

b. Find a formula for the volume of the tetrahedron S' using the fact that

$\{\text{volume of } S\} = (1/3) \cdot \{\text{area of base}\} \cdot \{\text{height}\}$

33. Let A be an $n \times n$ matrix. If $A^{-1} = \dfrac{1}{\det A} \operatorname{adj} A$ is computed, what should AA^{-1} be equal to in order to confirm that A^{-1} has been found correctly?

34. If a parallelogram fits inside a circle radius 1 and $\det A = 4$, where A is the matrix whose columns correspond to the edges of the parallelogram, does it seem like A and its determinant have been calculated correctly to correspond to the area of this parallelogram? Explain why or why not.

In Exercises 35–38, mark each statement as True or False **(T/F)**. Justify each answer.

35. **(T/F)** Two parallelograms with the same base and height have the same area.

36. **(T/F)** Applying a linear transformation to a region does not change its area.

37. **(T/F)** If A is an invertible $n \times n$ matrix, then $A^{-1} = \operatorname{adj} A$.

38. **(T/F)** Cramer's rule can only be used for invertible matrices.

T **39.** Test the inverse formula of Theorem 8 for a random 4×4 matrix A. Use your matrix program to compute the cofactors of the 3×3 submatrices, construct the adjugate, and set $B = (\text{adj } A)/(\det A)$. Then compute $B - \text{inv}(A)$, where $\text{inv}(A)$ is the inverse of A as computed by the matrix program. Use floating point arithmetic with the maximum possible number of decimal places. Report your results.

T **40.** Test Cramer's rule for a random 4×4 matrix A and a random 4×1 vector $\mathbf{b}$. Compute each entry in the solution of $A\mathbf{x} = \mathbf{b}$, and compare these entries with the entries in $A^{-1}\mathbf{b}$. Write the command (or keystrokes) for your matrix program that uses Cramer's rule to produce the second entry of $\mathbf{x}$.

T **41.** If your version of MATLAB has the `flops` command, use it to count the number of floating point operations to compute A^{-1} for a random 30×30 matrix. Compare this number with the number of flops needed to form $(\text{adj } A)/(\det A)$.

Solution to Practice Problem

The area of S is $\left|\det \begin{bmatrix} 1 & 5 \\ 3 & 1 \end{bmatrix}\right| = 14$, and $\det A = 2$. By Theorem 10, the area of the image of S under the mapping $\mathbf{x} \mapsto A\mathbf{x}$ is

$$|\det A| \cdot \{\text{area of } S\} = 2 \cdot 14 = 28$$

CHAPTER 3 PROJECTS

Chapter 3 projects are available online.

A. *Weighing Design*: This project develops the concept of weighing design and their corresponding matrices for use in weighing a few small, light objects.

B. *Jacobians*: This set of exercises examines how a particular determinant called the Jacobian may be used to allow us to change variables in double and triple integrals.

CHAPTER 3 SUPPLEMENTARY EXERCISES

In Exercises 1-15, mark each statement True or False **(T/F)**. Justify each answer. Assume that all matrices here are square.

1. **(T/F)** If A is a 2×2 matrix with a zero determinant, then one column of A is a multiple of the other.

2. **(T/F)** If two rows of a 3×3 matrix A are the same, then $\det A = 0$.

3. **(T/F)** If A is a 3×3 matrix, then $\det 5A = 5 \det A$.

4. **(T/F)** If A and B are $n \times n$ matrices, with $\det A = 2$ and $\det B = 3$, then $\det(A + B) = 5$.

5. **(T/F)** If A is $n \times n$ and $\det A = 2$, then $\det A^3 = 6$.

6. **(T/F)** If B is produced by interchanging two rows of A, then $\det B = \det A$.

7. **(T/F)** If B is produced by multiplying row 3 of A by 5, then $\det B = 5 \det A$.

8. **(T/F)** If B is formed by adding to one row of A a linear combination of the other rows, then $\det B = \det A$.

9. **(T/F)** $\det A^T = -\det A$.

10. **(T/F)** $\det(-A) = -\det A$.

11. **(T/F)** $\det A^TA \geq 0$.

12. **(T/F)** Any system of n linear equations in n variables can be solved by Cramer's rule.

13. **(T/F)** If $\mathbf{u}$ and $\mathbf{v}$ are in $\mathbb{R}^2$ and $\det[\,\mathbf{u} \quad \mathbf{v}\,] = 10$, then the area of the triangle in the plane with vertices at $\mathbf{0}$, $\mathbf{u}$, and $\mathbf{v}$ is 10.

14. **(T/F)** If $A^3 = 0$, then $\det A = 0$.

15. **(T/F)** If A is invertible, then $\det A^{-1} = \det A$.

Use row operations to show that the determinants in Exercises 16–18 are all zero.

16. $\begin{vmatrix} 12 & 13 & 14 \\ 15 & 16 & 17 \\ 18 & 19 & 20 \end{vmatrix}$

17. $\begin{vmatrix} 1 & a & b+c \\ 1 & b & a+c \\ 1 & c & a+b \end{vmatrix}$

18. $\begin{vmatrix} a & b & c \\ a+x & b+x & c+x \\ a+y & b+y & c+y \end{vmatrix}$

Compute the determinants in Exercises 19 and 20.

19. $\begin{vmatrix} 1 & 5 & 4 & 3 & 2 \\ 0 & 8 & 5 & 9 & 0 \\ 0 & 7 & 0 & 0 & 0 \\ 3 & 9 & 6 & 5 & 4 \\ 0 & 8 & 0 & 6 & 0 \end{vmatrix}$

virtually and synthesized using DSP. In Example 3 of Section 4.7, we see how signal processing can be used to add richness to virtual sounds.

Discrete-time signals and DSP have become significant tools in many industries and areas of research. Mathematically speaking, discrete-time signals can be viewed as vectors that are processed using linear transformations. The operations of adding, scaling, and applying linear transformations to signals is completely analogous to the same operations for vectors in $\mathbb{R}^n$. For this reason, the set of all possible signals, $\mathbb{S}$, is treated as a *vector space*. In Sections 4.7 and 4.8, we look at the vector space of discrete-time signals in more detail.

The focus of Chapter 4 is to extend the theory of vectors in $\mathbb{R}^n$ to include signals and other mathematical structures that behave like the vectors you are already familiar with. Later on in the text, you will see how other vector spaces and their corresponding linear transformations arise in engineering, physics, biology, and statistics.

The mathematical seeds planted in Chapters 1 and 2 germinate and begin to blossom in this chapter. The beauty and power of linear algebra will be seen more clearly when you view $\mathbb{R}^n$ as only one of a variety of vector spaces that arise naturally in applied problems.

Beginning with basic definitions in Section 4.1, the general vector space framework develops gradually throughout the chapter. A goal of Sections 4.5 and 4.6 is to demonstrate how closely other vector spaces resemble $\mathbb{R}^n$. Sections 4.7 and 4.8 apply the theory of this chapter to discrete-time signals, DSP, and difference equations—the mathematics underlying the digital revolution.

4.1 Vector Spaces and Subspaces

Much of the theory in Chapters 1 and 2 rested on certain simple and obvious algebraic properties of $\mathbb{R}^n$, listed in Section 1.3. In fact, many other mathematical systems have the same properties. The specific properties of interest are listed in the following definition.

DEFINITION

A **vector space** is a nonempty set V of objects, called *vectors*, on which are defined two operations, called *addition* and *multiplication by scalars* (real numbers), subject to the ten axioms (or rules) listed below.[1] The axioms must hold for all vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V and for all scalars c and d.

1. The sum of $\mathbf{u}$ and $\mathbf{v}$, denoted by $\mathbf{u} + \mathbf{v}$, is in V.
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
3. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.
4. There is a **zero** vector $\mathbf{0}$ in V such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$.
5. For each $\mathbf{u}$ in V, there is a vector $-\mathbf{u}$ in V such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.

[1] Technically, V is a *real vector space*. All of the theory in this chapter also holds for a *complex vector space* in which the scalars and matrix entries are complex numbers. We will look at this briefly in Chapter 5. Until then, all scalars and matrix entries are assumed to be real.

6. The scalar multiple of $\mathbf{u}$ by c, denoted by $c\mathbf{u}$, is in V.

7. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$.

8. $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$.

9. $c(d\mathbf{u}) = (cd)\mathbf{u}$.

10. $1\mathbf{u} = \mathbf{u}$.

Using only these axioms, one can show that the zero vector in Axiom 4 is unique, and the vector $-\mathbf{u}$, called the **negative** of $\mathbf{u}$, in Axiom 5 is unique for each $\mathbf{u}$ in V. See Exercises 33 and 34. Proofs of the following simple facts are also outlined in the exercises:

For each $\mathbf{u}$ in V and scalar c,

$$0\mathbf{u} = \mathbf{0} \tag{1}$$

$$c\mathbf{0} = \mathbf{0} \tag{2}$$

$$-\mathbf{u} = (-1)\mathbf{u} \tag{3}$$

EXAMPLE 1 The spaces $\mathbb{R}^n$, where $n \geq 1$, are the premier examples of vector spaces. The geometric intuition developed for $\mathbb{R}^3$ will help you understand and visualize many concepts throughout the chapter. ■

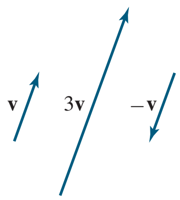


FIGURE 1

EXAMPLE 2 Let V be the set of all arrows (directed line segments) in three-dimensional space, with two arrows regarded as equal if they have the same length and point in the same direction. Define addition by the parallelogram rule (from Section 1.3), and for each $\mathbf{v}$ in V, define $c\mathbf{v}$ to be the arrow whose length is $|c|$ times the length of $\mathbf{v}$, pointing in the same direction as $\mathbf{v}$ if $c \geq 0$ and otherwise pointing in the opposite direction. (See Figure 1.) Show that V is a vector space. This space is a common model in physical problems for various forces.

SOLUTION The definition of V is geometric, using concepts of length and direction. No xyz-coordinate system is involved. An arrow of zero length is a single point and represents the zero vector. The negative of $\mathbf{v}$ is $(-1)\mathbf{v}$. So Axioms 1, 4, 5, 6, and 10 are evident. The rest are verified by geometry. For instance, see Figures 2 and 3. ■

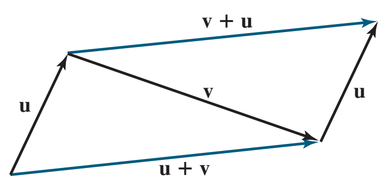


FIGURE 2 $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.

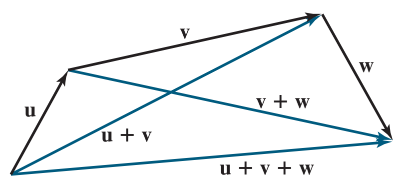


FIGURE 3 $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.

EXAMPLE 3 Let $\mathbb{S}$ be the space of all doubly infinite sequences of numbers (usually written in a row rather than a column):

$$\{y_k\} = (\ldots, y_{-2}, y_{-1}, y_0, y_1, y_2, \ldots)$$

If $\{z_k\}$ is another element of $\mathbb{S}$, then the sum $\{y_k\} + \{z_k\}$ is the sequence $\{y_k + z_k\}$ formed by adding corresponding terms of $\{y_k\}$ and $\{z_k\}$. The scalar multiple $c\,\{y_k\}$ is the sequence $\{cy_k\}$. The vector space axioms are verified in the same way as for $\mathbb{R}^n$.

Elements of $\mathbb{S}$ arise in engineering, for example, whenever a signal is measured (or sampled) at discrete times. A signal might be electrical, mechanical, optical, biological, audio, and so on. The digital signal processors mentioned in the chapter introduction use discrete (or digital) signals. For convenience, we will call $\mathbb{S}$ the space of (discrete-time) **signals**. A signal may be visualized by a graph as in Figure 4.

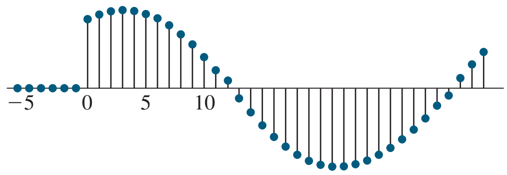


FIGURE 4 A discrete-time signal. ■

EXAMPLE 4 For $n \geq 0$, the set $\mathbb{P}_n$ of polynomials of degree at most n consists of all polynomials of the form

$$\mathbf{p}(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n \tag{4}$$

where the coefficients $a_0, \ldots, a_n$ and the variable t are real numbers. The *degree* of $\mathbf{p}$ is the highest power of t in (4) whose coefficient is not zero. If $\mathbf{p}(t) = a_0 \neq 0$, the degree of $\mathbf{p}$ is zero. If all the coefficients are zero, $\mathbf{p}$ is called the *zero polynomial*. The zero polynomial is included in $\mathbb{P}_n$ even though its degree, for technical reasons, is not defined.

If $\mathbf{p}$ is given by (4) and if $\mathbf{q}(t) = b_0 + b_1 t + \cdots + b_n t^n$, then the sum $\mathbf{p} + \mathbf{q}$ is defined by

$$\begin{aligned}(\mathbf{p} + \mathbf{q})(t) &= \mathbf{p}(t) + \mathbf{q}(t) \\ &= (a_0 + b_0) + (a_1 + b_1)t + \cdots + (a_n + b_n)t^n\end{aligned}$$

The scalar multiple $c\mathbf{p}$ is the polynomial defined by

$$(c\mathbf{p})(t) = c\mathbf{p}(t) = ca_0 + (ca_1)t + \cdots + (ca_n)t^n$$

These definitions satisfy Axioms 1 and 6 because $\mathbf{p} + \mathbf{q}$ and $c\mathbf{p}$ are polynomials of degree less than or equal to n. Axioms 2, 3, and 7–10 follow from properties of the real numbers. Clearly, the zero polynomial acts as the zero vector in Axiom 4. Finally, $(-1)\mathbf{p}$ acts as the negative of $\mathbf{p}$, so Axiom 5 is satisfied. Thus $\mathbb{P}_n$ is a vector space.

The vector spaces $\mathbb{P}_n$ for various n are used, for instance, in statistical trend analysis of data, discussed in Section 6.8. ■

EXAMPLE 5 Let V be the set of all real-valued functions defined on a set $\mathbb{D}$. (Typically, $\mathbb{D}$ is the set of real numbers or some interval on the real line.) Functions are added in the usual way: $\mathbf{f} + \mathbf{g}$ is the function whose value at t in the domain $\mathbb{D}$ is $\mathbf{f}(t) + \mathbf{g}(t)$. Likewise, for a scalar c and an $\mathbf{f}$ in V, the scalar multiple $c\mathbf{f}$ is the function whose value at t is $c\mathbf{f}(t)$. For instance, if $\mathbb{D} = \mathbb{R}$, $\mathbf{f}(t) = 1 + \sin 2t$, and $\mathbf{g}(t) = 2 + .5t$, then

$$(\mathbf{f} + \mathbf{g})(t) = 3 + \sin 2t + .5t \quad \text{and} \quad (2\mathbf{g})(t) = 4 + t$$

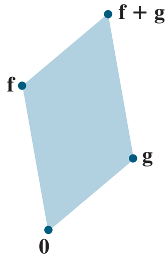


FIGURE 5
The sum of two vectors (functions).

Two functions in V are equal if and only if their values are equal for every t in $\mathbb{D}$. Hence the zero vector in V is the function that is identically zero, $\mathbf{f}(t) = 0$ for all t, and the negative of $\mathbf{f}$ is $(-1)\mathbf{f}$. Axioms 1 and 6 are obviously true, and the other axioms follow from properties of the real numbers, so V is a vector space. ■

It is important to think of each function in the vector space V of Example 5 as a single object, as just one "point" or vector in the vector space. The sum of two vectors $\mathbf{f}$ and $\mathbf{g}$ (functions in V, or elements of *any* vector space) can be visualized as in Figure 5, because this can help you carry over to a general vector space the geometric intuition you have developed while working with the vector space $\mathbb{R}^n$. See the *Study Guide* for help as you learn to adopt this more general point of view.

Subspaces

In many problems, a vector space consists of an appropriate subset of vectors from some larger vector space. In this case, only three of the ten vector space axioms need to be checked; the rest are automatically satisfied.

DEFINITION

A **subspace** of a vector space V is a subset H of V that has three properties:

a. The zero vector of V is in H.[2]
b. H is closed under vector addition. That is, for each $\mathbf{u}$ and $\mathbf{v}$ in H, the sum $\mathbf{u} + \mathbf{v}$ is in H.
c. H is closed under multiplication by scalars. That is, for each $\mathbf{u}$ in H and each scalar c, the vector $c\mathbf{u}$ is in H.

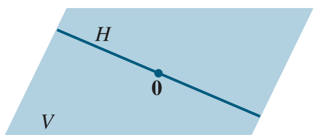


FIGURE 6
A subspace of V.

Properties (a), (b), and (c) guarantee that a subspace H of V is itself a *vector space*, under the vector space operations already defined in V. To verify this, note that properties (a), (b), and (c) are Axioms 1, 4, and 6. Axioms 2, 3, and 7–10 are automatically true in H because they apply to all elements of V, including those in H. Axiom 5 is also true in H, because if $\mathbf{u}$ is in H, then $(-1)\mathbf{u}$ is in H by property (c), and we know from equation (3) earlier in this section that $(-1)\mathbf{u}$ is the vector $-\mathbf{u}$ in Axiom 5.

So every subspace is a vector space. Conversely, every vector space is a subspace (of itself and possibly of other larger spaces). The term *subspace* is used when at least two vector spaces are in mind, with one inside the other, and the phrase *subspace of V* identifies V as the larger space. (See Figure 6.)

EXAMPLE 6 The set consisting of only the zero vector in a vector space V is a subspace of V, called the **zero subspace** and written as $\{\mathbf{0}\}$. ■

EXAMPLE 7 Let $\mathbb{P}$ be the set of all polynomials with real coefficients, with operations in $\mathbb{P}$ defined as for functions. Then $\mathbb{P}$ is a subspace of the space of all real-valued functions defined on $\mathbb{R}$. Also, for each $n \geq 0$, $\mathbb{P}_n$ is a subspace of $\mathbb{P}$, because $\mathbb{P}_n$ is a subset of $\mathbb{P}$ that contains the zero polynomial, the sum of two polynomials in $\mathbb{P}_n$ is also in $\mathbb{P}_n$, and a scalar multiple of a polynomial in $\mathbb{P}_n$ is also in $\mathbb{P}_n$. ■

[2] Some texts replace property (a) in this definition by the assumption that H is nonempty. Then (a) could be deduced from (c) and the fact that $0\mathbf{u} = \mathbf{0}$. But the best way to test for a subspace is to look first for the zero vector. If $\mathbf{0}$ is in H, then properties (b) and (c) must be checked. If $\mathbf{0}$ is *not* in H, then H cannot be a subspace and the other properties need not be checked.

EXAMPLE 8 The set of finitely supported signals $\mathbb{S}_f$ consists of the signals $\{y_k\}$, where only finitely many of the y_k are nonzero. Since the zero signal $\mathbf{0} = (\ldots, 0, 0, 0, \ldots)$ has no nonzero entries, it is clearly an element of $\mathbb{S}_f$. If two signals with finitely many nonzeros are added, the resulting signal will have finitely many nonzeros. Similarly if a signal with finitely many nonzeros is scaled, the result will still have finitely many nonzeros. Thus $\mathbb{S}_f$ is a subspace of $\mathbb{S}$, the discrete-time signals. See Figure 7.

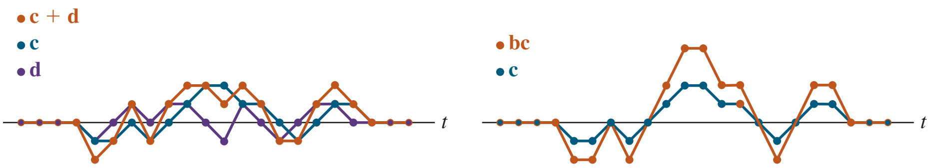


FIGURE 7 ■

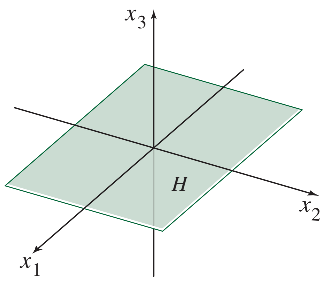


FIGURE 8
The x_1x_2-plane as a subspace of $\mathbb{R}^3$.

EXAMPLE 9 The vector space $\mathbb{R}^2$ is *not* a subspace of $\mathbb{R}^3$ because $\mathbb{R}^2$ is not even a subset of $\mathbb{R}^3$. (The vectors in $\mathbb{R}^3$ all have three entries, whereas the vectors in $\mathbb{R}^2$ have only two.) The set

$$H = \left\{ \begin{bmatrix} s \\ t \\ 0 \end{bmatrix} : s \text{ and } t \text{ are real} \right\}$$

is a subset of $\mathbb{R}^3$ that "looks" and "acts" like $\mathbb{R}^2$, although it is logically distinct from $\mathbb{R}^2$. See Figure 8. Show that H is a subspace of $\mathbb{R}^3$.

SOLUTION The zero vector is in H, and H is closed under vector addition and scalar multiplication because these operations on vectors in H always produce vectors whose third entries are zero (and so belong to H). Thus H is a subspace of $\mathbb{R}^3$. ■

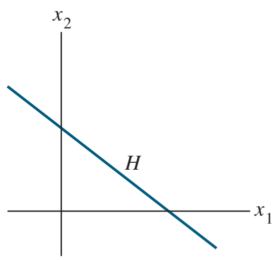


FIGURE 9
A line that is not a vector space.

EXAMPLE 10 A plane in $\mathbb{R}^3$ *not* through the origin is not a subspace of $\mathbb{R}^3$, because the plane does not contain the zero vector of $\mathbb{R}^3$. Similarly, a line in $\mathbb{R}^2$ *not* through the origin, such as in Figure 9, is *not* a subspace of $\mathbb{R}^2$. ■

A Subspace Spanned by a Set

The next example illustrates one of the most common ways of describing a subspace. As in Chapter 1, the term **linear combination** refers to any sum of scalar multiples of vectors, and Span $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ denotes the set of all vectors that can be written as linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_p$.

EXAMPLE 11 Given $\mathbf{v}_1$ and $\mathbf{v}_2$ in a vector space V, let $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$. Show that H is a subspace of V.

SOLUTION The zero vector is in H, since $\mathbf{0} = 0\mathbf{v}_1 + 0\mathbf{v}_2$. To show that H is closed under vector addition, take two arbitrary vectors in H, say,

$$\mathbf{u} = s_1\mathbf{v}_1 + s_2\mathbf{v}_2 \quad \text{and} \quad \mathbf{w} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2$$

By Axioms 2, 3, and 8 for the vector space V,

$$\begin{aligned} \mathbf{u} + \mathbf{w} &= (s_1\mathbf{v}_1 + s_2\mathbf{v}_2) + (t_1\mathbf{v}_1 + t_2\mathbf{v}_2) \\ &= (s_1 + t_1)\mathbf{v}_1 + (s_2 + t_2)\mathbf{v}_2 \end{aligned}$$

So $\mathbf{u} + \mathbf{w}$ is in H. Furthermore, if c is any scalar, then by Axioms 7 and 9,

$$c\mathbf{u} = c(s_1\mathbf{v}_1 + s_2\mathbf{v}_2) = (cs_1)\mathbf{v}_1 + (cs_2)\mathbf{v}_2$$

which shows that $c\mathbf{u}$ is in H and H is closed under scalar multiplication. Thus H is a subspace of V. ■

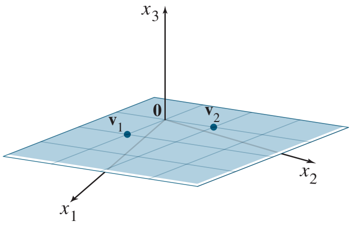


FIGURE 10
An example of a subspace.

In Section 4.5, you will see that every nonzero subspace of $\mathbb{R}^3$, other than $\mathbb{R}^3$ itself, is either Span $\{\mathbf{v}_1, \mathbf{v}_2\}$ for some linearly independent $\mathbf{v}_1$ and $\mathbf{v}_2$ or Span $\{\mathbf{v}\}$ for $\mathbf{v} \neq \mathbf{0}$. In the first case, the subspace is a plane through the origin; in the second case, it is a line through the origin. (See Figure 10.) It is helpful to keep these geometric pictures in mind, even for an abstract vector space.

The argument in Example 11 can easily be generalized to prove the following theorem.

THEOREM 1

If $\mathbf{v}_1, \dots, \mathbf{v}_p$ are in a vector space V, then Span $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is a subspace of V.

We call Span $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ **the subspace spanned** (or **generated**) by $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$. Given any subspace H of V, a **spanning** (or **generating**) **set** for H is a set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ in H such that $H = \text{Span}\,\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$.

The next example shows how to use Theorem 1.

EXAMPLE 12 Let H be the set of all vectors of the form $(a - 3b, b - a, a, b)$, where a and b are arbitrary scalars. That is, let $H = \{(a - 3b, b - a, a, b) : a \text{ and } b \text{ in } \mathbb{R}\}$. Show that H is a subspace of $\mathbb{R}^4$.

SOLUTION Write the vectors in H as column vectors. Then an arbitrary vector in H has the form

$$\begin{bmatrix} a - 3b \\ b - a \\ a \\ b \end{bmatrix} = a \underbrace{\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{v}_1} + b \underbrace{\begin{bmatrix} -3 \\ 1 \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{v}_2}$$

This calculation shows that $H = \text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2\}$, where $\mathbf{v}_1$ and $\mathbf{v}_2$ are the vectors indicated above. Thus H is a subspace of $\mathbb{R}^4$ by Theorem 1. ■

Example 12 illustrates a useful technique of expressing a subspace H as the set of linear combinations of some small collection of vectors. If $H = \text{Span}\,\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$, we can think of the vectors $\mathbf{v}_1, \dots, \mathbf{v}_p$ in the spanning set as "handles" that allow us to hold on to the subspace H. Calculations with the infinitely many vectors in H are often reduced to operations with the finite number of vectors in the spanning set.

EXAMPLE 13 For what value(s) of h will $\mathbf{y}$ be in the subspace of $\mathbb{R}^3$ spanned by $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, if

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 5 \\ -4 \\ -7 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} -4 \\ 3 \\ h \end{bmatrix}$$

SOLUTION This question is Practice Problem 2 in Section 1.3, written here with the term *subspace* rather than Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. The solution there shows that $\mathbf{y}$ is in Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ if and only if $h = 5$. That solution is worth reviewing now, along with Exercises 11–16 and 19–21 in Section 1.3. ■

Although many vector spaces in this chapter will be subspaces of $\mathbb{R}^n$, it is important to keep in mind that the abstract theory applies to other vector spaces as well. Vector spaces of functions arise in many applications, and they will receive more attention later.

Practice Problems

1. Show that the set H of all points in $\mathbb{R}^2$ of the form $(3s, 2 + 5s)$ is not a vector space, by showing that it is not closed under scalar multiplication. (Find a specific vector $\mathbf{u}$ in H and a scalar c such that $c\mathbf{u}$ is not in H.)
2. Let $W = \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$, where $\mathbf{v}_1, \dots, \mathbf{v}_p$ are in a vector space V. Show that $\mathbf{v}_k$ is in W for $1 \le k \le p$. [*Hint:* First write an equation that shows that $\mathbf{v}_1$ is in W. Then adjust your notation for the general case.]
3. An $n \times n$ matrix A is said to be symmetric if $A^T = A$. Let S be the set of all 3×3 symmetric matrices. Show that S is a subspace of $M_{3\times 3}$, the vector space of 3×3 matrices.

4.1 Exercises

1. Let V be the first quadrant in the xy-plane; that is, let
$$V = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \ge 0, y \ge 0 \right\}$$
   a. If $\mathbf{u}$ and $\mathbf{v}$ are in V, is $\mathbf{u} + \mathbf{v}$ in V? Why?
   b. Find a specific vector $\mathbf{u}$ in V and a specific scalar c such that $c\mathbf{u}$ is *not* in V. (This is enough to show that V is *not* a vector space.)

2. Let W be the union of the first and third quadrants in the xy-plane. That is, let $W = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : xy \ge 0 \right\}$.
   a. If $\mathbf{u}$ is in W and c is any scalar, is $c\mathbf{u}$ in W? Why?
   b. Find specific vectors $\mathbf{u}$ and $\mathbf{v}$ in W such that $\mathbf{u} + \mathbf{v}$ is not in W. (This is enough to show that W is *not* a vector space.)

3. Let H be the set of points inside and on the unit circle in the xy-plane. That is, let $H = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x^2 + y^2 \le 1 \right\}$. Find a specific example—two vectors or a vector and a scalar—to show that H is not a subspace of $\mathbb{R}^2$.

4. Construct a geometric figure that illustrates why a line in $\mathbb{R}^2$ *not* through the origin is not closed under vector addition.

In Exercises 5–8, determine if the given set is a subspace of $\mathbb{P}_n$ for an appropriate value of n. Justify your answers.

5. All polynomials of the form $\mathbf{p}(t) = at^2$, where a is in $\mathbb{R}$.

6. All polynomials of the form $\mathbf{p}(t) = a + t^2$, where a is in $\mathbb{R}$.

7. All polynomials of degree at most 3, with integers as coefficients.

8. All polynomials in $\mathbb{P}_n$ such that $\mathbf{p}(0) = 0$.

9. Let H be the set of all vectors of the form $\begin{bmatrix} s \\ 3s \\ 2s \end{bmatrix}$. Find a vector $\mathbf{v}$ in $\mathbb{R}^3$ such that $H = \text{Span}\{\mathbf{v}\}$. Why does this show that H is a subspace of $\mathbb{R}^3$?

10. Let H be the set of all vectors of the form $\begin{bmatrix} 2t \\ 0 \\ -t \end{bmatrix}$. Show that H is a subspace of $\mathbb{R}^3$. (Use the method of Exercise 9.)

11. Let W be the set of all vectors of the form $\begin{bmatrix} 6b + 7c \\ b \\ c \end{bmatrix}$, where b and c are arbitrary. Find vectors $\mathbf{u}$ and $\mathbf{v}$ such that $W = \text{Span}\ \mathbf{u}, \mathbf{v}$. Why does this show that W is a subspace of $\mathbb{R}^3$?

12. Let W be the set of all vectors of the form $\begin{bmatrix} s + 3t \\ s - t \\ 2s - t \\ 4t \end{bmatrix}$. Show that W is a subspace of $\mathbb{R}^4$. (Use the method of Exercise 11.)

13. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 4 \\ 2 \\ 6 \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$.

a. Is $\mathbf{w}$ in $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$? How many vectors are in $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$?

b. How many vectors are in Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$?

c. Is $\mathbf{w}$ in the subspace spanned by $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$? Why?

14. Let $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ be as in Exercise 13, and let $\mathbf{w} = \begin{bmatrix} 8 \\ 4 \\ 7 \end{bmatrix}$. Is $\mathbf{w}$ in the subspace spanned by $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$? Why?

In Exercises 15–18, let W be the set of all vectors of the form shown, where a, b, and c represent arbitrary real numbers. In each case, either find a set S of vectors that spans W or give an example to show that W is *not* a vector space.

15. $\begin{bmatrix} 3a + b \\ 4 \\ a - 5b \end{bmatrix}$

16. $\begin{bmatrix} -a + 1 \\ a - 6b \\ 2b + a \end{bmatrix}$

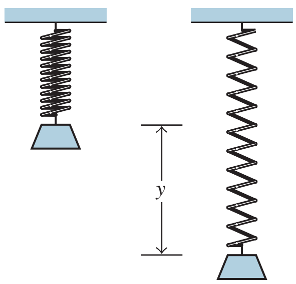
17. $\begin{bmatrix} a - b \\ b - c \\ c - a \\ b \end{bmatrix}$

18. $\begin{bmatrix} 4a + 3b \\ 0 \\ a + b + c \\ c - 2a \end{bmatrix}$

19. If a mass m is placed at the end of a spring, and if the mass is pulled downward and released, the mass–spring system will begin to oscillate. The displacement y of the mass from its resting position is given by a function of the form

$$y(t) = c_1 \cos \omega t + c_2 \sin \omega t \tag{5}$$

where ω is a constant that depends on the spring and the mass. (See the figure below.) Show that the set of all functions described in (5) (with ω fixed and c_1, c_2 arbitrary) is a vector space.



20. The set of all continuous real-valued functions defined on a closed interval $[a, b]$ in $\mathbb{R}$ is denoted by $C[a, b]$. This set is a subspace of the vector space of all real-valued functions defined on $[a, b]$.

a. What facts about continuous functions should be proved in order to demonstrate that $C[a, b]$ is indeed a subspace as claimed? (These facts are usually discussed in a calculus class.)

b. Show that $\{\mathbf{f} \text{ in } C[a, b] : \mathbf{f}(a) = \mathbf{f}(b)\}$ is a subspace of $C[a, b]$.

For fixed positive integers m and n, the set $M_{m \times n}$ of all $m \times n$ matrices is a vector space, under the usual operations of addition of matrices and multiplication by real scalars.

21. Determine if the set H of all matrices of the form $\begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$ is a subspace of $M_{2\times 2}$.

22. Let F be a fixed 3×2 matrix, and let H be the set of all matrices A in $M_{2\times 4}$ with the property that $FA = 0$ (the zero matrix in $M_{3\times 4}$). Determine if H is a subspace of $M_{2\times 4}$.

In Exercises 23–32, mark each statement True or False **(T/F)**. Justify each answer.

23. **(T/F)** If $\mathbf{f}$ is a function in the vector space V of all real-valued functions on $\mathbb{R}$ and if $\mathbf{f}(t) = 0$ for some t, then $\mathbf{f}$ is the zero vector in V.

24. **(T/F)** A vector is any element of a vector space.

25. **(T/F)** An arrow in three-dimensional space can be considered to be a vector.

26. **(T/F)** If $\mathbf{u}$ is a vector in a vector space V, then $(-1)\,\mathbf{u}$ is the same as the negative of $\mathbf{u}$.

27. **(T/F)** A subset H of a vector space V is a subspace of V if the zero vector is in H.

28. **(T/F)** A vector space is also a subspace.

29. **(T/F)** A subspace is also a vector space.

30. **(T/F)** $\mathbb{R}^2$ is a subspace of $\mathbb{R}^3$.

31. **(T/F)** The polynomials of degree two or less are a subspace of the polynomials of degree three or less.

32. **(T/F)** A subset H of a vector space V is a subspace of V if the following conditions are satisfied: (i) the zero vector of V is in H, (ii) $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u} + \mathbf{v}$ are in H, and (iii) c is a scalar and $c\mathbf{u}$ is in H.

Exercises 33–36 show how the axioms for a vector space V can be used to prove the elementary properties described after the definition of a vector space. Fill in the blanks with the appropriate axiom numbers. Because of Axiom 2, Axioms 4 and 5 imply, respectively, that $\mathbf{0} + \mathbf{u} = \mathbf{u}$ and $-\mathbf{u} + \mathbf{u} = \mathbf{0}$ for all $\mathbf{u}$.

33. Complete the following proof that the zero vector is unique. Suppose that $\mathbf{w}$ in V has the property that $\mathbf{u} + \mathbf{w} = \mathbf{w} + \mathbf{u} = \mathbf{u}$ for all $\mathbf{u}$ in V. In particular, $\mathbf{0} + \mathbf{w} = \mathbf{0}$. But $\mathbf{0} + \mathbf{w} = \mathbf{w}$, by Axiom _____. Hence $\mathbf{w} = \mathbf{0} + \mathbf{w} = \mathbf{0}$.

34. Complete the following proof that $-\mathbf{u}$ is the *unique vector* in V such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$. Suppose that $\mathbf{w}$ satisfies $\mathbf{u} + \mathbf{w} = \mathbf{0}$. Adding $-\mathbf{u}$ to both sides, we have

$$(-\mathbf{u}) + [\mathbf{u} + \mathbf{w}] = (-\mathbf{u}) + \mathbf{0}$$

$$[(-\mathbf{u}) + \mathbf{u}] + \mathbf{w} = (-\mathbf{u}) + \mathbf{0} \qquad \text{by Axiom \_\_\_\_\_ (a)}$$

$$\mathbf{0} + \mathbf{w} = (-\mathbf{u}) + \mathbf{0} \qquad \text{by Axiom \_\_\_\_\_ (b)}$$

$$\mathbf{w} = -\mathbf{u} \qquad \text{by Axiom \_\_\_\_\_ (c)}$$

35. Fill in the missing axiom numbers in the following proof that $0\mathbf{u} = \mathbf{0}$ for every $\mathbf{u}$ in V.

$$0\mathbf{u} = (0 + 0)\mathbf{u} = 0\mathbf{u} + 0\mathbf{u} \qquad \text{by Axiom \_\_\_\_\_ (a)}$$

Add the negative of $0\mathbf{u}$ to both sides:

$0\mathbf{u} + (-0\mathbf{u}) = [0\mathbf{u} + 0\mathbf{u}] + (-0\mathbf{u})$

$0\mathbf{u} + (-0\mathbf{u}) = 0\mathbf{u} + [0\mathbf{u} + (-0\mathbf{u})]$ by Axiom ____ (b)

$\mathbf{0} = 0\mathbf{u} + \mathbf{0}$ by Axiom ____ (c)

$\mathbf{0} = 0\mathbf{u}$ by Axiom ____ (d)

36. Fill in the missing axiom numbers in the following proof that $c\mathbf{0} = \mathbf{0}$ for every scalar c.

$c\mathbf{0} = c(\mathbf{0} + \mathbf{0})$ by Axiom ____ (a)

$= c\mathbf{0} + c\mathbf{0}$ by Axiom ____ (b)

Add the negative of $c\mathbf{0}$ to both sides:

$c\mathbf{0} + (-c\mathbf{0}) = [c\mathbf{0} + c\mathbf{0}] + (-c\mathbf{0})$

$c\mathbf{0} + (-c\mathbf{0}) = c\mathbf{0} + [c\mathbf{0} + (-c\mathbf{0})]$ by Axiom ____ (c)

$\mathbf{0} = c\mathbf{0} + \mathbf{0}$ by Axiom ____ (d)

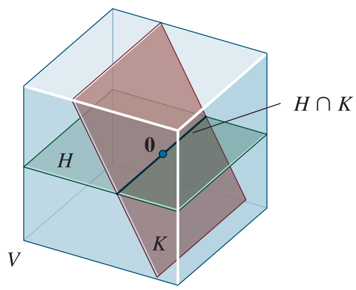
$\mathbf{0} = c\mathbf{0}$ by Axiom ____ (e)

37. Prove that $(-1)\mathbf{u} = -\mathbf{u}$. [*Hint:* Show that $\mathbf{u} + (-1)\mathbf{u} = \mathbf{0}$. Use some axioms and the results of Exercises 34 and 35.]

38. Suppose $c\mathbf{u} = \mathbf{0}$ for some nonzero scalar c. Show that $\mathbf{u} = \mathbf{0}$. Mention the axioms or properties you use.

39. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in a vector space V, and let H be any subspace of V that contains both $\mathbf{u}$ and $\mathbf{v}$. Explain why H also contains Span $\{\mathbf{u}, \mathbf{v}\}$. This shows that Span $\{\mathbf{u}, \mathbf{v}\}$ is the smallest subspace of V that contains both $\mathbf{u}$ and $\mathbf{v}$.

40. Let H and K be subspaces of a vector space V. The **intersection** of H and K, written as $H \cap K$, is the set of $\mathbf{v}$ in V that belong to both H and K. Show that $H \cap K$ is a subspace of V. (See the figure.) Give an example in $\mathbb{R}^2$ to show that the union of two subspaces is not, in general, a subspace.



41. Given subspaces H and K of a vector space V, the **sum** of H and K, written as $H + K$, is the set of all vectors in V that can be written as the sum of two vectors, one in H and the other in K; that is,

$$H + K = \{\mathbf{w} : \mathbf{w} = \mathbf{u} + \mathbf{v} \text{ for some } \mathbf{u} \text{ in } H \text{ and some } \mathbf{v} \text{ in } K\}$$

a. Show that $H + K$ is a subspace of V.

b. Show that H is a subspace of $H + K$ and K is a subspace of $H + K$.

42. Suppose $\mathbf{u}_1, \dots, \mathbf{u}_p$ and $\mathbf{v}_1, \dots, \mathbf{v}_q$ are vectors in a vector space V, and let

$$H = \text{Span}\{\mathbf{u}_1, \dots, \mathbf{u}_p\} \text{ and } K = \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_q\}$$

Show that $H + K = \text{Span}\{\mathbf{u}_1, \dots, \mathbf{u}_p, \mathbf{v}_1, \dots, \mathbf{v}_q\}$.

T 43. Show that $\mathbf{w}$ is in the subspace of $\mathbb{R}^4$ spanned by $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, where

$$\mathbf{w} = \begin{bmatrix} 6 \\ -7 \\ 8 \\ -9 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 7 \\ -6 \\ -5 \\ 4 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -3 \\ 2 \\ -1 \\ -4 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} -2 \\ 1 \\ 2 \\ -5 \end{bmatrix}$$

T 44. Determine if $\mathbf{y}$ is in the subspace of $\mathbb{R}^4$ spanned by the columns of A, where

$$\mathbf{y} = \begin{bmatrix} -4 \\ -8 \\ 6 \\ -5 \end{bmatrix}, \quad A = \begin{bmatrix} 3 & -5 & -9 \\ 8 & 7 & -6 \\ -5 & -8 & 3 \\ 2 & -2 & -9 \end{bmatrix}$$

T 45. The vector space $H = \text{Span}\{1, \cos^2 t, \cos^4 t, \cos^6 t\}$ contains at least two interesting functions that will be used in a later exercise:

$$\mathbf{f}(t) = 1 - 8\cos^2 t + 8\cos^4 t$$

$$\mathbf{g}(t) = -1 + 18\cos^2 t - 48\cos^4 t + 32\cos^6 t$$

Study the graph of $\mathbf{f}$ for $0 \le t \le 2\pi$, and guess a simple formula for $\mathbf{f}(t)$. Verify your conjecture by graphing the difference between $1 + \mathbf{f}(t)$ and your formula for $\mathbf{f}(t)$. (Hopefully, you will see the constant function 1.) Repeat for $\mathbf{g}$.

T 46. Repeat Exercise 45 for the functions

$$\mathbf{f}(t) = 3\sin t - 4\sin^3 t$$

$$\mathbf{g}(t) = 1 - 8\sin^2 t + 8\sin^4 t$$

$$\mathbf{h}(t) = 5\sin t - 20\sin^3 t + 16\sin^5 t$$

in the vector space Span $\{1, \sin t, \sin^2 t, \dots, \sin^5 t\}$.

Solutions to Practice Problems

1. Take any $\mathbf{u}$ in H—say, $\mathbf{u} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}$—and take any $c \neq 1$—say, $c = 2$. Then $c\mathbf{u} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}$. If this is in H, then there is some s such that

$$\begin{bmatrix} 3s \\ 2 + 5s \end{bmatrix} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}$$

That is, $s = 2$ and $s = 12/5$, which is impossible. So $2\mathbf{u}$ is not in H and H is not a vector space.

2. $\mathbf{v}_1 = 1\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_p$. This expresses $\mathbf{v}_1$ as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_p$, so $\mathbf{v}_1$ is in W. In general, $\mathbf{v}_k$ is in W because

$$\mathbf{v}_k = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_{k-1} + 1\mathbf{v}_k + 0\mathbf{v}_{k+1} + \cdots + 0\mathbf{v}_p$$

3. The subset S is a subspace of $M_{3\times 3}$ since it satisfies all three of the requirements listed in the definition of a subspace:

a. Observe that the $\mathbf{0}$ in $M_{3\times 3}$ is the 3×3 zero matrix and since $\mathbf{0}^T = \mathbf{0}$, the matrix $\mathbf{0}$ is symmetric and hence $\mathbf{0}$ is in S.

b. Let A and B in S. Notice that A and B are 3×3 symmetric matrices so $A^T = A$ and $B^T = B$. By the properties of transposes of matrices, $(A + B)^T = A^T + B^T = A + B$. Thus $A + B$ is symmetric and hence $A + B$ is in S.

c. Let A be in S and let c be a scalar. Since A is symmetric, by the properties of symmetric matrices, $(cA)^T = c(A^T) = cA$. Thus cA is also a symmetric matrix and hence cA is in S.

4.2 Null Spaces, Column Spaces, Row Spaces, and Linear Transformations

In applications of linear algebra, subspaces of $\mathbb{R}^n$ usually arise in one of two ways: (1) as the set of all solutions to a system of homogeneous linear equations or (2) as the set of all linear combinations of certain specified vectors. In this section, we compare and contrast these two descriptions of subspaces, allowing us to practice using the concept of a subspace. Actually, as you will soon discover, we have been working with subspaces ever since Section 1.3. The main new feature here is the terminology. The section concludes with a discussion of the kernel and range of a linear transformation.

The Null Space of a Matrix

Consider the following system of homogeneous equations:

$$\begin{aligned} x_1 - 3x_2 - 2x_3 &= 0 \\ -5x_1 + 9x_2 + x_3 &= 0 \end{aligned} \tag{1}$$

In matrix form, this system is written as $A\mathbf{x} = \mathbf{0}$, where

$$A = \begin{bmatrix} 1 & -3 & -2 \\ -5 & 9 & 1 \end{bmatrix} \tag{2}$$

Recall that the set of all $\mathbf{x}$ that satisfy (1) is called the **solution set** of the system (1). Often it is convenient to relate this set directly to the matrix A and the equation $A\mathbf{x} = \mathbf{0}$. We call the set of $\mathbf{x}$ that satisfy $A\mathbf{x} = \mathbf{0}$ the **null space** of the matrix A.

DEFINITION

The **null space** of an $m \times n$ matrix A, written as Nul A, is the set of all solutions of the homogeneous equation $A\mathbf{x} = \mathbf{0}$. In set notation,

$$\text{Nul } A = \{\mathbf{x} : \mathbf{x} \text{ is in } \mathbb{R}^n \text{ and } A\mathbf{x} = \mathbf{0}\}$$

A more dynamic description of Nul A is the set of all $\mathbf{x}$ in $\mathbb{R}^n$ that are mapped into the zero vector of $\mathbb{R}^m$ via the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$. See Figure 1.

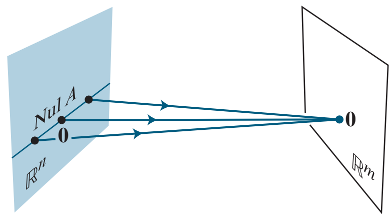


FIGURE 1

EXAMPLE 1 Let A be the matrix in (2), and let $\mathbf{u} = \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix}$. Determine if $\mathbf{u}$ belongs to the null space of A.

SOLUTION To test if $\mathbf{u}$ satisfies $A\mathbf{u} = \mathbf{0}$, simply compute

$$A\mathbf{u} = \begin{bmatrix} 1 & -3 & -2 \\ -5 & 9 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix} = \begin{bmatrix} 5 - 9 + 4 \\ -25 + 27 - 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Thus $\mathbf{u}$ is in Nul A. ■

The term *space* in *null space* is appropriate because the null space of a matrix is a vector space, as shown in the next theorem.

THEOREM 2

The null space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^n$. Equivalently, the set of all solutions to a system $A\mathbf{x} = \mathbf{0}$ of m homogeneous linear equations in n unknowns is a subspace of $\mathbb{R}^n$.

PROOF Certainly Nul A is a subset of $\mathbb{R}^n$ because A has n columns. We must show that Nul A satisfies the three properties of a subspace. Of course, $\mathbf{0}$ is in Nul A. Next, let $\mathbf{u}$ and $\mathbf{v}$ represent any two vectors in Nul A. Then

$$A\mathbf{u} = \mathbf{0} \quad \text{and} \quad A\mathbf{v} = \mathbf{0}$$

To show that $\mathbf{u} + \mathbf{v}$ is in Nul A, we must show that $A(\mathbf{u} + \mathbf{v}) = \mathbf{0}$. Using a property of matrix multiplication, compute

$$A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

Thus $\mathbf{u} + \mathbf{v}$ is in Nul A, and Nul A is closed under vector addition. Finally, if c is any scalar, then

$$A(c\mathbf{u}) = c(A\mathbf{u}) = c(\mathbf{0}) = \mathbf{0}$$

which shows that $c\mathbf{u}$ is in Nul A. Thus Nul A is a subspace of $\mathbb{R}^n$. ■

EXAMPLE 2 Let H be the set of all vectors in $\mathbb{R}^4$ whose coordinates a, b, c, d satisfy the equations $a - 2b + 5c = d$ and $c - a = b$. Show that H is a subspace of $\mathbb{R}^4$.

SOLUTION Rearrange the equations that describe the elements of H, and note that H is the set of all solutions of the following system of homogeneous linear equations:

$$\begin{aligned} a - 2b + 5c - d &= 0 \\ -a - b + c \phantom{{}-d} &= 0 \end{aligned}$$

By Theorem 2, H is a subspace of $\mathbb{R}^4$. ■

It is important that the linear equations defining the set H are homogeneous. Otherwise, the set of solutions will definitely *not* be a subspace (because the zero vector is not a solution of a nonhomogeneous system). Also, in some cases, the set of solutions could be empty.

An Explicit Description of Nul A

There is no obvious relation between vectors in Nul A and the entries in A. We say that Nul A is defined *implicitly*, because it is defined by a condition that must be checked. No explicit list or description of the elements in Nul A is given. However, *solving* the equation $A\mathbf{x} = \mathbf{0}$ amounts to producing an *explicit* description of Nul A. The next example reviews the procedure from Section 1.5.

EXAMPLE 3 Find a spanning set for the null space of the matrix

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$$

SOLUTION The first step is to find the general solution of $A\mathbf{x} = \mathbf{0}$ in terms of free variables. Row reduce the augmented matrix $[\, A \quad \mathbf{0} \,]$ to *reduced* echelon form in order to write the basic variables in terms of the free variables:

$$\begin{bmatrix} 1 & -2 & 0 & -1 & 3 & 0 \\ 0 & 0 & 1 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \begin{aligned} x_1 - 2x_2 \phantom{{}+x_3} - x_4 + 3x_5 &= 0 \\ x_3 + 2x_4 - 2x_5 &= 0 \\ 0 &= 0 \end{aligned}$$

The general solution is $x_1 = 2x_2 + x_4 - 3x_5$, $x_3 = -2x_4 + 2x_5$, with x_2, x_4, and x_5 free. Next, decompose the vector giving the general solution into a linear combination of vectors where *the weights are the free variables*. That is,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_4 + 2x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \underset{\substack{\uparrow \\ \mathbf{u}}}{\begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}} + x_4 \underset{\substack{\uparrow \\ \mathbf{v}}}{\begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}} + x_5 \underset{\substack{\uparrow \\ \mathbf{w}}}{\begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}}$$

$$= x_2\mathbf{u} + x_4\mathbf{v} + x_5\mathbf{w} \tag{3}$$

Every linear combination of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ is an element of Nul A and vice versa. Thus $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a spanning set for Nul A. ■

Two points should be made about the solution of Example 3 that apply to all problems of this type where Nul A contains nonzero vectors. We will use these facts later.

1. The spanning set produced by the method in Example 3 is automatically linearly independent because the free variables are the weights on the spanning vectors. For instance, look at the 2nd, 4th, and 5th entries in the solution vector in (3) and note that $x_2\mathbf{u} + x_4\mathbf{v} + x_5\mathbf{w}$ can be $\mathbf{0}$ only if the weights x_2, x_4, and x_5 are all zero.
2. When Nul A contains nonzero vectors, the number of vectors in the spanning set for Nul A equals the number of free variables in the equation $A\mathbf{x} = \mathbf{0}$.

The Column Space of a Matrix

Another important subspace associated with a matrix is its column space. Unlike the null space, the column space is defined explicitly via linear combinations.

DEFINITION

The **column space** of an $m \times n$ matrix A, written as Col A, is the set of all linear combinations of the columns of A. If $A = [\,\mathbf{a}_1 \ \cdots \ \mathbf{a}_n\,]$, then

$$\text{Col } A = \text{Span}\,\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$$

Since Span $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is a subspace, by Theorem 1, the next theorem follows from the definition of Col A and the fact that the columns of A are in $\mathbb{R}^m$.

THEOREM 3

The column space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^m$.

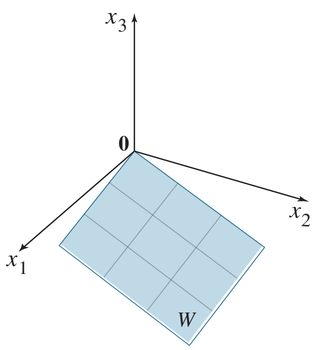
Note that a typical vector in Col A can be written as $A\mathbf{x}$ for some $\mathbf{x}$ because the notation $A\mathbf{x}$ stands for a linear combination of the columns of A. That is,

$$\boxed{\text{Col } A = \{\mathbf{b} : \mathbf{b} = A\mathbf{x} \text{ for some } \mathbf{x} \text{ in } \mathbb{R}^n\}}$$

The notation $A\mathbf{x}$ for vectors in Col A also shows that Col A is the *range* of the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$. We will return to this point of view at the end of the section.

EXAMPLE 4 Find a matrix A such that $W = \text{Col } A$.

$$W = \left\{ \begin{bmatrix} 6a - b \\ a + b \\ -7a \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\}$$



SOLUTION First, write W as a set of linear combinations.

$$W = \left\{ a\begin{bmatrix} 6 \\ 1 \\ -7 \end{bmatrix} + b\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\} = \text{Span}\left\{ \begin{bmatrix} 6 \\ 1 \\ -7 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

Second, use the vectors in the spanning set as the columns of A. Let $A = \begin{bmatrix} 6 & -1 \\ 1 & 1 \\ -7 & 0 \end{bmatrix}$.

Then $W = \text{Col } A$, as desired. ■

Recall from Theorem 4 in Section 1.4 that the columns of A span $\mathbb{R}^m$ if and only if the equation $A\mathbf{x} = \mathbf{b}$ has a solution for each $\mathbf{b}$. We can restate this fact as follows:

> The column space of an $m \times n$ matrix A is all of $\mathbb{R}^m$ if and only if the equation $A\mathbf{x} = \mathbf{b}$ has a solution for each $\mathbf{b}$ in $\mathbb{R}^m$.

The Row Space

If A is an $m \times n$ matrix, each row of A has n entries and thus can be identified with a vector in $\mathbb{R}^n$. The set of all linear combinations of the row vectors is called the **row space** of A and is denoted by Row A. Each row has n entries, so Row A is a subspace of $\mathbb{R}^n$. Since the rows of A are identified with the columns of A^T, we could also write Col A^T in place of Row A.

EXAMPLE 5 Let

$$A = \begin{bmatrix} -2 & -5 & 8 & 0 & -17 \\ 1 & 3 & -5 & 1 & 5 \\ 3 & 11 & -19 & 7 & 1 \\ 1 & 7 & -13 & 5 & -3 \end{bmatrix} \quad \text{and} \quad \begin{aligned} \mathbf{r}_1 &= (-2, -5, 8, 0, -17) \\ \mathbf{r}_2 &= (1, 3, -5, 1, 5) \\ \mathbf{r}_3 &= (3, 11, -19, 7, 1) \\ \mathbf{r}_4 &= (1, 7, -13, 5, -3) \end{aligned}$$

The row space of A is the subspace of $\mathbb{R}^5$ spanned by $\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4\}$. That is, Row $A =$ Span $\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4\}$. It is natural to write row vectors horizontally; however, they may also be written as column vectors if that is more convenient. ■

The Contrast Between Nul *A* and Col *A*

It is natural to wonder how the null space and column space of a matrix are related. In fact, the two spaces are quite dissimilar, as Examples 6–8 will show. Nevertheless, a surprising connection between the null space and column space will emerge in Section 4.5, after more theory is available.

EXAMPLE 6 Let

$$A = \begin{bmatrix} 2 & 4 & -2 & 1 \\ -2 & -5 & 7 & 3 \\ 3 & 7 & -8 & 6 \end{bmatrix}$$

a. If the column space of A is a subspace of $\mathbb{R}^k$, what is k?

b. If the null space of A is a subspace of $\mathbb{R}^k$, what is k?

SOLUTION

a. The columns of A each have three entries, so Col A is a subspace of $\mathbb{R}^k$, where $k = 3$.

b. A vector $\mathbf{x}$ such that $A\mathbf{x}$ is defined must have four entries, so Nul A is a subspace of $\mathbb{R}^k$, where $k = 4$. ■

When a matrix is not square, as in Example 6, the vectors in Nul A and Col A live in entirely different "universes." For example, no linear combination of vectors in $\mathbb{R}^3$ can produce a vector in $\mathbb{R}^4$. When A is square, Nul A and Col A do have the zero vector in common, and in special cases it is possible that some nonzero vectors belong to both Nul A and Col A.

EXAMPLE 7 With A as in Example 6, find a nonzero vector in Col A and a nonzero vector in Nul A.

SOLUTION It is easy to find a vector in Col A. Any column of A will do, say, $\begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix}$.

To find a nonzero vector in Nul A, row reduce the augmented matrix $[\,A \quad \mathbf{0}\,]$ and obtain

$$[\,A \quad \mathbf{0}\,] \sim \begin{bmatrix} 1 & 0 & 9 & 0 & 0 \\ 0 & 1 & -5 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Thus, if $\mathbf{x}$ satisfies $A\mathbf{x} = \mathbf{0}$, then $x_1 = -9x_3$, $x_2 = 5x_3$, $x_4 = 0$, and x_3 is free. Assigning a nonzero value to x_3—say, $x_3 = 1$—we obtain a vector in Nul A, namely, $\mathbf{x} = (-9, 5, 1, 0)$. ■

EXAMPLE 8 With A as in Example 6, let $\mathbf{u} = \begin{bmatrix} 3 \\ -2 \\ -1 \\ 0 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ -1 \\ 3 \end{bmatrix}$.

a. Determine if $\mathbf{u}$ is in Nul A. Could $\mathbf{u}$ be in Col A?

b. Determine if $\mathbf{v}$ is in Col A. Could $\mathbf{v}$ be in Nul A?

SOLUTION

a. An explicit description of Nul A is not needed here. Simply compute the product $A\mathbf{u}$.

$$A\mathbf{u} = \begin{bmatrix} 2 & 4 & -2 & 1 \\ -2 & -5 & 7 & 3 \\ 3 & 7 & -8 & 6 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -3 \\ 3 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Obviously, $\mathbf{u}$ is *not* a solution of $A\mathbf{x} = \mathbf{0}$, so $\mathbf{u}$ is not in Nul A. Also, with four entries, $\mathbf{u}$ could not possibly be in Col A, since Col A is a subspace of $\mathbb{R}^3$.

b. Reduce $[\,A \quad \mathbf{v}\,]$ to an echelon form.

$$[\,A \quad \mathbf{v}\,] = \begin{bmatrix} 2 & 4 & -2 & 1 & 3 \\ -2 & -5 & 7 & 3 & -1 \\ 3 & 7 & -8 & 6 & 3 \end{bmatrix} \sim \begin{bmatrix} 2 & 4 & -2 & 1 & 3 \\ 0 & 1 & -5 & -4 & -2 \\ 0 & 0 & 0 & 17 & 1 \end{bmatrix}$$

At this point, it is clear that the equation $A\mathbf{x} = \mathbf{v}$ is consistent, so $\mathbf{v}$ is in Col A. With only three entries, $\mathbf{v}$ could not possibly be in Nul A, since Nul A is a subspace of $\mathbb{R}^4$. ■

The table on page 241 summarizes what we have learned about Nul A and Col A. Item 8 is a restatement of Theorems 11 and 12(a) in Section 1.9.

Kernel and Range of a Linear Transformation

Subspaces of vector spaces other than $\mathbb{R}^n$ are often described in terms of a linear transformation instead of a matrix. To make this precise, we generalize the definition given in Section 1.8.

Contrast Between Nul A and Col A for an $m \times n$ Matrix A

Nul A	**Col A**
1. Nul A is a subspace of $\mathbb{R}^n$.	**1.** Col A is a subspace of $\mathbb{R}^m$.
2. Nul A is implicitly defined; that is, you are given only a condition ($A\mathbf{x} = \mathbf{0}$) that vectors in Nul A must satisfy.	**2.** Col A is explicitly defined; that is, you are told how to build vectors in Col A.
3. It takes time to find vectors in Nul A. Row operations on $[\,A \quad \mathbf{0}\,]$ are required.	**3.** It is easy to find vectors in Col A. The columns of A are displayed; others are formed from them.
4. There is no obvious relation between Nul A and the entries in A.	**4.** There is an obvious relation between Col A and the entries in A, since each column of A is in Col A.
5. A typical vector $\mathbf{v}$ in Nul A has the property that $A\mathbf{v} = \mathbf{0}$.	**5.** A typical vector $\mathbf{v}$ in Col A has the property that the equation $A\mathbf{x} = \mathbf{v}$ is consistent.
6. Given a specific vector $\mathbf{v}$, it is easy to tell if $\mathbf{v}$ is in Nul A. Just compute $A\mathbf{v}$.	**6.** Given a specific vector $\mathbf{v}$, it may take time to tell if $\mathbf{v}$ is in Col A. Row operations on $[\,A \quad \mathbf{v}\,]$ are required.
7. Nul $A = \{\mathbf{0}\}$ if and only if the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.	**7.** Col $A = \mathbb{R}^m$ if and only if the equation $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b}$ in $\mathbb{R}^m$.
8. Nul $A = \{\mathbf{0}\}$ if and only if the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.	**8.** Col $A = \mathbb{R}^m$ if and only if the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ *onto* $\mathbb{R}^m$.

DEFINITION

A **linear transformation** T from a vector space V into a vector space W is a rule that assigns to each vector $\mathbf{x}$ in V a unique vector $T(\mathbf{x})$ in W, such that

(i) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}$, $\mathbf{v}$ in V, and

(ii) $T(c\mathbf{u}) = cT(\mathbf{u})$ for all $\mathbf{u}$ in V and all scalars c.

The **kernel** (or **null space**) of such a T is the set of all $\mathbf{u}$ in V such that $T(\mathbf{u}) = \mathbf{0}$ (the zero vector in W). The **range** of T is the set of all vectors in W of the form $T(\mathbf{x})$ for some $\mathbf{x}$ in V. If T happens to arise as a matrix transformation—say, $T(\mathbf{x}) = A\mathbf{x}$ for some matrix A—then the kernel and the range of T are just the null space and the column space of A, as defined earlier.

It is not difficult to show that the kernel of T is a subspace of V. The proof is essentially the same as the one for Theorem 2. Also, the range of T is a subspace of W. See Figure 2 and Exercise 42.

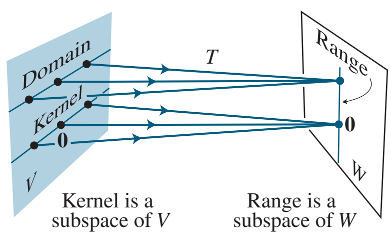


FIGURE 2 Subspaces associated with a linear transformation.

In applications, a subspace usually arises as either the kernel or the range of an appropriate linear transformation. For instance, the set of all solutions of a homogeneous linear differential equation turns out to be the kernel of a linear transformation. Typically, such a linear transformation is described in terms of one or more derivatives of a function. To explain this in any detail would take us too far afield at this point. So we consider only two examples. The first explains why the operation of differentiation is a linear transformation.

EXAMPLE 9 (Calculus required) Let V be the vector space of all real-valued functions f defined on an interval $[a, b]$ with the property that they are differentiable and their derivatives are continuous functions on $[a, b]$. Let W be the vector space $C[a, b]$ of all continuous functions on $[a, b]$, and let $D : V \to W$ be the transformation that changes f in V into its derivative f'. In calculus, two simple differentiation rules are

$$D(f + g) = D(f) + D(g) \quad \text{and} \quad D(cf) = cD(f)$$

That is, D is a linear transformation. It can be shown that the kernel of D is the set of constant functions on $[a, b]$ and the range of D is the set W of all continuous functions on $[a, b]$. ■

EXAMPLE 10 (Calculus required) The differential equation

$$y'' + \omega^2 y = 0 \tag{4}$$

where ω is a constant, is used to describe a variety of physical systems, such as the vibration of a weighted spring, the movement of a pendulum, and the voltage in an inductance-capacitance electrical circuit. The set of solutions of (4) is precisely the kernel of the linear transformation that maps a function $y = f(t)$ into the function $f''(t) + \omega^2 f(t)$. Finding an explicit description of this vector space is a problem in differential equations. The solution set turns out to be the space described in Exercise 19 in Section 4.1. ■

A common technique used in the stock market is technical analysis. Statistical trends gathered from stock-trading activity, such as price movement and volume, are analyzed. Technical analysts focus on patterns of stock-price movements, trading signals, and various other analytical charting tools to evaluate a security's strength or weakness. A moving average is a commonly used indicator in technical analysis. It smooths out price action by filtering out the effects from random price fluctuations. In the final example for this section, we examine the linear transformation that creates the two-day moving average from a "signal" of daily prices. We will look at moving average transformations that average over a longer period of time in Section 4.7.

EXAMPLE 11 Let $\{p_k\}$ in $\mathbb{S}$ represent the price of a stock that has been recorded daily over an extended period of time. Note that we can assume that $p_k = 0$ for k outside the time period under study. To create a two-day moving average, the mapping $M_2 : \mathbb{S} \to \mathbb{S}$ defined by $M_2(\{p_k\}) = \left\{\dfrac{p_k + p_{k-1}}{2}\right\}$ is applied to the data. Show that M_2 is a linear transformation and find its kernel.

SOLUTION To see that M_2 is a linear transformation, observe that for two signals $\{p_k\}$ and $\{q_k\}$ in $\mathbb{S}$ and any scalar c,

$$\begin{aligned} M_2(\{p_k\} + \{q_k\}) = M_2(\{p_k + q_k\}) &= \left\{\frac{p_k + q_k + p_{k-1} + q_{k-1}}{2}\right\} \\ &= \left\{\frac{p_k + p_{k-1}}{2}\right\} + \left\{\frac{q_k + q_{k-1}}{2}\right\} \\ &= M_2(\{p_k\}) + M_2(\{q_k\}) \end{aligned}$$

and

$$M_2(c\{p_k\}) = M_2(\{cp_k\}) = \left\{\frac{cp_k + cp_{k-1}}{2}\right\} = c\left\{\frac{p_k + p_{k-1}}{2}\right\} = cM_2(\{p_k\})$$

thus M_2 is a linear transformation.

To find the kernel of M_2, notice that $\{p_k\}$ is in the kernel if and only if $\dfrac{p_k + p_{k-1}}{2} = 0$ for all k, and hence $p_k = -p_{k-1}$. Since this relationship is true for all integers k, it can be applied recursively resulting in $p_k = -p_{k-1} = (-1)^2 p_{k-2} = (-1)^3 p_{k-3} \dots$. Working out from $k = 0$, any signal in the kernel can be written as $p_k = p_0(-1)^k$, a multiple of the alternating signal described by $\{(-1)^k\}$. Since the kernel of the two-day moving average function consists of all multiples of the alternating sequence, it smooths out daily fluctuations, without leveling out overall trends. (See Figure 3.)

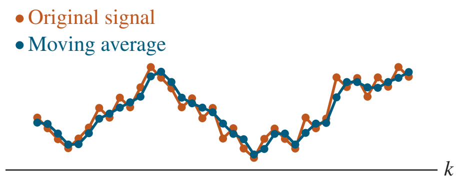


FIGURE 3

Practice Problems

1. Let $W = \left\{ \begin{bmatrix} a \\ b \\ c \end{bmatrix} : a - 3b - c = 0 \right\}$. Show in two different ways that W is a subspace of $\mathbb{R}^3$. (Use two theorems.)
2. Let $A = \begin{bmatrix} 7 & -3 & 5 \\ -4 & 1 & -5 \\ -5 & 2 & -4 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} 7 \\ 6 \\ -3 \end{bmatrix}$. Suppose you know that the equations $A\mathbf{x} = \mathbf{v}$ and $A\mathbf{x} = \mathbf{w}$ are both consistent. What can you say about the equation $A\mathbf{x} = \mathbf{v} + \mathbf{w}$?
3. Let A be an $n \times n$ matrix. If Col A = Nul A, show that Nul $A^2 = \mathbb{R}^n$.

4.2 Exercises

1. Determine if $\mathbf{w} = \begin{bmatrix} 1 \\ 3 \\ -4 \end{bmatrix}$ is in Nul A, where
$$A = \begin{bmatrix} 3 & -5 & -3 \\ 6 & -2 & 0 \\ -8 & 4 & 1 \end{bmatrix}.$$

2. Determine if $\mathbf{w} = \begin{bmatrix} 5 \\ -3 \\ 2 \end{bmatrix}$ is in Nul A, where
$$A = \begin{bmatrix} 5 & 21 & 19 \\ 13 & 23 & 2 \\ 8 & 14 & 1 \end{bmatrix}.$$

In Exercises 3–6, find an explicit description of Nul A by listing vectors that span the null space.

3. $A = \begin{bmatrix} 1 & 3 & 5 & 0 \\ 0 & 1 & 4 & -2 \end{bmatrix}$

4. $A = \begin{bmatrix} 1 & -6 & 4 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix}$

5. $A = \begin{bmatrix} 1 & -2 & 0 & 4 & 0 \\ 0 & 0 & 1 & -9 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

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

In Exercises 7–14, either use an appropriate theorem to show that the given set, W, is a vector space, or find a specific example to the contrary.

7. $\left\{ \begin{bmatrix} a \\ b \\ c \end{bmatrix} : a + b + c = 2 \right\}$

8. $\left\{ \begin{bmatrix} r \\ s \\ t \end{bmatrix} : 5r - 1 = s + 2t \right\}$

9. $\left\{ \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} : \begin{matrix} a - 2b = 4c \\ 2a = c + 3d \end{matrix} \right\}$

10. $\left\{ \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} : \begin{matrix} a + 3b = c \\ b + c + a = d \end{matrix} \right\}$

11. $\left\{ \begin{bmatrix} b - 2d \\ 5 + d \\ b + 3d \\ d \end{bmatrix} : b, d \text{ real} \right\}$

12. $\left\{ \begin{bmatrix} b - 5d \\ 2b \\ 2d + 1 \\ d \end{bmatrix} : b, d \text{ real} \right\}$

13. $\left\{ \begin{bmatrix} c - 6d \\ d \\ c \end{bmatrix} : c, d \text{ real} \right\}$

14. $\left\{ \begin{bmatrix} -a + 2b \\ a - 2b \\ 3a - 6b \end{bmatrix} : a, b \text{ real} \right\}$

In Exercises 15 and 16, find A such that the given set is Col A.

15. $\left\{ \begin{bmatrix} 2s + 3t \\ r + s - 2t \\ 4r + s \\ 3r - s - t \end{bmatrix} : r, s, t \text{ real} \right\}$

16. $\left\{ \begin{bmatrix} b - c \\ 2b + c + d \\ 5c - 4d \\ d \end{bmatrix} : b, c, d \text{ real} \right\}$

For the matrices in Exercises 17–20, (a) find k such that Nul A is a subspace of $\mathbb{R}^k$, and (b) find k such that Col A is a subspace of $\mathbb{R}^k$.

17. $A = \begin{bmatrix} 2 & -8 \\ -1 & 4 \\ 1 & -4 \end{bmatrix}$

18. $A = \begin{bmatrix} 8 & -3 & 0 & -1 \\ -3 & 0 & -1 & 8 \\ 0 & -1 & 8 & -3 \end{bmatrix}$

19. $A = \begin{bmatrix} 4 & 5 & -2 & 6 & 0 \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix}$

20. $A = \begin{bmatrix} 1 & -3 & 9 & 0 & -5 \end{bmatrix}$

21. With A as in Exercise 17, find a nonzero vector in Nul A, a nonzero vector in Col A, and a nonzero vector in Row A.

22. With A as in Exercise 3, find a nonzero vector in Nul A, a nonzero vector in Col A, and a nonzero vector in Row A.

23. Let $A = \begin{bmatrix} -6 & 12 \\ -3 & 6 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Determine if $\mathbf{w}$ is in Col A. Is $\mathbf{w}$ in Nul A?

24. Let $A = \begin{bmatrix} -8 & -2 & -9 \\ 6 & 4 & 8 \\ 4 & 0 & 4 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$. Determine if $\mathbf{w}$ is in Col A. Is $\mathbf{w}$ in Nul A?

In Exercises 25–38, A denotes an $m \times n$ matrix. Mark each statement True or False **(T/F)**. Justify each answer.

25. **(T/F)** The null space of A is the solution set of the equation $A\mathbf{x} = \mathbf{0}$.

26. **(T/F)** A null space is a vector space.

27. **(T/F)** The null space of an $m \times n$ matrix is in $\mathbb{R}^m$.

28. **(T/F)** The column space of an $m \times n$ matrix is in $\mathbb{R}^m$.

29. **(T/F)** The column space of A is the range of the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

30. **(T/F)** Col A is the set of all solutions of $A\mathbf{x} = \mathbf{b}$.

31. **(T/F)** If the equation $A\mathbf{x} = \mathbf{b}$ is consistent, then Col $A = \mathbb{R}^m$.

32. **(T/F)** Nul A is the kernel of the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

33. **(T/F)** The kernel of a linear transformation is a vector space.

34. **(T/F)** The range of a linear transformation is a vector space.

35. **(T/F)** Col A is the set of all vectors that can be written as $A\mathbf{x}$ for some $\mathbf{x}$.

36. **(T/F)** The set of all solutions of a homogeneous linear differential equation is the kernel of a linear transformation.

37. **(T/F)** The row space of A is the same as the column space of A^T.

38. **(T/F)** The null space of A is the same as the row space of A^T.

39. It can be shown that a solution of the system below is $x_1 = 3$, $x_2 = 2$, and $x_3 = -1$. Use this fact and the theory from this section to explain why another solution is $x_1 = 30$, $x_2 = 20$, and $x_3 = -10$. (Observe how the solutions are related, but make no other calculations.)

$$\begin{aligned} x_1 - 3x_2 - 3x_3 &= 0 \\ -2x_1 + 4x_2 + 2x_3 &= 0 \\ -x_1 + 5x_2 + 7x_3 &= 0 \end{aligned}$$

40. Consider the following two systems of equations:

$$\begin{aligned} 5x_1 + x_2 - 3x_3 &= 0 \\ -9x_1 + 2x_2 + 5x_3 &= 1 \\ 4x_1 + x_2 - 6x_3 &= 9 \end{aligned} \qquad \begin{aligned} 5x_1 + x_2 - 3x_3 &= 0 \\ -9x_1 + 2x_2 + 5x_3 &= 5 \\ 4x_1 + x_2 - 6x_3 &= 45 \end{aligned}$$

It can be shown that the first system has a solution. Use this fact and the theory from this section to explain why the second system must also have a solution. (Make no row operations.)

41. Prove Theorem 3 as follows: Given an $m \times n$ matrix A, an element in Col A has the form $A\mathbf{x}$ for some $\mathbf{x}$ in $\mathbb{R}^n$. Let $A\mathbf{x}$ and $A\mathbf{w}$ represent any two vectors in Col A.

a. Explain why the zero vector is in Col A.

b. Show that the vector $A\mathbf{x} + A\mathbf{w}$ is in Col A.

c. Given a scalar c, show that $c(A\mathbf{x})$ is in Col A.

42. Let $T : V \to W$ be a linear transformation from a vector space V into a vector space W. Prove that the range of T is a subspace of W. [*Hint:* Typical elements of the range have the form $T(\mathbf{x})$ and $T(\mathbf{w})$ for some $\mathbf{x}$, $\mathbf{w}$ in V.]

43. Define $T : \mathbb{P}_2 \to \mathbb{R}^2$ by $T(\mathbf{p}) = \begin{bmatrix} \mathbf{p}(0) \\ \mathbf{p}(1) \end{bmatrix}$. For instance, if $\mathbf{p}(t) = 3 + 5t + 7t^2$, then $T(\mathbf{p}) = \begin{bmatrix} 3 \\ 15 \end{bmatrix}$.

a. Show that T is a linear transformation. [*Hint:* For arbitrary polynomials $\mathbf{p}$, $\mathbf{q}$ in $\mathbb{P}_2$, compute $T(\mathbf{p} + \mathbf{q})$ and $T(c\mathbf{p})$.]

b. Find a polynomial $\mathbf{p}$ in $\mathbb{P}_2$ that spans the kernel of T, and describe the range of T.

44. Define a linear transformation $T : \mathbb{P}_2 \to \mathbb{R}^2$ by $T(\mathbf{p}) = \begin{bmatrix} \mathbf{p}(0) \\ \mathbf{p}(0) \end{bmatrix}$. Find polynomials $\mathbf{p}_1$ and $\mathbf{p}_2$ in $\mathbb{P}_2$ that span the kernel of T, and describe the range of T.

45. Let $M_{2\times 2}$ be the vector space of all 2×2 matrices, and define $T : M_{2\times 2} \to M_{2\times 2}$ by $T(A) = A + A^T$, where $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

a. Show that T is a linear transformation.

b. Let B be any element of $M_{2\times 2}$ such that $B^T = B$. Find an A in $M_{2\times 2}$ such that $T(A) = B$.

c. Show that the range of T is the set of B in $M_{2\times 2}$ with the property that $B^T = B$.

d. Describe the kernel of T.

STUDY GUIDE offers additional resources for mastering vector spaces, subspaces, and column row, and null spaces.

46. (*Calculus required*) Define $T : C[0, 1] \to C[0, 1]$ as follows: For $\mathbf{f}$ in $C[0, 1]$, let $T(\mathbf{f})$ be the antiderivative $\mathbf{F}$ of $\mathbf{f}$ such that $\mathbf{F}(0) = 0$. Show that T is a linear transformation, and describe the kernel of T. (See the notation in Exercise 20 of Section 4.1.)

47. Let V and W be vector spaces, and let $T : V \to W$ be a linear transformation. Given a subspace U of V, let $T(U)$ denote the set of all images of the form $T(\mathbf{x})$, where $\mathbf{x}$ is in U. Show that $T(U)$ is a subspace of W.

48. Given $T : V \to W$ as in Exercise 47, and given a subspace Z of W, let U be the set of all $\mathbf{x}$ in V such that $T(\mathbf{x})$ is in Z. Show that U is a subspace of V.

T 49. Determine whether $\mathbf{w}$ is in the column space of A, the null space of A, or both, where

$$\mathbf{w} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -3 \end{bmatrix}, \quad A = \begin{bmatrix} 7 & 6 & -4 & 1 \\ -5 & -1 & 0 & -2 \\ 9 & -11 & 7 & -3 \\ 19 & -9 & 7 & 1 \end{bmatrix}$$

T 50. Determine whether $\mathbf{w}$ is in the column space of A, the null space of A, or both, where

$$\mathbf{w} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \quad A = \begin{bmatrix} -8 & 5 & -2 & 0 \\ -5 & 2 & 1 & -2 \\ 10 & -8 & 6 & -3 \\ 3 & -2 & 1 & 0 \end{bmatrix}$$

T 51. Let $\mathbf{a}_1, \ldots, \mathbf{a}_5$ denote the columns of the matrix A, where

$$A = \begin{bmatrix} 5 & 1 & 2 & 2 & 0 \\ 3 & 3 & 2 & -1 & -12 \\ 8 & 4 & 4 & -5 & 12 \\ 2 & 1 & 1 & 0 & -2 \end{bmatrix}, \quad B = [\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_4\,]$$

a. Explain why $\mathbf{a}_3$ and $\mathbf{a}_5$ are in the column space of B.

b. Find a set of vectors that spans Nul A.

c. Let $T : \mathbb{R}^5 \to \mathbb{R}^4$ be defined by $T(\mathbf{x}) = A\mathbf{x}$. Explain why T is neither one-to-one nor onto.

T 52. Let $H = \text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2\}$ and $K = \text{Span}\,\{\mathbf{v}_3, \mathbf{v}_4\}$, where

$$\mathbf{v}_1 = \begin{bmatrix} 5 \\ 3 \\ 8 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}, \mathbf{v}_4 = \begin{bmatrix} 0 \\ -12 \\ -28 \end{bmatrix}.$$

Then H and K are subspaces of $\mathbb{R}^3$. In fact, H and K are planes in $\mathbb{R}^3$ through the origin, and they intersect in a line through $\mathbf{0}$. Find a nonzero vector $\mathbf{w}$ that generates that line. [*Hint:* $\mathbf{w}$ can be written as $c_1\mathbf{v}_1 + c_2\mathbf{v}_2$ and also as $c_3\mathbf{v}_3 + c_4\mathbf{v}_4$. To build $\mathbf{w}$, solve the equation $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = c_3\mathbf{v}_3 + c_4\mathbf{v}_4$ for the unknown c_j's.]

Solutions to Practice Problems

1. *First method:* W is a subspace of $\mathbb{R}^3$ by Theorem 2 because W is the set of all solutions to a system of homogeneous linear equations (where the system has only one equation). Equivalently, W is the null space of the 1×3 matrix $A = [\,1 \quad -3 \quad -1\,]$.
Second method: Solve the equation $a - 3b - c = 0$ for the leading variable a in terms of the free variables b and c. Any solution has the form $\begin{bmatrix} 3b+c \\ b \\ c \end{bmatrix}$, where b and c are arbitrary, and

$$\begin{bmatrix} 3b+c \\ b \\ c \end{bmatrix} = b\underbrace{\begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}}_{\uparrow\ \mathbf{v}_1} + c\underbrace{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}_{\uparrow\ \mathbf{v}_2}$$

This calculation shows that $W = \text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2\}$. Thus W is a subspace of $\mathbb{R}^3$ by Theorem 1. We could also solve the equation $a - 3b - c = 0$ for b or c and get alternative descriptions of W as a set of linear combinations of two vectors.

2. Both $\mathbf{v}$ and $\mathbf{w}$ are in Col A. Since Col A is a vector space, $\mathbf{v} + \mathbf{w}$ must be in Col A. That is, the equation $A\mathbf{x} = \mathbf{v} + \mathbf{w}$ is consistent.

3. Let $\mathbf{x}$ be any vector in $\mathbb{R}^n$. Notice $A\mathbf{x}$ is in Col A, since it is a linear combination of the columns of A. Since Col A = Nul A, the vector $A\mathbf{x}$ is also in Nul A. Hence $A^2\mathbf{x} = A(A\mathbf{x}) = \mathbf{0}$ establishing that every vector $\mathbf{x}$ from $\mathbb{R}^n$ is in Nul A^2.

4.3 Linearly Independent Sets; Bases

In this section we identify and study the subsets that span a vector space V or a subspace H as "efficiently" as possible. The key idea is that of linear independence, defined as in $\mathbb{R}^n$.

An indexed set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ in V is said to be **linearly independent** if the vector equation

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

has *only* the trivial solution, $c_1 = 0, \dots, c_p = 0$.[1]

The set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is said to be **linearly dependent** if (1) has a nontrivial solution, that is, if there are some weights, $c_1, \dots, c_p$, *not all zero*, such that (1) holds. In such a case, (1) is called a **linear dependence relation** among $\mathbf{v}_1, \dots, \mathbf{v}_p$.

Just as in $\mathbb{R}^n$, a set containing a single vector $\mathbf{v}$ is linearly independent if and only if $\mathbf{v} \neq \mathbf{0}$. Also, a set of two vectors is linearly dependent if and only if one of the vectors is a multiple of the other. And any set containing the zero vector is linearly dependent. The following theorem has the same proof as Theorem 7 in Section 1.7.

[1] It is convenient to use $c_1, \dots, c_p$ in (1) for the scalars instead of $x_1, \dots, x_p$, as we did previously.

THEOREM 4

An indexed set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ of two or more vectors, with $\mathbf{v}_1 \neq \mathbf{0}$, is linearly dependent if and only if some $\mathbf{v}_j$ (with $j > 1$) is a linear combination of the preceding vectors, $\mathbf{v}_1, \dots, \mathbf{v}_{j-1}$.

The main difference between linear dependence in $\mathbb{R}^n$ and in a general vector space is that when the vectors are not n-tuples, the homogeneous equation (1) usually cannot be written as a system of n linear equations. That is, the vectors cannot be made into the columns of a matrix A in order to study the equation $A\mathbf{x} = \mathbf{0}$. We must rely instead on the definition of linear dependence and on Theorem 4.

EXAMPLE 1 Let $\mathbf{p}_1(t) = 1$, $\mathbf{p}_2(t) = t$, and $\mathbf{p}_3(t) = 4 - t$. Then $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ is linearly dependent in $\mathbb{P}$ because $\mathbf{p}_3 = 4\mathbf{p}_1 - \mathbf{p}_2$. ■

EXAMPLE 2 The set $\{\sin t, \cos t\}$ is linearly independent in $C[0, 1]$, the space of all continuous functions on $0 \leq t \leq 1$, because $\sin t$ and $\cos t$ are not multiples of one another *as vectors in* $C[0, 1]$. That is, there is no scalar c such that $\cos t = c \cdot \sin t$ for all t in $[0, 1]$. (Look at the graphs of $\sin t$ and $\cos t$.) However, $\{\sin t \cos t, \sin 2t\}$ is linearly dependent because of the identity $\sin 2t = 2 \sin t \cos t$, for all t. ■

DEFINITION

Let H be a subspace of a vector space V. A set of vectors $\mathcal{B}$ in V is a **basis** for H if

(i) $\mathcal{B}$ is a linearly independent set, and

(ii) the subspace spanned by $\mathcal{B}$ coincides with H; that is,

$$H = \text{Span } \mathcal{B}$$

The definition of a basis applies to the case when $H = V$, because any vector space is a subspace of itself. Thus a basis of V is a linearly independent set that spans V. Observe that when $H \neq V$, condition (ii) includes the requirement that each of the vectors $\mathbf{b}$ in $\mathcal{B}$ must belong to H, because Span $\mathcal{B}$ contains every element in $\mathcal{B}$, as shown in Section 4.1.

EXAMPLE 3 Let A be an invertible $n \times n$ matrix—say, $A = [\, \mathbf{a}_1 \ \cdots \ \mathbf{a}_n \,]$. Then the columns of A form a basis for $\mathbb{R}^n$ because they are linearly independent and they span $\mathbb{R}^n$, by the Invertible Matrix Theorem. ■

EXAMPLE 4 Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be the columns of the $n \times n$ identity matrix, I_n. That is,

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

The set $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is called the **standard basis** for $\mathbb{R}^n$ (Figure 1). ■

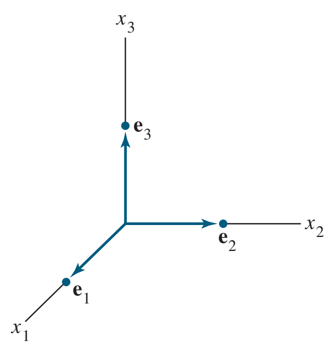


FIGURE 1
The standard basis for $\mathbb{R}^3$.

EXAMPLE 5 Let $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 0 \\ -6 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -4 \\ 1 \\ 7 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} -2 \\ 1 \\ 5 \end{bmatrix}$. Determine if $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a basis for $\mathbb{R}^3$.

SOLUTION Since there are exactly three vectors here in $\mathbb{R}^3$, we can use any of several methods to determine if the matrix $A = [\,\mathbf{v}_1 \;\; \mathbf{v}_2 \;\; \mathbf{v}_3\,]$ is invertible. For instance, two row replacements reveal that A has three pivot positions. Thus A is invertible. As in Example 3, the columns of A form a basis for $\mathbb{R}^3$. ■

EXAMPLE 6 Let $S = \{1, t, t^2, \ldots, t^n\}$. Verify that S is a basis for $\mathbb{P}_n$. This basis is called the **standard basis** for $\mathbb{P}_n$.

SOLUTION Certainly S spans $\mathbb{P}_n$. To show that S is linearly independent, suppose that $c_0, \ldots, c_n$ satisfy

$$c_0 1 + c_1 t + c_2 t^2 + \cdots + c_n t^n = \mathbf{0}(t) \tag{2}$$

This equality means that the polynomial on the left has the same values as the zero polynomial on the right. A fundamental theorem in algebra says that the only polynomial in $\mathbb{P}_n$ with more than n zeros is the zero polynomial. That is, equation (2) holds for all t only if $c_0 = \cdots = c_n = 0$. This proves that S is linearly independent and hence is a basis for $\mathbb{P}_n$. See Figure 2. ■

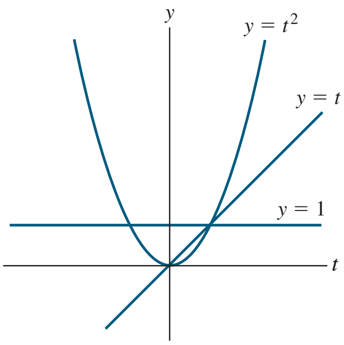


FIGURE 2
The standard basis for $\mathbb{P}_2$.

Problems involving linear independence and spanning in $\mathbb{P}_n$ are handled best by a technique to be discussed in Section 4.4.

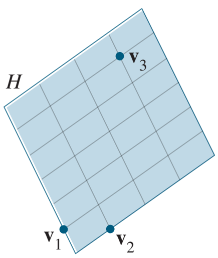
The Spanning Set Theorem

As we will see, a basis is an "efficient" spanning set that contains no unnecessary vectors. In fact, a basis can be constructed from a spanning set by discarding unneeded vectors.

EXAMPLE 7 Let

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 6 \\ 16 \\ -5 \end{bmatrix}, \quad \text{and} \quad H = \text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}.$$

Note that $\mathbf{v}_3 = 5\mathbf{v}_1 + 3\mathbf{v}_2$, and show that $\text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2\}$. Then find a basis for the subspace H.



SOLUTION Every vector in $\text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2\}$ belongs to H because

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + 0\mathbf{v}_3$$

Now let $\mathbf{x}$ be any vector in H—say, $\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3$. Since $\mathbf{v}_3 = 5\mathbf{v}_1 + 3\mathbf{v}_2$, we may substitute

$$\begin{aligned} \mathbf{x} &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3(5\mathbf{v}_1 + 3\mathbf{v}_2) \\ &= (c_1 + 5c_3)\mathbf{v}_1 + (c_2 + 3c_3)\mathbf{v}_2 \end{aligned}$$

Thus $\mathbf{x}$ is in $\text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2\}$, so every vector in H already belongs to $\text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2\}$. We conclude that H and $\text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2\}$ are actually the same set of vectors. It follows that $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis of H since $\{\mathbf{v}_1, \mathbf{v}_2\}$ is obviously linearly independent. ■

The next theorem generalizes Example 7.

THEOREM 5

The Spanning Set Theorem

Let $S = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ be a set in a vector space V, and let $H = \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$.

a. If one of the vectors in S—say, $\mathbf{v}_k$—is a linear combination of the remaining vectors in S, then the set formed from S by removing $\mathbf{v}_k$ still spans H.

b. If $H \neq \{\mathbf{0}\}$, some subset of S is a basis for H.

PROOF

a. By rearranging the list of vectors in S, if necessary, we may suppose that $\mathbf{v}_p$ is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_{p-1}$—say,

$$\mathbf{v}_p = a_1\mathbf{v}_1 + \cdots + a_{p-1}\mathbf{v}_{p-1} \tag{3}$$

Given any $\mathbf{x}$ in H, we may write

$$\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_{p-1}\mathbf{v}_{p-1} + c_p\mathbf{v}_p \tag{4}$$

for suitable scalars $c_1, \dots, c_p$. Substituting the expression for $\mathbf{v}_p$ from (3) into (4), it is easy to see that $\mathbf{x}$ is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_{p-1}$. Thus $\{\mathbf{v}_1, \dots, \mathbf{v}_{p-1}\}$ spans H, because $\mathbf{x}$ was an arbitrary element of H.

b. If the original spanning set S is linearly independent, then it is already a basis for H. Otherwise, one of the vectors in S depends on the others and can be deleted, by part (a). So long as there are two or more vectors in the spanning set, we can repeat this process until the spanning set is linearly independent and hence is a basis for H. If the spanning set is eventually reduced to one vector, that vector will be nonzero (and hence linearly independent) because $H \neq \{\mathbf{0}\}$. ■

Bases for Nul *A*, Col *A*, and Row *A*

We already know how to find vectors that span the null space of a matrix A. The discussion in Section 4.2 pointed out that our method always produces a linearly independent set when Nul A contains nonzero vectors. So, in this case, that method produces a *basis* for Nul A.

The next two examples describe a simple algorithm for finding a basis for the column space.

EXAMPLE 8 Find a basis for Col B, where

$$B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_5 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 0 & 2 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

SOLUTION Each nonpivot column of B is a linear combination of the pivot columns. In fact, $\mathbf{b}_2 = 4\mathbf{b}_1$ and $\mathbf{b}_4 = 2\mathbf{b}_1 - \mathbf{b}_3$. By the Spanning Set Theorem, we may discard $\mathbf{b}_2$ and $\mathbf{b}_4$, and $\{\mathbf{b}_1, \mathbf{b}_3, \mathbf{b}_5\}$ will still span Col B. Let

$$S = \{\mathbf{b}_1, \mathbf{b}_3, \mathbf{b}_5\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

Since $\mathbf{b}_1 \neq 0$ and no vector in S is a linear combination of the vectors that precede it, S is linearly independent (Theorem 4). Thus S is a basis for Col B. ■

What about a matrix A that is *not* in reduced echelon form? Recall that any linear dependence relationship among the columns of A can be expressed in the form $A\mathbf{x} = \mathbf{0}$, where $\mathbf{x}$ is a column of weights. (If some columns are not involved in a particular dependence relation, then their weights are zero.) When A is row reduced to a matrix B, the columns of B are often totally different from the columns of A. However, the equations $A\mathbf{x} = \mathbf{0}$ and $B\mathbf{x} = \mathbf{0}$ have exactly the same set of solutions. If $A = [\,\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_n\,]$ and $B = [\,\mathbf{b}_1 \quad \cdots \quad \mathbf{b}_n\,]$, then the vector equations

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{0} \quad \text{and} \quad x_1\mathbf{b}_1 + \cdots + x_n\mathbf{b}_n = \mathbf{0}$$

also have the same set of solutions. That is, the columns of A have *exactly the same linear dependence relationships* as the columns of B.

EXAMPLE 9 It can be shown that the matrix

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_5 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 0 & 2 & -1 \\ 3 & 12 & 1 & 5 & 5 \\ 2 & 8 & 1 & 3 & 2 \\ 5 & 20 & 2 & 8 & 8 \end{bmatrix}$$

is row equivalent to the matrix B in Example 8. Find a basis for Col A.

SOLUTION In Example 8 we saw that

$$\mathbf{b}_2 = 4\mathbf{b}_1 \quad \text{and} \quad \mathbf{b}_4 = 2\mathbf{b}_1 - \mathbf{b}_3$$

so we can expect that

$$\mathbf{a}_2 = 4\mathbf{a}_1 \quad \text{and} \quad \mathbf{a}_4 = 2\mathbf{a}_1 - \mathbf{a}_3$$

Check that this is indeed the case! Thus we may discard $\mathbf{a}_2$ and $\mathbf{a}_4$ when selecting a minimal spanning set for Col A. In fact, $\{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_5\}$ must be linearly independent because any linear dependence relationship among $\mathbf{a}_1$, $\mathbf{a}_3$, $\mathbf{a}_5$ would imply a linear dependence relationship among $\mathbf{b}_1$, $\mathbf{b}_3$, $\mathbf{b}_5$. But we know that $\{\mathbf{b}_1, \mathbf{b}_3, \mathbf{b}_5\}$ is a linearly independent set. Thus $\{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_5\}$ is a basis for Col A. The columns we have used for this basis are the pivot columns of A. ■

Examples 8 and 9 illustrate the following useful fact.

THEOREM 6

The pivot columns of a matrix A form a basis for Col A.

PROOF The general proof uses the arguments discussed above. Let B be the reduced echelon form of A. The set of pivot columns of B is linearly independent, for no vector in the set is a linear combination of the vectors that precede it. Since A is row equivalent to B, the pivot columns of A are linearly independent as well, because any linear dependence relation among the columns of A corresponds to a linear dependence relation among the columns of B. For this same reason, every nonpivot column of A is a linear combination of the pivot columns of A. Thus the nonpivot columns of A may be discarded from the spanning set for Col A, by the Spanning Set Theorem. This leaves the pivot columns of A as a basis for Col A. ■

Warning: The pivot columns of a matrix A are evident when A has been reduced only to *echelon* form. But, be careful to use the *pivot columns of A itself* for the basis of Col A. Row operations can change the column space of a matrix. The columns of an echelon form B of A are often not in the column space of A. For instance, the columns of matrix B in Example 8 all have zeros in their last entries, so they cannot span the column space of matrix A in Example 9.

In contrast, the following theorem establishes that row reduction does not change the row space of a matrix.

THEOREM 7 If two matrices A and B are row equivalent, then their row spaces are the same. If B is in echelon form, the nonzero rows of B form a basis for the row space of A as well as for that of B.

PROOF If B is obtained from A by row operations, the rows of B are linear combinations of the rows of A. It follows that any linear combination of the rows of B is automatically a linear combination of the rows of A. Thus the row space of B is contained in the row space of A. Since row operations are reversible, the same argument shows that the row space of A is a subset of the row space of B. So the two row spaces are the same. If B is in echelon form, its nonzero rows are linearly independent because no nonzero row is a linear combination of the nonzero rows below it. (Apply Theorem 4 to the nonzero rows of B in reverse order, with the first row last.) Thus the nonzero rows of B form a basis of the (common) row space of B and A. ■

EXAMPLE 10 Find a basis for the row space of the matrix A from Example 9.

SOLUTION To find a basis for the row space, recall that matrix A from Example 9 is row equivalent to matrix B from Example 8:

$$A = \begin{bmatrix} 1 & 4 & 0 & 2 & -1 \\ 3 & 12 & 1 & 5 & 5 \\ 2 & 8 & 1 & 3 & 2 \\ 5 & 20 & 2 & 8 & 8 \end{bmatrix} \sim B = \begin{bmatrix} 1 & 4 & 0 & 2 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

By Theorem 7, the first three rows of B form a basis for the row space of A (as well as for the row space of B). Thus

$$\text{Basis for Row } A : \{(1, 4, 0, 2, 0), (0, 0, 1, -1, 0), (0, 0, 0, 0, 1)\}$$

Observe that, unlike the basis for Col A, the bases for Row A and Nul A have no simple connection with the entries in A itself.[2]

Two Views of a Basis

When the Spanning Set Theorem is used, the deletion of vectors from a spanning set must stop when the set becomes linearly independent. If an additional vector is deleted,

[2] It is possible to find a basis for the row space Row A that uses rows of A. First form A^T, and then row reduce until the pivot columns of A^T are found. These pivot columns of A^T are rows of A, and they form a basis for the row space of A.

it will not be a linear combination of the remaining vectors, and hence the smaller set will no longer span V. Thus a basis is a spanning set that is as small as possible.

A basis is also a linearly independent set that is as large as possible. If S is a basis for V, and if S is enlarged by one vector—say, $\mathbf{w}$—from V, then the new set cannot be linearly independent, because S spans V, and $\mathbf{w}$ is therefore a linear combination of the elements in S.

EXAMPLE 11 The following three sets in $\mathbb{R}^3$ show how a linearly independent set can be enlarged to a basis and how further enlargement destroys the linear independence of the set. Also, a spanning set can be shrunk to a basis, but further shrinking destroys the spanning property.

$$\left\{\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}2\\3\\0\end{bmatrix}\right\} \qquad \left\{\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}2\\3\\0\end{bmatrix},\begin{bmatrix}4\\5\\6\end{bmatrix}\right\} \qquad \left\{\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}2\\3\\0\end{bmatrix},\begin{bmatrix}4\\5\\6\end{bmatrix},\begin{bmatrix}7\\8\\9\end{bmatrix}\right\}$$

Linearly independent but does not span $\mathbb{R}^3$ — A basis for $\mathbb{R}^3$ — Spans $\mathbb{R}^3$ but is linearly dependent ■

Practice Problems

1. Let $\mathbf{v}_1 = \begin{bmatrix}1\\-2\\3\end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix}-2\\7\\-9\end{bmatrix}$. Determine if $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for $\mathbb{R}^3$. Is $\{\mathbf{v}_1, \mathbf{v}_2\}$ a basis for $\mathbb{R}^2$?
2. Let $\mathbf{v}_1 = \begin{bmatrix}1\\-3\\4\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}6\\2\\-1\end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix}2\\-2\\3\end{bmatrix}$, and $\mathbf{v}_4 = \begin{bmatrix}-4\\-8\\9\end{bmatrix}$. Find a basis for the subspace W spanned by $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$.
3. Let $\mathbf{v}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}$, and $H = \left\{\begin{bmatrix}s\\s\\0\end{bmatrix} : s \text{ in } \mathbb{R}\right\}$. Then every vector in H is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$ because
$$\begin{bmatrix}s\\s\\0\end{bmatrix} = s\begin{bmatrix}1\\0\\0\end{bmatrix} + s\begin{bmatrix}0\\1\\0\end{bmatrix}$$
Is $\{\mathbf{v}_1, \mathbf{v}_2\}$ a basis for H?
4. Let V and W be vector spaces, let $T : V \to W$ and $U : V \to W$ be linear transformations, and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ be a basis for V. If $T(\mathbf{v}_j) = U(\mathbf{v}_j)$ for every value of j between 1 and p, show that $T(\mathbf{x}) = U(\mathbf{x})$ for every vector $\mathbf{x}$ in V.

STUDY GUIDE offers additional resources for mastering the concept of basis.

4.3 Exercises

Determine which sets in Exercises 1–8 are bases for $\mathbb{R}^3$. Of the sets that are *not* bases, determine which ones are linearly independent and which ones span $\mathbb{R}^3$. Justify your answers.

1. $\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}1\\1\\0\end{bmatrix},\begin{bmatrix}1\\1\\1\end{bmatrix}$

2. $\begin{bmatrix}1\\0\\1\end{bmatrix},\begin{bmatrix}0\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix}$

3. $\begin{bmatrix}1\\0\\-2\end{bmatrix},\begin{bmatrix}3\\2\\-4\end{bmatrix},\begin{bmatrix}-3\\-5\\1\end{bmatrix}$

4. $\begin{bmatrix}2\\-2\\1\end{bmatrix},\begin{bmatrix}1\\-3\\2\end{bmatrix},\begin{bmatrix}-7\\5\\4\end{bmatrix}$

5. $\begin{bmatrix}1\\-3\\0\end{bmatrix},\begin{bmatrix}-2\\9\\0\end{bmatrix},\begin{bmatrix}0\\0\\0\end{bmatrix},\begin{bmatrix}0\\-3\\5\end{bmatrix}$

6. $\begin{bmatrix}1\\2\\-3\end{bmatrix},\begin{bmatrix}-4\\-5\\6\end{bmatrix}$

7. $\begin{bmatrix} -2 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 6 \\ -1 \\ 5 \end{bmatrix}$

8. $\begin{bmatrix} 1 \\ -4 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ -5 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix}$

Find bases for the null spaces of the matrices given in Exercises 9 and 10. Refer to the remarks that follow Example 3 in Section 4.2.

9. $\begin{bmatrix} 1 & 0 & -3 & 2 \\ 0 & 1 & -5 & 4 \\ 3 & -2 & 1 & -2 \end{bmatrix}$

10. $\begin{bmatrix} 1 & 0 & -5 & 1 & 4 \\ -2 & 1 & 6 & -2 & -2 \\ 0 & 2 & -8 & 1 & 9 \end{bmatrix}$

11. Find a basis for the set of vectors in $\mathbb{R}^3$ in the plane $x + 4y - 5z = 0$. [Hint: Think of the equation as a "system" of homogeneous equations.]

12. Find a basis for the set of vectors in $\mathbb{R}^2$ on the line $y = 5x$.

In Exercises 13 and 14, assume that A is row equivalent to B. Find bases for Nul A, Col A, and Row A.

13. $A = \begin{bmatrix} -2 & 4 & -2 & -4 \\ 2 & -6 & -3 & 1 \\ -3 & 8 & 2 & -3 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 6 & 5 \\ 0 & 2 & 5 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

14. $A = \begin{bmatrix} 1 & 2 & -5 & 11 & -3 \\ 2 & 4 & -5 & 15 & 2 \\ 1 & 2 & 0 & 4 & 5 \\ 3 & 6 & -5 & 19 & -2 \end{bmatrix},$

$B = \begin{bmatrix} 1 & 2 & 0 & 4 & 5 \\ 0 & 0 & 5 & -7 & 8 \\ 0 & 0 & 0 & 0 & -9 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

In Exercises 15–18, find a basis for the space spanned by the given vectors, $\mathbf{v}_1, \ldots, \mathbf{v}_5$.

15. $\begin{bmatrix} 1 \\ 0 \\ -3 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} -3 \\ -4 \\ 1 \\ 6 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ -8 \\ 7 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -6 \\ 9 \end{bmatrix}$

16. $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 6 \\ -1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 5 \\ -3 \\ 3 \\ -4 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ -1 \\ 1 \end{bmatrix}$

T 17. $\begin{bmatrix} 8 \\ 9 \\ -3 \\ -6 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 1 \\ -4 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ -4 \\ -9 \\ 6 \\ -7 \end{bmatrix}, \begin{bmatrix} 6 \\ 8 \\ 4 \\ -7 \\ 10 \end{bmatrix}, \begin{bmatrix} -1 \\ 4 \\ 11 \\ -8 \\ -7 \end{bmatrix}$

T 18. $\begin{bmatrix} -8 \\ 7 \\ 6 \\ 5 \\ -7 \end{bmatrix}, \begin{bmatrix} 8 \\ -7 \\ -9 \\ -5 \\ 7 \end{bmatrix}, \begin{bmatrix} -8 \\ 7 \\ 4 \\ 5 \\ -7 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ 9 \\ 6 \\ -7 \end{bmatrix}, \begin{bmatrix} -9 \\ 3 \\ -4 \\ -1 \\ 0 \end{bmatrix}$

19. Let $\mathbf{v}_1 = \begin{bmatrix} 4 \\ -3 \\ 7 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 9 \\ -2 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 7 \\ 11 \\ 6 \end{bmatrix}$, and $H = \text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. It can be verified that $4\mathbf{v}_1 + 5\mathbf{v}_2 - 3\mathbf{v}_3 = \mathbf{0}$. Use this information to find a basis for H. There is more than one answer.

20. Let $\mathbf{v}_1 = \begin{bmatrix} 7 \\ 4 \\ -9 \\ -5 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 4 \\ -7 \\ 2 \\ 5 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 1 \\ -5 \\ 3 \\ 4 \end{bmatrix}$. It can be verified that $\mathbf{v}_1 - 3\mathbf{v}_2 + 5\mathbf{v}_3 = \mathbf{0}$. Use this information to find a basis for $H = \text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

In Exercises 21–32, mark each statement True or False **(T/F)**. Justify each answer.

21. **(T/F)** A single vector by itself is linearly dependent.

22. **(T/F)** A linearly independent set in a subspace H is a basis for H.

23. **(T/F)** If $H = \text{Span}\,\{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$, then $\{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$ is a basis for H.

24. **(T/F)** If a finite set S of nonzero vectors spans a vector space V, then some subset of S is a basis for V.

25. **(T/F)** The columns of an invertible $n \times n$ matrix form a basis for $\mathbb{R}^n$.

26. **(T/F)** A basis is a linearly independent set that is as large as possible.

27. **(T/F)** A basis is a spanning set that is as large as possible.

28. **(T/F)** The standard method for producing a spanning set for Nul A, described in Section 4.2, sometimes fails to produce a basis for Nul A.

29. **(T/F)** In some cases, the linear dependence relations among the columns of a matrix can be affected by certain elementary row operations on the matrix.

30. **(T/F)** If B is an echelon form of a matrix A, then the pivot columns of B form a basis for Col A.

31. **(T/F)** Row operations preserve the linear dependence relations among the rows of A.

32. **(T/F)** If A and B are row equivalent, then their row spaces are the same.

33. Suppose $\mathbb{R}^4 = \text{Span}\,\{\mathbf{v}_1, \ldots, \mathbf{v}_4\}$. Explain why $\{\mathbf{v}_1, \ldots, \mathbf{v}_4\}$ is a basis for $\mathbb{R}^4$.

34. Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a linearly independent set in $\mathbb{R}^n$. Explain why $\mathcal{B}$ must be a basis for $\mathbb{R}^n$.

35. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and let H be the set of vectors in $\mathbb{R}^3$ whose second and third entries are equal. Then every vector in H has a unique expansion as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, because

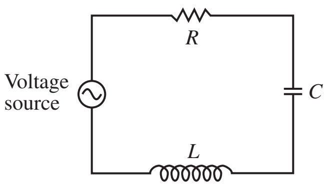
$$\begin{bmatrix} s \\ t \\ t \end{bmatrix} = s\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + (t-s)\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + s\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

for any s and t. Is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ a basis for H? Why or why not?

36. In the vector space of all real-valued functions, find a basis for the subspace spanned by $\{\sin t, \sin 2t, \sin t \cos t\}$.

37. Let V be the vector space of functions that describe the vibration of a mass–spring system. (Refer to Exercise 19 in Section 4.1.) Find a basis for V.

38. (*RLC circuit*) The circuit in the figure consists of a resistor (R ohms), an inductor (L henrys), a capacitor (C farads), and an initial voltage source. Let $b = R/(2L)$, and suppose R, L, and C have been selected so that b also equals $1/\sqrt{LC}$. (This is done, for instance, when the circuit is used in a voltmeter.) Let $v(t)$ be the voltage (in volts) at time t, measured across the capacitor. It can be shown that v is in the null space H of the linear transformation that maps $v(t)$ into $Lv''(t) + Rv'(t) + (1/C)v(t)$, and H consists of all functions of the form $v(t) = e^{-bt}(c_1 + c_2 t)$. Find a basis for H.



Exercises 39 and 40 show that every basis for $\mathbb{R}^n$ must contain exactly n vectors.

39. Let $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ be a set of k vectors in $\mathbb{R}^n$, with $k < n$. Use a theorem from Section 1.4 to explain why S cannot be a basis for $\mathbb{R}^n$.

40. Let $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ be a set of k vectors in $\mathbb{R}^n$, with $k > n$. Use a theorem from Chapter 1 to explain why S cannot be a basis for $\mathbb{R}^n$.

Exercises 41 and 42 reveal an important connection between linear independence and linear transformations and provide practice using the definition of linear dependence. Let V and W be vector spaces, let $T : V \to W$ be a linear transformation, and let $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ be a subset of V.

41. Show that if $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is linearly dependent in V, then the set of images, $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_p)\}$, is linearly dependent in W. This fact shows that if a linear transformation maps a set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ onto a linearly *independent* set $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_p)\}$, then the original set is linearly independent, too (because it cannot be linearly dependent).

42. Suppose that T is a one-to-one transformation, so that an equation $T(\mathbf{u}) = T(\mathbf{v})$ always implies $\mathbf{u} = \mathbf{v}$. Show that if the set of images $\{T(\mathbf{v}_1), \dots, T(\mathbf{v}_p)\}$ is linearly dependent, then $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is linearly dependent. This fact shows that *a one-to-one linear transformation maps a linearly independent set onto a linearly independent set* (because in this case the set of images cannot be linearly dependent).

43. Consider the polynomials $\mathbf{p}_1(t) = 1 + t^2$ and $\mathbf{p}_2(t) = 1 - t^2$. Is $\{\mathbf{p}_1, \mathbf{p}_2\}$ a linearly independent set in $\mathbb{P}_3$? Why or why not?

44. Consider the polynomials $\mathbf{p}_1(t) = 1 + t$, $\mathbf{p}_2(t) = 1 - t$, and $\mathbf{p}_3(t) = 2$ (for all t). By inspection, write a linear dependence relation among $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$. Then find a basis for Span $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$.

45. Let V be a vector space that contains a linearly independent set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$. Describe how to construct a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ in V such that $\{\mathbf{v}_1, \mathbf{v}_3\}$ is a basis for Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$.

T 46. Let $H = \text{Span}\,\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ and $K = \text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 3 \\ 0 \\ -1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 0 \\ 3 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 2 \\ -3 \\ 6 \\ -5 \end{bmatrix},$$

$$\mathbf{v}_1 = \begin{bmatrix} -4 \\ 3 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 9 \\ -4 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -1 \\ 7 \\ 6 \\ 5 \end{bmatrix}$$

Find bases for H, K, and $H + K$. (See Exercises 41 and 42 in Section 4.1.)

T 47. Show that $\{t, \sin t, \cos 2t, \sin t \cos t\}$ is a linearly independent set of functions defined on $\mathbb{R}$. Start by assuming that

$$c_1 t + c_2 \sin t + c_3 \cos 2t + c_4 \sin t \cos t = 0 \qquad (5)$$

Equation (5) must hold for all real t, so choose several specific values of t (say, $t = 0, .1, .2$) until you get a system of enough equations to determine that all the c_j must be zero.

T 48. Show that $\{1, \cos t, \cos^2 t, \dots, \cos^6 t\}$ is a linearly independent set of functions defined on $\mathbb{R}$. Use the method of Exercise 47. (This result will be needed in Exercise 54 in Section 4.5.)

Solutions to Practice Problems

1. Let $A = [\,\mathbf{v}_1 \quad \mathbf{v}_2\,]$. Row operations show that

$$A = \begin{bmatrix} 1 & -2 \\ -2 & 7 \\ 3 & -9 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 \\ 0 & 3 \\ 0 & 0 \end{bmatrix}$$

Not every row of A contains a pivot position. So the columns of A do not span $\mathbb{R}^3$, by Theorem 4 in Section 1.4. Hence $\{\mathbf{v}_1, \mathbf{v}_2\}$ is not a basis for $\mathbb{R}^3$. Since $\mathbf{v}_1$ and $\mathbf{v}_2$ are not in $\mathbb{R}^2$, they cannot possibly be a basis for $\mathbb{R}^2$. However, since $\mathbf{v}_1$ and $\mathbf{v}_2$ are obviously linearly independent, they are a basis for a subspace of $\mathbb{R}^3$, namely Span $\{\mathbf{v}_1, \mathbf{v}_2\}$.

2. Set up a matrix A whose column space is the space spanned by $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$, and then row reduce A to find its pivot columns.

$$A = \begin{bmatrix} 1 & 6 & 2 & -4 \\ -3 & 2 & -2 & -8 \\ 4 & -1 & 3 & 9 \end{bmatrix} \sim \begin{bmatrix} 1 & 6 & 2 & -4 \\ 0 & 20 & 4 & -20 \\ 0 & -25 & -5 & 25 \end{bmatrix} \sim \begin{bmatrix} 1 & 6 & 2 & -4 \\ 0 & 5 & 1 & -5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The first two columns of A are the pivot columns and hence form a basis of Col $A = W$. Hence $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for W. Note that the reduced echelon form of A is not needed in order to locate the pivot columns.

3. Neither $\mathbf{v}_1$ nor $\mathbf{v}_2$ is in H, so $\{\mathbf{v}_1, \mathbf{v}_2\}$ cannot be a basis for H. In fact, $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for the *plane* of all vectors of the form $(c_1, c_2, 0)$, but H is only a *line*.

4. Since $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is a basis for V, for any vector $\mathbf{x}$ in V, there exist scalars $c_1, \dots, c_p$ such that $\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$. Then since T and U are linear transformations

$$\begin{aligned} T(\mathbf{x}) &= T(c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p) = c_1T(\mathbf{v}_1) + \cdots + c_pT(\mathbf{v}_p) \\ &= c_1U(\mathbf{v}_1) + \cdots + c_pU(\mathbf{v}_p) = U(c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p) \\ &= U(\mathbf{x}) \end{aligned}$$

4.4 Coordinate Systems

An important reason for specifying a basis $\mathcal{B}$ for a vector space V is to impose a "coordinate system" on V. This section will show that if $\mathcal{B}$ contains n vectors, then the coordinate system will make V act like $\mathbb{R}^n$. If V is already $\mathbb{R}^n$ itself, then $\mathcal{B}$ will determine a coordinate system that gives a new "view" of V.

The existence of coordinate systems rests on the following fundamental result.

THEOREM 8

The Unique Representation Theorem

Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for a vector space V. Then for each $\mathbf{x}$ in V, there exists a unique set of scalars $c_1, \dots, c_n$ such that

$$\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_n\mathbf{b}_n \tag{1}$$

PROOF Since $\mathcal{B}$ spans V, there exist scalars such that (1) holds. Suppose $\mathbf{x}$ also has the representation

$$\mathbf{x} = d_1\mathbf{b}_1 + \cdots + d_n\mathbf{b}_n$$

for scalars $d_1, \dots, d_n$. Then, subtracting, we have

$$\mathbf{0} = \mathbf{x} - \mathbf{x} = (c_1 - d_1)\mathbf{b}_1 + \cdots + (c_n - d_n)\mathbf{b}_n \tag{2}$$

Since $\mathcal{B}$ is linearly independent, the weights in (2) must all be zero. That is, $c_j = d_j$ for $1 \le j \le n$. ■

DEFINITION

Suppose $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ is a basis for a vector space V and $\mathbf{x}$ is in V. The **coordinates of x relative to the basis $\mathcal{B}$** (or the **$\mathcal{B}$-coordinates of x**) are the weights $c_1, \dots, c_n$ such that $\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_n\mathbf{b}_n$.

If $c_1, \dots, c_n$ are the $\mathcal{B}$-coordinates of $\mathbf{x}$, then the vector in $\mathbb{R}^n$

$$[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

is the **coordinate vector of x (relative to $\mathcal{B}$)**, or the **$\mathcal{B}$-coordinate vector of x**. The mapping $\mathbf{x} \mapsto [\,\mathbf{x}\,]_{\mathcal{B}}$ is the **coordinate mapping (determined by $\mathcal{B}$)**.[1]

EXAMPLE 1 Consider a basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ for $\mathbb{R}^2$, where $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Suppose an $\mathbf{x}$ in $\mathbb{R}^2$ has the coordinate vector $[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$. Find $\mathbf{x}$.

SOLUTION The $\mathcal{B}$-coordinates of $\mathbf{x}$ tell how to build $\mathbf{x}$ from the vectors in $\mathcal{B}$. That is,

$$\mathbf{x} = (-2)\mathbf{b}_1 + 3\mathbf{b}_2 = (-2)\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$$ ■

EXAMPLE 2 The entries in the vector $\mathbf{x} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$ are the coordinates of $\mathbf{x}$ relative to the *standard basis* $\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2\}$, since

$$\begin{bmatrix} 1 \\ 6 \end{bmatrix} = 1\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 6\begin{bmatrix} 0 \\ 1 \end{bmatrix} = 1\mathbf{e}_1 + 6\mathbf{e}_2$$

If $\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2\}$, then $[\,\mathbf{x}\,]_{\mathcal{E}} = \mathbf{x}$. ■

A Graphical Interpretation of Coordinates

A coordinate system on a set consists of a one-to-one mapping of the points in the set into $\mathbb{R}^n$. For example, ordinary graph paper provides a coordinate system for the plane

[1] The concept of a coordinate mapping assumes that the basis $\mathcal{B}$ is an indexed set whose vectors are listed in some fixed preassigned order. This property makes the definition of $[\,\mathbf{x}\,]_{\mathcal{B}}$ unambiguous.

when one selects perpendicular axes and a unit of measurement on each axis. Figure 1 shows the standard basis $\{\mathbf{e}_1, \mathbf{e}_2\}$, the vectors $\mathbf{b}_1 (= \mathbf{e}_1)$ and $\mathbf{b}_2$ from Example 1, and the vector $\mathbf{x} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$. The coordinates 1 and 6 give the location of $\mathbf{x}$ relative to the standard basis: 1 unit in the $\mathbf{e}_1$ direction and 6 units in the $\mathbf{e}_2$ direction.

Figure 2 shows the vectors $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{x}$ from Figure 1. (Geometrically, the three vectors lie on a vertical line in both figures.) However, the standard coordinate grid was erased and replaced by a grid especially adapted to the basis $\mathcal{B}$ in Example 1. The coordinate vector $[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$ gives the location of $\mathbf{x}$ on this new coordinate system: -2 units in the $\mathbf{b}_1$ direction and 3 units in the $\mathbf{b}_2$ direction.

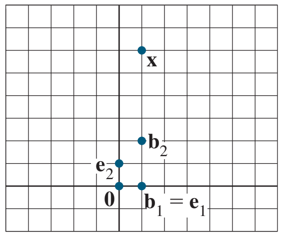


FIGURE 1 Standard graph paper.

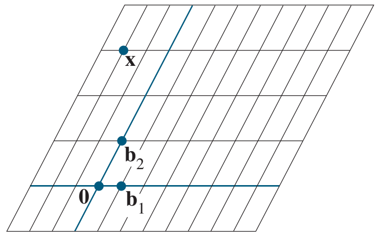
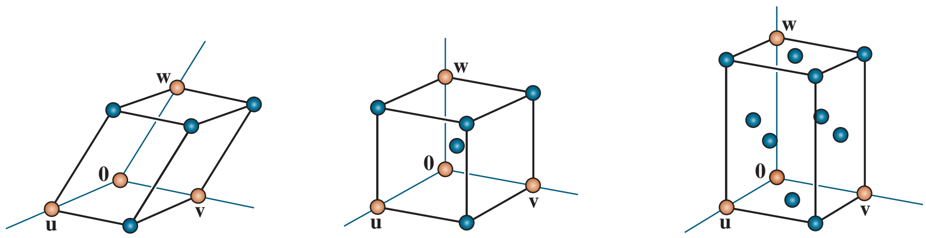


FIGURE 2 $\mathcal{B}$-graph paper.

EXAMPLE 3 In crystallography, the description of a crystal lattice is aided by choosing a basis $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ for $\mathbb{R}^3$ that corresponds to three adjacent edges of one "unit cell" of the crystal. An entire lattice is constructed by stacking together many copies of one cell. There are fourteen basic types of unit cells; three are displayed in Figure 3.[2]



(a) Simple monoclinic

(b) Body-centered cubic

(c) Face-centered orthorhombic

FIGURE 3 Examples of unit cells.

The coordinates of atoms within the crystal are given relative to the basis for the lattice. For instance,

$$\begin{bmatrix} 1/2 \\ 1/2 \\ 1 \end{bmatrix}$$

identifies the top face-centered atom in the cell in Figure 3(c). ■

---

[2] Adapted from *The Science and Engineering of Materials*, 4th Ed., by Donald R. Askeland (Boston: Prindle, Weber & Schmidt, © 2002), p. 36.

Coordinates in $\mathbb{R}^n$

When a basis $\mathcal{B}$ for $\mathbb{R}^n$ is fixed, the $\mathcal{B}$-coordinate vector of a specified $\mathbf{x}$ is easily found, as in the next example.

EXAMPLE 4 Let $\mathbf{b}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$, and $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$. Find the coordinate vector $[\,\mathbf{x}\,]_{\mathcal{B}}$ of $\mathbf{x}$ relative to $\mathcal{B}$.

SOLUTION The $\mathcal{B}$-coordinates c_1, c_2 of $\mathbf{x}$ satisfy

$$c_1 \underset{\mathbf{b}_1}{\begin{bmatrix} 2 \\ 1 \end{bmatrix}} + c_2 \underset{\mathbf{b}_2}{\begin{bmatrix} -1 \\ 1 \end{bmatrix}} = \underset{\mathbf{x}}{\begin{bmatrix} 4 \\ 5 \end{bmatrix}}$$

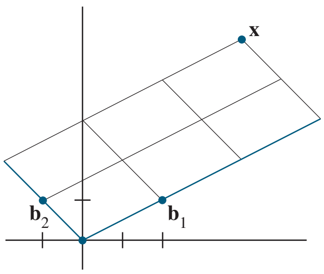


FIGURE 4
The $\mathcal{B}$-coordinate vector of $\mathbf{x}$ is $(3, 2)$.

or

$$\begin{array}{cc} \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \\ \mathbf{b}_1 \quad \mathbf{b}_2 \end{array} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \underset{\mathbf{x}}{\begin{bmatrix} 4 \\ 5 \end{bmatrix}} \tag{3}$$

This equation can be solved by row operations on an augmented matrix or by multiplying the vector $\mathbf{x}$ by the inverse of the matrix. In any case, the solution is $c_1 = 3$, $c_2 = 2$. Thus $\mathbf{x} = 3\mathbf{b}_1 + 2\mathbf{b}_2$, and

$$[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

See Figure 4. ■

The matrix in (3) changes the $\mathcal{B}$-coordinates of a vector $\mathbf{x}$ into the standard coordinates for $\mathbf{x}$. An analogous change of coordinates can be carried out in $\mathbb{R}^n$ for a basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$. Let

$$P_{\mathcal{B}} = [\,\mathbf{b}_1 \quad \mathbf{b}_2 \quad \cdots \quad \mathbf{b}_n\,]$$

Then the vector equation

$$\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n$$

is equivalent to

$$\boxed{\mathbf{x} = P_{\mathcal{B}}[\,\mathbf{x}\,]_{\mathcal{B}}} \tag{4}$$

We call $P_{\mathcal{B}}$ the **change-of-coordinates matrix** from $\mathcal{B}$ to the standard basis in $\mathbb{R}^n$. Left-multiplication by $P_{\mathcal{B}}$ transforms the coordinate vector $[\,\mathbf{x}\,]_{\mathcal{B}}$ into $\mathbf{x}$. The change-of-coordinates equation (4) is important and will be needed at several points in Chapters 5 and 7.

Since the columns of $P_{\mathcal{B}}$ form a basis for $\mathbb{R}^n$, $P_{\mathcal{B}}$ is invertible (by the Invertible Matrix Theorem). Left-multiplication by $P_{\mathcal{B}}^{-1}$ converts $\mathbf{x}$ into its $\mathcal{B}$-coordinate vector:

$$P_{\mathcal{B}}^{-1}\mathbf{x} = [\,\mathbf{x}\,]_{\mathcal{B}}$$

The correspondence $\mathbf{x} \mapsto [\,\mathbf{x}\,]_{\mathcal{B}}$, produced here by $P_{\mathcal{B}}^{-1}$, is the coordinate mapping mentioned earlier. Since $P_{\mathcal{B}}^{-1}$ is an invertible matrix, the coordinate mapping is a one-to-one linear transformation from $\mathbb{R}^n$ onto $\mathbb{R}^n$, by the Invertible Matrix Theorem. (See also Theorem 12 in Section 1.9.) This property of the coordinate mapping is also true in a general vector space that has a basis, as we shall see.

The Coordinate Mapping

Choosing a basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ for a vector space V introduces a coordinate system in V. The coordinate mapping $\mathbf{x} \mapsto [\,\mathbf{x}\,]_{\mathcal{B}}$ connects the possibly unfamiliar space V to the familiar space $\mathbb{R}^n$. See Figure 5. Points in V can now be identified by their new "names."

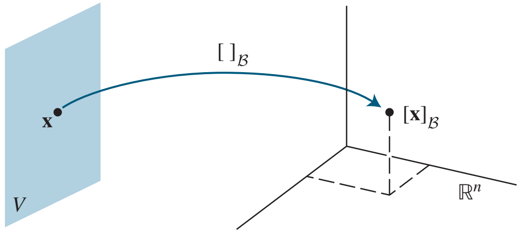


FIGURE 5 The coordinate mapping from V onto $\mathbb{R}^n$.

THEOREM 9

Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for a vector space V. Then the coordinate mapping $\mathbf{x} \mapsto [\,\mathbf{x}\,]_{\mathcal{B}}$ is a one-to-one linear transformation from V onto $\mathbb{R}^n$.

PROOF Take two typical vectors in V, say,

$$\mathbf{u} = c_1\mathbf{b}_1 + \cdots + c_n\mathbf{b}_n$$
$$\mathbf{w} = d_1\mathbf{b}_1 + \cdots + d_n\mathbf{b}_n$$

Then, using vector operations,

$$\mathbf{u} + \mathbf{w} = (c_1 + d_1)\mathbf{b}_1 + \cdots + (c_n + d_n)\mathbf{b}_n$$

It follows that

$$[\,\mathbf{u} + \mathbf{w}\,]_{\mathcal{B}} = \begin{bmatrix} c_1 + d_1 \\ \vdots \\ c_n + d_n \end{bmatrix} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} + \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix} = [\,\mathbf{u}\,]_{\mathcal{B}} + [\,\mathbf{w}\,]_{\mathcal{B}}$$

So the coordinate mapping preserves addition. If r is any scalar, then

$$r\mathbf{u} = r(c_1\mathbf{b}_1 + \cdots + c_n\mathbf{b}_n) = (rc_1)\mathbf{b}_1 + \cdots + (rc_n)\mathbf{b}_n$$

So

$$[\,r\mathbf{u}\,]_{\mathcal{B}} = \begin{bmatrix} rc_1 \\ \vdots \\ rc_n \end{bmatrix} = r\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = r[\,\mathbf{u}\,]_{\mathcal{B}}$$

Thus the coordinate mapping also preserves scalar multiplication and hence is a linear transformation. See Exercises 27 and 28 for verification that the coordinate mapping is one-to-one and maps V onto $\mathbb{R}^n$. ■

The linearity of the coordinate mapping extends to linear combinations, just as in Section 1.8. If $\mathbf{u}_1, \dots, \mathbf{u}_p$ are in V and if $c_1, \dots, c_p$ are scalars, then

$$[\,c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p\,]_{\mathcal{B}} = c_1[\,\mathbf{u}_1\,]_{\mathcal{B}} + \cdots + c_p[\,\mathbf{u}_p\,]_{\mathcal{B}} \tag{5}$$

In words, (5) says that the $\mathcal{B}$-coordinate vector of a linear combination of $\mathbf{u}_1, \ldots, \mathbf{u}_p$ is the *same* linear combination of their coordinate vectors.

The coordinate mapping in Theorem 9 is an important example of an *isomorphism* from V onto $\mathbb{R}^n$. In general, a one-to-one linear transformation from a vector space V onto a vector space W is called an **isomorphism** from V onto W (*iso* from the Greek for "the same," and *morph* from the Greek for "form" or "structure"). The notation and terminology for V and W may differ, but the two spaces are indistinguishable as vector spaces. *Every vector space calculation in V is accurately reproduced in W, and vice versa*. In particular, any real vector space with a basis of n vectors is indistinguishable from $\mathbb{R}^n$. See Exercises 29 and 30.

STUDY GUIDE offers additional resources about isomorphic vector spaces.

EXAMPLE 5 Let $\mathcal{B}$ be the standard basis of the space $\mathbb{P}_3$ of polynomials; that is, let $\mathcal{B} = \{1, t, t^2, t^3\}$. A typical element $\mathbf{p}$ of $\mathbb{P}_3$ has the form

$$\mathbf{p}(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3$$

Since $\mathbf{p}$ is already displayed as a linear combination of the standard basis vectors, we conclude that

$$[\,\mathbf{p}\,]_{\mathcal{B}} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Thus the coordinate mapping $\mathbf{p} \mapsto [\,\mathbf{p}\,]_{\mathcal{B}}$ is an isomorphism from $\mathbb{P}_3$ onto $\mathbb{R}^4$. All vector space operations in $\mathbb{P}_3$ correspond to operations in $\mathbb{R}^4$. ■

If we think of $\mathbb{P}_3$ and $\mathbb{R}^4$ as displays on two computer screens that are connected via the coordinate mapping, then every vector space operation in $\mathbb{P}_3$ on one screen is exactly duplicated by a corresponding vector operation in $\mathbb{R}^4$ on the other screen. The vectors on the $\mathbb{P}_3$ screen look different from those on the $\mathbb{R}^4$ screen, but they "act" as vectors in exactly the same way. See Figure 6.

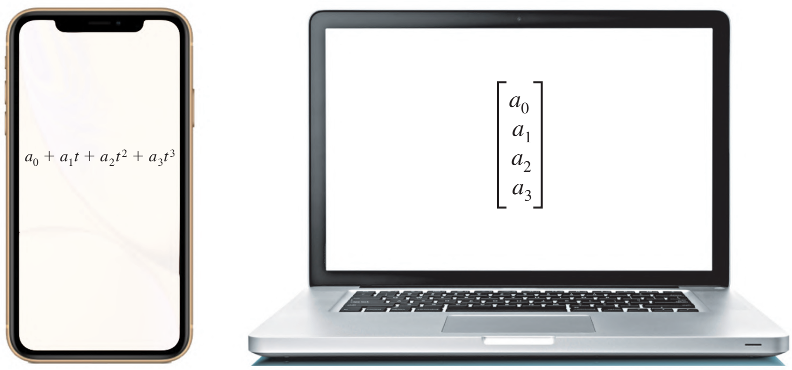


FIGURE 6 The space $\mathbb{P}_3$ is isomorphic to $\mathbb{R}^4$.

EXAMPLE 6 Use coordinate vectors to verify that the polynomials $1 + 2t^2$, $4 + t + 5t^2$, and $3 + 2t$ are linearly dependent in $\mathbb{P}_2$.

SOLUTION The coordinate mapping from Example 5 produces the coordinate vectors $(1, 0, 2)$, $(4, 1, 5)$, and $(3, 2, 0)$, respectively. Writing these vectors as the *columns* of a matrix A, we can determine their independence by row reducing the augmented matrix for $A\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} 1 & 4 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 2 & 5 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 4 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The columns of A are linearly dependent, so the corresponding polynomials are linearly dependent. In fact, it is easy to check that column 3 of A is 2 times column 2 minus 5 times column 1. The corresponding relation for the polynomials is

$$3 + 2t = 2(4 + t + 5t^2) - 5(1 + 2t^2) \qquad \blacksquare$$

The final example concerns a plane in $\mathbb{R}^3$ that is isomorphic to $\mathbb{R}^2$.

EXAMPLE 7 Let

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 3 \\ 12 \\ 7 \end{bmatrix},$$

and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$. Then $\mathcal{B}$ is a basis for $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$. Determine if $\mathbf{x}$ is in H, and if it is, find the coordinate vector of $\mathbf{x}$ relative to $\mathcal{B}$.

SOLUTION If $\mathbf{x}$ is in H, then the following vector equation is consistent:

$$c_1 \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 12 \\ 7 \end{bmatrix}$$

The scalars c_1 and c_2, if they exist, are the $\mathcal{B}$-coordinates of $\mathbf{x}$. Using row operations, we obtain

$$\begin{bmatrix} 3 & -1 & 3 \\ 6 & 0 & 12 \\ 2 & 1 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus $c_1 = 2$, $c_2 = 3$, and $[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. The coordinate system on H determined by $\mathcal{B}$ is shown in Figure 7. $\blacksquare$

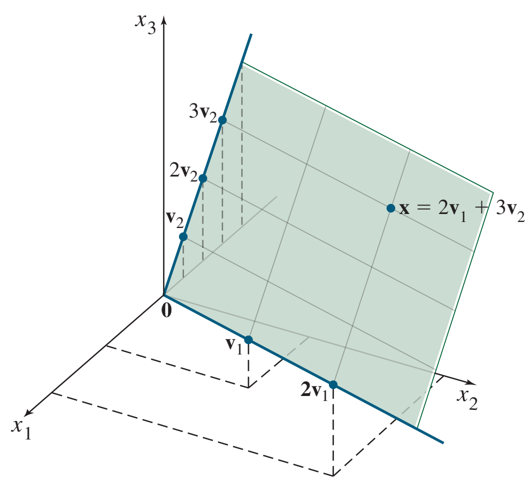


FIGURE 7 A coordinate system on a plane H in $\mathbb{R}^3$.

If a different basis for H were chosen, would the associated coordinate system also make H isomorphic to $\mathbb{R}^2$? Surely, this must be true. We shall prove it in the next section.

Practice Problems

1. Let $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} -3 \\ 4 \\ 0 \end{bmatrix}$, $\mathbf{b}_3 = \begin{bmatrix} 3 \\ -6 \\ 3 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} -8 \\ 2 \\ 3 \end{bmatrix}$.
   a. Show that the set $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ is a basis of $\mathbb{R}^3$.
   b. Find the change-of-coordinates matrix from $\mathcal{B}$ to the standard basis.
   c. Write the equation that relates $\mathbf{x}$ in $\mathbb{R}^3$ to $[\,\mathbf{x}\,]_{\mathcal{B}}$.
   d. Find $[\,\mathbf{x}\,]_{\mathcal{B}}$, for the $\mathbf{x}$ given above.
2. The set $\mathcal{B} = \{1+t, 1+t^2, t+t^2\}$ is a basis for $\mathbb{P}_2$. Find the coordinate vector of $\mathbf{p}(t) = 6 + 3t - t^2$ relative to $\mathcal{B}$.

4.4 Exercises

In Exercises 1–4, find the vector $\mathbf{x}$ determined by the given coordinate vector $[\,\mathbf{x}\,]_{\mathcal{B}}$ and the given basis $\mathcal{B}$.

1. $\mathcal{B} = \left\{\begin{bmatrix} 3 \\ -5 \end{bmatrix}, \begin{bmatrix} -4 \\ 6 \end{bmatrix}\right\}$, $[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$

2. $\mathcal{B} = \left\{\begin{bmatrix} 4 \\ 5 \end{bmatrix}, \begin{bmatrix} 6 \\ 7 \end{bmatrix}\right\}$, $[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} 8 \\ -5 \end{bmatrix}$

3. $\mathcal{B} = \left\{\begin{bmatrix} 1 \\ -8 \\ 6 \end{bmatrix}, \begin{bmatrix} 2 \\ -5 \\ 7 \end{bmatrix}, \begin{bmatrix} 3 \\ 9 \\ -4 \end{bmatrix}\right\}$, $[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} 2 \\ -3 \\ 0 \end{bmatrix}$

4. $\mathcal{B} = \left\{\begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ -5 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 \\ -7 \\ 3 \end{bmatrix}\right\}$, $[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} -4 \\ 8 \\ -7 \end{bmatrix}$

In Exercises 5–8, find the coordinate vector $[\,\mathbf{x}\,]_{\mathcal{B}}$ of $\mathbf{x}$ relative to the given basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$.

5. $\mathbf{b}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} 2 \\ -5 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$

6. $\mathbf{b}_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} 5 \\ -6 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$

7. $\mathbf{b}_1 = \begin{bmatrix} 1 \\ -1 \\ -3 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} -3 \\ 4 \\ 9 \end{bmatrix}$, $\mathbf{b}_3 = \begin{bmatrix} 2 \\ -2 \\ 4 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 8 \\ -9 \\ 6 \end{bmatrix}$

8. $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} 3 \\ 1 \\ 7 \end{bmatrix}$, $\mathbf{b}_3 = \begin{bmatrix} 1 \\ -1 \\ 5 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$

In Exercises 9 and 10, find the change-of-coordinates matrix from $\mathcal{B}$ to the standard basis in $\mathbb{R}^n$.

9. $\mathcal{B} = \left\{\begin{bmatrix} 2 \\ -9 \end{bmatrix}, \begin{bmatrix} 1 \\ 8 \end{bmatrix}\right\}$

10. $\mathcal{B} = \left\{\begin{bmatrix} 5 \\ -2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ -7 \\ 8 \end{bmatrix}\right\}$

In Exercises 11 and 12, use an inverse matrix to find $[\,\mathbf{x}\,]_{\mathcal{B}}$ for the given $\mathbf{x}$ and $\mathcal{B}$.

11. $\mathcal{B} = \left\{\begin{bmatrix} 3 \\ -5 \end{bmatrix}, \begin{bmatrix} -4 \\ 6 \end{bmatrix}\right\}$, $\mathbf{x} = \begin{bmatrix} 2 \\ -6 \end{bmatrix}$

12. $\mathcal{B} = \left\{\begin{bmatrix} 4 \\ 5 \end{bmatrix}, \begin{bmatrix} 6 \\ 7 \end{bmatrix}\right\}$, $\mathbf{x} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$

13. The set $\mathcal{B} = \{1+t^2, t+t^2, 1+2t+t^2\}$ is a basis for $\mathbb{P}_2$. Find the coordinate vector of $\mathbf{p}(t) = 1 + 4t + 7t^2$ relative to $\mathcal{B}$.

14. The set $\mathcal{B} = \{1-t^2, t-t^2, 2-2t+t^2\}$ is a basis for $\mathbb{P}_2$. Find the coordinate vector of $\mathbf{p}(t) = 3 + t - 6t^2$ relative to $\mathcal{B}$.

In Exercises 15–20, mark each statement True or False **(T/F)**. Justify each answer. Unless stated otherwise, $\mathcal{B}$ is a basis for a vector space V.

15. **(T/F)** If $\mathbf{x}$ is in V and if $\mathcal{B}$ contains n vectors, then the $\mathcal{B}$-coordinate vector of $\mathbf{x}$ is in $\mathbb{R}^n$.

16. **(T/F)** If $\mathcal{B}$ is the standard basis for $\mathbb{R}^n$, then the $\mathcal{B}$-coordinate vector of an $\mathbf{x}$ in $\mathbb{R}^n$ is $\mathbf{x}$ itself.

17. **(T/F)** If $P_{\mathcal{B}}$ is the change-of-coordinates matrix, then $[\mathbf{x}]_{\mathcal{B}} = P_{\mathcal{B}}\,\mathbf{x}$, for $\mathbf{x}$ in V.

18. **(T/F)** The correspondence $[\,\mathbf{x}\,]_{\mathcal{B}} \mapsto \mathbf{x}$ is called the coordinate mapping.

19. **(T/F)** The vector spaces $\mathbb{P}_3$ and $\mathbb{R}^3$ are isomorphic.

20. **(T/F)** In some cases, a plane in $\mathbb{R}^3$ can be isomorphic to $\mathbb{R}^2$.

21. The vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 2 \\ -8 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} -3 \\ 7 \end{bmatrix}$ span $\mathbb{R}^2$ but do not form a basis. Find two different ways to express $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$.

22. Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis for a vector space V. Explain why the $\mathcal{B}$-coordinate vectors of $\mathbf{b}_1, \ldots, \mathbf{b}_n$ are the columns $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of the $n \times n$ identity matrix.

23. Let S be a finite set in a vector space V with the property that every $\mathbf{x}$ in V has a unique representation as a linear combination of elements of S. Show that S is a basis of V.

24. Suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_4\}$ is a linearly dependent spanning set for a vector space V. Show that each $\mathbf{w}$ in V can be expressed in more than one way as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_4$. [*Hint:* Let $\mathbf{w} = k_1\mathbf{v}_1 + \cdots + k_4\mathbf{v}_4$ be an arbitrary vector in V. Use the linear dependence of $\{\mathbf{v}_1, \ldots, \mathbf{v}_4\}$ to produce another representation of $\mathbf{w}$ as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_4$.]

25. Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \begin{bmatrix} -3 \\ 7 \end{bmatrix} \right\}$. Since the coordinate mapping determined by $\mathcal{B}$ is a linear transformation from $\mathbb{R}^2$ into $\mathbb{R}^2$, this mapping must be implemented by some 2×2 matrix A. Find it. [*Hint:* Multiplication by A should transform a vector $\mathbf{x}$ into its coordinate vector $[\,\mathbf{x}\,]_{\mathcal{B}}$.]

26. Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis for $\mathbb{R}^n$. Produce a description of an $n \times n$ matrix A that implements the coordinate mapping $\mathbf{x} \mapsto [\,\mathbf{x}\,]_{\mathcal{B}}$. (See Exercise 25.)

Exercises 27–30 concern a vector space V, a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$, and the coordinate mapping $\mathbf{x} \mapsto [\,\mathbf{x}\,]_{\mathcal{B}}$.

27. Show that the coordinate mapping is one-to-one. [*Hint:* Suppose $[\,\mathbf{u}\,]_{\mathcal{B}} = [\,\mathbf{w}\,]_{\mathcal{B}}$ for some $\mathbf{u}$ and $\mathbf{w}$ in V, and show that $\mathbf{u} = \mathbf{w}$.]

28. Show that the coordinate mapping is *onto* $\mathbb{R}^n$. That is, given any $\mathbf{y}$ in $\mathbb{R}^n$, with entries $y_1, \ldots, y_n$, produce $\mathbf{u}$ in V such that $[\,\mathbf{u}\,]_{\mathcal{B}} = \mathbf{y}$.

29. Show that a subset $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ in V is linearly independent if and only if the set of coordinate vectors $\{[\,\mathbf{u}_1\,]_{\mathcal{B}}, \ldots, [\,\mathbf{u}_p\,]_{\mathcal{B}}\}$ is linearly independent in $\mathbb{R}^n$. [*Hint:* Since the coordinate mapping is one-to-one, the following equations have the same solutions, $c_1, \ldots, c_p$.]

$$c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p = \mathbf{0} \qquad \text{The zero vector in } V$$
$$[\,c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p\,]_{\mathcal{B}} = [\,\mathbf{0}\,]_{\mathcal{B}} \qquad \text{The zero vector in } \mathbb{R}^n$$

30. Given vectors $\mathbf{u}_1, \ldots, \mathbf{u}_p$, and $\mathbf{w}$ in V, show that $\mathbf{w}$ is a linear combination of $\mathbf{u}_1, \ldots, \mathbf{u}_p$ if and only if $[\,\mathbf{w}\,]_{\mathcal{B}}$ is a linear combination of the coordinate vectors $[\,\mathbf{u}_1\,]_{\mathcal{B}}, \ldots, [\,\mathbf{u}_p\,]_{\mathcal{B}}$.

In Exercises 31–34, use coordinate vectors to test the linear independence of the sets of polynomials. Explain your work.

31. $\{1 + 2t^3,\ 2 + t - 3t^2,\ -t + 2t^2 - t^3\}$

32. $\{1 - 2t^2 - t^3,\ t + 2t^3,\ 1 + t - 2t^2\}$

33. $\{(1 - t)^2,\ t - 2t^2 + t^3,\ (1 - t)^3\}$

34. $\{(2 - t)^3,\ (3 - t)^2,\ 1 + 6t - 5t^2 + t^3\}$

35. Use coordinate vectors to test whether the following sets of polynomials span $\mathbb{P}_2$. Justify your conclusions.

a. $\{1 - 3t + 5t^2, -3 + 5t - 7t^2, -4 + 5t - 6t^2, 1 - t^2\}$

b. $\{5t + t^2, 1 - 8t - 2t^2, -3 + 4t + 2t^2, 2 - 3t\}$

36. Let $\mathbf{p}_1(t) = 1 + t^2$, $\mathbf{p}_2(t) = t - 3t^2$, $\mathbf{p}_3(t) = 1 + t - 3t^2$.

a. Use coordinate vectors to show that these polynomials form a basis for $\mathbb{P}_2$.

b. Consider the basis $\mathcal{B} = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ for $\mathbb{P}_2$. Find $\mathbf{q}$ in $\mathbb{P}_2$, given that $[\mathbf{q}]_{\mathcal{B}} = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$.

In Exercises 37 and 38, determine whether the sets of polynomials form a basis for $\mathbb{P}_3$. Justify your conclusions.

T 37. $3 + 7t, 5 + t - 2t^3, t - 2t^2, 1 + 16t - 6t^2 + 2t^3$

T 38. $5 - 3t + 4t^2 + 2t^3, 9 + t + 8t^2 - 6t^3, 6 - 2t + 5t^2, t^3$

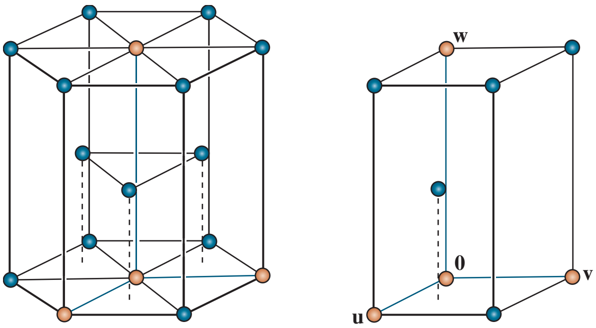
T 39. Let $H = \text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2\}$ and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$. Show that $\mathbf{x}$ is in H and find the $\mathcal{B}$-coordinate vector of $\mathbf{x}$, for

$$\mathbf{v}_1 = \begin{bmatrix} 11 \\ -5 \\ 10 \\ 7 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 14 \\ -8 \\ 13 \\ 10 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 19 \\ -13 \\ 18 \\ 15 \end{bmatrix}$$

T 40. Let $H = \text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. Show that $\mathcal{B}$ is a basis for H and $\mathbf{x}$ is in H, and find the $\mathcal{B}$-coordinate vector of $\mathbf{x}$, for

$$\mathbf{v}_1 = \begin{bmatrix} -6 \\ 4 \\ -9 \\ 4 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 8 \\ -3 \\ 7 \\ -3 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} -9 \\ 5 \\ -8 \\ 3 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 4 \\ 7 \\ -8 \\ 3 \end{bmatrix}$$

Exercises 41 and 42 concern the crystal lattice for titanium, which has the hexagonal structure shown on the left in the accompanying figure. The vectors $\begin{bmatrix} 2.6 \\ -1.5 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 4.8 \end{bmatrix}$ in $\mathbb{R}^3$ form a basis for the unit cell shown on the right. The numbers here are Ångstrom units (1 Å $= 10^{-8}$ cm). In alloys of titanium, some additional atoms may be in the unit cell at the *octahedral* and *tetrahedral* sites (so named because of the geometric objects formed by atoms at these locations).



The hexagonal close-packed lattice and its unit cell.

41. One of the octahedral sites is $\begin{bmatrix} 1/2 \\ 1/4 \\ 1/6 \end{bmatrix}$, relative to the lattice basis. Determine the coordinates of this site relative to the standard basis of $\mathbb{R}^3$.

42. One of the tetrahedral sites is $\begin{bmatrix} 1/2 \\ 1/2 \\ 1/3 \end{bmatrix}$. Determine the coordinates of this site relative to the standard basis of $\mathbb{R}^3$.

Solutions to Practice Problems

1. a. It is evident that the matrix $P_{\mathcal{B}} = [\,\mathbf{b}_1 \quad \mathbf{b}_2 \quad \mathbf{b}_3\,]$ is row-equivalent to the identity matrix. By the Invertible Matrix Theorem, $P_{\mathcal{B}}$ is invertible and its columns form a basis for $\mathbb{R}^3$.

b. From part (a), the change-of-coordinates matrix is $P_{\mathcal{B}} = \begin{bmatrix} 1 & -3 & 3 \\ 0 & 4 & -6 \\ 0 & 0 & 3 \end{bmatrix}$.

c. $\mathbf{x} = P_{\mathcal{B}}[\,\mathbf{x}\,]_{\mathcal{B}}$

d. To solve the equation in (c), it is probably easier to row reduce an augmented matrix than to compute $P_{\mathcal{B}}^{-1}$:

$$\underbrace{\begin{bmatrix} 1 & -3 & 3 & -8 \\ 0 & 4 & -6 & 2 \\ 0 & 0 & 3 & 3 \end{bmatrix}}_{P_{\mathcal{B}} \qquad\quad \mathbf{x}} \sim \underbrace{\begin{bmatrix} 1 & 0 & 0 & -5 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix}}_{I \qquad\quad [\,\mathbf{x}\,]_{\mathcal{B}}}$$

Hence

$$[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} -5 \\ 2 \\ 1 \end{bmatrix}$$

2. The coordinates of $\mathbf{p}(t) = 6 + 3t - t^2$ with respect to $\mathcal{B}$ satisfy

$$c_1(1 + t) + c_2(1 + t^2) + c_3(t + t^2) = 6 + 3t - t^2$$

Equating coefficients of like powers of t, we have

$$\begin{aligned} c_1 + c_2 \qquad\quad &= 6 \\ c_1 \qquad\quad + c_3 &= 3 \\ c_2 + c_3 &= -1 \end{aligned}$$

Solving, we find that $c_1 = 5$, $c_2 = 1$, $c_3 = -2$, and $[\,\mathbf{p}\,]_{\mathcal{B}} = \begin{bmatrix} 5 \\ 1 \\ -2 \end{bmatrix}$.

4.5 The Dimension of a Vector Space

Theorem 9 in Section 4.4 implies that a vector space V with a basis $\mathcal{B}$ containing n vectors is isomorphic to $\mathbb{R}^n$. This section shows that this number n is an intrinsic property (called the dimension) of the space V that does not depend on the particular choice of basis. The discussion of dimension will give additional insight into properties of bases.

The first theorem generalizes a well-known result about the vector space $\mathbb{R}^n$.

THEOREM 10

If a vector space V has a basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$, then any set in V containing more than n vectors must be linearly dependent.

PROOF Let $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ be a set in V with more than n vectors. The coordinate vectors $[\,\mathbf{u}_1\,]_\mathcal{B}, \dots, [\,\mathbf{u}_p\,]_\mathcal{B}$ form a linearly dependent set in $\mathbb{R}^n$, because there are more vectors (p) than entries (n) in each vector. So there exist scalars $c_1, \dots, c_p$, not all zero, such that

$$c_1[\,\mathbf{u}_1\,]_\mathcal{B} + \cdots + c_p[\,\mathbf{u}_p\,]_\mathcal{B} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

The zero vector in $\mathbb{R}^n$

Since the coordinate mapping is a linear transformation,

$$\left[\,c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p\,\right]_\mathcal{B} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

The zero vector on the right displays the n weights needed to build the vector $c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ from the basis vectors in $\mathcal{B}$. That is, $c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p = 0\mathbf{b}_1 + \cdots + 0\mathbf{b}_n = \mathbf{0}$. Since the c_i are not all zero, $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is linearly dependent.[1] ∎

Theorem 10 implies that if a vector space V has a basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$, then each linearly independent set in V has no more than n vectors.

THEOREM 11

If a vector space V has a basis of n vectors, then every basis of V must consist of exactly n vectors.

PROOF Let $\mathcal{B}_1$ be a basis of n vectors and $\mathcal{B}_2$ be any other basis (of V). Since $\mathcal{B}_1$ is a basis and $\mathcal{B}_2$ is linearly independent, $\mathcal{B}_2$ has no more than n vectors, by Theorem 10. Also, since $\mathcal{B}_2$ is a basis and $\mathcal{B}_1$ is linearly independent, $\mathcal{B}_2$ has at least n vectors. Thus $\mathcal{B}_2$ consists of exactly n vectors. ∎

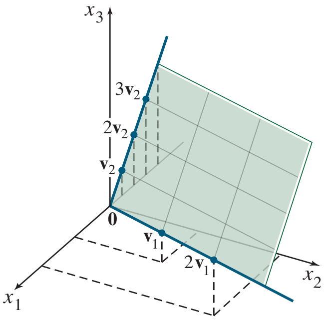
[1] Theorem 10 also applies to infinite sets in V. An infinite set is said to be linearly dependent if some finite subset is linearly dependent; otherwise, the set is linearly independent. If S is an infinite set in V, take any subset $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ of S, with $p > n$. The proof above shows that this subset is linearly dependent and hence so is S.

If a nonzero vector space V is spanned by a finite set S, then a subset of S is a basis for V, by the Spanning Set Theorem. In this case, Theorem 11 ensures that the following definition makes sense.

DEFINITION

If a vector space V is spanned by a finite set, then V is said to be **finite-dimensional**, and the **dimension** of V, written as $\dim V$, is the number of vectors in a basis for V. The dimension of the zero vector space $\{\mathbf{0}\}$ is defined to be zero. If V is not spanned by a finite set, then V is said to be **infinite-dimensional**.

EXAMPLE 1 The standard basis for $\mathbb{R}^n$ contains n vectors, so $\dim \mathbb{R}^n = n$. The standard polynomial basis $\{1, t, t^2\}$ shows that $\dim \mathbb{P}_2 = 3$. In general, $\dim \mathbb{P}_n = n + 1$. The space $\mathbb{P}$ of all polynomials is infinite-dimensional. ■



EXAMPLE 2 Let $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$, where $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$. Then H is the plane studied in Example 7 in Section 4.4. A basis for H is $\{\mathbf{v}_1, \mathbf{v}_2\}$, since $\mathbf{v}_1$ and $\mathbf{v}_2$ are not multiples and hence are linearly independent. Thus $\dim H = 2$. ■

EXAMPLE 3 Find the dimension of the subspace

$$H = \left\{ \begin{bmatrix} a - 3b + 6c \\ 5a + 4d \\ b - 2c - d \\ 5d \end{bmatrix} : a, b, c, d \text{ in } \mathbb{R} \right\}$$

SOLUTION It is easy to see that H is the set of all linear combinations of the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 5 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -3 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 6 \\ 0 \\ -2 \\ 0 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} 0 \\ 4 \\ -1 \\ 5 \end{bmatrix}$$

Clearly, $\mathbf{v}_1 \neq \mathbf{0}$, $\mathbf{v}_2$ is not a multiple of $\mathbf{v}_1$, but $\mathbf{v}_3$ is a multiple of $\mathbf{v}_2$. By the Spanning Set Theorem, we may discard $\mathbf{v}_3$ and still have a set that spans H. Finally, $\mathbf{v}_4$ is not a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. So $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\}$ is linearly independent (by Theorem 4 in Section 4.3) and hence is a basis for H. Thus $\dim H = 3$. ■

EXAMPLE 4 The subspaces of $\mathbb{R}^3$ can be classified by dimension. See Figure 1.

0-dimensional subspaces. Only the zero subspace.

1-dimensional subspaces. Any subspace spanned by a single nonzero vector. Such subspaces are lines through the origin.

2-dimensional subspaces. Any subspace spanned by two linearly independent vectors. Such subspaces are planes through the origin.

3-dimensional subspaces. Only $\mathbb{R}^3$ itself. Any three linearly independent vectors in $\mathbb{R}^3$ span all of $\mathbb{R}^3$, by the Invertible Matrix Theorem. ■

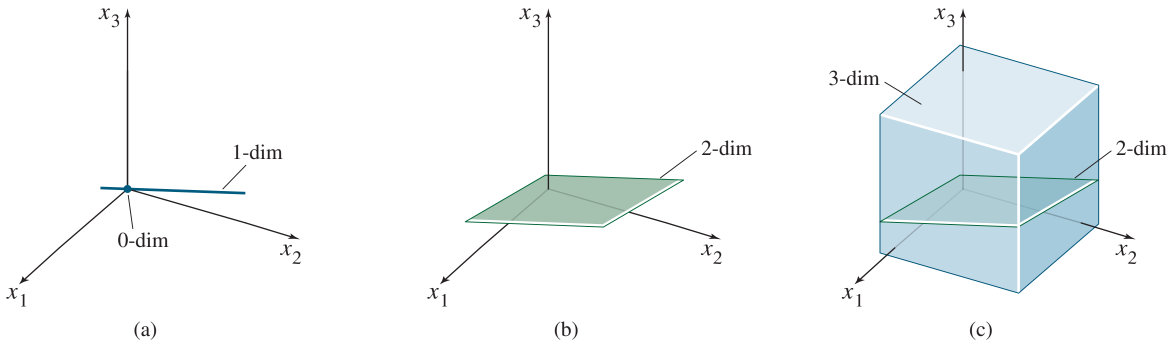


FIGURE 1 Sample subspaces of $\mathbb{R}^3$.

Subspaces of a Finite-Dimensional Space

The next theorem is a natural counterpart to the Spanning Set Theorem.

THEOREM 12

Let H be a subspace of a finite-dimensional vector space V. Any linearly independent set in H can be expanded, if necessary, to a basis for H. Also, H is finite-dimensional and

$$\dim H \leq \dim V$$

PROOF If $H = \{\mathbf{0}\}$, then certainly $\dim H = 0 \leq \dim V$. Otherwise, let $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be any linearly independent set in H. If S spans H, then S is a basis for H. Otherwise, there is some $\mathbf{u}_{k+1}$ in H that is not in Span S. But then $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}\}$ will be linearly independent, because no vector in the set can be a linear combination of vectors that precede it (by Theorem 4).

So long as the new set does not span H, we can continue this process of expanding S to a larger linearly independent set in H. But the number of vectors in a linearly independent expansion of S can never exceed the dimension of V, by Theorem 10. So eventually the expansion of S will span H and hence will be a basis for H, and $\dim H \leq \dim V$. ■

When the dimension of a vector space or subspace is known, the search for a basis is simplified by the next theorem. It says that if a set has the right number of elements, then one has only to show either that the set is linearly independent or that it spans the space. The theorem is of critical importance in numerous applied problems (involving differential equations or difference equations, for example) where linear independence is much easier to verify than spanning.

THEOREM 13

The Basis Theorem

Let V be a p-dimensional vector space, $p \geq 1$. Any linearly independent set of exactly p elements in V is automatically a basis for V. Any set of exactly p elements that spans V is automatically a basis for V.

PROOF By Theorem 12, a linearly independent set S of p elements can be extended to a basis for V. But that basis must contain exactly p elements, since $\dim V = p$. So S must already be a basis for V. Now suppose that S has p elements and spans V. Since V is nonzero, the Spanning Set Theorem implies that a subset S' of S is a basis of V. Since $\dim V = p$, S' must contain p vectors. Hence $S = S'$. ■

The Dimensions of Nul A, Col A, and Row A

Since the dimensions of the null space and column space of an $m \times n$ matrix are referred to frequently, they have specific names:

DEFINITION

The **rank** of an $m \times n$ matrix A is the dimension of the column space and the **nullity** of A is the dimension of the null space.

The pivot columns of a matrix A form a basis for Col A, so the rank of A is just the number of pivot columns. Since a basis for Row A can be found by taking the pivot rows from the row reduced echelon form of A, the dimension of Row A is also equal to the rank of A.

The nullity of A might seem to require more work, since finding a basis for Nul A usually takes more time than finding a basis for Col A. There is a shortcut: Let A be an $m \times n$ matrix, and suppose the equation $A\mathbf{x} = \mathbf{0}$ has k free variables. From Section 4.2, we know that the standard method of finding a spanning set for Nul A will produce exactly k linearly independent vectors—say, $\mathbf{u}_1, \ldots, \mathbf{u}_k$ – one for each free variable. So $\mathbf{u}_1, \ldots, \mathbf{u}_k$ is a basis for Nul A, and the number of free variables determines the size of the basis.

To summarize these facts for future reference:

The rank of an $m \times n$ matrix A is the number of pivot columns and the nullity of A is the number of free variables. Since the dimension of the row space is the number of pivot rows, it is also equal to the rank of A.

Putting these observations together results in the rank theorem.

THEOREM 14

The Rank Theorem

The dimensions of the column space and the null space of an $m \times n$ matrix A satisfy the equation

$$\text{rank } A + \text{nullity } A = \text{number of columns in } A$$

PROOF By Theorem 6 in Section 4.3, rank A is the number of pivot columns in A. The nullity of A equals the number of free variables in the equation $A\mathbf{x} = \mathbf{0}$. Expressed another way, the nullity of A is the number of columns of A that are *not* pivot columns. (It is the number of these columns, not the columns themselves, that is related to Nul A.) Obviously,

$$\left\{\begin{array}{c}\text{number of}\\ \text{pivot columns}\end{array}\right\} + \left\{\begin{array}{c}\text{number of}\\ \text{nonpivot columns}\end{array}\right\} = \left\{\begin{array}{c}\text{number of}\\ \text{columns}\end{array}\right\}$$

This proves the theorem. ■

EXAMPLE 5 Find the nullity and rank of

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$$

SOLUTION Row reduce the augmented matrix $[A \ \ \mathbf{0}]$ to echelon form:

$$B = \begin{bmatrix} 1 & -2 & 2 & 3 & -1 & 0 \\ 0 & 0 & 1 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

There are three free variables: x_2, x_4, and x_5. Hence the nullity of A is 3. Also, the rank of A is 2 because A has two pivot columns. ■

The ideas behind Theorem 14 are visible in the calculations in Example 5. The two pivot positions in B, an echelon form of A, determine the basic variables and identify the basis vectors for Col A and those for Row A.

EXAMPLE 6

a. If A is a 7×9 matrix with nullity 2, what is the rank of A?

b. Could a 6×9 matrix have nullity 2?

SOLUTION

a. Since A has 9 columns, $(\text{rank } A) + 2 = 9$, and hence rank $A = 7$.

b. No. If a 6×9 matrix, call it B, had a two-dimensional null space, it would have to have rank 7, by the Rank Theorem. But the columns of B are vectors in $\mathbb{R}^6$, and so the dimension of Col B cannot exceed 6; that is, rank B cannot exceed 6. ■

The next example provides a nice way to visualize the subspaces we have been studying. In Chapter 6, we will learn that Row A and Nul A have only the zero vector in common and are actually perpendicular to each other. The same fact applies to Row A^T $(= \text{Col} A)$ and Nul A^T. So Figure 2, which accompanies Example 7, creates a good mental image for the general case.

EXAMPLE 7 Let $A = \begin{bmatrix} 3 & 0 & -1 \\ 3 & 0 & -1 \\ 4 & 0 & 5 \end{bmatrix}$. It is readily checked that Nul A is the x_2-axis, Row A is the x_1x_3-plane, Col A is the plane whose equation is $x_1 - x_2 = 0$, and Nul A^T is the set of all multiples of $(1, -1, 0)$. Figure 2 shows Nul A and Row A in the domain of the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$; the range of this mapping, Col A, is shown in a separate copy of $\mathbb{R}^3$, along with Nul A^T. ■

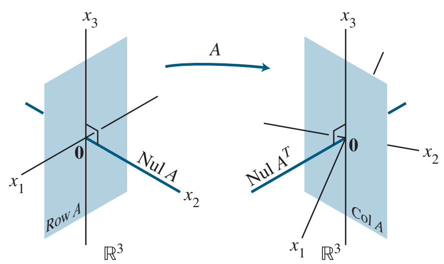


FIGURE 2 Subspaces determined by a matrix A.

Applications to Systems of Equations

The Rank Theorem is a powerful tool for processing information about systems of linear equations. The next example simulates the way a real-life problem using linear equations might be stated, without explicit mention of linear algebra terms such as matrix, subspace, and dimension.

EXAMPLE 8 A scientist has found two solutions to a homogeneous system of 40 equations in 42 variables. The two solutions are not multiples, and all other solutions can be constructed by adding together appropriate multiples of these two solutions. Can the scientist be *certain* that an associated nonhomogeneous system (with the same coefficients) has a solution?

SOLUTION Yes. Let A be the 40×42 coefficient matrix of the system. The given information implies that the two solutions are linearly independent and span Nul A. So nullity $A = 2$. By the Rank Theorem, rank $A = 42 - 2 = 40$. Since $\mathbb{R}^{40}$ is the only subspace of $\mathbb{R}^{40}$ whose dimension is 40, Col A must be all of $\mathbb{R}^{40}$. This means that every nonhomogeneous equation $A\mathbf{x} = \mathbf{b}$ has a solution. ■

Rank and the Invertible Matrix Theorem

The various vector space concepts associated with a matrix provide several more statements for the Invertible Matrix Theorem. The new statements listed here follow those in the original Invertible Matrix Theorem in Section 2.3 and other theorems in the text where statements have been added to it.

THEOREM

The Invertible Matrix Theorem (continued)

Let A be an $n \times n$ matrix. Then the following statements are each equivalent to the statement that A is an invertible matrix.

m. The columns of A form a basis of $\mathbb{R}^n$.
n. Col $A = \mathbb{R}^n$
o. rank $A = n$
p. nullity $A = 0$
q. Nul $A = \{\mathbf{0}\}$

PROOF Statement (m) is logically equivalent to statements (e) and (h) regarding linear independence and spanning. The other five statements are linked to the earlier ones of the theorem by the following chain of almost trivial implications:

$$(g) \Rightarrow (n) \Rightarrow (o) \Rightarrow (p) \Rightarrow (q) \Rightarrow (d)$$

Statement (g), which says that the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$, implies (n), because Col A is precisely the set of all $\mathbf{b}$ such that the equation $A\mathbf{x} = \mathbf{b}$ is consistent. The implication (n) $\Rightarrow$ (o) follows from the definitions of dimension and rank. If the rank of A is n, the number of columns of A, then nullity $A = 0$, by the Rank Theorem, and so Nul $A = \{\mathbf{0}\}$. Thus (o) $\Rightarrow$ (p) $\Rightarrow$ (q). Also, (q) implies that the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, which is statement (d). Since statements (d) and (g) are already known to be equivalent to the statement that A is invertible, the proof is complete. ■

We have refrained from adding to the Invertible Matrix Theorem obvious statements about the row space of A, because the row space is the column space of A^T. Recall from statement (l) of the Invertible Matrix Theorem that A is invertible if and only if A^T is invertible. Hence every statement in the Invertible Matrix Theorem can also be stated for A^T. To do so would double the length of the theorem and produce a list of more than 30 statements!

Numerical Notes

Many algorithms discussed in this text are useful for understanding concepts and making simple computations by hand. However, the algorithms are often unsuitable for large-scale problems in real life.

Rank determination is a good example. It would seem easy to reduce a matrix to echelon form and count the pivots. But unless exact arithmetic is performed on a matrix whose entries are specified exactly, row operations can change the apparent rank of a matrix. For instance, if the value of x in the matrix $\begin{bmatrix} 5 & 7 \\ 5 & x \end{bmatrix}$ is not stored exactly as 7 in a computer, then the rank may be 1 or 2, depending on whether the computer treats $x - 7$ as zero.

In practical applications, the effective rank of a matrix A is often determined from the singular value decomposition of A, to be discussed in Section 7.4. This decomposition is also a reliable source of bases for Col A, Row A, Nul A, and Nul A^T.

Practice Problems

1. Decide whether each statement is True or False, and give a reason for each answer. Here V is a nonzero finite-dimensional vector space.
   a. If dim $V = p$ and if S is a linearly dependent subset of V, then S contains more than p vectors.
   b. If S spans V and if T is a subset of V that contains more vectors than S, then T is linearly dependent.
2. Let H and K be subspaces of a vector space V. In Section 4.1, Exercise 40, it is established that $H \cap K$ is also a subspace of V. Prove dim $(H \cap K) \leq$ dim H.

4.5 Exercises

For each subspace in Exercises 1–8, (a) find a basis, and (b) state the dimension.

1. $\left\{ \begin{bmatrix} s-2t \\ s+t \\ 3t \end{bmatrix} : s, t \text{ in } \mathbb{R} \right\}$

2. $\left\{ \begin{bmatrix} 5s \\ -t \\ -7s \end{bmatrix} : s, t \in \mathbb{R} \right\}$

3. $\left\{ \begin{bmatrix} 2c \\ a-b \\ b-3c \\ a+2b \end{bmatrix} : a, b, c \text{ in } \mathbb{R} \right\}$

4. $\left\{ \begin{bmatrix} a+b \\ 2a \\ 3a-b \\ -b \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\}$

5. $\left\{ \begin{bmatrix} a-4b-2c \\ 2a+5b-4c \\ -a+2c \\ -3a+7b+6c \end{bmatrix} : a, b, c \text{ in } \mathbb{R} \right\}$

6. $\left\{ \begin{bmatrix} 3a+6b-c \\ 6a-2b-2c \\ -9a+5b+3c \\ -3a+b+c \end{bmatrix} : a, b, c \text{ in } \mathbb{R} \right\}$

7. $\{(a, b, c) : a - 3b + c = 0, b - 2c = 0, 2b - c = 0\}$

8. $\{(a, b, c, d) : a - 3b + c = 0\}$

In Exercises 9 and 10, find the dimension of the subspace spanned by the given vectors.

9. $\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 9 \\ 4 \\ -2 \end{bmatrix}, \begin{bmatrix} -7 \\ -3 \\ 1 \end{bmatrix}$

10. $\begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 4 \\ 1 \end{bmatrix}, \begin{bmatrix} -8 \\ 6 \\ 5 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 7 \end{bmatrix}$

Determine the dimensions of Nul A, Col A, and Row A for the matrices shown in Exercises 11–16.

11. $A = \begin{bmatrix} 1 & -6 & 9 & 0 & -2 \\ 0 & 1 & 2 & -4 & 5 \\ 0 & 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

12. $A = \begin{bmatrix} 1 & 3 & -4 & 2 & -1 & 6 \\ 0 & 0 & 1 & -3 & 7 & 0 \\ 0 & 0 & 0 & 1 & 4 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

13. $A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 0 & 0 & 1 & -6 \end{bmatrix}$

14. $A = \begin{bmatrix} 3 & 4 \\ -6 & 10 \end{bmatrix}$

15. $A = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 4 & 7 \\ 0 & 0 & 5 \end{bmatrix}$

16. $A = \begin{bmatrix} 1 & 4 & -1 \\ 0 & 7 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

In Exercises 17–26, V is a vector space and A is an $m \times n$ matrix. Mark each statement True or False **(T/F)**. Justify each answer.

17. **(T/F)** The number of pivot columns of a matrix equals the dimension of its column space.

18. **(T/F)** The number of variables in the equation $A\mathbf{x} = \mathbf{0}$ equals the nullity A.

19. **(T/F)** A plane in $\mathbb{R}^3$ is a two-dimensional subspace of $\mathbb{R}^3$.

20. **(T/F)** The dimension of the vector space $\mathbb{P}_4$ is 4.

21. **(T/F)** The dimension of the vector space of signals, $\mathbb{S}$, is 10.

22. **(T/F)** The dimensions of the row space and the column space of A are the same, even if A is not square.

23. **(T/F)** If B is any echelon form of A, then the pivot columns of B form a basis for the column space of A.

24. **(T/F)** The nullity of A is the number of columns of A that are not pivot columns.

25. **(T/F)** If a set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ spans a finite-dimensional vector space V and if T is a set of more than p vectors in V, then T is linearly dependent.

26. **(T/F)** A vector space is infinite-dimensional if it is spanned by an infinite set.

27. The first four Hermite polynomials are 1, $2t$, $-2 + 4t^2$, and $-12t + 8t^3$. These polynomials arise naturally in the study of certain important differential equations in mathematical physics.[2] Show that the first four Hermite polynomials form a basis of $\mathbb{P}_3$.

28. The first four Laguerre polynomials are 1, $1 - t$, $2 - 4t + t^2$, and $6 - 18t + 9t^2 - t^3$. Show that these polynomials form a basis of $\mathbb{P}_3$.

29. Let $\mathcal{B}$ be the basis of $\mathbb{P}_3$ consisting of the Hermite polynomials in Exercise 27, and let $\mathbf{p}(t) = 7 - 12t - 8t^2 + 12t^3$. Find the coordinate vector of $\mathbf{p}$ relative to $\mathcal{B}$.

30. Let $\mathcal{B}$ be the basis of $\mathbb{P}_2$ consisting of the first three Laguerre polynomials listed in Exercise 28, and let $\mathbf{p}(t) = 7 - 8t + 3t^2$. Find the coordinate vector of $\mathbf{p}$ relative to $\mathcal{B}$.

31. Let S be a subset of an n-dimensional vector space V, and suppose S contains fewer than n vectors. Explain why S cannot span V.

32. Let H be an n-dimensional subspace of an n-dimensional vector space V. Show that $H = V$.

33. If a 4×7 matrix A has rank 4, find nullity A, rank A, and rank A^T.

34. If a 6×3 matrix A has rank 3, find nullity A, rank A, and rank A^T.

35. Suppose a 5×9 matrix A has four pivot columns. Is Col $A = \mathbb{R}^5$? Is Nul $A = \mathbb{R}^4$? Explain your answers.

36. Suppose a 5×6 matrix A has four pivot columns. What is nullity A? Is Col $A = \mathbb{R}^4$? Why or why not?

37. If the nullity of a 5×6 matrix A is 4, what are the dimensions of the column and row spaces of A?

38. If the nullity of a 7×6 matrix A is 5, what are the dimensions of the column and row spaces of A?

39. If A is a 7×5 matrix, what is the largest possible rank of A? If A is a 5×7 matrix, what is the largest possible rank of A? Explain your answers.

40. If A is a 4×3 matrix, what is the largest possible dimension of the row space of A? If A is a 3×4 matrix, what is the largest possible dimension of the row space of A? Explain.

41. Explain why the space $\mathbb{P}$ of all polynomials is an infinite-dimensional space.

42. Show that the space $C(\mathbb{R})$ of all continuous functions defined on the real line is an infinite-dimensional space.

In Exercises 43–48, V is a nonzero finite-dimensional vector space, and the vectors listed belong to V. Mark each statement True or False **(T/F)**. Justify each answer. (These questions are more difficult than those in Exercises 17–26.)

[2] See *Introduction to Functional Analysis*, 2nd ed., by A. E. Taylor and David C. Lay (New York: John Wiley & Sons, 1980), pp. 92–93. Other sets of polynomials are discussed there, too.

43. **(T/F)** If there exists a set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ that spans V, then $\dim V \le p$.

44. **(T/F)** If there exists a linearly dependent set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ in V, then $\dim V \le p$.

45. **(T/F)** If there exists a linearly independent set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ in V, then $\dim V \ge p$.

46. **(T/F)** If $\dim V = p$, then there exists a spanning set of $p+1$ vectors in V.

47. **(T/F)** If every set of p elements in V fails to span V, then $\dim V > p$.

48. **(T/F)** If $p \ge 2$ and $\dim V = p$, then every set of $p-1$ nonzero vectors is linearly independent.

49. Justify the following equality: $\dim\ \text{Row}\ A + \text{nullity}\ A = n$, the number of columns of A

50. Justify the following equality: $\dim\ \text{Row}\ A + \text{nullity}\ A^T = m$, the number of rows of A

Exercises 51 and 52 concern finite-dimensional vector spaces V and W and a linear transformation $T : V \to W$.

51. Let H be a nonzero subspace of V, and let $T(H)$ be the set of images of vectors in H. Then $T(H)$ is a subspace of W, by Exercise 47 in Section 4.2. Prove that $\dim T(H) \le \dim H$.

52. Let H be a nonzero subspace of V, and suppose T is a one-to-one (linear) mapping of V into W. Prove that $\dim T(H) = \dim H$. If T happens to be a one-to-one mapping of V *onto* W, then $\dim V = \dim W$. Isomorphic finite-dimensional vector spaces have the same dimension.

T **53.** According to Theorem 12, a linearly independent set $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ can be expanded to a basis for $\mathbb{R}^n$. One way to do this is to create $A = [\,\mathbf{v}_1 \ \cdots \ \mathbf{v}_k \ \ \mathbf{e}_1 \ \cdots \ \mathbf{e}_n\,]$, with $\mathbf{e}_1, \dots, \mathbf{e}_n$ the columns of the identity matrix; the pivot columns of A form a basis for $\mathbb{R}^n$.

a. Use the method described to extend the following vectors to a basis for $\mathbb{R}^5$:

$$\mathbf{v}_1 = \begin{bmatrix} -9 \\ -7 \\ 8 \\ -5 \\ 7 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 9 \\ 4 \\ 1 \\ 6 \\ -7 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 6 \\ 7 \\ -8 \\ 5 \\ -7 \end{bmatrix}$$

b. Explain why the method works in general: Why are the original vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ included in the basis found for Col A? Why is Col $A = \mathbb{R}^n$?

T **54.** Let $\mathcal{B} = \{1, \cos t, \cos^2 t, \dots, \cos^6 t\}$ and $\mathcal{C} = \{1, \cos t, \cos 2t, \dots, \cos 6t\}$. Assume the following trigonometric identities (see Exercise 45 in Section 4.1).

$$\begin{aligned}
\cos 2t &= -1 + 2\cos^2 t \\
\cos 3t &= -3\cos t + 4\cos^3 t \\
\cos 4t &= 1 - 8\cos^2 t + 8\cos^4 t \\
\cos 5t &= 5\cos t - 20\cos^3 t + 16\cos^5 t \\
\cos 6t &= -1 + 18\cos^2 t - 48\cos^4 t + 32\cos^6 t
\end{aligned}$$

Let H be the subspace of functions spanned by the functions in $\mathcal{B}$. Then $\mathcal{B}$ is a basis for H, by Exercise 48 in Section 4.3.

a. Write the $\mathcal{B}$-coordinate vectors of the vectors in $\mathcal{C}$, and use them to show that $\mathcal{C}$ is a linearly independent set in H.

b. Explain why $\mathcal{C}$ is a basis for H.

Solutions to Practice Problems

1. a. False. Consider the set $\{\mathbf{0}\}$.

b. True. By the Spanning Set Theorem, S contains a basis for V; call that basis S'. Then T will contain more vectors than S'. By Theorem 10, T is linearly dependent.

2. Let $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ be a basis for $H \cap K$. Notice $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is a linearly independent subset of H, hence by Theorem 12, $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ can be expanded, if necessary, to a basis for H. Since the dimension of a subspace is just the number of vectors in a basis, it follows that $\dim (H \cap K) = p \le \dim H$.

4.6 Change of Basis

When a basis $\mathcal{B}$ is chosen for an n-dimensional vector space V, the associated coordinate mapping onto $\mathbb{R}^n$ provides a coordinate system for V. Each $\mathbf{x}$ in V is identified uniquely by its $\mathcal{B}$-coordinate vector $[\,\mathbf{x}\,]_{\mathcal{B}}$.[1]

[1] Think of $[\,\mathbf{x}\,]_{\mathcal{B}}$ as a name for $\mathbf{x}$ that lists the weights used to build $\mathbf{x}$ as a linear combination of the basis vectors in $\mathcal{B}$.

In some applications, a problem is described initially using a basis $\mathcal{B}$, but the problem's solution is aided by changing $\mathcal{B}$ to a new basis $\mathcal{C}$. (Examples will be given in Chapters 5 and 7.) Each vector is assigned a new $\mathcal{C}$-coordinate vector. In this section, we study how $[\,\mathbf{x}\,]_{\mathcal{C}}$ and $[\,\mathbf{x}\,]_{\mathcal{B}}$ are related for each $\mathbf{x}$ in V.

To visualize the problem, consider the two coordinate systems in Figure 1. In Figure 1(a), $\mathbf{x} = 3\mathbf{b}_1 + \mathbf{b}_2$, while in Figure 1(b), the same $\mathbf{x}$ is shown as $\mathbf{x} = 6\mathbf{c}_1 + 4\mathbf{c}_2$. That is,

$$[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \quad \text{and} \quad [\,\mathbf{x}\,]_{\mathcal{C}} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

Our problem is to find the connection between the two coordinate vectors. Example 1 shows how to do this, provided we know how $\mathbf{b}_1$ and $\mathbf{b}_2$ are formed from $\mathbf{c}_1$ and $\mathbf{c}_2$.

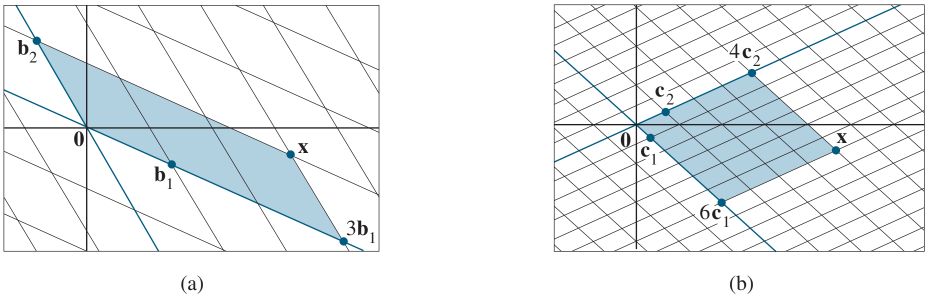


FIGURE 1 Two coordinate systems for the same vector space.

EXAMPLE 1 Consider two bases $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$ for a vector space V, such that

$$\mathbf{b}_1 = 4\mathbf{c}_1 + \mathbf{c}_2 \quad \text{and} \quad \mathbf{b}_2 = -6\mathbf{c}_1 + \mathbf{c}_2 \tag{1}$$

Suppose

$$\mathbf{x} = 3\mathbf{b}_1 + \mathbf{b}_2 \tag{2}$$

That is, suppose $[\,\mathbf{x}\,]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$. Find $[\,\mathbf{x}\,]_{\mathcal{C}}$.

SOLUTION Apply the coordinate mapping determined by $\mathcal{C}$ to $\mathbf{x}$ in (2). Since the coordinate mapping is a linear transformation,

$$\begin{aligned} [\,\mathbf{x}\,]_{\mathcal{C}} &= [\,3\mathbf{b}_1 + \mathbf{b}_2\,]_{\mathcal{C}} \\ &= 3[\,\mathbf{b}_1\,]_{\mathcal{C}} + [\,\mathbf{b}_2\,]_{\mathcal{C}} \end{aligned}$$

We can write this vector equation as a matrix equation, using the vectors in the linear combination as the columns of a matrix:

$$[\,\mathbf{x}\,]_{\mathcal{C}} = \begin{bmatrix} [\,\mathbf{b}_1\,]_{\mathcal{C}} & [\,\mathbf{b}_2\,]_{\mathcal{C}} \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} \tag{3}$$

This formula gives $[\,\mathbf{x}\,]_{\mathcal{C}}$, once we know the columns of the matrix. From (1),

$$[\,\mathbf{b}_1\,]_{\mathcal{C}} = \begin{bmatrix} 4 \\ 1 \end{bmatrix} \quad \text{and} \quad [\,\mathbf{b}_2\,]_{\mathcal{C}} = \begin{bmatrix} -6 \\ 1 \end{bmatrix}$$

Thus (3) provides the solution:

$$[\,\mathbf{x}\,]_{\mathcal{C}} = \begin{bmatrix} 4 & -6 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

The $\mathcal{C}$-coordinates of $\mathbf{x}$ match those of the $\mathbf{x}$ in Figure 1. ∎

The argument used to derive formula (3) can be generalized to yield the following result. (See Exercises 17 and 18.)

THEOREM 15

Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_n\}$ be bases of a vector space V. Then there is a unique $n \times n$ matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ such that

$$[\,\mathbf{x}\,]_{\mathcal{C}} = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}[\,\mathbf{x}\,]_{\mathcal{B}} \tag{4}$$

The columns of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ are the $\mathcal{C}$-coordinate vectors of the vectors in the basis $\mathcal{B}$. That is,

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} [\mathbf{b}_1]_{\mathcal{C}} & [\mathbf{b}_2]_{\mathcal{C}} & \cdots & [\mathbf{b}_n]_{\mathcal{C}} \end{bmatrix} \tag{5}$$

The matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ in Theorem 15 is called the **change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$**. Multiplication by $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ converts $\mathcal{B}$-coordinates into $\mathcal{C}$-coordinates.[2] Figure 2 illustrates the change-of-coordinates equation (4).

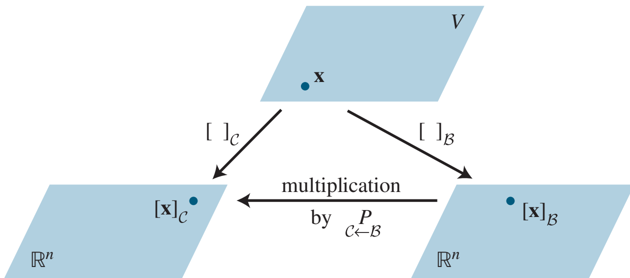


FIGURE 2 Two coordinate systems for V.

The columns of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ are linearly independent because they are the coordinate vectors of the linearly independent set $\mathcal{B}$. (See Exercise 29 in Section 4.4.) Since $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ is square, it must be invertible, by the Invertible Matrix Theorem. Left-multiplying both sides of equation (4) by $(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P})^{-1}$ yields

$$(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P})^{-1}[\,\mathbf{x}\,]_{\mathcal{C}} = [\,\mathbf{x}\,]_{\mathcal{B}}$$

Thus $(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P})^{-1}$ is the matrix that converts $\mathcal{C}$-coordinates into $\mathcal{B}$-coordinates. That is,

$$\boxed{(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P})^{-1} = \underset{\mathcal{B}\leftarrow\mathcal{C}}{P}} \tag{6}$$

Change of Basis in $\mathbb{R}^n$

If $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ and $\mathcal{E}$ is the *standard basis* $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ in $\mathbb{R}^n$, then $[\mathbf{b}_1]_{\mathcal{E}} = \mathbf{b}_1$, and likewise for the other vectors in $\mathcal{B}$. In this case, $\underset{\mathcal{E}\leftarrow\mathcal{B}}{P}$ is the same as the change-of-coordinates matrix $P_{\mathcal{B}}$ introduced in Section 4.4, namely

$$P_{\mathcal{B}} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix}$$

[2] To remember how to construct the matrix, think of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}[\,\mathbf{x}\,]_{\mathcal{B}}$ as a linear combination of the columns of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$. The matrix-vector product is a $\mathcal{C}$-coordinate vector, so the columns of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ should be $\mathcal{C}$-coordinate vectors, too.

To change coordinates between two nonstandard bases in $\mathbb{R}^n$, we need Theorem 15. The theorem shows that to solve the change-of-basis problem, we need the coordinate vectors of the old basis relative to the new basis.

EXAMPLE 2 Let $\mathbf{b}_1 = \begin{bmatrix} -9 \\ 1 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} -5 \\ -1 \end{bmatrix}$, $\mathbf{c}_1 = \begin{bmatrix} 1 \\ -4 \end{bmatrix}$, $\mathbf{c}_2 = \begin{bmatrix} 3 \\ -5 \end{bmatrix}$, and consider the bases for $\mathbb{R}^2$ given by $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$. Find the change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$.

SOLUTION The matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ involves the $\mathcal{C}$-coordinate vectors of $\mathbf{b}_1$ and $\mathbf{b}_2$. Let $[\,\mathbf{b}_1\,]_{\mathcal{C}} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $[\,\mathbf{b}_2\,]_{\mathcal{C}} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. Then, by definition,

$$\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{b}_1 \quad \text{and} \quad \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{b}_2$$

To solve both systems simultaneously, augment the coefficient matrix with $\mathbf{b}_1$ *and* $\mathbf{b}_2$, and row reduce:

$$\left[\begin{array}{cc:cc} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{b}_1 & \mathbf{b}_2 \end{array}\right] = \left[\begin{array}{cc:cc} 1 & 3 & -9 & -5 \\ -4 & -5 & 1 & -1 \end{array}\right] \sim \left[\begin{array}{cc:cc} 1 & 0 & 6 & 4 \\ 0 & 1 & -5 & -3 \end{array}\right] \tag{7}$$

Thus

$$[\,\mathbf{b}_1\,]_{\mathcal{C}} = \begin{bmatrix} 6 \\ -5 \end{bmatrix} \quad \text{and} \quad [\,\mathbf{b}_2\,]_{\mathcal{C}} = \begin{bmatrix} 4 \\ -3 \end{bmatrix}$$

The desired change-of-coordinates matrix is therefore

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} [\,\mathbf{b}_1\,]_{\mathcal{C}} & [\,\mathbf{b}_2\,]_{\mathcal{C}} \end{bmatrix} = \begin{bmatrix} 6 & 4 \\ -5 & -3 \end{bmatrix}$$

Observe that the matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ in Example 2 already appeared in (7). This is not surprising because the first column of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ results from row reducing $\left[\begin{array}{cc:c} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{b}_1 \end{array}\right]$ to $\left[\begin{array}{c:c} I & [\,\mathbf{b}_1\,]_{\mathcal{C}} \end{array}\right]$, and similarly for the second column of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$. Thus

$$\boxed{\left[\begin{array}{cc:cc} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{b}_1 & \mathbf{b}_2 \end{array}\right] \sim \left[\begin{array}{c:c} I & \underset{\mathcal{C}\leftarrow\mathcal{B}}{P} \end{array}\right]}$$

An analogous procedure works for finding the change-of-coordinates matrix between any two bases in $\mathbb{R}^n$.

EXAMPLE 3 Let $\mathbf{b}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$, $\mathbf{c}_1 = \begin{bmatrix} -7 \\ 9 \end{bmatrix}$, $\mathbf{c}_2 = \begin{bmatrix} -5 \\ 7 \end{bmatrix}$, and consider the bases for $\mathbb{R}^2$ given by $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$.

a. Find the change-of-coordinates matrix from $\mathcal{C}$ to $\mathcal{B}$.

b. Find the change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$.

SOLUTION

a. Notice that $\underset{\mathcal{B}\leftarrow\mathcal{C}}{P}$ is needed rather than $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$, and compute

$$\left[\begin{array}{cc:cc} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{c}_1 & \mathbf{c}_2 \end{array}\right] = \left[\begin{array}{cc:cc} 1 & -2 & -7 & -5 \\ -3 & 4 & 9 & 7 \end{array}\right] \sim \left[\begin{array}{cc:cc} 1 & 0 & 5 & 3 \\ 0 & 1 & 6 & 4 \end{array}\right]$$

So

$$\underset{\mathcal{B}\leftarrow\mathcal{C}}{P} = \begin{bmatrix} 5 & 3 \\ 6 & 4 \end{bmatrix}$$

b. By part (a) and property (6) (with $\mathcal{B}$ and $\mathcal{C}$ interchanged),

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = (\underset{\mathcal{B}\leftarrow\mathcal{C}}{P})^{-1} = \frac{1}{2}\begin{bmatrix} 4 & -3 \\ -6 & 5 \end{bmatrix} = \begin{bmatrix} 2 & -3/2 \\ -3 & 5/2 \end{bmatrix} \quad \blacksquare$$

Another description of the change-of-coordinates matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ uses the change-of-coordinate matrices $P_{\mathcal{B}}$ and $P_{\mathcal{C}}$ that convert $\mathcal{B}$-coordinates and $\mathcal{C}$-coordinates, respectively, into standard coordinates. Recall that for each $\mathbf{x}$ in $\mathbb{R}^n$,

$$P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}} = \mathbf{x}, \quad P_{\mathcal{C}}[\mathbf{x}]_{\mathcal{C}} = \mathbf{x}, \quad \text{and} \quad [\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{C}}^{-1}\mathbf{x}$$

Thus

$$[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{C}}^{-1}\mathbf{x} = P_{\mathcal{C}}^{-1}P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$$

In $\mathbb{R}^n$, the change-of-coordinates matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ may be computed as $P_{\mathcal{C}}^{-1}P_{\mathcal{B}}$. Actually, for matrices larger than 2×2, an algorithm analogous to the one in Example 3 is faster than computing $P_{\mathcal{C}}^{-1}$ and then $P_{\mathcal{C}}^{-1}P_{\mathcal{B}}$. See Exercise 22 in Section 2.2.

Practice Problems

1. Let $\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2\}$ and $\mathcal{G} = \{\mathbf{g}_1, \mathbf{g}_2\}$ be bases for a vector space V, and let P be a matrix whose columns are $[\,\mathbf{f}_1\,]_{\mathcal{G}}$ and $[\,\mathbf{f}_2\,]_{\mathcal{G}}$. Which of the following equations is satisfied by P for all $\mathbf{v}$ in V?

   (i) $[\,\mathbf{v}\,]_{\mathcal{F}} = P[\,\mathbf{v}\,]_{\mathcal{G}}$ (ii) $[\,\mathbf{v}\,]_{\mathcal{G}} = P[\,\mathbf{v}\,]_{\mathcal{F}}$
2. Let $\mathcal{B}$ and $\mathcal{C}$ be as in Example 1. Use the results of that example to find the change-of-coordinates matrix from $\mathcal{C}$ to $\mathcal{B}$.

4.6 Exercises

1. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$ be bases for a vector space V, and suppose $\mathbf{b}_1 = 6\mathbf{c}_1 - 2\mathbf{c}_2$ and $\mathbf{b}_2 = 9\mathbf{c}_1 - 4\mathbf{c}_2$.
   a. Find the change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$.
   b. Find $[\,\mathbf{x}\,]_{\mathcal{C}}$ for $\mathbf{x} = -3\mathbf{b}_1 + 2\mathbf{b}_2$. Use part (a).

2. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$ be bases for a vector space V, and suppose $\mathbf{b}_1 = -\mathbf{c}_1 + 4\mathbf{c}_2$ and $\mathbf{b}_2 = 5\mathbf{c}_1 - 3\mathbf{c}_2$.
   a. Find the change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$.
   b. Find $[\,\mathbf{x}\,]_{\mathcal{C}}$ for $\mathbf{x} = 5\mathbf{b}_1 + 3\mathbf{b}_2$.

3. Let $\mathcal{U} = \{\mathbf{u}_1, \mathbf{u}_2\}$ and $\mathcal{W} = \{\mathbf{w}_1, \mathbf{w}_2\}$ be bases for V, and let P be a matrix whose columns are $[\mathbf{u}_1]_{\mathcal{W}}$ and $[\mathbf{u}_2]_{\mathcal{W}}$. Which of the following equations is satisfied by P for all $\mathbf{x}$ in V?

   (i) $[\,\mathbf{x}\,]_{\mathcal{U}} = P[\,\mathbf{x}\,]_{\mathcal{W}}$ (ii) $[\,\mathbf{x}\,]_{\mathcal{W}} = P[\,\mathbf{x}\,]_{\mathcal{U}}$

4. Let $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ and $\mathcal{D} = \{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$ be bases for V, and let $P = [\,[\mathbf{d}_1]_{\mathcal{A}} \quad [\mathbf{d}_2]_{\mathcal{A}} \quad [\mathbf{d}_3]_{\mathcal{A}}\,]$. Which of the following equations is satisfied by P for all $\mathbf{x}$ in V?

   (i) $[\,\mathbf{x}\,]_{\mathcal{A}} = P[\,\mathbf{x}\,]_{\mathcal{D}}$ (ii) $[\,\mathbf{x}\,]_{\mathcal{D}} = P[\,\mathbf{x}\,]_{\mathcal{A}}$

5. Let $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ and $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ be bases for a vector space V, and suppose $\mathbf{a}_1 = 4\mathbf{b}_1 - \mathbf{b}_2$, $\mathbf{a}_2 = -\mathbf{b}_1 + \mathbf{b}_2 + \mathbf{b}_3$, and $\mathbf{a}_3 = \mathbf{b}_2 - 2\mathbf{b}_3$.
   a. Find the change-of-coordinates matrix from $\mathcal{A}$ to $\mathcal{B}$.
   b. Find $[\,\mathbf{x}\,]_{\mathcal{B}}$ for $\mathbf{x} = 3\mathbf{a}_1 + 4\mathbf{a}_2 + \mathbf{a}_3$.

6. Let $\mathcal{D} = \{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$ and $\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$ be bases for a vector space V, and suppose $\mathbf{f}_1 = 2\mathbf{d}_1 - \mathbf{d}_2 + \mathbf{d}_3$, $\mathbf{f}_2 = 3\mathbf{d}_2 + \mathbf{d}_3$, and $\mathbf{f}_3 = -3\mathbf{d}_1 + 2\mathbf{d}_3$.
   a. Find the change-of-coordinates matrix from $\mathcal{F}$ to $\mathcal{D}$.
   b. Find $[\,\mathbf{x}\,]_{\mathcal{D}}$ for $\mathbf{x} = \mathbf{f}_1 - 2\mathbf{f}_2 + 2\mathbf{f}_3$.

In Exercises 7–10, let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$ be bases for $\mathbb{R}^2$. In each exercise, find the change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$ and the change-of-coordinates matrix from $\mathcal{C}$ to $\mathcal{B}$.

7. $\mathbf{b}_1 = \begin{bmatrix} 7 \\ 5 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} -3 \\ -1 \end{bmatrix}, \mathbf{c}_1 = \begin{bmatrix} 1 \\ -5 \end{bmatrix}, \mathbf{c}_2 = \begin{bmatrix} -2 \\ 2 \end{bmatrix}$

8. $\mathbf{b}_1 = \begin{bmatrix} -3 \\ 1 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} -4 \\ 1 \end{bmatrix}, \mathbf{c}_1 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}, \mathbf{c}_2 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$

9. $\mathbf{b}_1 = \begin{bmatrix} -6 \\ -1 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \mathbf{c}_1 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \mathbf{c}_2 = \begin{bmatrix} 6 \\ -2 \end{bmatrix}$

10. $\mathbf{b}_1 = \begin{bmatrix} 8 \\ -3 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 3 \\ -1 \end{bmatrix}, \mathbf{c}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \mathbf{c}_2 = \begin{bmatrix} 7 \\ 3 \end{bmatrix}$

In Exercises 11–14, $\mathcal{B}$ and $\mathcal{C}$ are bases for a vector space V. Mark each statement True or False **(T/F)**. Justify each answer.

11. **(T/F)** The columns of the change-of-coordinates matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ are $\mathcal{B}$-coordinate vectors of the vectors in $\mathcal{C}$.

12. **(T/F)** The columns of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ are linearly independent.

13. **(T/F)** If $V = \mathbb{R}^n$ and $\mathcal{C}$ is the *standard* basis for V, then $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ is the same as the change-of-coordinates matrix $P_{\mathcal{B}}$ introduced in Section 4.4.

14. **(T/F)** If $V = \mathbb{R}^2$, $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$, and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$, then row reduction of $[\,\mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{b}_1 \quad \mathbf{b}_2\,]$ to $[\,I \quad P\,]$ produces a matrix P that satisfies $[\,\mathbf{x}\,]_{\mathcal{B}} = P[\,\mathbf{x}\,]_{\mathcal{C}}$ for all $\mathbf{x}$ in V.

15. In $\mathbb{P}_2$, find the change-of-coordinates matrix from the basis $\mathcal{B} = \{1 - 2t + t^2, 3 - 5t + 4t^2, 2t + 3t^2\}$ to the standard basis $\mathcal{C} = \{1, t, t^2\}$. Then find the $\mathcal{B}$-coordinate vector for $-1 + 2t$.

16. In $\mathbb{P}_2$, find the change-of-coordinates matrix from the basis $\mathcal{B} = \{1 - 3t^2, 2 + t - 5t^2, 1 + 2t\}$ to the standard basis. Then write t^2 as a linear combination of the polynomials in $\mathcal{B}$.

Exercises 17 and 18 provide a proof of Theorem 15. Fill in a justification for each step.

17. Given $\mathbf{v}$ in V, there exist scalars $x_1, \dots, x_n$, such that

$$\mathbf{v} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots + x_n\mathbf{b}_n$$

because (a) ______. Apply the coordinate mapping determined by the basis $\mathcal{C}$, and obtain

$$[\mathbf{v}]_{\mathcal{C}} = x_1[\mathbf{b}_1]_{\mathcal{C}} + x_2[\mathbf{b}_2]_{\mathcal{C}} + \cdots + x_n[\mathbf{b}_n]_{\mathcal{C}}$$

because (b) ______. This equation may be written in the form

$$[\,\mathbf{v}\,]_{\mathcal{C}} = \big[\,[\,\mathbf{b}_1\,]_{\mathcal{C}} \quad [\,\mathbf{b}_2\,]_{\mathcal{C}} \quad \cdots \quad [\,\mathbf{b}_n\,]_{\mathcal{C}}\,\big] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \tag{8}$$

by the definition of (c) ______. This shows that the matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ shown in (5) satisfies $[\mathbf{v}]_{\mathcal{C}} = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}[\mathbf{v}]_{\mathcal{B}}$ for each $\mathbf{v}$ in V, because the vector on the right side of (8) is (d) ______.

18. Suppose Q is any matrix such that

$$[\mathbf{v}]_{\mathcal{C}} = Q[\mathbf{v}]_{\mathcal{B}} \quad \text{for each } \mathbf{v} \text{ in } V \tag{9}$$

Set $\mathbf{v} = \mathbf{b}_1$ in (9). Then (9) shows that $[\mathbf{b}_1]_{\mathcal{C}}$ is the first column of Q because (a) ______. Similarly, for $k = 2, \dots, n$, the kth column of Q is (b) ______ because (c) ______. This shows that the matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ defined by (5) in Theorem 15 is the only matrix that satisfies condition (4).

T 19. Let $\mathcal{B} = \{\mathbf{x}_0, \dots, \mathbf{x}_6\}$ and $\mathcal{C} = \{\mathbf{y}_0, \dots, \mathbf{y}_6\}$, where $\mathbf{x}_k$ is the function $\cos^k t$ and $\mathbf{y}_k$ is the function $\cos kt$. Exercise 54 in Section 4.5 showed that both $\mathcal{B}$ and $\mathcal{C}$ are bases for the vector space $H = \text{Span}\{\mathbf{x}_0, \dots, \mathbf{x}_6\}$.

a. Set $P = \big[\,[\,\mathbf{y}_0\,]_{\mathcal{B}} \quad \cdots \quad [\,\mathbf{y}_6\,]_{\mathcal{B}}\,\big]$, and calculate P^{-1}.

b. Explain why the columns of P^{-1} are the $\mathcal{C}$-coordinate vectors of $\mathbf{x}_0, \dots, \mathbf{x}_6$. Then use these coordinate vectors to write trigonometric identities that express powers of $\cos t$ in terms of the functions in $\mathcal{C}$.

See the *Study Guide*.

T 20. (*Calculus required*)[3] Recall from calculus that integrals such as

$$\int (5\cos^3 t - 6\cos^4 t + 5\cos^5 t - 12\cos^6 t)\,dt \tag{10}$$

are tedious to compute. (The usual method is to apply integration by parts repeatedly and use the half-angle formula.) Use the matrix P or P^{-1} from Exercise 19 to transform (10); then compute the integral.

T 21. Let

$$P = \begin{bmatrix} 1 & 2 & -1 \\ -3 & -5 & 0 \\ 4 & 6 & 1 \end{bmatrix},$$

$$\mathbf{v}_1 = \begin{bmatrix} -2 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -8 \\ 5 \\ 2 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -7 \\ 2 \\ 6 \end{bmatrix}$$

a. Find a basis $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ for $\mathbb{R}^3$ such that P is the change-of-coordinates matrix from $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ to the basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. [*Hint:* What do the columns of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ represent?]

b. Find a basis $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ for $\mathbb{R}^3$ such that P is the change-of-coordinates matrix from $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ to $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$.

T 22. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$, $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$, and $\mathcal{D} = \{\mathbf{d}_1, \mathbf{d}_2\}$ be bases for a two-dimensional vector space.

a. Write an equation that relates the matrices $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$, $\underset{\mathcal{D}\leftarrow\mathcal{C}}{P}$, and $\underset{\mathcal{D}\leftarrow\mathcal{B}}{P}$. Justify your result.

b. Use a matrix program either to help you find the equation or to check the equation you write. Work with three bases for $\mathbb{R}^2$. (See Exercises 7–10.)

[3] The idea for Exercises 19 and 20 and five related exercises in earlier sections came from a paper by Jack W. Rogers, Jr., of Auburn University, presented at a meeting of the International Linear Algebra Society, August 1995. See "Applications of Linear Algebra in Calculus," *American Mathematical Monthly* **104** (1), 1997.

Solutions to Practice Problems

1. Since the columns of P are $\mathcal{G}$-coordinate vectors, a vector of the form $P\mathbf{x}$ must be a $\mathcal{G}$-coordinate vector. Thus P satisfies equation (ii).
2. The coordinate vectors found in Example 1 show that

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \left[\,[\,\mathbf{b}_1\,]_{\mathcal{C}} \quad [\,\mathbf{b}_2\,]_{\mathcal{C}}\,\right] = \begin{bmatrix} 4 & -6 \\ 1 & 1 \end{bmatrix}$$

Hence

$$\underset{\mathcal{B}\leftarrow\mathcal{C}}{P} = (\underset{\mathcal{C}\leftarrow\mathcal{B}}{P})^{-1} = \frac{1}{10}\begin{bmatrix} 1 & 6 \\ -1 & 4 \end{bmatrix} = \begin{bmatrix} .1 & .6 \\ -.1 & .4 \end{bmatrix}$$

4.7 Digital Signal Processing

Introduction

In the space of just a few decades, digital signal processing (DSP) has led to a dramatic shift in how data is collected, processed, and synthesized. DSP models unify the approach to dealing with data that was previously viewed as unrelated. From stock market analysis to telecommunications and computer science, the data collected over time can be viewed as discrete-time signals and DSP used to store and process the data for more efficient and effective use. Not only do digital signals arise in electrical and control systems engineering, but discrete-data sequences are also generated in biology, physics, economics, demography, and many other areas, wherever a process is measured, or *sampled*, at discrete time intervals. In this section, we will explore the properties of the discrete-time signal space, $\mathbb{S}$, and some of its subspaces, as well as how linear transformations can be used to process, filter, and synthesize the data contained in signals.

Discrete-Time Signals

The vector space $\mathbb{S}$ of discrete-time signals was introduced in Section 4.1. A **signal** in $\mathbb{S}$ is an infinite sequence of numbers, $\{y_k\}$, where the subscripts k range over all integers. Table 1 shows several examples of signals.

TABLE 1 Examples of Signals

Signals			
Name	**Symbol**	**Vector**	**Formal Description**
delta	δ	$(\ldots,0,0,0,1,0,0,0,\ldots)$	$\{d_k\}$, where $d_k = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}$
unit step	υ	$(\ldots,0,0,0,1,1,1,1,\ldots)$	$\{u_k\}$, where $u_k = \begin{cases} 1 & \text{if } k \geq 0 \\ 0 & \text{if } k < 0 \end{cases}$
constant	χ	$(\ldots,1,1,1,1,1,1,1,\ldots)$	$\{c_k\}$, where $c_k = 1$
alternating	α	$(\ldots,-1,1,-1,1,-1,1,-1,\ldots)$	$\{a_k\}$, where $a_k = (-1)^k$
Fibonacci	F	$(\ldots,2,-1,1,0,1,1,2,\ldots)$	$\{f_k\}$, where $f_k = \begin{cases} 0 & \text{if } k = 0 \\ 1 & \text{if } k = 1 \\ f_{k-1}+f_{k-2} & \text{if } k > 1 \\ f_{k+2}-f_{k+1} & \text{if } k < 0 \end{cases}$
exponential	ϵ_c	$(\ldots,c^{-2},c^{-1},c^0,c^1,c^2,\ldots)$ $\uparrow$ $k=0$	$\{e_k\}$, where $e_k = c^k$

system) because it describes the changes in a system as time passes.

The 18% juvenile survival rate in the Lamberson stage matrix is the entry affected most by the amount of old-growth forest available. Actually, 60% of the juveniles normally survive to leave the nest, but in the Willow Creek region of California studied by Lamberson and his colleagues, only 30% of the juveniles that left the nest were able to find new home ranges. The rest perished during the search process.

A significant reason for the failure of owls to find new home ranges is the increasing fragmentation of old-growth timber stands due to clear-cutting of scattered areas on the old-growth land. When an owl leaves the protective canopy of the forest and crosses a clear-cut area, the risk of attack by predators increases dramatically. Section 5.6 will show that the model described in the chapter introduction predicts the eventual demise of the spotted owl, but that if 50% of the juveniles who survive to leave the nest also find new home ranges, then the owl population will thrive.

The goal of this chapter is to dissect the action of a linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ into elements that are easily visualized. All matrices in the chapter are square. The main applications described here are to discrete dynamical systems, differential equations, and Markov chains. However, the basic concepts—eigenvectors and eigenvalues—are useful throughout pure and applied mathematics, and they appear in settings far more general than we consider here. Eigenvalues are also used to study differential equations and *continuous* dynamical systems, they provide critical information in engineering design, and they arise naturally in fields such as physics and chemistry.

5.1 Eigenvectors and Eigenvalues

Although a transformation $\mathbf{x} \mapsto A\mathbf{x}$ may move vectors in a variety of directions, it often happens that there are special vectors on which the action of A is quite simple.

EXAMPLE 1 Let $A = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. The images of $\mathbf{u}$ and $\mathbf{v}$ under multiplication by A are shown in Figure 1. In fact, $A\mathbf{v}$ is just $2\mathbf{v}$. So A only "stretches" or dilates $\mathbf{v}$. ■

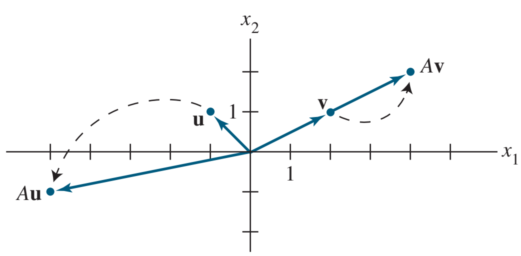


FIGURE 1 Effects of multiplication by A.

This section studies equations such as

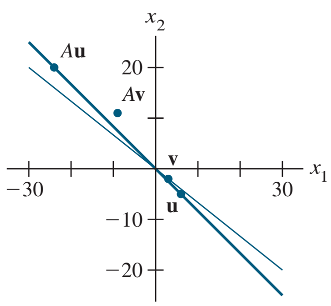
$$A\mathbf{x} = 2\mathbf{x} \quad \text{or} \quad A\mathbf{x} = -4\mathbf{x}$$

where special vectors are transformed by A into scalar multiples of themselves.

DEFINITION

An **eigenvector** of an $n \times n$ matrix A is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ. A scalar λ is called an **eigenvalue** of A if there is a nontrivial solution $\mathbf{x}$ of $A\mathbf{x} = \lambda\mathbf{x}$; such an $\mathbf{x}$ is called an *eigenvector corresponding to* λ.[1]

It is easy to determine if a given vector is an eigenvector of a matrix. See Example 2. It is also easy to decide if a specified scalar is an eigenvalue. See Example 3.



$A\mathbf{u} = -4\mathbf{u}$, but $A\mathbf{v} \neq \lambda\mathbf{v}$.

EXAMPLE 2 Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$. Are $\mathbf{u}$ and $\mathbf{v}$ eigenvectors of A?

SOLUTION

$$A\mathbf{u} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}\begin{bmatrix} 6 \\ -5 \end{bmatrix} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = -4\begin{bmatrix} 6 \\ -5 \end{bmatrix} = -4\mathbf{u}$$

$$A\mathbf{v} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}\begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 11 \end{bmatrix} \neq \lambda\begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

Thus $\mathbf{u}$ is an eigenvector corresponding to an eigenvalue (-4), but $\mathbf{v}$ is not an eigenvector of A, because $A\mathbf{v}$ is not a multiple of $\mathbf{v}$. ■

EXAMPLE 3 Show that 7 is an eigenvalue of matrix A in Example 2, and find the corresponding eigenvectors.

SOLUTION The scalar 7 is an eigenvalue of A if and only if the equation

$$A\mathbf{x} = 7\mathbf{x} \tag{1}$$

has a nontrivial solution. But (1) is equivalent to $A\mathbf{x} - 7\mathbf{x} = \mathbf{0}$, or

$$(A - 7I)\mathbf{x} = \mathbf{0} \tag{2}$$

To solve this homogeneous equation, form the matrix

$$A - 7I = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} - \begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} -6 & 6 \\ 5 & -5 \end{bmatrix}$$

The columns of $A - 7I$ are obviously linearly dependent, so (2) has nontrivial solutions. Thus 7 *is* an eigenvalue of A. To find the corresponding eigenvectors, use row operations:

$$\begin{bmatrix} -6 & 6 & 0 \\ 5 & -5 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The general solution has the form $x_2\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Each vector of this form with $x_2 \neq 0$ is an eigenvector corresponding to $\lambda = 7$. ■

[1] Note that an eigenvector must be *nonzero*, by definition, but an eigenvalue may be zero. The case in which the number 0 is an eigenvalue is discussed after Example 5.

Warning: Although row reduction was used in Example 3 to find eigen*vectors*, it cannot be used to find eigen*values*. An echelon form of a matrix A usually does *not* display the eigenvalues of A.

The equivalence of equations (1) and (2) obviously holds for any λ in place of $\lambda = 7$. Thus λ is an eigenvalue of an $n \times n$ matrix A if and only if the equation

$$(A - \lambda I)\mathbf{x} = \mathbf{0} \tag{3}$$

has a nontrivial solution. The set of *all* solutions of (3) is just the null space of the matrix $A - \lambda I$. So this set is a *subspace* of $\mathbb{R}^n$ and is called the **eigenspace** of A corresponding to λ. The eigenspace consists of the zero vector and all the eigenvectors corresponding to λ.

Example 3 shows that for matrix A in Example 2, the eigenspace corresponding to $\lambda = 7$ consists of *all* multiples of (1, 1), which is the line through (1, 1) and the origin. From Example 2, you can check that the eigenspace corresponding to $\lambda = -4$ is the line through $(6, -5)$. These eigenspaces are shown in Figure 2, along with eigenvectors (1, 1) and $(3/2, -5/4)$ and the geometric action of the transformation $\mathbf{x} \mapsto A\mathbf{x}$ on each eigenspace.

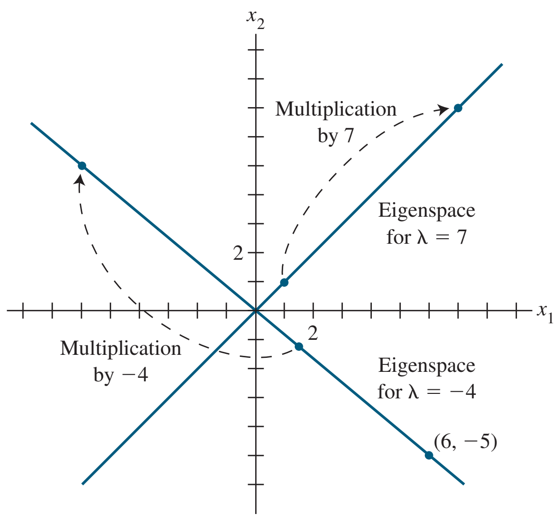


FIGURE 2 Eigenspaces for $\lambda = -4$ and $\lambda = 7$.

EXAMPLE 4 Let $A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$. An eigenvalue of A is 2. Find a basis for the corresponding eigenspace.

SOLUTION Form

$$A - 2I = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix} - \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 6 \\ 2 & -1 & 6 \\ 2 & -1 & 6 \end{bmatrix}$$

and row reduce the augmented matrix for $(A - 2I)\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} 2 & -1 & 6 & 0 \\ 2 & -1 & 6 & 0 \\ 2 & -1 & 6 & 0 \end{bmatrix} \sim \begin{bmatrix} 2 & -1 & 6 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

At this point, it is clear that 2 is indeed an eigenvalue of A because the equation $(A - 2I)\mathbf{x} = \mathbf{0}$ has free variables. The general solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} 1/2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}, \quad x_2 \text{ and } x_3 \text{ free}$$

The eigenspace, shown in Figure 3, is a two-dimensional subspace of $\mathbb{R}^3$. A basis is

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} \right\}$$

■

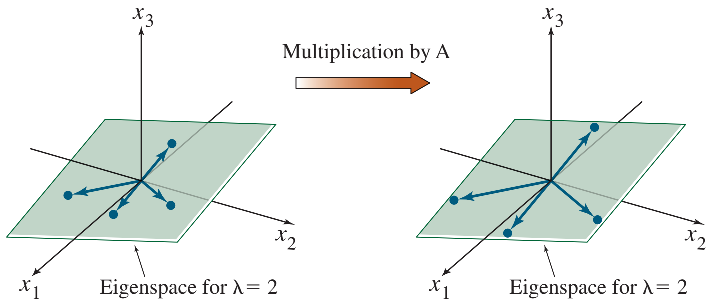


FIGURE 3 A acts as a dilation on the eigenspace.

Reasonable Answers

Remember that once you find a potential eigenvector $\mathbf{v}$, you can easily check your answer: just find $A\mathbf{v}$ and see if it is a multiple of $\mathbf{v}$. For example, to check whether $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector of $A = \begin{bmatrix} 1 & 2 \\ -1 & -2 \end{bmatrix}$, notice $A\mathbf{v} = \begin{bmatrix} 3 \\ -3 \end{bmatrix}$, which is not a multiple of $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, establishing that $\mathbf{v}$ is not an eigenvector. It turns out we had a sign error. The vector $\mathbf{u} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ is a correct eigenvector for A since $A\mathbf{u} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} = -1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -1\,\mathbf{u}$.

Numerical Notes

Example 4 shows a good method for manual computation of eigenvectors in simple cases when an eigenvalue is known. Using a matrix program and row reduction to find an eigenspace (for a specified eigenvalue) usually works, too, but this is not entirely reliable. Roundoff error can lead occasionally to a reduced echelon form with the wrong number of pivots. The best computer programs compute approximations for eigenvalues and eigenvectors simultaneously, to any desired degree of accuracy, for matrices that are not too large. The size of matrices that can be analyzed increases each year as computing power and software improve.

The following theorem describes one of the few special cases in which eigenvalues can be found precisely. Calculation of eigenvalues will also be discussed in Section 5.2.

THEOREM 1

The eigenvalues of a triangular matrix are the entries on its main diagonal.

PROOF For simplicity, consider the 3×3 case. If A is upper triangular, then $A - \lambda I$ has the form

$$\begin{aligned} A - \lambda I &= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} \\ &= \begin{bmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ 0 & a_{22} - \lambda & a_{23} \\ 0 & 0 & a_{33} - \lambda \end{bmatrix} \end{aligned}$$

The scalar λ is an eigenvalue of A if and only if the equation $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has a nontrivial solution, that is, if and only if the equation has a free variable. Because of the zero entries in $A - \lambda I$, it is easy to see that $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has a free variable if and only if at least one of the entries on the diagonal of $A - \lambda I$ is zero. This happens if and only if λ equals one of the entries a_{11}, a_{22}, a_{33} in A. For the case in which A is lower triangular, see Exercise 36. ■

EXAMPLE 5 Let $A = \begin{bmatrix} 3 & 6 & -8 \\ 0 & 0 & 6 \\ 0 & 0 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 0 & 0 \\ -2 & 1 & 0 \\ 5 & 3 & 4 \end{bmatrix}$. The eigenvalues of A are 3, 0, and 2. The eigenvalues of B are 4 and 1. ■

What does it mean for a matrix A to have an eigenvalue of 0, such as in Example 5? This happens if and only if the equation

$$A\mathbf{x} = 0\mathbf{x} \tag{4}$$

has a nontrivial solution. But (4) is equivalent to $A\mathbf{x} = \mathbf{0}$, which has a nontrivial solution if and only if A is not invertible. Thus 0 *is an eigenvalue of A if and only if A is not invertible*. This fact will be added to the Invertible Matrix Theorem in Section 5.2.

The following important theorem will be needed later. Its proof illustrates a typical calculation with eigenvectors. One way to prove the statement "If P then Q" is to show that P and the negation of Q leads to a contradiction. This strategy is used in the proof of the theorem.

THEOREM 2

If $\mathbf{v}_1, \dots, \mathbf{v}_r$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1, \dots, \lambda_r$ of an $n \times n$ matrix A, then the set $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ is linearly independent.

PROOF Suppose $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ is linearly dependent. Since $\mathbf{v}_1$ is nonzero, Theorem 7 in Section 1.7 says that one of the vectors in the set is a linear combination of the preceding vectors. Let p be the least index such that $\mathbf{v}_{p+1}$ is a linear combination of the preceding (linearly independent) vectors. Then there exist scalars $c_1, \dots, c_p$ such that

$$c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{v}_{p+1} \tag{5}$$

Multiplying both sides of (5) by A and using the fact that $A\mathbf{v}_k = \lambda_k \mathbf{v}_k$ for each k, we obtain

$$\begin{aligned} c_1 A\mathbf{v}_1 + \cdots + c_p A\mathbf{v}_p &= A\mathbf{v}_{p+1} \\ c_1 \lambda_1 \mathbf{v}_1 + \cdots + c_p \lambda_p \mathbf{v}_p &= \lambda_{p+1} \mathbf{v}_{p+1} \end{aligned} \tag{6}$$

Multiplying both sides of (5) by λ_{p+1} and subtracting the result from (6), we have

$$c_1(\lambda_1 - \lambda_{p+1})\mathbf{v}_1 + \cdots + c_p(\lambda_p - \lambda_{p+1})\mathbf{v}_p = \mathbf{0} \tag{7}$$

Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly independent, the weights in (7) are all zero. But none of the factors $\lambda_i - \lambda_{p+1}$ are zero, because the eigenvalues are distinct. Hence $c_i = 0$ for $i = 1, \ldots, p$. But then (5) says that $\mathbf{v}_{p+1} = \mathbf{0}$, which is impossible. Hence $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ cannot be linearly dependent and therefore must be linearly independent. ■

Eigenvectors and Difference Equations

This section concludes by showing how to construct solutions of the first-order difference equation discussed in the chapter introductory example:

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \quad (k = 0, 1, 2, \ldots) \tag{8}$$

If A is an $n \times n$ matrix, then (8) is a *recursive* description of a sequence $\{\mathbf{x}_k\}$ in $\mathbb{R}^n$. A **solution** of (8) is an explicit description of $\{\mathbf{x}_k\}$ whose formula for each $\mathbf{x}_k$ does not depend directly on A or on the preceding terms in the sequence other than the initial term $\mathbf{x}_0$.

The simplest way to build a solution of (8) is to take an eigenvector $\mathbf{x}_0$ and its corresponding eigenvalue λ and let

$$\mathbf{x}_k = \lambda^k \mathbf{x}_0 \quad (k = 1, 2, \ldots) \tag{9}$$

This sequence is a solution because

$$A\mathbf{x}_k = A(\lambda^k \mathbf{x}_0) = \lambda^k (A\mathbf{x}_0) = \lambda^k (\lambda \mathbf{x}_0) = \lambda^{k+1} \mathbf{x}_0 = \mathbf{x}_{k+1}$$

Linear combinations of solutions in the form of equation (9) are solutions, too! See Exercise 41.

Practice Problems

1. Is 5 an eigenvalue of $A = \begin{bmatrix} 6 & -3 & 1 \\ 3 & 0 & 5 \\ 2 & 2 & 6 \end{bmatrix}$?
2. If $\mathbf{x}$ is an eigenvector of A corresponding to λ, what is $A^3\mathbf{x}$?
3. Suppose that $\mathbf{b}_1$ and $\mathbf{b}_2$ are eigenvectors corresponding to distinct eigenvalues λ_1 and λ_2, respectively, and suppose that $\mathbf{b}_3$ and $\mathbf{b}_4$ are linearly independent eigenvectors corresponding to a third distinct eigenvalue λ_3. Does it necessarily follow that $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4\}$ is a linearly independent set? [*Hint:* Consider the equation $c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + (c_3\mathbf{b}_3 + c_4\mathbf{b}_4) = \mathbf{0}$.]
4. If A is an $n \times n$ matrix and λ is an eigenvalue of A, show that 2λ is an eigenvalue of $2A$.

5.1 Exercises

1. Is $\lambda = 2$ an eigenvalue of $\begin{bmatrix} 3 & 2 \\ 3 & 8 \end{bmatrix}$? Why or why not?

2. Is $\lambda = -2$ an eigenvalue of $\begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix}$? Why or why not?

3. Is $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$ an eigenvector of $\begin{bmatrix} -3 & 1 \\ -3 & 8 \end{bmatrix}$? If so, find the eigenvalue.

4. Is $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ an eigenvector of $\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$? If so, find the eigenvalue.

5. Is $\begin{bmatrix} 4 \\ -3 \\ 1 \end{bmatrix}$ an eigenvector of $\begin{bmatrix} 3 & 7 & 9 \\ -4 & -5 & 1 \\ 2 & 4 & 4 \end{bmatrix}$? If so, find the eigenvalue.

6. Is $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ an eigenvector of $\begin{bmatrix} 2 & 6 & 7 \\ 3 & 2 & 7 \\ 5 & 6 & 4 \end{bmatrix}$? If so, find the eigenvalue.

7. Is $\lambda = 4$ an eigenvalue of $\begin{bmatrix} 3 & 0 & -1 \\ 2 & 3 & 1 \\ -3 & 4 & 5 \end{bmatrix}$? If so, find one corresponding eigenvector.

8. Is $\lambda = 3$ an eigenvalue of $\begin{bmatrix} 1 & 2 & 2 \\ 3 & -2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$? If so, find one corresponding eigenvector.

In Exercises 9–16, find a basis for the eigenspace corresponding to each listed eigenvalue.

9. $A = \begin{bmatrix} 9 & 0 \\ 2 & 3 \end{bmatrix}, \lambda = 3, 9$

10. $A = \begin{bmatrix} 14 & -4 \\ 16 & -2 \end{bmatrix}, \lambda = 6$

11. $A = \begin{bmatrix} 4 & -2 \\ -3 & 9 \end{bmatrix}, \lambda = 10$

12. $A = \begin{bmatrix} 1 & 4 \\ 3 & 2 \end{bmatrix}, \lambda = -2, 5$

13. $A = \begin{bmatrix} 4 & 0 & 1 \\ -2 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}, \lambda = 1, 2, 3$

14. $A = \begin{bmatrix} 3 & -1 & 3 \\ -1 & 3 & 3 \\ 6 & 6 & 2 \end{bmatrix}, \lambda = -4$

15. $A = \begin{bmatrix} 8 & 3 & -4 \\ -1 & 4 & 4 \\ 2 & 6 & -1 \end{bmatrix}, \lambda = 7$

16. $A = \begin{bmatrix} 3 & 0 & 2 & 0 \\ 1 & 3 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}, \lambda = 4$

Find the eigenvalues of the matrices in Exercises 17 and 18.

17. $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 5 \\ 0 & 0 & -1 \end{bmatrix}$

18. $\begin{bmatrix} 8 & 0 & 0 \\ -7 & 0 & 0 \\ 6 & -5 & -4 \end{bmatrix}$

19. For $A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}$, find one eigenvalue, with no calculation. Justify your answer.

20. Without calculation, find one eigenvalue and two linearly independent eigenvectors of $A = \begin{bmatrix} 4 & 4 & -4 \\ 4 & 4 & -4 \\ 4 & 4 & -4 \end{bmatrix}$. Justify your answer.

In Exercises 21–30, A is an $n \times n$ matrix. Mark each statement True or False **(T/F)**. Justify each answer.

21. **(T/F)** If $A\mathbf{x} = \lambda\mathbf{x}$ for some vector $\mathbf{x}$, then λ is an eigenvalue of A.

22. **(T/F)** If $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ, then $\mathbf{x}$ is an eigenvector of A.

23. **(T/F)** A matrix A is invertible if and only if 0 is an eigenvalue of A.

24. **(T/F)** A number c is an eigenvalue of A if and only if the equation $(A - cI)\mathbf{x} = 0$ has a nontrivial solution.

25. **(T/F)** Finding an eigenvector of A may be difficult, but checking whether a given vector is in fact an eigenvector is easy.

26. **(T/F)** To find the eigenvalues of A, reduce A to echelon form.

27. **(T/F)** If $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent eigenvectors, then they correspond to distinct eigenvalues.

28. **(T/F)** The eigenvalues of a matrix are on its main diagonal.

29. **(T/F)** If $\mathbf{v}$ is an eigenvector with eigenvalue 2, then $2\mathbf{v}$ is an eigenvector with eigenvalue 4.

30. **(T/F)** An eigenspace of A is a null space of a certain matrix.

31. Explain why a 2×2 matrix can have at most two distinct eigenvalues. Explain why an $n \times n$ matrix can have at most n distinct eigenvalues.

32. Construct an example of a 2×2 matrix with only one distinct eigenvalue.

33. Let λ be an eigenvalue of an invertible matrix A. Show that λ^{-1} is an eigenvalue of A^{-1}. [*Hint:* Suppose a nonzero $\mathbf{x}$ satisfies $A\mathbf{x} = \lambda\mathbf{x}$.]

34. Show that if A^2 is the zero matrix, then the only eigenvalue of A is 0.

35. Show that λ is an eigenvalue of A if and only if λ is an eigenvalue of A^T. [*Hint:* Find out how $A - \lambda I$ and $A^T - \lambda I$ are related.]

36. Use Exercise 35 to complete the proof of Theorem 1 for the case when A is lower triangular.

37. Consider an $n \times n$ matrix A with the property that the row sums all equal the same number s. Show that s is an eigenvalue of A. [*Hint:* Find an eigenvector.]

38. Consider an $n \times n$ matrix A with the property that the column sums all equal the same number s. Show that s is an eigenvalue of A. [*Hint:* Use Exercises 35 and 37.]

In Exercises 39 and 40, let A be the matrix of the linear transformation T. Without writing A, find an eigenvalue of A and describe the eigenspace.

39. T is the transformation on $\mathbb{R}^2$ that reflects points across some line through the origin.

40. T is the transformation on $\mathbb{R}^3$ that rotates points about some line through the origin.

41. Let $\mathbf{u}$ and $\mathbf{v}$ be eigenvectors of a matrix A, with corresponding eigenvalues λ and μ, and let c_1 and c_2 be scalars. Define

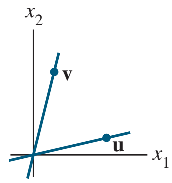
$$\mathbf{x}_k = c_1\lambda^k\mathbf{u} + c_2\mu^k\mathbf{v} \quad (k = 0, 1, 2, \ldots)$$

a. What is $\mathbf{x}_{k+1}$, by definition?

b. Compute $A\mathbf{x}_k$ from the formula for $\mathbf{x}_k$, and show that $A\mathbf{x}_k = \mathbf{x}_{k+1}$. This calculation will prove that the sequence $\{\mathbf{x}_k\}$ defined above satisfies the difference equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$ $(k = 0, 1, 2, \ldots)$.

42. Describe how you might try to build a solution of a difference equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$ $(k = 0, 1, 2, \ldots)$ if you were given the initial $\mathbf{x}_0$ and this vector did not happen to be an eigenvector of A. [*Hint:* How might you relate $\mathbf{x}_0$ to eigenvectors of A?]

43. Let $\mathbf{u}$ and $\mathbf{v}$ be the vectors shown in the figure, and suppose $\mathbf{u}$ and $\mathbf{v}$ are eigenvectors of a 2×2 matrix A that correspond to eigenvalues 2 and 3, respectively. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation given by $T(\mathbf{x}) = A\mathbf{x}$ for each $\mathbf{x}$ in $\mathbb{R}^2$, and let $\mathbf{w} = \mathbf{u} + \mathbf{v}$. Make a copy of the figure, and on the same coordinate system, carefully plot the vectors $T(\mathbf{u})$, $T(\mathbf{v})$, and $T(\mathbf{w})$.



44. Repeat Exercise 43, assuming $\mathbf{u}$ and $\mathbf{v}$ are eigenvectors of A that correspond to eigenvalues -1 and 3, respectively.

T In Exercises 45–48, use a matrix program to find the eigenvalues of the matrix. Then use the method of Example 4 with a row reduction routine to produce a basis for each eigenspace.

45. $$\begin{bmatrix} 8 & -10 & -5 \\ 2 & 17 & 2 \\ -9 & -18 & 4 \end{bmatrix}$$

46. $$\begin{bmatrix} 9 & -4 & -2 & -4 \\ -56 & 32 & -28 & 44 \\ -14 & -14 & 6 & -14 \\ 42 & -33 & 21 & -45 \end{bmatrix}$$

47. $$\begin{bmatrix} 4 & -9 & -7 & 8 & 2 \\ -7 & -9 & 0 & 7 & 14 \\ 5 & 10 & 5 & -5 & -10 \\ -2 & 3 & 7 & 0 & 4 \\ -3 & -13 & -7 & 10 & 11 \end{bmatrix}$$

48. $$\begin{bmatrix} -4 & -4 & 20 & -8 & -1 \\ 14 & 12 & 46 & 18 & 2 \\ 6 & 4 & -18 & 8 & 1 \\ 11 & 7 & -37 & 17 & 2 \\ 18 & 12 & -60 & 24 & 5 \end{bmatrix}$$

Solutions to Practice Problems

1. The number 5 is an eigenvalue of A if and only if the equation $(A - 5I)\mathbf{x} = \mathbf{0}$ has a nontrivial solution. Form

$$A - 5I = \begin{bmatrix} 6 & -3 & 1 \\ 3 & 0 & 5 \\ 2 & 2 & 6 \end{bmatrix} - \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix} = \begin{bmatrix} 1 & -3 & 1 \\ 3 & -5 & 5 \\ 2 & 2 & 1 \end{bmatrix}$$

and row reduce the augmented matrix:

$$\begin{bmatrix} 1 & -3 & 1 & 0 \\ 3 & -5 & 5 & 0 \\ 2 & 2 & 1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 1 & 0 \\ 0 & 4 & 2 & 0 \\ 0 & 8 & -1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 1 & 0 \\ 0 & 4 & 2 & 0 \\ 0 & 0 & -5 & 0 \end{bmatrix}$$

Solutions to Practice Problems (Continued)

At this point, it is clear that the homogeneous system has no free variables. Thus $A - 5I$ is an invertible matrix, which means that 5 is *not* an eigenvalue of A.

2. If $\mathbf{x}$ is an eigenvector of A corresponding to λ, then $A\mathbf{x} = \lambda\mathbf{x}$ and so

$$A^2\mathbf{x} = A(\lambda\mathbf{x}) = \lambda A\mathbf{x} = \lambda^2\mathbf{x}$$

Again, $A^3\mathbf{x} = A(A^2\mathbf{x}) = A(\lambda^2\mathbf{x}) = \lambda^2 A\mathbf{x} = \lambda^3\mathbf{x}$. The general pattern, $A^k\mathbf{x} = \lambda^k\mathbf{x}$, is proved by induction.

3. Yes. Suppose $c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + (c_3\mathbf{b}_3 + c_4\mathbf{b}_4) = \mathbf{0}$. Since any linear combination of eigenvectors corresponding to the same eigenvalue is in the eigenspace for that eigenvalue, $c_3\mathbf{b}_3 + c_4\mathbf{b}_4$ is either $\mathbf{0}$ or an eigenvector for λ_3. If $c_3\mathbf{b}_3 + c_4\mathbf{b}_4$ were an eigenvector for λ_3, then by Theorem 2, $\{\mathbf{b}_1, \mathbf{b}_2, c_3\mathbf{b}_3 + c_4\mathbf{b}_4\}$ would be a linearly independent set, which would force $c_1 = c_2 = 0$ and $c_3\mathbf{b}_3 + c_4\mathbf{b}_4 = \mathbf{0}$, contradicting that $c_3\mathbf{b}_3 + c_4\mathbf{b}_4$ is an eigenvector. Thus $c_3\mathbf{b}_3 + c_4\mathbf{b}_4$ must be $\mathbf{0}$, implying that $c_1\mathbf{b}_1 + c_2\mathbf{b}_2 = \mathbf{0}$ also. By Theorem 2, $\{\mathbf{b}_1, \mathbf{b}_2\}$ is a linearly independent set so $c_1 = c_2 = 0$. Moreover, $\{\mathbf{b}_3, \mathbf{b}_4\}$ is a linearly independent set so $c_3 = c_4 = 0$. Since all of the coefficients c_1, c_2, c_3, and c_4 must be zero, it follows that $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4\}$ is a linearly independent set.

4. Since λ is an eigenvalue of A, there is a nonzero vector $\mathbf{x}$ in $\mathbb{R}^n$ such that $A\mathbf{x} = \lambda\mathbf{x}$. Multiplying both sides of this equation by 2 results in the equation $2(A\mathbf{x}) = 2(\lambda\mathbf{x})$. Thus $(2A)\mathbf{x} = (2\lambda)\mathbf{x}$ and hence 2λ is an eigenvalue of $2A$.

5.2 The Characteristic Equation

Useful information about the eigenvalues of a square matrix A is encoded in a special scalar equation called the characteristic equation of A. A simple example will lead to the general case.

EXAMPLE 1 Find the eigenvalues of $A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$.

SOLUTION We must find all scalars λ such that the matrix equation

$$(A - \lambda I)\mathbf{x} = \mathbf{0}$$

has a nontrivial solution. By the Invertible Matrix Theorem in Section 2.3, this problem is equivalent to finding all λ such that the matrix $A - \lambda I$ is *not* invertible, where

$$A - \lambda I = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{bmatrix}$$

By Theorem 4 in Section 2.2, this matrix fails to be invertible precisely when its determinant is zero. So the eigenvalues of A are the solutions of the equation

$$\det(A - \lambda I) = \det \begin{bmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{bmatrix} = 0$$

Recall that

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

So

$$\begin{aligned}\det(A-\lambda I) &= (2-\lambda)(-6-\lambda)-(3)(3)\\ &= -12+6\lambda-2\lambda+\lambda^2-9\\ &= \lambda^2+4\lambda-21\\ &= (\lambda-3)(\lambda+7)\end{aligned}$$

If $\det(A-\lambda I)=0$, then $\lambda=3$ or $\lambda=-7$. So the eigenvalues of A are 3 and -7. ■

Determinants

The determinant in Example 1 transformed the matrix equation $(A-\lambda I)\mathbf{x}=\mathbf{0}$, which involves *two* unknowns λ and $\mathbf{x}$, into the scalar equation $\lambda^2+4\lambda-21=0$, which involves only *one* unknown. The same idea works for $n\times n$ matrices.

Before turning to larger matrices, recall from Section 3.1 that the matrix A_{ij} is obtained from A by deleting the ith row and jth column. The determinant of an $n\times n$ matrix A can be computed by an expansion across any row or down any column. The expansion across the ith row is given by

$$\det A=(-1)^{i+1}a_{i1}\det A_{i1}+(-1)^{i+2}a_{i2}\det A_{i2}+\cdots+(-1)^{i+n}a_{in}\det A_{in}$$

The expansion down the jth column is given by

$$\det A=(-1)^{1+j}a_{1j}\det A_{1j}+(-1)^{2+j}a_{2j}\det A_{2j}+\cdots+(-1)^{n+j}a_{nj}\det A_{nj}$$

EXAMPLE 2 Compute the determinant of

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

SOLUTION Any row or column can be chosen for the expansion. For example, expanding down the first column of A results in

$$\begin{aligned}\det A &= a_{11}\det A_{11}-a_{21}\det A_{21}+a_{31}\det A_{31}\\ &= 2\det\begin{bmatrix}0 & -1\\ 2 & 1\end{bmatrix}-4\det\begin{bmatrix}3 & 1\\ 2 & 1\end{bmatrix}+0\det\begin{bmatrix}3 & 1\\ 0 & -1\end{bmatrix}\\ &= 2(0-(-2))-4(3-2)+0(-3-0)=0\end{aligned}$$

The next theorem lists facts from Sections 3.1 and 3.2 and is included here for convenient reference. ■

THEOREM 3

Properties of Determinants

Let A and B be $n\times n$ matrices.

a. A is invertible if and only if $\det A\neq 0$.

b. $\det AB=(\det A)(\det B)$.

c. $\det A^T=\det A$.

d. If A is triangular, then $\det A$ is the product of the entries on the main diagonal of A.

e. A row replacement operation on A does not change the determinant. A row interchange changes the sign of the determinant. A row scaling also scales the determinant by the same scalar factor.

Recall that A is invertible if and only if the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Notice that the number 0 is an eigenvalue of A if and only if there *is* a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = 0\mathbf{x} = \mathbf{0}$, which happens if and only if $0 = \det(A - 0I) = \det A$. Hence A is invertible if and only if 0 is *not* an eigenvalue.

THEOREM

The Invertible Matrix Theorem (continued)

Let A be an $n \times n$ matrix. Then A is invertible if and only if

r. The number 0 is *not* an eigenvalue of A.

The Characteristic Equation

Theorem 3(a) shows how to determine when a matrix of the form $A - \lambda I$ is *not* invertible. The scalar equation $\det(A - \lambda I) = 0$ is called the **characteristic equation** of A, and the argument in Example 1 justifies the following fact.

A scalar λ is an eigenvalue of an $n \times n$ matrix A if and only if λ satisfies the characteristic equation

$$\det(A - \lambda I) = 0$$

EXAMPLE 3 Find the characteristic equation of

$$A = \begin{bmatrix} 5 & -2 & 6 & -1 \\ 0 & 3 & -8 & 0 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

SOLUTION Form $A - \lambda I$, and use Theorem 3(d):

$$\begin{aligned} \det(A - \lambda I) &= \det \begin{bmatrix} 5-\lambda & -2 & 6 & -1 \\ 0 & 3-\lambda & -8 & 0 \\ 0 & 0 & 5-\lambda & 4 \\ 0 & 0 & 0 & 1-\lambda \end{bmatrix} \\ &= (5-\lambda)(3-\lambda)(5-\lambda)(1-\lambda) \end{aligned}$$

The characteristic equation is

$$(5-\lambda)^2(3-\lambda)(1-\lambda) = 0$$

or

$$(\lambda - 5)^2(\lambda - 3)(\lambda - 1) = 0$$

Expanding the product, we can also write

$$\lambda^4 - 14\lambda^3 + 68\lambda^2 - 130\lambda + 75 = 0$$ ■

Reasonable Answers

If you want to verify λ is an eigenvalue of A, row reduce $A - \lambda I$. If you get a pivot in every column, something is amiss—the scalar λ is not an eigenvalue of A. Looking back at Example 3, notice that $A - 5I$, $A - 3I$, and $A - I$ all have at least one column without a pivot; however, if λ is chosen to be any number other than 5, 3, or 1, the matrix $A - \lambda I$ has a pivot in every column.

In Examples 1 and 3, det $(A - \lambda I)$ is a polynomial in λ. It can be shown that if A is an $n \times n$ matrix, then det $(A - \lambda I)$ is a polynomial of degree n called the **characteristic polynomial** of A.

The eigenvalue 5 in Example 3 is said to have *multiplicity* 2 because $(\lambda - 5)$ occurs two times as a factor of the characteristic polynomial. In general, the (**algebraic**) **multiplicity** of an eigenvalue λ is its multiplicity as a root of the characteristic equation.

EXAMPLE 4 The characteristic polynomial of a 6×6 matrix is $\lambda^6 - 4\lambda^5 - 12\lambda^4$. Find the eigenvalues and their multiplicities.

SOLUTION Factor the polynomial

$$\lambda^6 - 4\lambda^5 - 12\lambda^4 = \lambda^4(\lambda^2 - 4\lambda - 12) = \lambda^4(\lambda - 6)(\lambda + 2)$$

The eigenvalues are 0 (multiplicity 4), 6 (multiplicity 1), and −2 (multiplicity 1). ■

We could also list the eigenvalues in Example 4 as 0, 0, 0, 0, 6, and −2, so that the eigenvalues are repeated according to their multiplicities.

Because the characteristic equation for an $n \times n$ matrix involves an nth-degree polynomial, the equation has exactly n roots, counting multiplicities, provided complex roots are allowed. Such complex roots, called *complex eigenvalues*, will be discussed in Section 5.5. Until then, we consider only real eigenvalues, and scalars will continue to be real numbers.

The characteristic equation is important for theoretical purposes. In practical work, however, eigenvalues of any matrix larger than 2×2 should be found by a computer, unless the matrix is triangular or has other special properties. Although a 3×3 characteristic polynomial is easy to compute by hand, factoring it can be difficult (unless the matrix is carefully chosen). See the Numerical Notes at the end of this section.

STUDY GUIDE has advice on how to factor a polynomial.

Similarity

The next theorem illustrates one use of the characteristic polynomial, and it provides the foundation for several iterative methods that *approximate* eigenvalues. If A and B are $n \times n$ matrices, then A **is similar to** B if there is an invertible matrix P such that $P^{-1}AP = B$, or, equivalently, $A = PBP^{-1}$. Writing Q for P^{-1}, we have $Q^{-1}BQ = A$. So B is also similar to A, and we say simply that A and B **are similar**. Changing A into $P^{-1}AP$ is called a **similarity transformation**.

THEOREM 4

If $n \times n$ matrices A and B are similar, then they have the same characteristic polynomial and hence the same eigenvalues (with the same multiplicities).

PROOF If $B = P^{-1}AP$, then

$$B - \lambda I = P^{-1}AP - \lambda P^{-1}P = P^{-1}(AP - \lambda P) = P^{-1}(A - \lambda I)P$$

Using the multiplicative property (b) in Theorem 3, we compute

$$\begin{aligned} \det(B - \lambda I) &= \det[P^{-1}(A - \lambda I)P] \\ &= \det(P^{-1}) \cdot \det(A - \lambda I) \cdot \det(P) \end{aligned} \tag{1}$$

Since $\det(P^{-1}) \cdot \det(P) = \det(P^{-1}P) = \det I = 1$, we see from equation (1) that $\det(B - \lambda I) = \det(A - \lambda I)$. ■

***Warnings*:**

1. The matrices

$$\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} \text{ and } \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

are not similar even though they have the same eigenvalues.

2. Similarity is not the same as row equivalence. (If A is row equivalent to B, then $B = EA$ for some invertible matrix E.) Row operations on a matrix usually change its eigenvalues.

Application to Dynamical Systems

Eigenvalues and eigenvectors hold the key to the discrete evolution of a dynamical system, as mentioned in the chapter introduction.

EXAMPLE 5 Let $A = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix}$. Analyze the long-term behavior (as k increases) of the dynamical system defined by $\mathbf{x}_{k+1} = A\mathbf{x}_k$ $(k = 0, 1, 2, \ldots)$, with $\mathbf{x}_0 = \begin{bmatrix} .6 \\ .4 \end{bmatrix}$.

SOLUTION The first step is to find the eigenvalues of A and a basis for each eigenspace. The characteristic equation for A is

$$\begin{aligned} 0 = \det \begin{bmatrix} .95 - \lambda & .03 \\ .05 & .97 - \lambda \end{bmatrix} &= (.95 - \lambda)(.97 - \lambda) - (.03)(.05) \\ = \lambda^2 - 1.92\lambda + .92 \end{aligned}$$

By the quadratic formula

$$\begin{aligned} \lambda &= \frac{1.92 \pm \sqrt{(-1.92)^2 - 4(.92)}}{2} = \frac{1.92 \pm \sqrt{.0064}}{2} \\ &= \frac{1.92 \pm .08}{2} = 1 \quad \text{or} \quad .92 \end{aligned}$$

It is readily checked that eigenvectors corresponding to $\lambda = 1$ and $\lambda = .92$ are multiples of

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 5 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

respectively.

The next step is to write the given $\mathbf{x}_0$ in terms of $\mathbf{v}_1$ and $\mathbf{v}_2$. This can be done because $\{\mathbf{v}_1, \mathbf{v}_2\}$ is obviously a basis for $\mathbb{R}^2$. (Why?) So there exist weights c_1 and c_2 such that

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = [\,\mathbf{v}_1 \quad \mathbf{v}_2\,]\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \tag{2}$$

In fact,

$$\begin{aligned} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} &= [\,\mathbf{v}_1 \quad \mathbf{v}_2\,]^{-1}\mathbf{x}_0 = \begin{bmatrix} 3 & 1 \\ 5 & -1 \end{bmatrix}^{-1}\begin{bmatrix} .60 \\ .40 \end{bmatrix} \\ &= \frac{1}{-8}\begin{bmatrix} -1 & -1 \\ -5 & 3 \end{bmatrix}\begin{bmatrix} .60 \\ .40 \end{bmatrix} = \begin{bmatrix} .125 \\ .225 \end{bmatrix} \end{aligned} \tag{3}$$

Because $\mathbf{v}_1$ and $\mathbf{v}_2$ in (3) are eigenvectors of A, with $A\mathbf{v}_1 = \mathbf{v}_1$ and $A\mathbf{v}_2 = .92\mathbf{v}_2$, we easily compute each $\mathbf{x}_k$:

$$\begin{aligned} \mathbf{x}_1 = A\mathbf{x}_0 &= c_1A\mathbf{v}_1 + c_2A\mathbf{v}_2 && \text{Using linearity of } \mathbf{x} \mapsto A\mathbf{x} \\ &= c_1\mathbf{v}_1 + c_2(.92)\mathbf{v}_2 && \mathbf{v}_1 \text{ and } \mathbf{v}_2 \text{ are eigenvectors.} \\ \mathbf{x}_2 = A\mathbf{x}_1 &= c_1A\mathbf{v}_1 + c_2(.92)A\mathbf{v}_2 \\ &= c_1\mathbf{v}_1 + c_2(.92)^2\mathbf{v}_2 \end{aligned}$$

and so on. In general,

$$\mathbf{x}_k = c_1\mathbf{v}_1 + c_2(.92)^k\mathbf{v}_2 \quad (k = 0, 1, 2, \dots)$$

Using c_1 and c_2 from (4),

$$\mathbf{x}_k = .125\begin{bmatrix} 3 \\ 5 \end{bmatrix} + .225(.92)^k\begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad (k = 0, 1, 2, \dots) \tag{4}$$

This explicit formula for $\mathbf{x}_k$ gives the solution of the difference equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$. As $k \to \infty$, $(.92)^k$ tends to zero and $\mathbf{x}_k$ tends to $\begin{bmatrix} .375 \\ .625 \end{bmatrix} = .125\mathbf{v}_1$. ■

The calculations in Example 5 have an interesting application to a Markov chain discussed in Section 5.9. Those who read that section may recognize that matrix A in Example 5 above is the same as the migration matrix M in Section 5.9, $\mathbf{x}_0$ is the initial population distribution between city and suburbs, and $\mathbf{x}_k$ represents the population distribution after k years.

Numerical Notes

1. Computer software such as Mathematica and Maple can use symbolic calculations to find the characteristic polynomial of a moderate-sized matrix. But there is no formula or finite algorithm to solve the characteristic equation of a general $n \times n$ matrix for $n \geq 5$.

Numerical Notes (Continued)

2. The best numerical methods for finding eigenvalues avoid the characteristic polynomial entirely. In fact, MATLAB finds the characteristic polynomial of a matrix A by first computing the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A and then expanding the product $(\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)$.
3. Several common algorithms for estimating the eigenvalues of a matrix A are based on Theorem 4. The powerful *QR algorithm* is discussed in the exercises. Another technique, called *Jacobi's method*, works when $A = A^T$ and computes a sequence of matrices of the form

$$A_1 = A \quad \text{and} \quad A_{k+1} = P_k^{-1} A_k P_k \quad (k = 1, 2, \ldots)$$

Each matrix in the sequence is similar to A and so has the same eigenvalues as A. The nondiagonal entries of A_{k+1} tend to zero as k increases, and the diagonal entries tend to approach the eigenvalues of A.
4. Other methods of estimating eigenvalues are discussed in Section 5.8.

Practice Problem

Find the characteristic equation and eigenvalues of $A = \begin{bmatrix} 1 & -4 \\ 4 & 2 \end{bmatrix}$.

5.2 Exercises

Find the characteristic polynomial and the eigenvalues of the matrices in Exercises 1–8.

1. $\begin{bmatrix} 2 & 7 \\ 7 & 2 \end{bmatrix}$ **2.** $\begin{bmatrix} 8 & 4 \\ 4 & 8 \end{bmatrix}$

3. $\begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix}$ **4.** $\begin{bmatrix} 5 & -5 \\ -2 & 3 \end{bmatrix}$

5. $\begin{bmatrix} 2 & 1 \\ -1 & 4 \end{bmatrix}$ **6.** $\begin{bmatrix} 1 & -4 \\ 4 & 6 \end{bmatrix}$

7. $\begin{bmatrix} 5 & 3 \\ -4 & 4 \end{bmatrix}$ **8.** $\begin{bmatrix} 7 & -2 \\ 2 & 3 \end{bmatrix}$

Exercises 9–14 require techniques from Section 3.1. Find the characteristic polynomial of each matrix using expansion across a row or down a column. [*Note:* Finding the characteristic polynomial of a 3×3 matrix is not easy to do with just row operations, because the variable λ is involved.]

9. $\begin{bmatrix} 1 & 0 & -1 \\ 2 & 3 & -1 \\ 0 & 6 & 0 \end{bmatrix}$ **10.** $\begin{bmatrix} 0 & 3 & 1 \\ 3 & 0 & 2 \\ 1 & 2 & 0 \end{bmatrix}$

11. $\begin{bmatrix} 6 & 0 & 0 \\ 5 & 4 & 3 \\ 1 & 0 & 2 \end{bmatrix}$ **12.** $\begin{bmatrix} 1 & 0 & 1 \\ -3 & 6 & 1 \\ 0 & 0 & 4 \end{bmatrix}$

13. $\begin{bmatrix} 6 & -2 & 0 \\ -2 & 9 & 0 \\ 5 & 8 & 3 \end{bmatrix}$ **14.** $\begin{bmatrix} 3 & -2 & 3 \\ 0 & -1 & 0 \\ 6 & 7 & -4 \end{bmatrix}$

For the matrices in Exercises 15–17, list the eigenvalues, repeated according to their multiplicities.

15. $\begin{bmatrix} 7 & -5 & 3 & 0 \\ 0 & 3 & 7 & -5 \\ 0 & 0 & 5 & -3 \\ 0 & 0 & 0 & 7 \end{bmatrix}$ **16.** $\begin{bmatrix} 5 & 0 & 0 & 0 \\ 8 & -4 & 0 & 0 \\ 0 & 7 & 1 & 0 \\ 1 & -5 & 2 & 1 \end{bmatrix}$

17. $\begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ -5 & 1 & 0 & 0 & 0 \\ 3 & 8 & 0 & 0 & 0 \\ 0 & -7 & 2 & 1 & 0 \\ -4 & 1 & 9 & -2 & 3 \end{bmatrix}$

18. It can be shown that the algebraic multiplicity of an eigenvalue λ is always greater than or equal to the dimension of the eigenspace corresponding to λ. Find h in the matrix A below such that the eigenspace for $\lambda = 6$ is two-dimensional:

$$A = \begin{bmatrix} 6 & 3 & 9 & -5 \\ 0 & 9 & h & 2 \\ 0 & 0 & 6 & 8 \\ 0 & 0 & 0 & 7 \end{bmatrix}$$

19. Let A be an $n \times n$ matrix, and suppose A has n real eigenvalues, $\lambda_1, \ldots, \lambda_n$, repeated according to multiplicities, so that
$\det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda)$

Explain why $\det A$ is the product of the n eigenvalues of A. (This result is true for any square matrix when complex eigenvalues are considered.)

20. Use a property of determinants to show that A and A^T have the same characteristic polynomial.

In Exercises 21–30, A and B are $n \times n$ matrices. Mark each statement True or False **(T/F)**. Justify each answer.

21. **(T/F)** If 0 is an eigenvalue of A, then A is invertible.

22. **(T/F)** The zero vector is in the eigenspace of A associated with an eigenvalue λ.

23. **(T/F)** The matrix A and its transpose, A^T, have different sets of eigenvalues.

24. **(T/F)** The matrices A and $B^{-1}AB$ have the same sets of eigenvalues for every invertible matrix B.

25. **(T/F)** If 2 is an eigenvalue of A, then $A - 2I$ is not invertible.

26. **(T/F)** If two matrices have the same set of eigenvalues, then they are similar.

27. **(T/F)** If $\lambda + 5$ is a factor of the characteristic polynomial of A, then 5 is an eigenvalue of A.

28. **(T/F)** The multiplicity of a root r of the characteristic equation of A is called the algebraic multiplicity of r as an eigenvalue of A.

29. **(T/F)** The eigenvalue of the $n \times n$ identity matrix is 1 with algebraic multiplicity n.

30. **(T/F)** The matrix A can have more than n eigenvalues.

A widely used method for estimating eigenvalues of a general matrix A is the *QR algorithm*. Under suitable conditions, this algorithm produces a sequence of matrices, all similar to A, that become almost upper triangular, with diagonal entries that approach the eigenvalues of A. The main idea is to factor A (or another matrix similar to A) in the form $A = Q_1R_1$, where $Q_1^T = Q_1^{-1}$ and R_1 is upper triangular. The factors are interchanged to form $A_1 = R_1Q_1$, which is again factored as $A_1 = Q_2R_2$; then to form $A_2 = R_2Q_2$, and so on. The similarity of $A, A_1, \dots$ follows from the more general result in Exercise 31.

31. Show that if $A = QR$ with Q invertible, then A is similar to $A_1 = RQ$.

32. Show that if A and B are similar, then $\det A = \det B$.

T 33. Construct a random integer-valued 4×4 matrix A, and verify that A and A^T have the same characteristic polynomial (the same eigenvalues with the same multiplicities). Do A and A^T have the same eigenvectors? Make the same analysis of a 5×5 matrix. Report the matrices and your conclusions.

T 34. Construct a random integer-valued 4×4 matrix A.

a. Reduce A to echelon form U with no row scaling, and compute $\det A$. (If A happens to be singular, start over with a new random matrix.)

b. Compute the eigenvalues of A and the product of these eigenvalues (as accurately as possible).

c. List the matrix A, and, to four decimal places, list the pivots in U and the eigenvalues of A. Compute $\det A$ with your matrix program, and compare it with the products you found in (a) and (b).

T 35. Let $A = \begin{bmatrix} -6 & 28 & 21 \\ 4 & -15 & -12 \\ -8 & a & 25 \end{bmatrix}$. For each value of a in the set $\{32, 31.9, 31.8, 32.1, 32.2\}$, compute the characteristic polynomial of A and the eigenvalues. In each case, create a graph of the characteristic polynomial $p(t) = \det(A - tI)$ for $0 \le t \le 3$. If possible, construct all graphs on one coordinate system. Describe how the graphs reveal the changes in the eigenvalues as a changes.

Solution to Practice Problem

The characteristic equation is

$$0 = \det(A - \lambda I) = \det \begin{bmatrix} 1-\lambda & -4 \\ 4 & 2-\lambda \end{bmatrix}$$

$$= (1-\lambda)(2-\lambda) - (-4)(4) = \lambda^2 - 3\lambda + 18$$

From the quadratic formula,

$$\lambda = \frac{3 \pm \sqrt{(-3)^2 - 4(18)}}{2} = \frac{3 \pm \sqrt{-63}}{2}$$

It is clear that the characteristic equation has no real solutions, so A has no real eigenvalues. The matrix A is acting on the real vector space $\mathbb{R}^2$, and there is no nonzero vector $\mathbf{v}$ in $\mathbb{R}^2$ such that $A\mathbf{v} = \lambda\mathbf{v}$ for some scalar λ.

5.3 Diagonalization

In many cases, the eigenvalue–eigenvector information contained within a matrix A can be displayed in a useful factorization of the form $A = PDP^{-1}$ where D is a diagonal matrix. In this section, the factorization enables us to compute A^k quickly for large values of k, a fundamental idea in several applications of linear algebra. Later, in Sections 5.6 and 5.7, the factorization will be used to analyze (and *decouple*) dynamical systems.

The following example illustrates that powers of a diagonal matrix are easy to compute.

EXAMPLE 1 If $D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$, then $D^2 = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix}$ and

$$D^3 = DD^2 = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix} = \begin{bmatrix} 5^3 & 0 \\ 0 & 3^3 \end{bmatrix}$$

In general,

$$D^k = \begin{bmatrix} 5^k & 0 \\ 0 & 3^k \end{bmatrix} \quad \text{for } k \geq 1$$

■

If $A = PDP^{-1}$ for some invertible P and diagonal D, then A^k is also easy to compute, as the next example shows.

EXAMPLE 2 Let $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. Find a formula for A^k, given that $A = PDP^{-1}$, where

$$P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$$

SOLUTION The standard formula for the inverse of a 2×2 matrix yields

$$P^{-1} = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

Then, by associativity of matrix multiplication,

$$A^2 = (PDP^{-1})(PDP^{-1}) = PD\underbrace{(P^{-1}P)}_{I}DP^{-1} = PDDP^{-1}$$

$$= PD^2P^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

Again,

$$A^3 = (PDP^{-1})A^2 = (PD\underbrace{P^{-1})P}_{I}D^2P^{-1} = PDD^2P^{-1} = PD^3P^{-1}$$

In general, for $k \geq 1$,

$$A^k = PD^kP^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} 5^k & 0 \\ 0 & 3^k \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \cdot 5^k - 3^k & 5^k - 3^k \\ 2 \cdot 3^k - 2 \cdot 5^k & 2 \cdot 3^k - 5^k \end{bmatrix}$$

■

A square matrix A is said to be **diagonalizable** if A is similar to a diagonal matrix, that is, if $A = PDP^{-1}$ for some invertible matrix P and some diagonal matrix D. The next theorem gives a characterization of diagonalizable matrices and tells how to construct a suitable factorization.

THEOREM 5

The Diagonalization Theorem

An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

In fact, $A = PDP^{-1}$, with D a diagonal matrix, if and only if the columns of P are n linearly independent eigenvectors of A. In this case, the diagonal entries of D are eigenvalues of A that correspond, respectively, to the eigenvectors in P.

In other words, A is diagonalizable if and only if there are enough eigenvectors to form a basis of $\mathbb{R}^n$. We call such a basis an **eigenvector basis** of $\mathbb{R}^n$.

PROOF First, observe that if P is any $n \times n$ matrix with columns $\mathbf{v}_1, \dots, \mathbf{v}_n$, and if D is any diagonal matrix with diagonal entries $\lambda_1, \dots, \lambda_n$, then

$$AP = A[\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n\,] = [\,A\mathbf{v}_1 \quad A\mathbf{v}_2 \quad \cdots \quad A\mathbf{v}_n\,] \tag{1}$$

while

$$PD = P\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = [\,\lambda_1\mathbf{v}_1 \quad \lambda_2\mathbf{v}_2 \quad \cdots \quad \lambda_n\mathbf{v}_n\,] \tag{2}$$

Now suppose A is diagonalizable and $A = PDP^{-1}$. Then right-multiplying this relation by P, we have $AP = PD$. In this case, equations (1) and (2) imply that

$$[\,A\mathbf{v}_1 \quad A\mathbf{v}_2 \quad \cdots \quad A\mathbf{v}_n\,] = [\,\lambda_1\mathbf{v}_1 \quad \lambda_2\mathbf{v}_2 \quad \cdots \quad \lambda_n\mathbf{v}_n\,] \tag{3}$$

Equating columns, we find that

$$A\mathbf{v}_1 = \lambda_1\mathbf{v}_1, \quad A\mathbf{v}_2 = \lambda_2\mathbf{v}_2, \quad \dots, \quad A\mathbf{v}_n = \lambda_n\mathbf{v}_n \tag{4}$$

Since P is invertible, its columns $\mathbf{v}_1, \dots, \mathbf{v}_n$ must be linearly independent. Also, since these columns are nonzero, the equations in (4) show that $\lambda_1, \dots, \lambda_n$ are eigenvalues and $\mathbf{v}_1, \dots, \mathbf{v}_n$ are corresponding eigenvectors. This argument proves the "only if" parts of the first and second statements, along with the third statement, of the theorem.

Finally, given any n eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$, use them to construct the columns of P and use corresponding eigenvalues $\lambda_1, \dots, \lambda_n$ to construct D. By equations (1)–(3), $AP = PD$. This is true without any condition on the eigenvectors. If, in fact, the eigenvectors are linearly independent, then P is invertible (by the Invertible Matrix Theorem), and $AP = PD$ implies that $A = PDP^{-1}$. ■

Diagonalizing Matrices

EXAMPLE 3 Diagonalize the following matrix, if possible.

$$A = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix}$$

That is, find an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$.

SOLUTION There are four steps to implement the description in Theorem 5.

Step 1. Find the eigenvalues of A. As mentioned in Section 5.2, the mechanics of this step are appropriate for a computer when the matrix is larger than 2×2. To avoid unnecessary distractions, the text will usually supply information needed for this step. In the present case, the characteristic equation turns out to involve a cubic polynomial that can be factored:

$$\begin{aligned} 0 = \det(A - \lambda I) &= -\lambda^3 - 3\lambda^2 + 4 \\ &= -(\lambda - 1)(\lambda + 2)^2 \end{aligned}$$

The eigenvalues are $\lambda = 1$ and $\lambda = -2$.

Step 2. Find three linearly independent eigenvectors of A. *Three* vectors are needed because A is a 3×3 matrix. This is the critical step. If it fails, then Theorem 5 says that A cannot be diagonalized. The method in Section 5.1 produces a basis for each eigenspace:

$$\text{Basis for } \lambda = 1: \quad \mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

$$\text{Basis for } \lambda = -2: \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

You can check that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly independent set.

Step 3. Construct P from the vectors in step 2. The vectors may be listed in any order. Using the order chosen in step 2, form

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Step 4. Construct D from the corresponding eigenvalues. In this step, it is essential that the order of the eigenvalues matches the order chosen for the columns of P. Use the eigenvalue $\lambda = -2$ twice, once for each of the eigenvectors corresponding to $\lambda = -2$:

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

It is a good idea to check that P and D really work. To avoid computing P^{-1}, simply verify that $AP = PD$. This is equivalent to $A = PDP^{-1}$ when P is invertible. (However, be sure that P is invertible!) Compute

$$AP = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 \\ -1 & -2 & 0 \\ 1 & 0 & -2 \end{bmatrix}$$

$$PD = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 \\ -1 & -2 & 0 \\ 1 & 0 & -2 \end{bmatrix} \quad ■$$

EXAMPLE 4 Diagonalize the following matrix, if possible.

$$A = \begin{bmatrix} 2 & 4 & 3 \\ -4 & -6 & -3 \\ 3 & 3 & 1 \end{bmatrix}$$

SOLUTION The characteristic equation of A turns out to be exactly the same as that in Example 3:

$$0 = \det(A - \lambda I) = -\lambda^3 - 3\lambda^2 + 4 = -(\lambda - 1)(\lambda + 2)^2$$

The eigenvalues are $\lambda = 1$ and $\lambda = -2$. However, it is easy to verify that each eigenspace is only one-dimensional:

$$\text{Basis for } \lambda = 1: \qquad \mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

$$\text{Basis for } \lambda = -2: \qquad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

There are no other eigenvalues, and every eigenvector of A is a multiple of either $\mathbf{v}_1$ or $\mathbf{v}_2$. Hence it is impossible to construct a basis of $\mathbb{R}^3$ using eigenvectors of A. By Theorem 5, A is *not* diagonalizable. ■

The following theorem provides a *sufficient* condition for a matrix to be diagonalizable.

THEOREM 6 An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

PROOF Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be eigenvectors corresponding to the n distinct eigenvalues of a matrix A. Then $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is linearly independent, by Theorem 2 in Section 5.1. Hence A is diagonalizable, by Theorem 5. ■

It is not *necessary* for an $n \times n$ matrix to have n distinct eigenvalues in order to be diagonalizable. The 3×3 matrix in Example 3 is diagonalizable even though it has only two distinct eigenvalues.

EXAMPLE 5 Determine if the following matrix is diagonalizable.

$$A = \begin{bmatrix} 5 & -8 & 1 \\ 0 & 0 & 7 \\ 0 & 0 & -2 \end{bmatrix}$$

SOLUTION This is easy! Since the matrix is triangular, its eigenvalues are obviously 5, 0, and -2. Since A is a 3×3 matrix with three distinct eigenvalues, A is diagonalizable. ■

Matrices Whose Eigenvalues Are Not Distinct

If an $n \times n$ matrix A has n distinct eigenvalues, with corresponding eigenvectors $\mathbf{v}_1, \dots,$ $\mathbf{v}_n$, and if $P = [\,\mathbf{v}_1 \;\; \cdots \;\; \mathbf{v}_n\,]$, then P is automatically invertible because its columns are linearly independent, by Theorem 2. When A is diagonalizable but has fewer than n distinct eigenvalues, it is still possible to build P in a way that makes P automatically invertible, as the next theorem shows.[1]

THEOREM 7

Let A be an $n \times n$ matrix whose distinct eigenvalues are $\lambda_1, \dots, \lambda_p$.

a. For $1 \le k \le p$, the dimension of the eigenspace for λ_k is less than or equal to the multiplicity of the eigenvalue λ_k.
b. The matrix A is diagonalizable if and only if the sum of the dimensions of the eigenspaces equals n, and this happens if and only if (*i*) the characteristic polynomial factors completely into linear factors and (*ii*) the dimension of the eigenspace for each λ_k equals the multiplicity of λ_k.
c. If A is diagonalizable and $\mathcal{B}_k$ is a basis for the eigenspace corresponding to λ_k for each k, then the total collection of vectors in the sets $\mathcal{B}_1, \dots, \mathcal{B}_p$ forms an eigenvector basis for $\mathbb{R}^n$.

EXAMPLE 6 Diagonalize the following matrix, if possible.

$$A = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 1 & 4 & -3 & 0 \\ -1 & -2 & 0 & -3 \end{bmatrix}$$

SOLUTION Since A is a triangular matrix, the eigenvalues are 5 and -3, each with multiplicity 2. Using the method in Section 5.1, we find a basis for each eigenspace.

$$\text{Basis for } \lambda = 5: \quad \mathbf{v}_1 = \begin{bmatrix} -8 \\ 4 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} -16 \\ 4 \\ 0 \\ 1 \end{bmatrix}$$

$$\text{Basis for } \lambda = -3: \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

The set $\{\mathbf{v}_1, \dots, \mathbf{v}_4\}$ is linearly independent, by Theorem 7. So the matrix $P = [\,\mathbf{v}_1 \ \cdots \ \mathbf{v}_4\,]$ is invertible, and $A = PDP^{-1}$, where

$$P = \begin{bmatrix} -8 & -16 & 0 & 0 \\ 4 & 4 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & -3 \end{bmatrix}$$

■

[1] The proof of Theorem 7 is somewhat lengthy but not difficult. For instance, see S. Friedberg, A. Insel, and L. Spence, *Linear Algebra*, 4th ed. (Englewood Cliffs, NJ: Prentice-Hall, 2002), Section 5.2.

Practice Problems

1. Compute A^8, where $A = \begin{bmatrix} 4 & -3 \\ 2 & -1 \end{bmatrix}$.
2. Let $A = \begin{bmatrix} -3 & 12 \\ -2 & 7 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, and $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Suppose you are told that $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors of A. Use this information to diagonalize A.
3. Let A be a 4×4 matrix with eigenvalues 5, 3, and -2, and suppose you know that the eigenspace for $\lambda = 3$ is two-dimensional. Do you have enough information to determine if A is diagonalizable?

5.3 Exercises

In Exercises 1 and 2, let $A = PDP^{-1}$ and compute A^4.

1. $P = \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$, $D = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$

2. $P = \begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}$, $D = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

In Exercises 3 and 4, use the factorization $A = PDP^{-1}$ to compute A^k, where k represents an arbitrary positive integer.

3. $\begin{bmatrix} a & 0 \\ 3(a-b) & b \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}\begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}$

4. $\begin{bmatrix} 15 & -36 \\ 6 & -15 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} -3 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} -1 & 3 \\ 1 & -2 \end{bmatrix}$

In Exercises 5 and 6, the matrix A is factored in the form PDP^{-1}. Use the Diagonalization Theorem to find the eigenvalues of A and a basis for each eigenspace.

5. $\begin{bmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix} =$
$\begin{bmatrix} 1 & 1 & 2 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \end{bmatrix}\begin{bmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1/4 & 1/2 & -3/4 \\ 1/4 & -1/2 & 1/4 \end{bmatrix}$

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

Diagonalize the matrices in Exercises 7–20, if possible. The eigenvalues for Exercises 11–16 are as follows: (11) $\lambda = 1, 2, 3$; (12) $\lambda = 1, 4$; (13) $\lambda = 5, 1$; (14) $\lambda = 3, 4$; (15) $\lambda = 3, 1$; (16) $\lambda = 2, 1$. For Exercise 18, one eigenvalue is $\lambda = 5$ and one eigenvector is $(-2, 1, 2)$.

7. $\begin{bmatrix} 1 & 0 \\ 6 & -1 \end{bmatrix}$

8. $\begin{bmatrix} 5 & 1 \\ 0 & 5 \end{bmatrix}$

9. $\begin{bmatrix} 3 & -1 \\ 1 & 5 \end{bmatrix}$

10. $\begin{bmatrix} 3 & 6 \\ 4 & 1 \end{bmatrix}$

11. $\begin{bmatrix} -1 & 4 & -2 \\ -3 & 4 & 0 \\ -3 & 1 & 3 \end{bmatrix}$

12. $\begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix}$

13. $\begin{bmatrix} 2 & 2 & -1 \\ 1 & 3 & -1 \\ -1 & -2 & 2 \end{bmatrix}$

14. $\begin{bmatrix} 4 & 0 & 2 \\ 2 & 3 & 4 \\ 0 & 0 & 3 \end{bmatrix}$

15. $\begin{bmatrix} -7 & 24 & -16 \\ -2 & 7 & -4 \\ 2 & -6 & 5 \end{bmatrix}$

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

17. $\begin{bmatrix} 4 & 0 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix}$

18. $\begin{bmatrix} -7 & -16 & 4 \\ 6 & 13 & -2 \\ 12 & 16 & 1 \end{bmatrix}$

19. $\begin{bmatrix} 5 & -3 & 0 & 9 \\ 0 & 3 & 1 & -2 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$

20. $\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 2 \end{bmatrix}$

In Exercises 21–28, A, P, and D are $n \times n$ matrices. Mark each statement True or False **(T/F)**. Justify each answer. (Study Theorems 5 and 6 and the examples in this section carefully before you try these exercises.)

21. **(T/F)** A is diagonalizable if $A = PDP^{-1}$ for some matrix D and some invertible matrix P.

22. **(T/F)** If $\mathbb{R}^n$ has a basis of eigenvectors of A, then A is diagonalizable.

23. **(T/F)** A is diagonalizable if and only if A has n eigenvalues, counting multiplicities.

24. **(T/F)** If A is diagonalizable, then A is invertible.

25. **(T/F)** A is diagonalizable if A has n eigenvectors.

26. **(T/F)** If A is diagonalizable, then A has n distinct eigenvalues.

27. **(T/F)** If $AP = PD$, with D diagonal, then the nonzero columns of P must be eigenvectors of A.

28. **(T/F)** If A is invertible, then A is diagonalizable.

29. A is a 5×5 matrix with two eigenvalues. One eigenspace is three-dimensional, and the other eigenspace is two-dimensional. Is A diagonalizable? Why?

30. A is a 3×3 matrix with two eigenvalues. Each eigenspace is one-dimensional. Is A diagonalizable? Why?

31. A is a 4×4 matrix with three eigenvalues. One eigenspace is one-dimensional, and one of the other eigenspaces is two-dimensional. Is it possible that A is *not* diagonalizable? Justify your answer.

32. A is a 7×7 matrix with three eigenvalues. One eigenspace is two-dimensional, and one of the other eigenspaces is three-dimensional. Is it possible that A is *not* diagonalizable? Justify your answer.

33. Show that if A is both diagonalizable and invertible, then so is A^{-1}.

34. Show that if A has n linearly independent eigenvectors, then so does A^T. [*Hint:* Use the Diagonalization Theorem.]

35. A factorization $A = PDP^{-1}$ is not unique. Demonstrate this for the matrix A in Example 2. With $D_1 = \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}$, use the information in Example 2 to find a matrix P_1 such that $A = P_1 D_1 P_1^{-1}$.

36. With A and D as in Example 2, find an invertible P_2 unequal to the P in Example 2, such that $A = P_2 D P_2^{-1}$.

37. Construct a nonzero 2×2 matrix that is invertible but not diagonalizable.

38. Construct a nondiagonal 2×2 matrix that is diagonalizable but not invertible.

Diagonalize the matrices in Exercises 39–42. Use your matrix program's eigenvalue command to find the eigenvalues, and then compute bases for the eigenspaces as in Section 5.1.

T **39.** $\begin{bmatrix} -6 & 4 & 0 & 9 \\ -3 & 0 & 1 & 6 \\ -1 & -2 & 1 & 0 \\ -4 & 4 & 0 & 7 \end{bmatrix}$

T **40.** $\begin{bmatrix} 0 & 13 & 8 & 4 \\ 4 & 9 & 8 & 4 \\ 8 & 6 & 12 & 8 \\ 0 & 5 & 0 & -4 \end{bmatrix}$

T **41.** $\begin{bmatrix} 11 & -6 & 4 & -10 & -4 \\ -3 & 5 & -2 & 4 & 1 \\ -8 & 12 & -3 & 12 & 4 \\ 1 & 6 & -2 & 3 & -1 \\ 8 & -18 & 8 & -14 & -1 \end{bmatrix}$

T **42.** $\begin{bmatrix} 4 & 4 & 2 & 3 & -2 \\ 0 & 1 & -2 & -2 & 2 \\ 6 & 12 & 11 & 2 & -4 \\ 9 & 20 & 10 & 10 & -6 \\ 15 & 28 & 14 & 5 & -3 \end{bmatrix}$

Solutions to Practice Problems

1. $\det(A - \lambda I) = \lambda^2 - 3\lambda + 2 = (\lambda - 2)(\lambda - 1)$. The eigenvalues are 2 and 1, and the corresponding eigenvectors are $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Next, form

$$P = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad \text{and} \quad P^{-1} = \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix}$$

Since $A = PDP^{-1}$,

$$\begin{aligned} A^8 = PD^8P^{-1} &= \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2^8 & 0 \\ 0 & 1^8 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 256 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 766 & -765 \\ 510 & -509 \end{bmatrix} \end{aligned}$$

2. Compute $A\mathbf{v}_1 = \begin{bmatrix} -3 & 12 \\ -2 & 7 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 1 \cdot \mathbf{v}_1$, and

$$A\mathbf{v}_2 = \begin{bmatrix} -3 & 12 \\ -2 & 7 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix} = 3 \cdot \mathbf{v}_2$$

So, $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors for the eigenvalues 1 and 3, respectively. Thus

$$A = PDP^{-1}, \quad \text{where} \quad P = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$$

3. Yes, A is diagonalizable. There is a basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ for the eigenspace corresponding to $\lambda = 3$. In addition, there will be at least one eigenvector for $\lambda = 5$ and one for $\lambda = -2$. Call them $\mathbf{v}_3$ and $\mathbf{v}_4$. Then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is linearly independent by Theorem 2 and Practice Problem 3 in Section 5.1. There can be no additional eigenvectors that are linearly independent from $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$, because the vectors are all in $\mathbb{R}^4$. Hence the eigenspaces for $\lambda = 5$ and $\lambda = -2$ are both one-dimensional. It follows that A is diagonalizable by Theorem 7(b).

STUDY GUIDE has advice on mastering eigenvalues and eigenspaces.

5.4 Eigenvectors and Linear Transformations

In this section, we will look at eigenvalues and eigenvectors of linear transformations $T : V \to V$, where V is any vector space. In the case where V is a finite dimensional vector space and there is a basis for V consisting of eigenvectors of T, we will see how to represent the transformation T as left multiplication by a diagonal matrix.

Eigenvectors of Linear Transformations

Previously, we looked at a variety of vector spaces including the discrete-time signal space, $\mathbb{S}$, and the set of polynomials, $\mathbb{P}$. Eigenvalues and eigenvectors can be defined for linear transformations from any vector space to itself.

DEFINITION

Let V be a vector space. An **eigenvector** of a linear transformation $T : V \to V$ is a nonzero vector $\mathbf{x}$ in V such that $T(\mathbf{x}) = \lambda\mathbf{x}$ for some scalar λ. A scalar λ is called an **eigenvalue** of T if there is a nontrivial solution $\mathbf{x}$ of $T(\mathbf{x}) = \lambda\mathbf{x}$; such an $\mathbf{x}$ is called an **eigenvector** corresponding to λ.

EXAMPLE 1 The sinusoidal signals were studied in detail in Sections 4.7 and 4.8. Consider the signal defined by $\{s_k\} = \left\{\cos\left(\frac{k\pi}{2}\right)\right\}$, where k ranges over all integers. The left double-shift linear transformation D is defined by $D(\{x_k\}) = \{x_{k+2}\}$. Show that $\{s_k\}$ is an eigenvector of D and determine the associated eigenvalue.

SOLUTION The trigonometric formula $\cos(\theta + \pi) = -\cos(\theta)$ is useful here. Set $\{y_k\} = D(\{s_k\})$ and observe that

$$y_k = s_{k+2} = \cos\left(\frac{(k+2)\pi}{2}\right) = \cos\left(\frac{k\pi}{2} + \pi\right) = -\cos\left(\frac{k\pi}{2}\right) = -s_k$$

and so $D(\{s_k\}) = \{-s_k\} = -\{s_k\}$. This establishes that $\{s_k\}$ is an eigenvector of D with eigenvalue -1. ■

In Figure 1, different values for the frequency, f, are chosen to graph a section of the sinusoidal signals $\left\{\cos\left(\frac{fk\pi}{4}\right)\right\}$ and $D\left(\left\{\cos\left(\frac{fk\pi}{4}\right)\right\}\right)$. Setting $f = 2$ illustrates the eigenvector for D established in Example 1. What is the relationship in the patterns of the dots that signifies an eigenvector relationship between the original signal and the transformed signal? Which other choices of the frequency, f, create a signal that is an eigenvector for D? What are the associated eigenvalues? In Figure 1, the graph on the left illustrates the sinusoidal signal with $f = 1$ and the graph on the right illustrates the sinusoidal signal with $f = 2$.

5.7 Applications to Differential Equations

This section describes continuous analogues of the difference equations studied in Section 5.6. In many applied problems, several quantities are varying continuously in time, and they are related by a system of differential equations:

$$\begin{aligned} x_1' &= a_{11}x_1 + \cdots + a_{1n}x_n \\ x_2' &= a_{21}x_1 + \cdots + a_{2n}x_n \\ &\vdots \\ x_n' &= a_{n1}x_1 + \cdots + a_{nn}x_n \end{aligned}$$

Here $x_1, \dots, x_n$ are differentiable functions of t, with derivatives $x_1', \dots, x_n'$, and the a_{ij} are constants. The crucial feature of this system is that it is *linear*. To see this, write the system as a matrix differential equation

$$\mathbf{x}'(t) = A\mathbf{x}(t) \tag{1}$$

where

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \quad \mathbf{x}'(t) = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix}, \quad \text{and} \quad A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

A **solution** of equation (1) is a vector-valued function that satisfies (1) for all t in some interval of real numbers, such as $t \geq 0$.

Equation (1) is *linear* because both differentiation of functions and multiplication of vectors by a matrix are linear transformations. Thus, if $\mathbf{u}$ and $\mathbf{v}$ are solutions of $\mathbf{x}' = A\mathbf{x}$, then $c\mathbf{u} + d\mathbf{v}$ is also a solution, because

$$\begin{aligned} (c\mathbf{u} + d\mathbf{v})' &= c\mathbf{u}' + d\mathbf{v}' \\ &= cA\mathbf{u} + dA\mathbf{v} = A(c\mathbf{u} + d\mathbf{v}) \end{aligned}$$

(Engineers call this property *superposition* of solutions.) Also, the identically zero function is a (trivial) solution of (1). In the terminology of Chapter 4, the set of all solutions of (1) is a *subspace* of the set of all continuous functions with values in $\mathbb{R}^n$.

Standard texts on differential equations show that there always exists what is called a **fundamental set of solutions** to (1). If A is $n \times n$, then there are n linearly independent functions in a fundamental set, and each solution of (1) is a unique linear combination of these n functions. That is, a fundamental set of solutions is a *basis* for the set of all solutions of (1), and the solution set is an n-dimensional vector space of functions. If a vector $\mathbf{x}_0$ is specified, then the **initial value problem** is to construct the (unique) function $\mathbf{x}$ such that $\mathbf{x}' = A\mathbf{x}$ and $\mathbf{x}(0) = \mathbf{x}_0$.

When A is a diagonal matrix, the solutions of (1) can be produced by elementary calculus. For instance, consider

$$\begin{bmatrix} x_1'(t) \\ x_2'(t) \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & -5 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \tag{2}$$

that is,

$$\begin{aligned} x_1'(t) &= 3x_1(t) \\ x_2'(t) &= -5x_2(t) \end{aligned} \tag{3}$$

The system (2) is said to be *decoupled* because each derivative of a function depends only on the function itself, not on some combination or "coupling" of both $x_1(t)$ and $x_2(t)$. From calculus, the solutions of (3) are $x_1(t) = c_1e^{3t}$ and $x_2(t) = c_2e^{-5t}$, for any constants c_1 and c_2. Each solution of equation (2) can be written in the form

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} c_1e^{3t} \\ c_2e^{-5t} \end{bmatrix} = c_1\begin{bmatrix} 1 \\ 0 \end{bmatrix}e^{3t} + c_2\begin{bmatrix} 0 \\ 1 \end{bmatrix}e^{-5t}$$

This example suggests that for the general equation $\mathbf{x}' = A\mathbf{x}$, a solution might be a linear combination of functions of the form

$$\mathbf{x}(t) = \mathbf{v}e^{\lambda t} \tag{4}$$

for some scalar λ and some fixed nonzero vector $\mathbf{v}$. [If $\mathbf{v} = \mathbf{0}$, the function $\mathbf{x}(t)$ is identically zero and hence satisfies $\mathbf{x}' = A\mathbf{x}$.] Observe that

$$\begin{aligned} \mathbf{x}'(t) &= \lambda\mathbf{v}e^{\lambda t} && \text{By calculus, since } \mathbf{v} \text{ is a constant vector} \\ A\mathbf{x}(t) &= A\mathbf{v}e^{\lambda t} && \text{Multiplying both sides of (4) by } A \end{aligned}$$

Since $e^{\lambda t}$ is never zero, $\mathbf{x}'(t)$ will equal $A\mathbf{x}(t)$ if and only if $\lambda\mathbf{v} = A\mathbf{v}$, that is, if and only if λ is an eigenvalue of A and $\mathbf{v}$ is a corresponding eigenvector. Thus each eigenvalue–eigenvector pair provides a solution (4) of $\mathbf{x}' = A\mathbf{x}$. Such solutions are sometimes called *eigenfunctions* of the differential equation. Eigenfunctions provide the key to solving systems of differential equations.

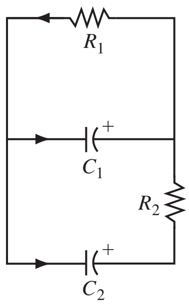


FIGURE 1

EXAMPLE 1 The circuit in Figure 1 can be described by the differential equation

$$\begin{bmatrix} x_1'(t) \\ x_2'(t) \end{bmatrix} = \begin{bmatrix} -(1/R_1 + 1/R_2)/C_1 & 1/(R_2C_1) \\ 1/(R_2C_2) & -1/(R_2C_2) \end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

where $x_1(t)$ and $x_2(t)$ are the voltages across the two capacitors at time t. Suppose resistor R_1 is 1 ohm, R_2 is 2 ohms, capacitor C_1 is 1 farad, and C_2 is .5 farad, and suppose there is an initial charge of 5 volts on capacitor C_1 and 4 volts on capacitor C_2. Find formulas for $x_1(t)$ and $x_2(t)$ that describe how the voltages change over time.

SOLUTION Let A denote the matrix displayed above, and let $\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$. For the data given, $A = \begin{bmatrix} -1.5 & .5 \\ 1 & -1 \end{bmatrix}$, and $\mathbf{x}(0) = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$. The eigenvalues of A are $\lambda_1 = -.5$ and $\lambda_2 = -2$, with corresponding eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

The eigenfunctions $\mathbf{x}_1(t) = \mathbf{v}_1e^{\lambda_1 t}$ and $\mathbf{x}_2(t) = \mathbf{v}_2e^{\lambda_2 t}$ both satisfy $\mathbf{x}' = A\mathbf{x}$, and so does any linear combination of $\mathbf{x}_1$ and $\mathbf{x}_2$. Set

$$\mathbf{x}(t) = c_1\mathbf{v}_1e^{\lambda_1 t} + c_2\mathbf{v}_2e^{\lambda_2 t} = c_1\begin{bmatrix} 1 \\ 2 \end{bmatrix}e^{-.5t} + c_2\begin{bmatrix} -1 \\ 1 \end{bmatrix}e^{-2t}$$

and note that $\mathbf{x}(0) = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$. Since $\mathbf{v}_1$ and $\mathbf{v}_2$ are obviously linearly independent and hence span $\mathbb{R}^2$, c_1 and c_2 can be found to make $\mathbf{x}(0)$ equal to $\mathbf{x}_0$. In fact, the equation

$$c_1\underset{\mathbf{v}_1}{\underset{\uparrow}{\begin{bmatrix} 1 \\ 2 \end{bmatrix}}} + c_2\underset{\mathbf{v}_2}{\underset{\uparrow}{\begin{bmatrix} -1 \\ 1 \end{bmatrix}}} = \underset{\mathbf{x}_0}{\underset{\uparrow}{\begin{bmatrix} 5 \\ 4 \end{bmatrix}}}$$

leads easily to $c_1 = 3$ and $c_2 = -2$. Thus the desired solution of the differential equation $\mathbf{x}' = A\mathbf{x}$ is

$$\mathbf{x}(t) = 3\begin{bmatrix} 1 \\ 2 \end{bmatrix} e^{-.5t} - 2\begin{bmatrix} -1 \\ 1 \end{bmatrix} e^{-2t}$$

or

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 3e^{-.5t} + 2e^{-2t} \\ 6e^{-.5t} - 2e^{-2t} \end{bmatrix}$$

Figure 2 shows the graph, or *trajectory*, of $\mathbf{x}(t)$, for $t \geq 0$, along with trajectories for some other initial points. The trajectories of the two eigenfunctions $\mathbf{x}_1$ and $\mathbf{x}_2$ lie in the eigenspaces of A.

The functions $\mathbf{x}_1$ and $\mathbf{x}_2$ both decay to zero as $t \to \infty$, but the values of $\mathbf{x}_2$ decay faster because its exponent is more negative. The entries in the corresponding eigenvector $\mathbf{v}_2$ show that the voltages across the capacitors will decay to zero as rapidly as possible if the initial voltages are equal in magnitude but opposite in sign. ■

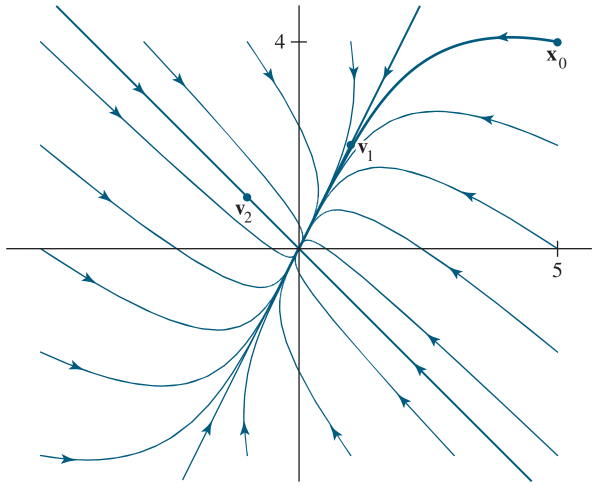


FIGURE 2 The origin as an attractor.

In Figure 2, the origin is called an **attractor**, or **sink**, of the dynamical system because all trajectories are drawn into the origin. The direction of greatest attraction is along the trajectory of the eigenfunction $\mathbf{x}_2$ (along the line through $\mathbf{0}$ and $\mathbf{v}_2$) corresponding to the more negative eigenvalue, $\lambda = -2$. Trajectories that begin at points not on this line become asymptotic to the line through $\mathbf{0}$ and $\mathbf{v}_1$ because their components in the $\mathbf{v}_2$ direction decay so rapidly.

If the eigenvalues in Example 1 were positive instead of negative, the corresponding trajectories would be similar in shape, but the trajectories would be traversed *away* from the origin. In such a case, the origin is called a **repeller**, or **source**, of the dynamical system, and the direction of greatest repulsion is the line containing the trajectory of the eigenfunction corresponding to the more positive eigenvalue.

EXAMPLE 2 Suppose a particle is moving in a planar force field and its position vector $\mathbf{x}$ satisfies $\mathbf{x}' = A\mathbf{x}$ and $\mathbf{x}(0) = \mathbf{x}_0$, where

$$A = \begin{bmatrix} 4 & -5 \\ -2 & 1 \end{bmatrix}, \qquad \mathbf{x}_0 = \begin{bmatrix} 2.9 \\ 2.6 \end{bmatrix}$$

Solve this initial value problem for $t \geq 0$, and sketch the trajectory of the particle.

SOLUTION The eigenvalues of A turn out to be $\lambda_1 = 6$ and $\lambda_2 = -1$, with corresponding eigenvectors $\mathbf{v}_1 = (-5, 2)$ and $\mathbf{v}_2 = (1, 1)$. For any constants c_1 and c_2, the function

$$\mathbf{x}(t) = c_1\mathbf{v}_1 e^{\lambda_1 t} + c_2\mathbf{v}_2 e^{\lambda_2 t} = c_1\begin{bmatrix} -5 \\ 2 \end{bmatrix} e^{6t} + c_2\begin{bmatrix} 1 \\ 1 \end{bmatrix} e^{-t}$$

is a solution of $\mathbf{x}' = A\mathbf{x}$. We want c_1 and c_2 to satisfy $\mathbf{x}(0) = \mathbf{x}_0$, that is,

$$c_1\begin{bmatrix} -5 \\ 2 \end{bmatrix} + c_2\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2.9 \\ 2.6 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} -5 & 1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 2.9 \\ 2.6 \end{bmatrix}$$

Calculations show that $c_1 = -3/70$ and $c_2 = 188/70$, and so the desired function is

$$\mathbf{x}(t) = \frac{-3}{70}\begin{bmatrix} -5 \\ 2 \end{bmatrix} e^{6t} + \frac{188}{70}\begin{bmatrix} 1 \\ 1 \end{bmatrix} e^{-t}$$

Trajectories of $\mathbf{x}$ and other solutions are shown in Figure 3. ■

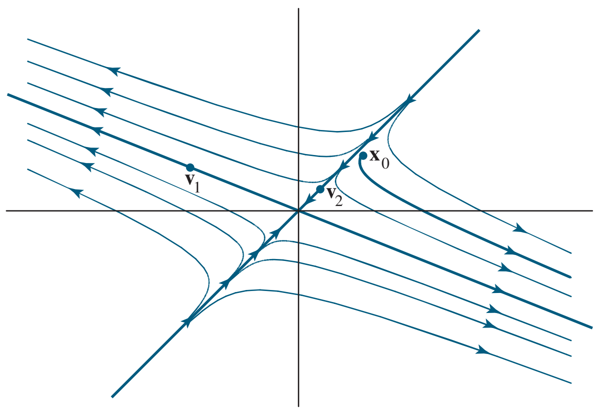


FIGURE 3 The origin as a saddle point.

In Figure 3, the origin is called a **saddle point** of the dynamical system because some trajectories approach the origin at first and then change direction and move away from the origin. A saddle point arises whenever the matrix A has both positive and negative eigenvalues. The direction of greatest repulsion is the line through $\mathbf{v}_1$ and $\mathbf{0}$, corresponding to the positive eigenvalue. The direction of greatest attraction is the line through $\mathbf{v}_2$ and $\mathbf{0}$, corresponding to the negative eigenvalue.

Decoupling a Dynamical System

The following discussion shows that the method of Examples 1 and 2 produces a fundamental set of solutions for any dynamical system described by $\mathbf{x}' = A\mathbf{x}$ when A is $n \times n$ and has n linearly independent eigenvectors, that is, when A is diagonalizable. Suppose the eigenfunctions for A are

$$\mathbf{v}_1 e^{\lambda_1 t}, \ldots, \mathbf{v}_n e^{\lambda_n t}$$

with $\mathbf{v}_1, \ldots, \mathbf{v}_n$ linearly independent eigenvectors. Let $P = [\,\mathbf{v}_1 \;\; \cdots \;\; \mathbf{v}_n\,]$, and let D be the diagonal matrix with entries $\lambda_1, \ldots, \lambda_n$, so that $A = PDP^{-1}$. Now make a *change of variable*, defining a new function $\mathbf{y}$ by

$$\mathbf{y}(t) = P^{-1}\mathbf{x}(t) \quad \text{or, equivalently,} \quad \mathbf{x}(t) = P\mathbf{y}(t)$$

The equation $\mathbf{x}(t) = P\mathbf{y}(t)$ says that $\mathbf{y}(t)$ is the coordinate vector of $\mathbf{x}(t)$ relative to the eigenvector basis. Substitution of $P\mathbf{y}$ for $\mathbf{x}$ in the equation $\mathbf{x}' = A\mathbf{x}$ gives

$$\frac{d}{dt}(P\mathbf{y}) = A(P\mathbf{y}) = (PDP^{-1})P\mathbf{y} = PD\mathbf{y} \tag{5}$$

Since P is a constant matrix, the left side of (5) is $P\mathbf{y}'$. Left-multiply both sides of (5) by P^{-1} and obtain $\mathbf{y}' = D\mathbf{y}$, or

$$\begin{bmatrix} y_1'(t) \\ y_2'(t) \\ \vdots \\ y_n'(t) \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{bmatrix}$$

The change of variable from $\mathbf{x}$ to $\mathbf{y}$ has *decoupled* the system of differential equations, because the derivative of each scalar function y_k depends only on y_k. (Review the analogous change of variables in Section 5.6.) Since $y_1' = \lambda_1 y_1$, we have $y_1(t) = c_1 e^{\lambda_1 t}$, with similar formulas for $y_2, \ldots, y_n$. Thus

$$\mathbf{y}(t) = \begin{bmatrix} c_1 e^{\lambda_1 t} \\ \vdots \\ c_n e^{\lambda_n t} \end{bmatrix}, \quad \text{where} \quad \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \mathbf{y}(0) = P^{-1}\mathbf{x}(0) = P^{-1}\mathbf{x}_0$$

To obtain the general solution $\mathbf{x}$ of the original system, compute

$$\begin{aligned} \mathbf{x}(t) &= P\mathbf{y}(t) = [\, \mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n \,]\, \mathbf{y}(t) \\ &= c_1 \mathbf{v}_1 e^{\lambda_1 t} + \cdots + c_n \mathbf{v}_n e^{\lambda_n t} \end{aligned}$$

This is the eigenfunction expansion constructed as in Example 1.

Complex Eigenvalues

In the next example, a real matrix A has a pair of complex eigenvalues λ and $\overline{\lambda}$, with associated complex eigenvectors $\mathbf{v}$ and $\overline{\mathbf{v}}$. (Recall from Section 5.5 that for a real matrix, complex eigenvalues and associated eigenvectors come in conjugate pairs.) So two solutions of $\mathbf{x}' = A\mathbf{x}$ are

$$\mathbf{x}_1(t) = \mathbf{v}e^{\lambda t} \quad \text{and} \quad \mathbf{x}_2(t) = \overline{\mathbf{v}}e^{\overline{\lambda}t} \tag{6}$$

It can be shown that $\mathbf{x}_2(t) = \overline{\mathbf{x}_1(t)}$ by using a power series representation for the complex exponential function. Although the complex eigenfunctions $\mathbf{x}_1$ and $\mathbf{x}_2$ are convenient for some calculations (particularly in electrical engineering), real functions are more appropriate for many purposes. Fortunately, the real and imaginary parts of $\mathbf{x}_1$ are (real) solutions of $\mathbf{x}' = A\mathbf{x}$, because they are linear combinations of the solutions in (6):

$$\operatorname{Re}(\mathbf{v}e^{\lambda t}) = \frac{1}{2}[\, \mathbf{x}_1(t) + \overline{\mathbf{x}_1(t)} \,], \qquad \operatorname{Im}(\mathbf{v}e^{\lambda t}) = \frac{1}{2i}[\, \mathbf{x}_1(t) - \overline{\mathbf{x}_1(t)} \,]$$

To understand the nature of $\operatorname{Re}(\mathbf{v}e^{\lambda t})$, recall from calculus that for any number x, the exponential function e^x can be computed from the power series:

$$e^x = 1 + x + \frac{1}{2!}x^2 + \cdots + \frac{1}{n!}x^n + \cdots$$

This series can be used to define $e^{\lambda t}$ when λ is complex:

$$e^{\lambda t} = 1 + (\lambda t) + \frac{1}{2!}(\lambda t)^2 + \cdots + \frac{1}{n!}(\lambda t)^n + \cdots$$

By writing $\lambda = a + bi$ (with a and b real), and using similar power series for the cosine and sine functions, one can show that

$$e^{(a+bi)t} = e^{at}e^{ibt} = e^{at}(\cos bt + i \sin bt) \tag{7}$$

Hence

$$\begin{aligned} \mathbf{v}e^{\lambda t} &= (\operatorname{Re}\mathbf{v} + i \operatorname{Im}\mathbf{v})\,(e^{at})(\cos bt + i \sin bt) \\ &= [\,(\operatorname{Re}\mathbf{v})\cos bt - (\operatorname{Im}\mathbf{v})\sin bt\,]e^{at} \\ &\quad + i[\,(\operatorname{Re}\mathbf{v})\sin bt + (\operatorname{Im}\mathbf{v})\cos bt\,]e^{at} \end{aligned}$$

So two real solutions of $\mathbf{x}' = A\mathbf{x}$ are

$$\begin{aligned} \mathbf{y}_1(t) &= \operatorname{Re}\mathbf{x}_1(t) = [\,(\operatorname{Re}\mathbf{v})\cos bt - (\operatorname{Im}\mathbf{v})\sin bt\,]\,e^{at} \\ \mathbf{y}_2(t) &= \operatorname{Im}\mathbf{x}_1(t) = [\,(\operatorname{Re}\mathbf{v})\sin bt + (\operatorname{Im}\mathbf{v})\cos bt\,]\,e^{at} \end{aligned}$$

It can be shown that $\mathbf{y}_1$ and $\mathbf{y}_2$ are linearly independent functions (when $b \neq 0$).[1]

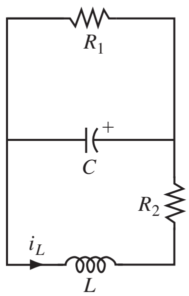


FIGURE 4

EXAMPLE 3 The circuit in Figure 4 can be described by the equation

$$\begin{bmatrix} i_L' \\ v_C' \end{bmatrix} = \begin{bmatrix} -R_2/L & -1/L \\ 1/C & -1/(R_1C) \end{bmatrix} \begin{bmatrix} i_L \\ v_C \end{bmatrix}$$

where i_L is the current passing through the inductor L and v_C is the voltage drop across the capacitor C. Suppose R_1 is 5 ohms, R_2 is .8 ohm, C is .1 farad, and L is .4 henry. Find formulas for i_L and v_C, if the initial current through the inductor is 3 amperes and the initial voltage across the capacitor is 3 volts.

SOLUTION For the data given, $A = \begin{bmatrix} -2 & -2.5 \\ 10 & -2 \end{bmatrix}$ and $\mathbf{x}_0 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$. The method discussed in Section 5.5 produces the eigenvalue $\lambda = -2 + 5i$ and the corresponding eigenvector $\mathbf{v}_1 = \begin{bmatrix} i \\ 2 \end{bmatrix}$. The complex solutions of $\mathbf{x}' = A\mathbf{x}$ are complex linear combinations of

$$\mathbf{x}_1(t) = \begin{bmatrix} i \\ 2 \end{bmatrix} e^{(-2+5i)t} \quad \text{and} \quad \mathbf{x}_2(t) = \begin{bmatrix} -i \\ 2 \end{bmatrix} e^{(-2-5i)t}$$

Next, use equation (7) to write

$$\mathbf{x}_1(t) = \begin{bmatrix} i \\ 2 \end{bmatrix} e^{-2t}(\cos 5t + i \sin 5t)$$

The real and imaginary parts of $\mathbf{x}_1$ provide real solutions:

$$\mathbf{y}_1(t) = \begin{bmatrix} -\sin 5t \\ 2\cos 5t \end{bmatrix} e^{-2t}, \qquad \mathbf{y}_2(t) = \begin{bmatrix} \cos 5t \\ 2\sin 5t \end{bmatrix} e^{-2t}$$

[1] Since $\mathbf{x}_2(t)$ is the complex conjugate of $\mathbf{x}_1(t)$, the real and imaginary parts of $\mathbf{x}_2(t)$ are $\mathbf{y}_1(t)$ and $-\mathbf{y}_2(t)$, respectively. Thus one can use either $\mathbf{x}_1(t)$ or $\mathbf{x}_2(t)$, but not both, to produce two real linearly independent solutions of $\mathbf{x}' = A\mathbf{x}$.

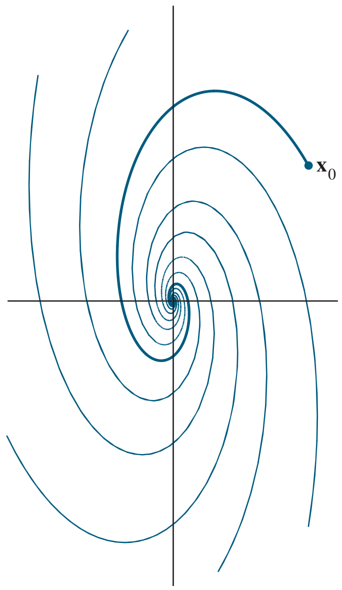


FIGURE 5
The origin as a spiral point.

Since $\mathbf{y}_1$ and $\mathbf{y}_2$ are linearly independent functions, they form a basis for the two-dimensional real vector space of solutions of $\mathbf{x}' = A\mathbf{x}$. Thus the general solution is

$$\mathbf{x}(t) = c_1\begin{bmatrix} -\sin 5t \\ 2\cos 5t \end{bmatrix} e^{-2t} + c_2\begin{bmatrix} \cos 5t \\ 2\sin 5t \end{bmatrix} e^{-2t}$$

To satisfy $\mathbf{x}(0) = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$, we need $c_1\begin{bmatrix} 0 \\ 2 \end{bmatrix} + c_2\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$, which leads to $c_1 = 1.5$ and $c_2 = 3$. Thus

$$\mathbf{x}(t) = 1.5\begin{bmatrix} -\sin 5t \\ 2\cos 5t \end{bmatrix} e^{-2t} + 3\begin{bmatrix} \cos 5t \\ 2\sin 5t \end{bmatrix} e^{-2t}$$

or

$$\begin{bmatrix} i_L(t) \\ v_C(t) \end{bmatrix} = \begin{bmatrix} -1.5\sin 5t + 3\cos 5t \\ 3\cos 5t + 6\sin 5t \end{bmatrix} e^{-2t}$$

See Figure 5. ■

In Figure 5, the origin is called a **spiral point** of the dynamical system. The rotation is caused by the sine and cosine functions that arise from a complex eigenvalue. The trajectories spiral inward because the factor e^{-2t} tends to zero. Recall that -2 is the real part of the eigenvalue in Example 3. When A has a complex eigenvalue with positive real part, the trajectories spiral outward. If the real part of the eigenvalue is zero, the trajectories form ellipses around the origin.

Practice Problems

A real 3×3 matrix A has eigenvalues $-.5$, $.2 + .3i$, and $.2 - .3i$, with corresponding eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1+2i \\ 4i \\ 2 \end{bmatrix}, \quad \text{and} \quad \mathbf{v}_3 = \begin{bmatrix} 1-2i \\ -4i \\ 2 \end{bmatrix}$$

1. Is A diagonalizable as $A = PDP^{-1}$, using complex matrices?
2. Write the general solution of $\mathbf{x}' = A\mathbf{x}$ using complex eigenfunctions, and then find the general real solution.
3. Describe the shapes of typical trajectories.

5.7 Exercises

1. A particle moving in a planar force field has a position vector $\mathbf{x}$ that satisfies $\mathbf{x}' = A\mathbf{x}$. The 2×2 matrix A has eigenvalues 4 and 2, with corresponding eigenvectors $\mathbf{v}_1 = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Find the position of the particle at time t, assuming that $\mathbf{x}(0) = \begin{bmatrix} -6 \\ 1 \end{bmatrix}$.
2. Let A be a 2×2 matrix with eigenvalues -3 and -1 and corresponding eigenvectors $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Let $\mathbf{x}(t)$ be the position of a particle at time t. Solve the initial value problem $\mathbf{x}' = A\mathbf{x}$, $\mathbf{x}(0) = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.

In Exercises 3–6, solve the initial value problem $\mathbf{x}'(t) = A\mathbf{x}(t)$ for $t \geq 0$, with $\mathbf{x}(0) = (3, 2)$. Classify the nature of the origin as an attractor, repeller, or saddle point of the dynamical system described by $\mathbf{x}' = A\mathbf{x}$. Find the directions of greatest attraction and/or repulsion. When the origin is a saddle point, sketch typical trajectories.

3. $A = \begin{bmatrix} 2 & 3 \\ -1 & -2 \end{bmatrix}$
4. $A = \begin{bmatrix} -2 & -5 \\ 1 & 4 \end{bmatrix}$
5. $A = \begin{bmatrix} 2 & -4 \\ 5 & -7 \end{bmatrix}$
6. $A = \begin{bmatrix} 7 & -3 \\ 5 & -1 \end{bmatrix}$

In Exercises 7 and 8, make a change of variable that decouples the equation $\mathbf{x}' = A\mathbf{x}$. Write the equation $\mathbf{x}(t) = P\mathbf{y}(t)$ and show the calculation that leads to the uncoupled system $\mathbf{y}' = D\mathbf{y}$, specifying P and D.

7. A as in Exercise 5 **8.** A as in Exercise 6

In Exercises 9–18, construct the general solution of $\mathbf{x}' = A\mathbf{x}$ involving complex eigenfunctions and then obtain the general real solution. Describe the shapes of typical trajectories.

9. $A = \begin{bmatrix} -3 & 2 \\ -1 & -1 \end{bmatrix}$ **10.** $A = \begin{bmatrix} 3 & 1 \\ -2 & 1 \end{bmatrix}$

11. $A = \begin{bmatrix} -3 & -9 \\ 2 & 3 \end{bmatrix}$ **12.** $A = \begin{bmatrix} -7 & 10 \\ -4 & 5 \end{bmatrix}$

13. $A = \begin{bmatrix} 4 & -3 \\ 6 & -2 \end{bmatrix}$ **14.** $A = \begin{bmatrix} -3 & 2 \\ -9 & 3 \end{bmatrix}$

T **15.** $A = \begin{bmatrix} -8 & -12 & -6 \\ 2 & 1 & 2 \\ 7 & 12 & 5 \end{bmatrix}$

T **16.** $A = \begin{bmatrix} -6 & -11 & 16 \\ 2 & 5 & -4 \\ -4 & -5 & 10 \end{bmatrix}$

T **17.** $A = \begin{bmatrix} 30 & 64 & 23 \\ -11 & -23 & -9 \\ 6 & 15 & 4 \end{bmatrix}$

T **18.** $A = \begin{bmatrix} 53 & -30 & -2 \\ 90 & -52 & -3 \\ 20 & -10 & 2 \end{bmatrix}$

T **19.** Find formulas for the voltages v_1 and v_2 (as functions of time t) for the circuit in Example 1, assuming that $R_1 = 1/5$ ohm, $R_2 = 1/3$ ohm, $C_1 = 4$ farads, $C_2 = 3$ farads, and the initial charge on each capacitor is 4 volts.

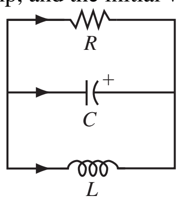
T **20.** Find formulas for the voltages v_1 and v_2 for the circuit in Example 1, assuming that $R_1 = 1/15$ ohm, $R_2 = 1/3$ ohm, $C_1 = 9$ farads, $C_2 = 2$ farads, and the initial charge on each capacitor is 3 volts.

T **21.** Find formulas for the current i_L and the voltage v_C for the circuit in Example 3, assuming that $R_1 = 1$ ohm, $R_2 = .125$ ohm, $C = .2$ farad, $L = .125$ henry, the initial current is 0 amp, and the initial voltage is 15 volts.

T **22.** The circuit in the figure is described by the equation

$$\begin{bmatrix} i_L' \\ v_C' \end{bmatrix} = \begin{bmatrix} 0 & 1/L \\ -1/C & -1/(RC) \end{bmatrix} \begin{bmatrix} i_L \\ v_C \end{bmatrix}$$

where i_L is the current through the inductor L and v_C is the voltage drop across the capacitor C. Find formulas for i_L and v_C when $R = .5$ ohm, $C = 2.5$ farads, $L = .5$ henry, the initial current is 0 amp, and the initial voltage is 12 volts.



Solutions to Practice Problems

1. Yes, the 3×3 matrix is diagonalizable because it has three distinct eigenvalues. Theorem 2 in Section 5.1 and Theorem 6 in Section 5.3 are valid when complex scalars are used. (The proofs are essentially the same as for real scalars.)

2. The general solution has the form

$$\mathbf{x}(t) = c_1 \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} e^{-.5t} + c_2 \begin{bmatrix} 1+2i \\ 4i \\ 2 \end{bmatrix} e^{(.2+.3i)t} + c_3 \begin{bmatrix} 1-2i \\ -4i \\ 2 \end{bmatrix} e^{(.2-.3i)t}$$

The scalars c_1, c_2, and c_3 here can be any complex numbers. The first term in $\mathbf{x}(t)$ is real, provided c_1 is real. Two more real solutions can be produced using the real and imaginary parts of the second term in $\mathbf{x}(t)$:

$$\begin{bmatrix} 1+2i \\ 4i \\ 2 \end{bmatrix} e^{.2t}(\cos .3t + i \sin .3t)$$

The general real solution has the following form, with *real* scalars c_1, c_2, and c_3:

$$c_1 \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} e^{-.5t} + c_2 \begin{bmatrix} \cos .3t - 2\sin .3t \\ -4\sin .3t \\ 2\cos .3t \end{bmatrix} e^{.2t} + c_3 \begin{bmatrix} \sin .3t + 2\cos .3t \\ 4\cos .3t \\ 2\sin .3t \end{bmatrix} e^{.2t}$$

3. Any solution with $c_2 = c_3 = 0$ is attracted to the origin because of the negative exponential factor. Other solutions have components that grow without bound, and the trajectories spiral outward.

Be careful not to mistake this problem for one in Section 5.6. There the condition for attraction toward **0** was that an eigenvalue be less than 1 in magnitude, to make $|\lambda|^k \to 0$. Here the real part of the eigenvalue must be negative, to make $e^{\lambda t} \to 0$.

5.8 Iterative Estimates for Eigenvalues

In scientific applications of linear algebra, eigenvalues are seldom known precisely. Fortunately, a close numerical approximation is usually quite satisfactory. In fact, some applications require only a rough approximation to the largest eigenvalue. The first algorithm described below can work well for this case. Also, it provides a foundation for a more powerful method that can give fast estimates for other eigenvalues as well.

The Power Method

The power method applies to an $n \times n$ matrix A with a **strictly dominant eigenvalue** λ_1, which means that λ_1 must be larger in absolute value than all the other eigenvalues. In this case, the power method produces a scalar sequence that approaches λ_1 and a vector sequence that approaches a corresponding eigenvector. The background for the method rests on the eigenvector decomposition used at the beginning of Section 5.6.

Assume for simplicity that A is diagonalizable and $\mathbb{R}^n$ has a basis of eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$, arranged so their corresponding eigenvalues $\lambda_1, \dots, \lambda_n$ decrease in size, with the strictly dominant eigenvalue first. That is,

$$|\lambda_1| \underbrace{>}_{\text{Strictly larger}} |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n| \tag{1}$$

As we saw in equation (2) of Section 5.6, if $\mathbf{x}$ in $\mathbb{R}^n$ is written as $\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$, then

$$A^k\mathbf{x} = c_1(\lambda_1)^k\mathbf{v}_1 + c_2(\lambda_2)^k\mathbf{v}_2 + \cdots + c_n(\lambda_n)^k\mathbf{v}_n \quad (k = 1, 2, \dots)$$

Assume $c_1 \neq 0$. Then, dividing by $(\lambda_1)^k$,

$$\frac{1}{(\lambda_1)^k}A^k\mathbf{x} = c_1\mathbf{v}_1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^k\mathbf{v}_2 + \cdots + c_n\left(\frac{\lambda_n}{\lambda_1}\right)^k\mathbf{v}_n \quad (k = 1, 2, \dots) \tag{2}$$

From inequality (1), the fractions $\lambda_2/\lambda_1, \dots, \lambda_n/\lambda_1$ are all less than 1 in magnitude and so their powers go to zero. Hence

$$(\lambda_1)^{-k}A^k\mathbf{x} \to c_1\mathbf{v}_1 \quad \text{as } k \to \infty \tag{3}$$

Thus, for large k, a scalar multiple of $A^k\mathbf{x}$ determines almost the same *direction* as the eigenvector $c_1\mathbf{v}_1$. Since positive scalar multiples do not change the direction of a vector, $A^k\mathbf{x}$ itself points almost in the same direction as $\mathbf{v}_1$ or $-\mathbf{v}_1$, provided $c_1 \neq 0$.

EXAMPLE 1 Let $A = \begin{bmatrix} 1.8 & .8 \\ .2 & 1.2 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} -.5 \\ 1 \end{bmatrix}$. Then A has eigenvalues 2 and 1, and the eigenspace for $\lambda_1 = 2$ is the line through **0** and $\mathbf{v}_1$. For $k = 0, \dots, 8$, compute $A^k\mathbf{x}$ and construct the line through **0** and $A^k\mathbf{x}$. What happens as k increases?

In order to find an approximate solution to an inconsistent system of equations that has no actual solution, a well-defined notion of nearness is needed. Section 6.1 introduces the concepts of distance and orthogonality in a vector space. Sections 6.2 and 6.3 show how orthogonality can be used to identify the point within a subspace W that is nearest to a point $\mathbf{y}$ lying outside of W. By taking W to be the column space of a matrix, Section 6.5 develops a method for producing approximate ("least-squares") solutions for inconsistent linear systems, an important technique in machine learning, which is discussed in Sections 6.6 and 6.8.

Section 6.4 provides another opportunity to see orthogonal projections at work, creating a matrix factorization widely used in numerical linear algebra. The remaining sections examine some of the many least-squares problems that arise in applications, including those in vector spaces more general than $\mathbb{R}^n$.

6.1 Inner Product, Length, and Orthogonality

Geometric concepts of length, distance, and perpendicularity, which are well known for $\mathbb{R}^2$ and $\mathbb{R}^3$, are defined here for $\mathbb{R}^n$. These concepts provide powerful geometric tools for solving many applied problems, including the least-squares problems mentioned above. All three notions are defined in terms of the inner product of two vectors.

The Inner Product

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^n$, then we regard $\mathbf{u}$ and $\mathbf{v}$ as $n \times 1$ matrices. The transpose $\mathbf{u}^T$ is a $1 \times n$ matrix, and the matrix product $\mathbf{u}^T\mathbf{v}$ is a 1×1 matrix, which we write as a single real number (a scalar) without brackets. The number $\mathbf{u}^T\mathbf{v}$ is called the **inner product** of $\mathbf{u}$ and $\mathbf{v}$, and often it is written as $\mathbf{u}\cdot\mathbf{v}$. This inner product, mentioned in the exercises for Section 2.1, is also referred to as a **dot product**. If

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

then the inner product of $\mathbf{u}$ and $\mathbf{v}$ is

$$\begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$$

EXAMPLE 1 Compute $\mathbf{u}\cdot\mathbf{v}$ and $\mathbf{v}\cdot\mathbf{u}$ for $\mathbf{u} = \begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ -3 \end{bmatrix}$.

SOLUTION

$$\mathbf{u}\cdot\mathbf{v} = \mathbf{u}^T\mathbf{v} = \begin{bmatrix} 2 & -5 & -1 \end{bmatrix}\begin{bmatrix} 3 \\ 2 \\ -3 \end{bmatrix} = (2)(3) + (-5)(2) + (-1)(-3) = -1$$

$$\mathbf{v}\cdot\mathbf{u} = \mathbf{v}^T\mathbf{u} = \begin{bmatrix} 3 & 2 & -3 \end{bmatrix}\begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix} = (3)(2) + (2)(-5) + (-3)(-1) = -1$$

It is clear from the calculations in Example 1 why $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$. This commutativity of the inner product holds in general. The following properties of the inner product are easily deduced from properties of the transpose operation in Section 2.1. (See Exercises 29 and 30 at the end of this section.)

THEOREM 1

Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{R}^n$, and let c be a scalar. Then

a. $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$

b. $(\mathbf{u}+\mathbf{v})\cdot\mathbf{w} = \mathbf{u}\cdot\mathbf{w} + \mathbf{v}\cdot\mathbf{w}$

c. $(c\mathbf{u})\cdot\mathbf{v} = c(\mathbf{u}\cdot\mathbf{v}) = \mathbf{u}\cdot(c\mathbf{v})$

d. $\mathbf{u}\cdot\mathbf{u} \geq 0$, and $\mathbf{u}\cdot\mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$

Properties (b) and (c) can be combined several times to produce the following useful rule:

$$(c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p)\cdot\mathbf{w} = c_1(\mathbf{u}_1\cdot\mathbf{w}) + \cdots + c_p(\mathbf{u}_p\cdot\mathbf{w})$$

The Length of a Vector

If $\mathbf{v}$ is in $\mathbb{R}^n$, with entries $v_1, \dots, v_n$, then the square root of $\mathbf{v}\cdot\mathbf{v}$ is defined because $\mathbf{v}\cdot\mathbf{v}$ is nonnegative.

DEFINITION

The **length** (or **norm**) of $\mathbf{v}$ is the nonnegative scalar $\|\mathbf{v}\|$ defined by

$$\|\mathbf{v}\| = \sqrt{\mathbf{v}\cdot\mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}, \quad \text{and} \quad \|\mathbf{v}\|^2 = \mathbf{v}\cdot\mathbf{v}$$

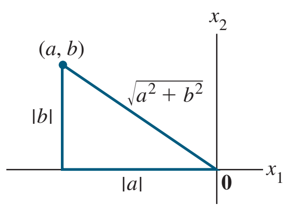


FIGURE 1
Interpretation of $\|\mathbf{v}\|$ as length.

Suppose $\mathbf{v}$ is in $\mathbb{R}^2$, say, $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$. If we identify $\mathbf{v}$ with a geometric point in the plane, as usual, then $\|\mathbf{v}\|$ coincides with the standard notion of the length of the line segment from the origin to $\mathbf{v}$. This follows from the Pythagorean Theorem applied to a triangle such as the one in Figure 1.

A similar calculation with the diagonal of a rectangular box shows that the definition of length of a vector $\mathbf{v}$ in $\mathbb{R}^3$ coincides with the usual notion of length.

For any scalar c, the length of $c\mathbf{v}$ is $|c|$ times the length of $\mathbf{v}$. That is,

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

(To see this, compute $\|c\mathbf{v}\|^2 = (c\mathbf{v})\cdot(c\mathbf{v}) = c^2\mathbf{v}\cdot\mathbf{v} = c^2\|\mathbf{v}\|^2$ and take square roots.)

A vector whose length is 1 is called a **unit vector**. If we *divide* a nonzero vector $\mathbf{v}$ by its length—that is, multiply by $1/\|\mathbf{v}\|$—we obtain a unit vector $\mathbf{u}$ because the length of $\mathbf{u}$ is $(1/\|\mathbf{v}\|)\|\mathbf{v}\|$. The process of creating $\mathbf{u}$ from $\mathbf{v}$ is sometimes called **normalizing** $\mathbf{v}$, and we say that $\mathbf{u}$ is *in the same direction* as $\mathbf{v}$.

Several examples that follow use the space-saving notation for (column) vectors.

EXAMPLE 2 Let $\mathbf{v} = (1, -2, 2, 0)$. Find a unit vector $\mathbf{u}$ in the same direction as $\mathbf{v}$.

SOLUTION First, compute the length of $\mathbf{v}$:

$$\|\mathbf{v}\|^2 = \mathbf{v}\cdot\mathbf{v} = (1)^2 + (-2)^2 + (2)^2 + (0)^2 = 9$$
$$\|\mathbf{v}\| = \sqrt{9} = 3$$

Then, multiply $\mathbf{v}$ by $1/\|\mathbf{v}\|$ to obtain

$$\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v} = \frac{1}{3}\mathbf{v} = \frac{1}{3}\begin{bmatrix} 1 \\ -2 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \\ 0 \end{bmatrix}$$

To check that $\|\mathbf{u}\| = 1$, it suffices to show that $\|\mathbf{u}\|^2 = 1$.

$$\begin{aligned}\|\mathbf{u}\|^2 = \mathbf{u}\cdot\mathbf{u} &= \left(\tfrac{1}{3}\right)^2 + \left(-\tfrac{2}{3}\right)^2 + \left(\tfrac{2}{3}\right)^2 + (0)^2 \\ &= \tfrac{1}{9} + \tfrac{4}{9} + \tfrac{4}{9} + 0 = 1\end{aligned}$$

■

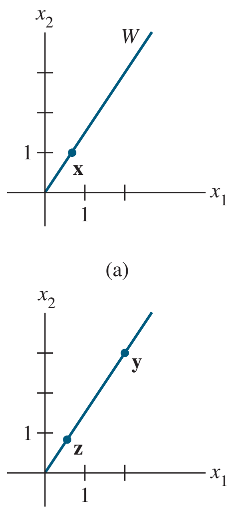


FIGURE 2
Normalizing a vector to produce a unit vector.

EXAMPLE 3 Let W be the subspace of $\mathbb{R}^2$ spanned by $\mathbf{x} = (\frac{2}{3}, 1)$. Find a unit vector $\mathbf{z}$ that is a basis for W.

SOLUTION W consists of all multiples of $\mathbf{x}$, as in Figure 2(a). Any nonzero vector in W is a basis for W. To simplify the calculation, "scale" $\mathbf{x}$ to eliminate fractions. That is, multiply $\mathbf{x}$ by 3 to get

$$\mathbf{y} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

Now compute $\|\mathbf{y}\|^2 = 2^2 + 3^2 = 13$, $\|\mathbf{y}\| = \sqrt{13}$, and normalize $\mathbf{y}$ to get

$$\mathbf{z} = \frac{1}{\sqrt{13}}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{13} \\ 3/\sqrt{13} \end{bmatrix}$$

See Figure 2(b). Another unit vector is $(-2/\sqrt{13}, -3/\sqrt{13})$. ■

Distance in $\mathbb{R}^n$

We are ready now to describe how close one vector is to another. Recall that if a and b are real numbers, the distance on the number line between a and b is the number $|a - b|$. Two examples are shown in Figure 3. This definition of distance in $\mathbb{R}$ has a direct analogue in $\mathbb{R}^n$.

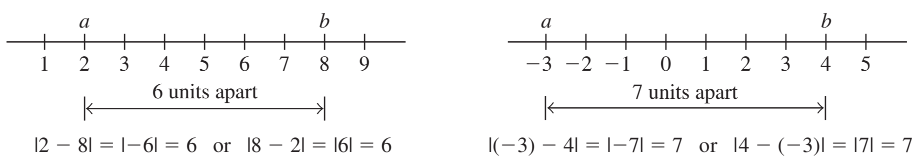


FIGURE 3 Distances in $\mathbb{R}$.

DEFINITION

For $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, the **distance between u and v**, written as $\text{dist}(\mathbf{u}, \mathbf{v})$, is the length of the vector $\mathbf{u} - \mathbf{v}$. That is,

$$\text{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

In $\mathbb{R}^2$ and $\mathbb{R}^3$, this definition of distance coincides with the usual formulas for the Euclidean distance between two points, as the next two examples show.

EXAMPLE 4 Compute the distance between the vectors $\mathbf{u} = (7, 1)$ and $\mathbf{v} = (3, 2)$.

SOLUTION Calculate

$$\mathbf{u} - \mathbf{v} = \begin{bmatrix} 7 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$$

$$\|\mathbf{u} - \mathbf{v}\| = \sqrt{4^2 + (-1)^2} = \sqrt{17}$$

The vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u} - \mathbf{v}$ are shown in Figure 4. When the vector $\mathbf{u} - \mathbf{v}$ is added to $\mathbf{v}$, the result is $\mathbf{u}$. Notice that the parallelogram in Figure 4 shows that the distance from $\mathbf{u}$ to $\mathbf{v}$ is the same as the distance from $\mathbf{u} - \mathbf{v}$ to $\mathbf{0}$. ■

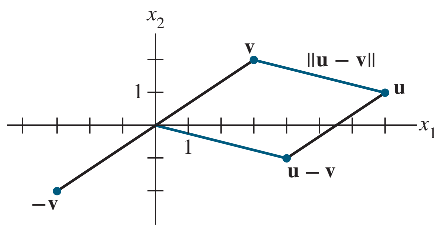


FIGURE 4 The distance between $\mathbf{u}$ and $\mathbf{v}$ is the length of $\mathbf{u} - \mathbf{v}$.

EXAMPLE 5 If $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$, then

$$\begin{aligned} \text{dist}(\mathbf{u}, \mathbf{v}) &= \|\mathbf{u} - \mathbf{v}\| = \sqrt{(\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v})} \\ &= \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2} \end{aligned}$$ ■

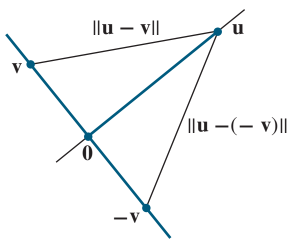


FIGURE 5

Orthogonal Vectors

The rest of this chapter depends on the fact that the concept of perpendicular lines in ordinary Euclidean geometry has an analogue in $\mathbb{R}^n$.

Consider $\mathbb{R}^2$ or $\mathbb{R}^3$ and two lines through the origin determined by vectors $\mathbf{u}$ and $\mathbf{v}$. The two lines shown in Figure 5 are geometrically perpendicular if and only if the distance from $\mathbf{u}$ to $\mathbf{v}$ is the same as the distance from $\mathbf{u}$ to $-\mathbf{v}$. This is the same as requiring the squares of the distances to be the same. Now

$$\begin{aligned} [\,\text{dist}(\mathbf{u}, -\mathbf{v})\,]^2 &= \|\mathbf{u} - (-\mathbf{v})\|^2 = \|\mathbf{u} + \mathbf{v}\|^2 \\ &= (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) \\ &= \mathbf{u} \cdot (\mathbf{u} + \mathbf{v}) + \mathbf{v} \cdot (\mathbf{u} + \mathbf{v}) && \text{Theorem 1(b)} \\ &= \mathbf{u} \cdot \mathbf{u} + \mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v} && \text{Theorem 1(a), (b)} \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\mathbf{u} \cdot \mathbf{v} && \text{Theorem 1(a)} \qquad (1) \end{aligned}$$

The same calculations with $\mathbf{v}$ and $-\mathbf{v}$ interchanged show that

$$\begin{aligned}[\text{dist}(\mathbf{u},\mathbf{v})]^2 &= \|\mathbf{u}\|^2 + \|-\mathbf{v}\|^2 + 2\mathbf{u}\cdot(-\mathbf{v}) \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\mathbf{u}\cdot\mathbf{v}\end{aligned}$$

The two squared distances are equal if and only if $2\mathbf{u}\cdot\mathbf{v} = -2\mathbf{u}\cdot\mathbf{v}$, which happens if and only if $\mathbf{u}\cdot\mathbf{v} = 0$.

This calculation shows that when vectors $\mathbf{u}$ and $\mathbf{v}$ are identified with geometric points, the corresponding lines through the points and the origin are perpendicular if and only if $\mathbf{u}\cdot\mathbf{v} = 0$. The following definition generalizes to $\mathbb{R}^n$ this notion of perpendicularity (or *orthogonality*, as it is commonly called in linear algebra).

DEFINITION

Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are **orthogonal** (to each other) if $\mathbf{u}\cdot\mathbf{v} = 0$.

Observe that the zero vector is orthogonal to every vector in $\mathbb{R}^n$ because $\mathbf{0}^T\mathbf{v} = 0$ for all $\mathbf{v}$.

The next theorem provides a useful fact about orthogonal vectors. The proof follows immediately from the calculation in (1) and the definition of orthogonality. The right triangle shown in Figure 6 provides a visualization of the lengths that appear in the theorem.

THEOREM 2

The Pythagorean Theorem

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

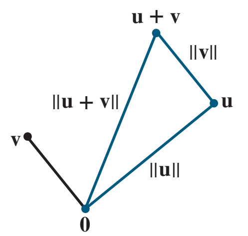


FIGURE 6

Orthogonal Complements

To provide practice using inner products, we introduce a concept here that will be of use in Section 6.3 and elsewhere in the chapter. If a vector $\mathbf{z}$ is orthogonal to every vector in a subspace W of $\mathbb{R}^n$, then $\mathbf{z}$ is said to be **orthogonal to** W. The set of all vectors $\mathbf{z}$ that are orthogonal to W is called the **orthogonal complement** of W and is denoted by $W^\perp$ (and read as "W perpendicular" or simply "W perp").

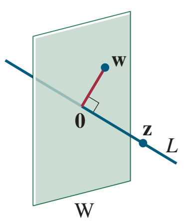


FIGURE 7
A plane and line through $\mathbf{0}$ as orthogonal complements.

EXAMPLE 6 Let W be a plane through the origin in $\mathbb{R}^3$, and let L be the line through the origin and perpendicular to W. If $\mathbf{z}$ and $\mathbf{w}$ are nonzero, $\mathbf{z}$ is on L, and $\mathbf{w}$ is in W, then the line segment from $\mathbf{0}$ to $\mathbf{z}$ is perpendicular to the line segment from $\mathbf{0}$ to $\mathbf{w}$; that is, $\mathbf{z}\cdot\mathbf{w} = 0$. See Figure 7. So each vector on L is orthogonal to every $\mathbf{w}$ in W. In fact, L consists of *all* vectors that are orthogonal to the $\mathbf{w}$'s in W, and W consists of all vectors orthogonal to the $\mathbf{z}$'s in L. That is,

$$L = W^\perp \quad \text{and} \quad W = L^\perp \qquad \blacksquare$$

The following two facts about $W^\perp$, with W a subspace of $\mathbb{R}^n$, are needed later in the chapter. Proofs are suggested in Exercises 37 and 38. Exercises 35–39 provide excellent practice using properties of the inner product.

1. A vector $\mathbf{x}$ is in $W^\perp$ if and only if $\mathbf{x}$ is orthogonal to every vector in a set that spans W.
2. $W^\perp$ is a subspace of $\mathbb{R}^n$.

The next theorem and Exercise 39 verify the claims made in Section 4.5 concerning the subspaces shown in Figure 8.

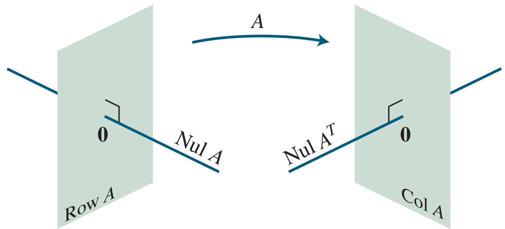


FIGURE 8 The fundamental subspaces determined by an $m \times n$ matrix A.

Remark: A common way to prove that two sets, say S and T, are equal is to show that S is a subset of T and T is a subset of S. The proof of the next theorem that Nul $A =$ (Row $A)^{\perp}$ is established by showing that Nul A is a subset of (Row $A)^{\perp}$ and (Row $A)^{\perp}$ is a subset of Nul A. That is, an arbitrary element $\mathbf{x}$ in Nul A is shown to be in (Row $A)^{\perp}$, and then an arbitrary element $\mathbf{x}$ in (Row $A)^{\perp}$ is shown to be in Nul A.

THEOREM 3

Let A be an $m \times n$ matrix. The orthogonal complement of the row space of A is the null space of A, and the orthogonal complement of the column space of A is the null space of A^T:

$$(\text{Row } A)^{\perp} = \text{Nul } A \quad \text{and} \quad (\text{Col } A)^{\perp} = \text{Nul } A^T$$

PROOF The row–column rule for computing $A\mathbf{x}$ shows that if $\mathbf{x}$ is in Nul A, then $\mathbf{x}$ is orthogonal to each row of A (with the rows treated as vectors in $\mathbb{R}^n$). Since the rows of A span the row space, $\mathbf{x}$ is orthogonal to Row A. Conversely, if $\mathbf{x}$ is orthogonal to Row A, then $\mathbf{x}$ is certainly orthogonal to each row of A, and hence $A\mathbf{x} = \mathbf{0}$. This proves the first statement of the theorem. Since this statement is true for any matrix, it is true for A^T. That is, the orthogonal complement of the row space of A^T is the null space of A^T. This proves the second statement, because Row $A^T =$ Col A.

Angles in $\mathbb{R}^2$ and $\mathbb{R}^3$ (Optional)

If $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors in either $\mathbb{R}^2$ or $\mathbb{R}^3$, then there is a nice connection between their inner product and the angle ϑ between the two line segments from the origin to the points identified with $\mathbf{u}$ and $\mathbf{v}$. The formula is

$$\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|\,\|\mathbf{v}\|\cos\vartheta \tag{2}$$

To verify this formula for vectors in $\mathbb{R}^2$, consider the triangle shown in Figure 9, with sides of lengths $\|\mathbf{u}\|$, $\|\mathbf{v}\|$, and $\|\mathbf{u}-\mathbf{v}\|$. By the law of cosines,

$$\|\mathbf{u}-\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\,\|\mathbf{v}\|\cos\vartheta$$

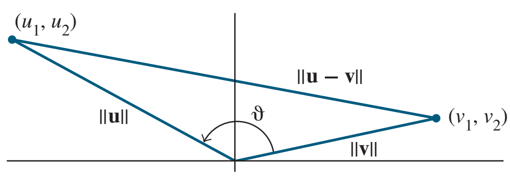


FIGURE 9 The angle between two vectors.

which can be rearranged to produce

$$\begin{aligned}\|\mathbf{u}\|\,\|\mathbf{v}\|\cos\vartheta &= \frac{1}{2}\left[\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - \|\mathbf{u}-\mathbf{v}\|^2\right] \\ &= \frac{1}{2}\left[u_1^2 + u_2^2 + v_1^2 + v_2^2 - (u_1 - v_1)^2 - (u_2 - v_2)^2\right] \\ &= u_1 v_1 + u_2 v_2 \\ &= \mathbf{u}\cdot\mathbf{v}\end{aligned}$$

The verification for $\mathbb{R}^3$ is similar. When $n > 3$, formula (2) may be used to *define* the angle between two vectors in $\mathbb{R}^n$. In statistics, for instance, the value of $\cos\vartheta$ defined by (2) for suitable vectors **u** and **v** is what statisticians call a *correlation coefficient*.

Practice Problems

1. Let $\mathbf{a} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$. Compute $\dfrac{\mathbf{a}\cdot\mathbf{b}}{\mathbf{a}\cdot\mathbf{a}}$ and $\left(\dfrac{\mathbf{a}\cdot\mathbf{b}}{\mathbf{a}\cdot\mathbf{a}}\right)\mathbf{a}$.

2. Let $\mathbf{c} = \begin{bmatrix} 4/3 \\ -1 \\ 2/3 \end{bmatrix}$ and $\mathbf{d} = \begin{bmatrix} 5 \\ 6 \\ -1 \end{bmatrix}$.

a. Find a unit vector **u** in the direction of **c**.

b. Show that **d** is orthogonal to **c**.

c. Use the results of (a) and (b) to explain why **d** must be orthogonal to the unit vector **u**.

3. Let W be a subspace of R^n. Exercise 38 establishes that $W^\perp$ is also a subspace of R^n. Prove that $\dim W + \dim W^\perp = n$.

6.1 Exercises

Compute the quantities in Exercises 1–8 using the vectors

$$\mathbf{u} = \begin{bmatrix} -1 \\ 2 \end{bmatrix},\quad \mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix},\quad \mathbf{w} = \begin{bmatrix} 3 \\ -1 \\ -5 \end{bmatrix},\quad \mathbf{x} = \begin{bmatrix} 6 \\ -2 \\ 3 \end{bmatrix}$$

1. $\mathbf{u}\cdot\mathbf{u}$, $\mathbf{v}\cdot\mathbf{u}$, and $\dfrac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}$

2. $\mathbf{w}\cdot\mathbf{w}$, $\mathbf{x}\cdot\mathbf{w}$, and $\dfrac{\mathbf{x}\cdot\mathbf{w}}{\mathbf{w}\cdot\mathbf{w}}$

3. $\dfrac{1}{\mathbf{w}\cdot\mathbf{w}}\mathbf{w}$

4. $\dfrac{1}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$

5. $\left(\dfrac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v}$

6. $\left(\dfrac{\mathbf{x}\cdot\mathbf{w}}{\mathbf{x}\cdot\mathbf{x}}\right)\mathbf{x}$

7. $\|\mathbf{w}\|$

8. $\|\mathbf{x}\|$

In Exercises 9–12, find a unit vector in the direction of the given vector.

9. $\begin{bmatrix} -30 \\ 40 \end{bmatrix}$

10. $\begin{bmatrix} 3 \\ 6 \\ -3 \end{bmatrix}$

11. $\begin{bmatrix} 2/9 \\ 1/3 \\ 1 \end{bmatrix}$

12. $\begin{bmatrix} 8/3 \\ 1 \end{bmatrix}$

13. Find the distance between $\mathbf{x} = \begin{bmatrix} 10 \\ -3 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} -1 \\ -5 \end{bmatrix}$.

14. Find the distance between $\mathbf{u} = \begin{bmatrix} 0 \\ -1 \\ 3 \end{bmatrix}$ and $\mathbf{z} = \begin{bmatrix} -7 \\ -5 \\ 7 \end{bmatrix}$.

Determine which pairs of vectors in Exercises 15–18 are orthogonal.

15. $\mathbf{a} = \begin{bmatrix} 8 \\ -5 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -2 \\ -3 \end{bmatrix}$

16. $\mathbf{x} = \begin{bmatrix} 4 \\ -2 \\ 5 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 11 \\ -1 \\ -9 \end{bmatrix}$

17. $\mathbf{u} = \begin{bmatrix} 3 \\ 2 \\ -5 \\ 0 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -4 \\ 1 \\ -2 \\ 6 \end{bmatrix}$

18. $\mathbf{w} = \begin{bmatrix} 3 \\ -6 \\ 7 \\ 8 \end{bmatrix}$, $\mathbf{z} = \begin{bmatrix} -9 \\ 6 \\ 17 \\ -7 \end{bmatrix}$

In Exercises 19–28, all vectors are in $\mathbb{R}^n$. Mark each statement True or False (**T/F**). Justify each answer.

19. **(T/F)** $\mathbf{v}\cdot\mathbf{v} = \|\mathbf{v}\|^2$.

20. **(T/F)** $\mathbf{u}\cdot\mathbf{v} - \mathbf{v}\cdot\mathbf{u} = 0$.

21. **(T/F)** If the distance from $\mathbf{u}$ to $\mathbf{v}$ equals the distance from $\mathbf{u}$ to $-\mathbf{v}$, then $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

22. **(T/F)** If $\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 = \|\mathbf{u}+\mathbf{v}\|^2$, then $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

23. **(T/F)** If vectors $\mathbf{v}_1, \dots, \mathbf{v}_p$ span a subspace W and if $\mathbf{x}$ is orthogonal to each $\mathbf{v}_j$ for $j = 1, \dots, p$, then $\mathbf{x}$ is in $W^\perp$.

24. **(T/F)** If $\mathbf{x}$ is orthogonal to every vector in a subspace W then $\mathbf{x}$ is in $W^\perp$.

25. **(T/F)** For any scalar c, $\|c\mathbf{v}\| = c\|\mathbf{v}\|$.

26. **(T/F)** For any scalar c, $\mathbf{u}\cdot(c\mathbf{v}) = c(\mathbf{u}\cdot\mathbf{v})$.

27. **(T/F)** For a square matrix A, vectors in Col A are orthogonal to vectors in Nul A.

28. **(T/F)** For an $m \times n$ matrix A, vectors in the null space of A are orthogonal to vectors in the row space of A.

29. Use the transpose definition of the inner product to verify parts (b) and (c) of Theorem 1. Mention the appropriate facts from Chapter 2.

30. Let $\mathbf{u} = (u_1, u_2, u_3)$. Explain why $\mathbf{u}\cdot\mathbf{u} \geq 0$. When is $\mathbf{u}\cdot\mathbf{u} = 0$?

31. Let $\mathbf{u} = \begin{bmatrix} 3 \\ -4 \\ -1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -8 \\ -7 \\ 4 \end{bmatrix}$. Compute and compare $\mathbf{u}\cdot\mathbf{v}$, $\|\mathbf{u}\|^2$, $\|\mathbf{v}\|^2$, and $\|\mathbf{u}+\mathbf{v}\|^2$. Do not use the Pythagorean Theorem.

32. Verify the *parallelogram law* for vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$:

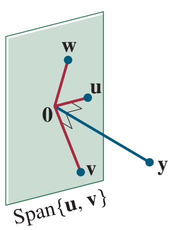
$$\|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2 = 2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2$$

33. Let $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$. Describe the set H of vectors $\begin{bmatrix} x \\ y \end{bmatrix}$ that are orthogonal to $\mathbf{v}$. [*Hint:* Consider $\mathbf{v} = \mathbf{0}$ and $\mathbf{v} \neq \mathbf{0}$.]

34. Let $\mathbf{u} = \begin{bmatrix} 5 \\ -6 \\ 7 \end{bmatrix}$, and let W be the set of all $\mathbf{x}$ in $\mathbb{R}^3$ such that $\mathbf{u}\cdot\mathbf{x} = 0$. What theorem in Chapter 4 can be used to show that W is a subspace of $\mathbb{R}^3$? Describe W in geometric language.

35. Suppose a vector $\mathbf{y}$ is orthogonal to vectors $\mathbf{u}$ and $\mathbf{v}$. Show that $\mathbf{y}$ is orthogonal to the vector $\mathbf{u}+\mathbf{v}$.

36. Suppose $\mathbf{y}$ is orthogonal to $\mathbf{u}$ and $\mathbf{v}$. Show that $\mathbf{y}$ is orthogonal to every $\mathbf{w}$ in Span $\{\mathbf{u}, \mathbf{v}\}$. [*Hint:* An arbitrary $\mathbf{w}$ in Span $\{\mathbf{u}, \mathbf{v}\}$ has the form $\mathbf{w} = c_1\mathbf{u} + c_2\mathbf{v}$. Show that $\mathbf{y}$ is orthogonal to such a vector $\mathbf{w}$.]



37. Let $W = \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$. Show that if $\mathbf{x}$ is orthogonal to each $\mathbf{v}_j$, for $1 \leq j \leq p$, then $\mathbf{x}$ is orthogonal to every vector in W.

38. Let W be a subspace of $\mathbb{R}^n$, and let $W^\perp$ be the set of all vectors orthogonal to W. Show that $W^\perp$ is a subspace of $\mathbb{R}^n$ using the following steps.

a. Take $\mathbf{z}$ in $W^\perp$, and let $\mathbf{u}$ represent any element of W. Then $\mathbf{z}\cdot\mathbf{u} = 0$. Take any scalar c and show that $c\mathbf{z}$ is orthogonal to $\mathbf{u}$. (Since $\mathbf{u}$ was an arbitrary element of W, this will show that $c\mathbf{z}$ is in $W^\perp$.)

b. Take $\mathbf{z}_1$ and $\mathbf{z}_2$ in $W^\perp$, and let $\mathbf{u}$ be any element of W. Show that $\mathbf{z}_1 + \mathbf{z}_2$ is orthogonal to $\mathbf{u}$. What can you conclude about $\mathbf{z}_1 + \mathbf{z}_2$? Why?

c. Finish the proof that $W^\perp$ is a subspace of $\mathbb{R}^n$.

39. Show that if $\mathbf{x}$ is in both W and $W^\perp$, then $\mathbf{x} = \mathbf{0}$.

T **40.** Construct a pair $\mathbf{u}$, $\mathbf{v}$ of random vectors in $\mathbb{R}^4$, and let

$$A = \begin{bmatrix} .5 & .5 & .5 & .5 \\ .5 & .5 & -.5 & -.5 \\ .5 & -.5 & .5 & -.5 \\ .5 & -.5 & -.5 & .5 \end{bmatrix}$$

a. Denote the columns of A by $\mathbf{a}_1, \dots, \mathbf{a}_4$. Compute the length of each column, and compute $\mathbf{a}_1\cdot\mathbf{a}_2$, $\mathbf{a}_1\cdot\mathbf{a}_3$, $\mathbf{a}_1\cdot\mathbf{a}_4$, $\mathbf{a}_2\cdot\mathbf{a}_3$, $\mathbf{a}_2\cdot\mathbf{a}_4$, and $\mathbf{a}_3\cdot\mathbf{a}_4$.

b. Compute and compare the lengths of $\mathbf{u}$, $A\mathbf{u}$, $\mathbf{v}$, and $A\mathbf{v}$.

c. Use equation (2) in this section to compute the cosine of the angle between $\mathbf{u}$ and $\mathbf{v}$. Compare this with the cosine of the angle between $A\mathbf{u}$ and $A\mathbf{v}$.

d. Repeat parts (b) and (c) for two other pairs of random vectors. What do you conjecture about the effect of A on vectors?

T **41.** Generate random vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{v}$ in $\mathbb{R}^4$ with integer entries (and $\mathbf{v} \neq \mathbf{0}$), and compute the quantities

$$\left(\frac{\mathbf{x}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v},\ \left(\frac{\mathbf{y}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v},\ \frac{(\mathbf{x}+\mathbf{y})\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v},\ \frac{(10\mathbf{x})\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v}$$

Repeat the computations with new random vectors $\mathbf{x}$ and $\mathbf{y}$. What do you conjecture about the mapping $\mathbf{x} \mapsto T(\mathbf{x}) = \left(\frac{\mathbf{x}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v}$ (for $\mathbf{v} \neq \mathbf{0}$)? Verify your conjecture algebraically.

T **42.** Let $A = \begin{bmatrix} -6 & 3 & -27 & -33 & -13 \\ 6 & -5 & 25 & 28 & 14 \\ 8 & -6 & 34 & 38 & 18 \\ 12 & -10 & 50 & 41 & 23 \\ 14 & -21 & 49 & 29 & 33 \end{bmatrix}$. Construct a matrix N whose columns form a basis for Nul A, and construct a matrix R whose *rows* form a basis for Row A (see Section 4.6 for details). Perform a matrix computation with N and R that illustrates a fact from Theorem 3.

Solutions to Practice Problems

1. $\mathbf{a}\cdot\mathbf{b} = 7$, $\mathbf{a}\cdot\mathbf{a} = 5$. Hence $\dfrac{\mathbf{a}\cdot\mathbf{b}}{\mathbf{a}\cdot\mathbf{a}} = \dfrac{7}{5}$, and $\left(\dfrac{\mathbf{a}\cdot\mathbf{b}}{\mathbf{a}\cdot\mathbf{a}}\right)\mathbf{a} = \dfrac{7}{5}\mathbf{a} = \begin{bmatrix} -14/5 \\ 7/5 \end{bmatrix}$.

2. a. Scale $\mathbf{c}$, multiplying by 3 to get $\mathbf{y} = \begin{bmatrix} 4 \\ -3 \\ 2 \end{bmatrix}$. Compute $\|\mathbf{y}\|^2 = 29$ and $\|\mathbf{y}\| = \sqrt{29}$. The unit vector in the direction of both $\mathbf{c}$ and $\mathbf{y}$ is $\mathbf{u} = \dfrac{1}{\|\mathbf{y}\|}\mathbf{y} = \begin{bmatrix} 4/\sqrt{29} \\ -3/\sqrt{29} \\ 2/\sqrt{29} \end{bmatrix}$.

b. $\mathbf{d}$ is orthogonal to $\mathbf{c}$, because

$$\mathbf{d}\cdot\mathbf{c} = \begin{bmatrix} 5 \\ 6 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 4/3 \\ -1 \\ 2/3 \end{bmatrix} = \frac{20}{3} - 6 - \frac{2}{3} = 0$$

c. $\mathbf{d}$ is orthogonal to $\mathbf{u}$, because $\mathbf{u}$ has the form $k\mathbf{c}$ for some k, and

$$\mathbf{d}\cdot\mathbf{u} = \mathbf{d}\cdot(k\mathbf{c}) = k(\mathbf{d}\cdot\mathbf{c}) = k(0) = 0$$

3. If $W \neq \{\mathbf{0}\}$, let $\{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ be a basis for W, where $1 \leq p \leq n$. Let A be the $p \times n$ matrix having rows $\mathbf{b}_1^T, \dots, \mathbf{b}_p^T$. It follows that W is the row space of A. Theorem 3 implies that $W^\perp = (\text{Row } A)^\perp = \text{Nul } A$ and hence $\dim W^\perp = \dim \text{Nul } A$. Thus, $\dim W + \dim W^\perp = \dim \text{Row } A + \dim \text{Nul } A = \text{rank } A + \dim \text{Nul } A = n$, by the Rank Theorem. If $W = \{\mathbf{0}\}$, then $W^\perp = \mathbb{R}^n$, and the result follows.

6.2 Orthogonal Sets

A set of vectors $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is said to be an **orthogonal set** if each pair of distinct vectors from the set is orthogonal, that is, if $\mathbf{u}_i\cdot\mathbf{u}_j = 0$ whenever $i \neq j$.

EXAMPLE 1 Show that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set, where

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}$$

SOLUTION Consider the three possible pairs of distinct vectors, namely $\{\mathbf{u}_1, \mathbf{u}_2\}$, $\{\mathbf{u}_1, \mathbf{u}_3\}$, and $\{\mathbf{u}_2, \mathbf{u}_3\}$.

$$\mathbf{u}_1\cdot\mathbf{u}_2 = 3(-1) + 1(2) + 1(1) = 0$$
$$\mathbf{u}_1\cdot\mathbf{u}_3 = 3\left(-\tfrac{1}{2}\right) + 1(-2) + 1\left(\tfrac{7}{2}\right) = 0$$
$$\mathbf{u}_2\cdot\mathbf{u}_3 = -1\left(-\tfrac{1}{2}\right) + 2(-2) + 1\left(\tfrac{7}{2}\right) = 0$$

Each pair of distinct vectors is orthogonal, and so $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set. See Figure 1; the three line segments are mutually perpendicular. ■

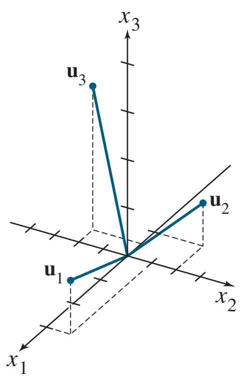


FIGURE 1

THEOREM 4

If $S = \{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent and hence is a basis for the subspace spanned by S.

PROOF If $\mathbf{0} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ for some scalars $c_1, \dots, c_p$, then

$$\begin{aligned} 0 = \mathbf{0}\cdot\mathbf{u}_1 &= (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p)\cdot\mathbf{u}_1 \\ &= (c_1\mathbf{u}_1)\cdot\mathbf{u}_1 + (c_2\mathbf{u}_2)\cdot\mathbf{u}_1 + \cdots + (c_p\mathbf{u}_p)\cdot\mathbf{u}_1 \\ &= c_1(\mathbf{u}_1\cdot\mathbf{u}_1) + c_2(\mathbf{u}_2\cdot\mathbf{u}_1) + \cdots + c_p(\mathbf{u}_p\cdot\mathbf{u}_1) \\ &= c_1(\mathbf{u}_1\cdot\mathbf{u}_1) \end{aligned}$$

because $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2, \dots, \mathbf{u}_p$. Since $\mathbf{u}_1$ is nonzero, $\mathbf{u}_1\cdot\mathbf{u}_1$ is not zero and so $c_1 = 0$. Similarly, $c_2, \dots, c_p$ must be zero. Thus S is linearly independent. ■

DEFINITION

An **orthogonal basis** for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

The next theorem suggests why an orthogonal basis is much nicer than other bases. The weights in a linear combination can be computed easily.

THEOREM 5

Let $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\mathbf{y}$ in W, the weights in the linear combination

$$\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$$

are given by

$$c_j = \frac{\mathbf{y}\cdot\mathbf{u}_j}{\mathbf{u}_j\cdot\mathbf{u}_j} \qquad (j = 1, \dots, p)$$

PROOF As in the preceding proof, the orthogonality of $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ shows that

$$\mathbf{y}\cdot\mathbf{u}_1 = (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p)\cdot\mathbf{u}_1 = c_1(\mathbf{u}_1\cdot\mathbf{u}_1)$$

Since $\mathbf{u}_1\cdot\mathbf{u}_1$ is not zero, the equation above can be solved for c_1. To find c_j for $j = 2, \dots, p$, compute $\mathbf{y}\cdot\mathbf{u}_j$ and solve for c_j. ■

EXAMPLE 2 The set $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ in Example 1 is an orthogonal basis for $\mathbb{R}^3$. Express the vector $\mathbf{y} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}$ as a linear combination of the vectors in S.

SOLUTION Compute

$$\begin{aligned} &\mathbf{y}\cdot\mathbf{u}_1 = 11, \qquad &&\mathbf{y}\cdot\mathbf{u}_2 = -12, \qquad &&\mathbf{y}\cdot\mathbf{u}_3 = -33 \\ &\mathbf{u}_1\cdot\mathbf{u}_1 = 11, &&\mathbf{u}_2\cdot\mathbf{u}_2 = 6, &&\mathbf{u}_3\cdot\mathbf{u}_3 = 33/2 \end{aligned}$$

By Theorem 5,

$$\begin{aligned} \mathbf{y} &= \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 + \frac{\mathbf{y}\cdot\mathbf{u}_3}{\mathbf{u}_3\cdot\mathbf{u}_3}\mathbf{u}_3 \\ &= \frac{11}{11}\mathbf{u}_1 + \frac{-12}{6}\mathbf{u}_2 + \frac{-33}{33/2}\mathbf{u}_3 \\ &= \mathbf{u}_1 - 2\mathbf{u}_2 - 2\mathbf{u}_3 \end{aligned}$$

■

Notice how easy it is to compute the weights needed to build $\mathbf{y}$ from an orthogonal basis. If the basis were not orthogonal, it would be necessary to solve a system of linear equations in order to find the weights, as in Chapter 1.

We turn next to a construction that will become a key step in many calculations involving orthogonality, and it will lead to a geometric interpretation of Theorem 5.

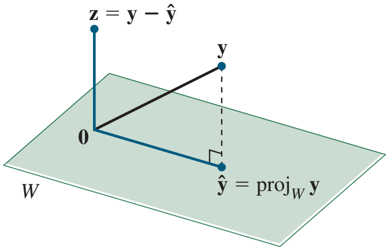


FIGURE 2
Finding α to make $\mathbf{y} - \hat{\mathbf{y}}$ orthogonal to $\mathbf{u}$.

An Orthogonal Projection

Given a nonzero vector $\mathbf{u}$ in $\mathbb{R}^n$, consider the problem of decomposing a vector $\mathbf{y}$ in $\mathbb{R}^n$ into the sum of two vectors, one a multiple of $\mathbf{u}$ and the other orthogonal to $\mathbf{u}$. We wish to write

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \tag{1}$$

where $\hat{\mathbf{y}} = \alpha\mathbf{u}$ for some scalar α and $\mathbf{z}$ is some vector orthogonal to $\mathbf{u}$. See Figure 2. Given any scalar α, let $\mathbf{z} = \mathbf{y} - \alpha\mathbf{u}$, so that (1) is satisfied. Then $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to $\mathbf{u}$ if and only if

$$0 = (\mathbf{y} - \alpha\mathbf{u})\cdot\mathbf{u} = \mathbf{y}\cdot\mathbf{u} - (\alpha\mathbf{u})\cdot\mathbf{u} = \mathbf{y}\cdot\mathbf{u} - \alpha(\mathbf{u}\cdot\mathbf{u})$$

That is, (1) is satisfied with $\mathbf{z}$ orthogonal to $\mathbf{u}$ if and only if $\alpha = \dfrac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}$ and $\hat{\mathbf{y}} = \dfrac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$. The vector $\hat{\mathbf{y}}$ is called the **orthogonal projection of y onto u**, and the vector $\mathbf{z}$ is called the **component of y orthogonal to u**.

If c is any nonzero scalar and if $\mathbf{u}$ is replaced by $c\mathbf{u}$ in the definition of $\hat{\mathbf{y}}$, then the orthogonal projection of $\mathbf{y}$ onto $c\mathbf{u}$ is exactly the same as the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$ (Exercise 39). Hence this projection is determined by the *subspace* L spanned by $\mathbf{u}$ (the line through $\mathbf{u}$ and $\mathbf{0}$). Sometimes $\hat{\mathbf{y}}$ is denoted by $\text{proj}_L\,\mathbf{y}$ and is called the **orthogonal projection of y onto** L. That is,

$$\boxed{\hat{\mathbf{y}} = \text{proj}_L\,\mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}} \tag{2}$$

EXAMPLE 3 Let $\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$. Find the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$. Then write $\mathbf{y}$ as the sum of two orthogonal vectors, one in Span $\{\mathbf{u}\}$ and one orthogonal to $\mathbf{u}$.

SOLUTION Compute

$$\mathbf{y}\cdot\mathbf{u} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}\cdot\begin{bmatrix} 4 \\ 2 \end{bmatrix} = 40$$
$$\mathbf{u}\cdot\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}\cdot\begin{bmatrix} 4 \\ 2 \end{bmatrix} = 20$$

The orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$ is

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u} = \frac{40}{20}\mathbf{u} = 2\begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

and the component of $\mathbf{y}$ orthogonal to $\mathbf{u}$ is

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

The sum of these two vectors is $\mathbf{y}$. That is,

$$\underset{\substack{\uparrow \\ \mathbf{y}}}{\begin{bmatrix} 7 \\ 6 \end{bmatrix}} = \underset{\substack{\uparrow \\ \hat{\mathbf{y}}}}{\begin{bmatrix} 8 \\ 4 \end{bmatrix}} + \underset{\substack{\uparrow \\ (\mathbf{y} - \hat{\mathbf{y}})}}{\begin{bmatrix} -1 \\ 2 \end{bmatrix}}$$

This decomposition of $\mathbf{y}$ is illustrated in Figure 3. *Note:* If the calculations above are correct, then $\{\hat{\mathbf{y}}, \mathbf{y} - \hat{\mathbf{y}}\}$ will be an orthogonal set. As a check, compute

$$\hat{\mathbf{y}}\cdot(\mathbf{y} - \hat{\mathbf{y}}) = \begin{bmatrix} 8 \\ 4 \end{bmatrix}\cdot\begin{bmatrix} -1 \\ 2 \end{bmatrix} = -8 + 8 = 0$$

■

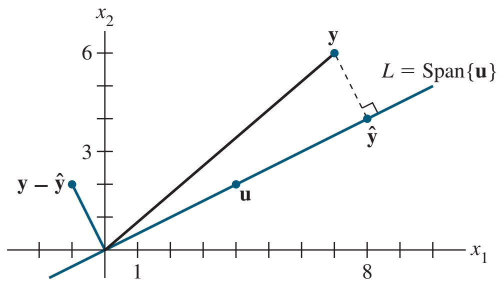


FIGURE 3 The orthogonal projection of $\mathbf{y}$ onto a line L through the origin.

Since the line segment in Figure 3 between $\mathbf{y}$ and $\hat{\mathbf{y}}$ is perpendicular to L, by construction of $\hat{\mathbf{y}}$, the point identified with $\hat{\mathbf{y}}$ is the closest point of L to $\mathbf{y}$. (This can be proved from geometry. We will assume this for $\mathbb{R}^2$ now and prove it for $\mathbb{R}^n$ in Section 6.3.)

EXAMPLE 4 Find the distance in Figure 3 from $\mathbf{y}$ to L.

SOLUTION The distance from $\mathbf{y}$ to L is the length of the perpendicular line segment from $\mathbf{y}$ to the orthogonal projection $\hat{\mathbf{y}}$. This length equals the length of $\mathbf{y} - \hat{\mathbf{y}}$. Thus the distance is

$$\|\mathbf{y} - \hat{\mathbf{y}}\| = \sqrt{(-1)^2 + 2^2} = \sqrt{5}$$ ■

A Geometric Interpretation of Theorem 5

The formula for the orthogonal projection $\hat{\mathbf{y}}$ in (2) has the same appearance as each of the terms in Theorem 5. Thus Theorem 5 decomposes a vector $\mathbf{y}$ into a sum of orthogonal projections onto one-dimensional subspaces.

It is easy to visualize the case in which $W = \mathbb{R}^2 = \text{Span}\,\{\mathbf{u}_1, \mathbf{u}_2\}$, with $\mathbf{u}_1$ and $\mathbf{u}_2$ orthogonal. Any $\mathbf{y}$ in $\mathbb{R}^2$ can be written in the form

$$\mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 \tag{3}$$

The first term in (3) is the projection of $\mathbf{y}$ onto the subspace spanned by $\mathbf{u}_1$ (the line through $\mathbf{u}_1$ and the origin), and the second term is the projection of $\mathbf{y}$ onto the subspace spanned by $\mathbf{u}_2$. Thus (3) expresses $\mathbf{y}$ as the sum of its projections onto the (orthogonal) axes determined by $\mathbf{u}_1$ and $\mathbf{u}_2$. See Figure 4.

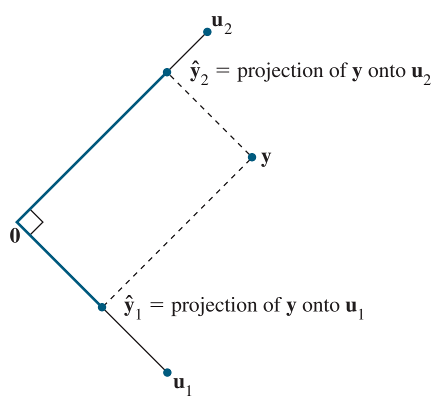


FIGURE 4 A vector decomposed into the sum of two projections.

Theorem 5 decomposes each $\mathbf{y}$ in Span $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ into the sum of p projections onto one-dimensional subspaces that are mutually orthogonal.

Decomposing a Force into Component Forces

The decomposition in Figure 4 can occur in physics when some sort of force is applied to an object. Choosing an appropriate coordinate system allows the force to be represented by a vector $\mathbf{y}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$. Often the problem involves some particular direction of interest, which is represented by another vector $\mathbf{u}$. For instance, if the object is moving in a straight line when the force is applied, the vector $\mathbf{u}$ might point in the direction of movement, as in Figure 5. A key step in the problem is to decompose the force into a component in the direction of $\mathbf{u}$ and a component orthogonal to $\mathbf{u}$. The calculations would be analogous to those previously made in Example 3.

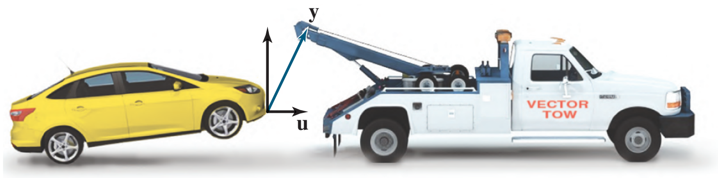


FIGURE 5

Orthonormal Sets

A set $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an **orthonormal set** if it is an orthogonal set of unit vectors. If W is the subspace spanned by such a set, then $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an **orthonormal basis** for W, since the set is automatically linearly independent, by Theorem 4.

The simplest example of an orthonormal set is the standard basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ for $\mathbb{R}^n$. Any nonempty subset of $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is orthonormal, too. Here is a more complicated example.

EXAMPLE 5 Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis of $\mathbb{R}^3$, where

$$\mathbf{v}_1 = \begin{bmatrix} 3/\sqrt{11} \\ 1/\sqrt{11} \\ 1/\sqrt{11} \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -1/\sqrt{66} \\ -4/\sqrt{66} \\ 7/\sqrt{66} \end{bmatrix}$$

SOLUTION Compute

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = -3/\sqrt{66} + 2/\sqrt{66} + 1/\sqrt{66} = 0$$
$$\mathbf{v}_1 \cdot \mathbf{v}_3 = -3/\sqrt{726} - 4/\sqrt{726} + 7/\sqrt{726} = 0$$
$$\mathbf{v}_2 \cdot \mathbf{v}_3 = 1/\sqrt{396} - 8/\sqrt{396} + 7/\sqrt{396} = 0$$

Thus $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal set. Also,

$$\mathbf{v}_1 \cdot \mathbf{v}_1 = 9/11 + 1/11 + 1/11 = 1$$
$$\mathbf{v}_2 \cdot \mathbf{v}_2 = 1/6 + 4/6 + 1/6 = 1$$
$$\mathbf{v}_3 \cdot \mathbf{v}_3 = 1/66 + 16/66 + 49/66 = 1$$

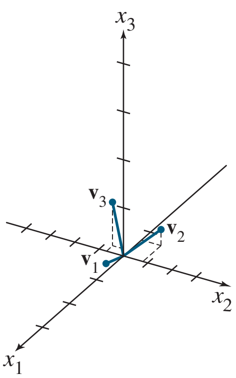


FIGURE 6

which shows that $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are unit vectors. Thus $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal set. Since the set is linearly independent, its three vectors form a basis for $\mathbb{R}^3$. See Figure 6.

When the vectors in an orthogonal set of nonzero vectors are *normalized* to have unit length, the new vectors will still be orthogonal, and hence the new set will be an orthonormal set. See Exercise 40. It is easy to check that the vectors in Figure 6 (Example 5) are simply the unit vectors in the directions of the vectors in Figure 1 (Example 1).

Matrices whose columns form an orthonormal set are important in applications and in computer algorithms for matrix computations. Their main properties are given in Theorems 6 and 7.

THEOREM 6 An $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I$.

PROOF To simplify notation, we suppose that U has only three columns, each a vector in $\mathbb{R}^m$. The proof of the general case is essentially the same. Let $U = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3\,]$ and compute

$$U^T U = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \mathbf{u}_3^T \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \mathbf{u}_1^T\mathbf{u}_3 \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \mathbf{u}_2^T\mathbf{u}_3 \\ \mathbf{u}_3^T\mathbf{u}_1 & \mathbf{u}_3^T\mathbf{u}_2 & \mathbf{u}_3^T\mathbf{u}_3 \end{bmatrix} \tag{4}$$

The entries in the matrix at the right are inner products, using transpose notation. The columns of U are orthogonal if and only if

$$\mathbf{u}_1^T\mathbf{u}_2 = \mathbf{u}_2^T\mathbf{u}_1 = 0, \quad \mathbf{u}_1^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_1 = 0, \quad \mathbf{u}_2^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_2 = 0 \tag{5}$$

The columns of U all have unit length if and only if

$$\mathbf{u}_1^T\mathbf{u}_1 = 1, \quad \mathbf{u}_2^T\mathbf{u}_2 = 1, \quad \mathbf{u}_3^T\mathbf{u}_3 = 1 \tag{6}$$

The theorem follows immediately from (4)–(6). ■

THEOREM 7 Let U be an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}$ and $\mathbf{y}$ be in $\mathbb{R}^n$. Then

a. $\|U\mathbf{x}\| = \|\mathbf{x}\|$
b. $(U\mathbf{x})\cdot(U\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$
c. $(U\mathbf{x})\cdot(U\mathbf{y}) = 0$ if and only if $\mathbf{x}\cdot\mathbf{y} = 0$

Properties (a) and (c) say that the linear mapping $\mathbf{x} \mapsto U\mathbf{x}$ preserves lengths and orthogonality. These properties are crucial for many computer algorithms. See Exercise 33 for the proof of Theorem 7.

EXAMPLE 6 Let $U = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix}$. Notice that U has orthonormal columns and

$$U^T U = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 2/3 & -2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Verify that $\|U\mathbf{x}\| = \|\mathbf{x}\|$.

SOLUTION

$$U\mathbf{x} = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$$

$$\|U\mathbf{x}\| = \sqrt{9+1+1} = \sqrt{11}$$

$$\|\mathbf{x}\| = \sqrt{2+9} = \sqrt{11}$$

Theorems 6 and 7 are particularly useful when applied to *square* matrices. An **orthogonal matrix** is a square invertible matrix U such that $U^{-1} = U^T$. By Theorem 6, such a matrix has orthonormal columns.[1] It is easy to see that any *square* matrix with orthonormal columns is an orthogonal matrix. Surprisingly, such a matrix must have orthonormal *rows*, too. See Exercises 35 and 36. Orthogonal matrices will appear frequently in Chapter 7.

EXAMPLE 7 The matrix

$$U = \begin{bmatrix} 3/\sqrt{11} & -1/\sqrt{6} & -1/\sqrt{66} \\ 1/\sqrt{11} & 2/\sqrt{6} & -4/\sqrt{66} \\ 1/\sqrt{11} & 1/\sqrt{6} & 7/\sqrt{66} \end{bmatrix}$$

is an orthogonal matrix because it is square and because its columns are orthonormal, by Example 5. Verify that the rows are orthonormal, too!

Practice Problems

1. Let $\mathbf{u}_1 = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$. Show that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthonormal basis for $\mathbb{R}^2$.
2. Let $\mathbf{y}$ and L be as in Example 3 and Figure 3. Compute the orthogonal projection $\hat{\mathbf{y}}$ of $\mathbf{y}$ onto L using $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ instead of the $\mathbf{u}$ in Example 3.
3. Let U and $\mathbf{x}$ be as in Example 6, and let $\mathbf{y} = \begin{bmatrix} -3\sqrt{2} \\ 6 \end{bmatrix}$. Verify that $U\mathbf{x}\cdot U\mathbf{y} = \mathbf{x}\cdot\mathbf{y}$.
4. Let U be an $n \times n$ matrix with orthonormal columns. Show that det $U = \pm 1$.

6.2 Exercises

In Exercises 1–6, determine which sets of vectors are orthogonal.

1. $\begin{bmatrix} -1 \\ 4 \\ -3 \end{bmatrix}, \begin{bmatrix} 5 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ -4 \\ -7 \end{bmatrix}$

2. $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -5 \\ -2 \\ 1 \end{bmatrix}$

3. $\begin{bmatrix} 2 \\ -7 \\ -1 \end{bmatrix}, \begin{bmatrix} -6 \\ -3 \\ 9 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}$

4. $\begin{bmatrix} 2 \\ -5 \\ -3 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ 6 \end{bmatrix}$

5. $\begin{bmatrix} 3 \\ -2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 8 \\ 7 \\ 0 \end{bmatrix}$

6. $\begin{bmatrix} 5 \\ -4 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} -4 \\ 1 \\ -3 \\ 8 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 5 \\ -1 \end{bmatrix}$

In Exercises 7–10, show that $\{\mathbf{u}_1, \mathbf{u}_2\}$ or $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal basis for $\mathbb{R}^2$ or $\mathbb{R}^3$, respectively. Then express $\mathbf{x}$ as a linear combination of the $\mathbf{u}$'s.

7. $\mathbf{u}_1 = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} 9 \\ -7 \end{bmatrix}$

[1] A better name might be *orthonormal matrix*, and this term is found in some statistics texts. However, *orthogonal matrix* is the standard term in linear algebra.

8. $\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 6 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$

9. $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 1 \\ -4 \\ 1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 4 \\ 2 \\ 4 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} 6 \\ 4 \\ -2 \end{bmatrix}$

10. $\mathbf{u}_1 = \begin{bmatrix} 4 \\ -4 \\ 0 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} 3 \\ -4 \\ 7 \end{bmatrix}$

11. Compute the orthogonal projection of $\begin{bmatrix} 1 \\ 7 \end{bmatrix}$ onto the line through $\begin{bmatrix} -4 \\ 2 \end{bmatrix}$ and the origin.

12. Compute the orthogonal projection of $\begin{bmatrix} -3 \\ 4 \end{bmatrix}$ onto the line through $\begin{bmatrix} 1 \\ -3 \end{bmatrix}$ and the origin.

13. Let $\mathbf{y} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 4 \\ -7 \end{bmatrix}$. Write $\mathbf{y}$ as the sum of two orthogonal vectors, one in Span $\{\mathbf{u}\}$ and one orthogonal to $\mathbf{u}$.

14. Let $\mathbf{y} = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 6 \\ 1 \end{bmatrix}$. Write $\mathbf{y}$ as the sum of a vector in Span $\{\mathbf{u}\}$ and a vector orthogonal to $\mathbf{u}$.

15. Let $\mathbf{y} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 8 \\ 6 \end{bmatrix}$. Compute the distance from $\mathbf{y}$ to the line through $\mathbf{u}$ and the origin.

16. Let $\mathbf{y} = \begin{bmatrix} -1 \\ 7 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$. Compute the distance from $\mathbf{y}$ to the line through $\mathbf{u}$ and the origin.

In Exercises 17–22, determine which sets of vectors are orthonormal. If a set is only orthogonal, normalize the vectors to produce an orthonormal set.

17. $\begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \begin{bmatrix} -1/2 \\ 0 \\ 1/2 \end{bmatrix}$

18. $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$

19. $\begin{bmatrix} -.6 \\ .8 \end{bmatrix}, \begin{bmatrix} .8 \\ .6 \end{bmatrix}$

20. $\begin{bmatrix} 4/3 \\ 7/3 \\ 4/3 \end{bmatrix}, \begin{bmatrix} 7/3 \\ -4/3 \\ 0 \end{bmatrix}$

21. $\begin{bmatrix} 1/\sqrt{10} \\ 3/\sqrt{20} \\ 3/\sqrt{20} \end{bmatrix}, \begin{bmatrix} 3/\sqrt{10} \\ -1/\sqrt{20} \\ -1/\sqrt{20} \end{bmatrix}, \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$

22. $\begin{bmatrix} 1/\sqrt{18} \\ 4/\sqrt{18} \\ 1/\sqrt{18} \end{bmatrix}, \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix}, \begin{bmatrix} -2/3 \\ 1/3 \\ -2/3 \end{bmatrix}$

In Exercises 23–32, all vectors are in $\mathbb{R}^n$. Mark each statement True or False **(T/F)**. Justify each answer.

23. **(T/F)** Not every linearly independent set in $\mathbb{R}^n$ is an orthogonal set.

24. **(T/F)** Not every orthogonal set in $\mathbb{R}^n$ is linearly independent.

25. **(T/F)** If $\mathbf{y}$ is a linear combination of nonzero vectors from an orthogonal set, then the weights in the linear combination can be computed without row operations on a matrix.

26. **(T/F)** If a set $S = \{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ has the property that $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ whenever $i \neq j$, then S is an orthonormal set.

27. **(T/F)** If the vectors in an orthogonal set of nonzero vectors are normalized, then some of the new vectors may not be orthogonal.

28. **(T/F)** If the columns of an $m \times n$ matrix A are orthonormal, then the linear mapping $\mathbf{x} \mapsto A\mathbf{x}$ preserves lengths.

29. **(T/F)** A matrix with orthonormal columns is an orthogonal matrix.

30. **(T/F)** The orthogonal projection of $\mathbf{y}$ onto $\mathbf{v}$ is the same as the orthogonal projection of $\mathbf{y}$ onto $c\mathbf{v}$ whenever $c \neq 0$.

31. **(T/F)** If L is a line through $\mathbf{0}$ and if $\hat{\mathbf{y}}$ is the orthogonal projection of $\mathbf{y}$ onto L, then $\|\hat{\mathbf{y}}\|$ gives the distance from $\mathbf{y}$ to L.

32. **(T/F)** An orthogonal matrix is invertible.

33. Prove Theorem 7. [*Hint:* For (a), compute $\|U\mathbf{x}\|^2$, or prove (b) first.]

34. Suppose W is a subspace of $\mathbb{R}^n$ spanned by n nonzero orthogonal vectors. Explain why $W = \mathbb{R}^n$.

35. Let U be a square matrix with orthonormal columns. Explain why U is invertible. (Mention the theorems you use.)

36. Let U be an $n \times n$ orthogonal matrix. Show that the rows of U form an orthonormal basis of $\mathbb{R}^n$.

37. Let U and V be $n \times n$ orthogonal matrices. Explain why UV is an orthogonal matrix. [That is, explain why UV is invertible and its inverse is $(UV)^T$.]

38. Let U be an orthogonal matrix, and construct V by interchanging some of the columns of U. Explain why V is an orthogonal matrix.

39. Show that the orthogonal projection of a vector $\mathbf{y}$ onto a line L through the origin in $\mathbb{R}^2$ does not depend on the choice of the nonzero $\mathbf{u}$ in L used in the formula for $\hat{\mathbf{y}}$. To do this, suppose $\mathbf{y}$ and $\mathbf{u}$ are given and $\hat{\mathbf{y}}$ has been computed by formula (2) in this section. Replace $\mathbf{u}$ in that formula by $c\mathbf{u}$, where c is an unspecified nonzero scalar. Show that the new formula gives the same $\hat{\mathbf{y}}$.

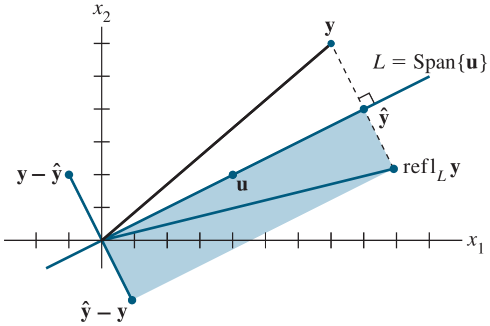
40. Let $\{\mathbf{v}_1, \mathbf{v}_2\}$ be an orthogonal set of nonzero vectors, and let c_1, c_2 be any nonzero scalars. Show that $\{c_1\mathbf{v}_1, c_2\mathbf{v}_2\}$ is also an orthogonal set. Since orthogonality of a set is defined in terms of pairs of vectors, this shows that if the vectors in an orthogonal set are normalized, the new set will still be orthogonal.

41. Given $\mathbf{u} \neq \mathbf{0}$ in $\mathbb{R}^n$, let $L = \text{Span}\{\mathbf{u}\}$. Show that the mapping $\mathbf{x} \mapsto \text{proj}_L\, \mathbf{x}$ is a linear transformation.

42. Given $\mathbf{u} \neq \mathbf{0}$ in $\mathbb{R}^n$, let $L = \text{Span}\{\mathbf{u}\}$. For $\mathbf{y}$ in $\mathbb{R}^n$, the **reflection of y in** L is the point $\text{refl}_L\, \mathbf{y}$ defined by

$$\text{refl}_L\, \mathbf{y} = 2\ \text{proj}_L\, \mathbf{y} - \mathbf{y}$$

See the figure, which shows that $\text{refl}_L\, \mathbf{y}$ is the sum of $\hat{\mathbf{y}} = \text{proj}_L\, \mathbf{y}$ and $\hat{\mathbf{y}} - \mathbf{y}$. Show that the mapping $\mathbf{y} \mapsto \text{refl}_L\, \mathbf{y}$ is a linear transformation.



The reflection of $\mathbf{y}$ in a line through the origin.

T 43. Show that the columns of the matrix A are orthogonal by making an appropriate matrix calculation. State the calculation you use.

$$A = \begin{bmatrix} -6 & -3 & 6 & 1 \\ -1 & 2 & 1 & -6 \\ 3 & 6 & 3 & -2 \\ 6 & -3 & 6 & -1 \\ 2 & -1 & 2 & 3 \\ -3 & 6 & 3 & 2 \\ -2 & -1 & 2 & -3 \\ 1 & 2 & 1 & 6 \end{bmatrix}$$

T 44. In parts (a)–(d), let U be the matrix formed by normalizing each column of the matrix A in Exercise 43.

a. Compute U^TU and UU^T. How do they differ?

b. Generate a random vector $\mathbf{y}$ in $\mathbb{R}^8$, and compute $\mathbf{p} = UU^T\mathbf{y}$ and $\mathbf{z} = \mathbf{y} - \mathbf{p}$. Explain why $\mathbf{p}$ is in Col A. Verify that $\mathbf{z}$ is orthogonal to $\mathbf{p}$.

c. Verify that $\mathbf{z}$ is orthogonal to each column of U.

d. Notice that $\mathbf{y} = \mathbf{p} + \mathbf{z}$, with $\mathbf{p}$ in Col A. Explain why $\mathbf{z}$ is in $(\text{Col}\, A)^{\perp}$. (The significance of this decomposition of $\mathbf{y}$ will be explained in the next section.)

Solutions to Practice Problems

1. The vectors are orthogonal because

$$\mathbf{u}_1 \cdot \mathbf{u}_2 = -2/5 + 2/5 = 0$$

They are unit vectors because

$$\|\mathbf{u}_1\|^2 = (-1/\sqrt{5})^2 + (2/\sqrt{5})^2 = 1/5 + 4/5 = 1$$
$$\|\mathbf{u}_2\|^2 = (2/\sqrt{5})^2 + (1/\sqrt{5})^2 = 4/5 + 1/5 = 1$$

In particular, the set $\{\mathbf{u}_1, \mathbf{u}_2\}$ is linearly independent, and hence is a basis for $\mathbb{R}^2$ since there are two vectors in the set.

2. When $\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$,

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u} = \frac{20}{5}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = 4\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

This is the same $\hat{\mathbf{y}}$ found in Example 3. The orthogonal projection does not depend on the $\mathbf{u}$ chosen on the line. See Exercise 39.

3. $U\mathbf{y} = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} \begin{bmatrix} -3\sqrt{2} \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \\ -7 \\ 2 \end{bmatrix}$

Also, from Example 6, $\mathbf{x} = \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix}$ and $U\mathbf{x} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$. Hence

$$U\mathbf{x}\cdot U\mathbf{y} = 3 + 7 + 2 = 12, \quad \text{and} \quad \mathbf{x}\cdot\mathbf{y} = -6 + 18 = 12$$

STUDY GUIDE offers additional resources for mastering the concepts around an orthogonal basis.

4. Since U is an $n \times n$ matrix with orthonormal columns, by Theorem 6, $U^T U = I$. Taking the determinant of the left side of this equation, and applying Theorems 5 and 6 from Section 3.2 results in $\det U^T U = (\det U^T)(\det U) = (\det U)(\det U) = (\det U)^2$. Recall $\det I = 1$. Putting the two sides of the equation back together results in $(\det U)^2 = 1$ and hence $\det U = \pm 1$.

6.3 Orthogonal Projections

The orthogonal projection of a point in $\mathbb{R}^2$ onto a line through the origin has an important analogue in $\mathbb{R}^n$. Given a vector $\mathbf{y}$ and a subspace W in $\mathbb{R}^n$, there is a vector $\hat{\mathbf{y}}$ in W such that (1) $\hat{\mathbf{y}}$ is the unique vector in W for which $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to W, and (2) $\hat{\mathbf{y}}$ is the unique vector in W closest to $\mathbf{y}$. See Figure 1. These two properties of $\hat{\mathbf{y}}$ provide the key to finding least-squares solutions of linear systems.

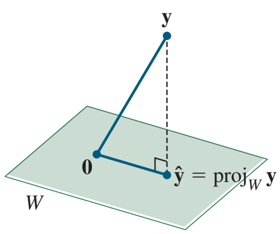


FIGURE 1

To prepare for the first theorem, observe that whenever a vector $\mathbf{y}$ is written as a linear combination of vectors $\mathbf{u}_1, \dots, \mathbf{u}_n$ in $\mathbb{R}^n$, the terms in the sum for $\mathbf{y}$ can be grouped into two parts so that $\mathbf{y}$ can be written as

$$\mathbf{y} = \mathbf{z}_1 + \mathbf{z}_2$$

where $\mathbf{z}_1$ is a linear combination of some of the $\mathbf{u}_i$ and $\mathbf{z}_2$ is a linear combination of the rest of the $\mathbf{u}_i$. This idea is particularly useful when $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ is an orthogonal basis. Recall from Section 6.1 that $W^\perp$ denotes the set of all vectors orthogonal to a subspace W.

EXAMPLE 1 Let $\{\mathbf{u}_1, \dots, \mathbf{u}_5\}$ be an orthogonal basis for $\mathbb{R}^5$ and let

$$\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_5\mathbf{u}_5$$

Consider the subspace $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, and write $\mathbf{y}$ as the sum of a vector $\mathbf{z}_1$ in W and a vector $\mathbf{z}_2$ in $W^\perp$.

SOLUTION Write

$$\mathbf{y} = \underbrace{c_1\mathbf{u}_1 + c_2\mathbf{u}_2}_{\mathbf{z}_1} + \underbrace{c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5}_{\mathbf{z}_2}$$

where $\quad \mathbf{z}_1 = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 \quad$ is in $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$

and $\quad \mathbf{z}_2 = c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5 \quad$ is in $\text{Span}\{\mathbf{u}_3, \mathbf{u}_4, \mathbf{u}_5\}$.

To show that $\mathbf{z}_2$ is in $W^\perp$, it suffices to show that $\mathbf{z}_2$ is orthogonal to the vectors in the basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ for W. (See Section 6.1.) Using properties of the inner product, compute

$$\begin{aligned}\mathbf{z}_2 \cdot \mathbf{u}_1 &= (c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5) \cdot \mathbf{u}_1 \\ &= c_3\mathbf{u}_3 \cdot \mathbf{u}_1 + c_4\mathbf{u}_4 \cdot \mathbf{u}_1 + c_5\mathbf{u}_5 \cdot \mathbf{u}_1 \\ &= 0\end{aligned}$$

because $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_3$, $\mathbf{u}_4$, and $\mathbf{u}_5$. A similar calculation shows that $\mathbf{z}_2 \cdot \mathbf{u}_2 = 0$. Thus $\mathbf{z}_2$ is in $W^\perp$. ■

The next theorem shows that the decomposition $\mathbf{y} = \mathbf{z}_1 + \mathbf{z}_2$ in Example 1 can be computed without having an orthogonal basis for $\mathbb{R}^n$. It is enough to have an orthogonal basis only for W.

THEOREM 8

The Orthogonal Decomposition Theorem

Let W be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \tag{1}$$

where $\hat{\mathbf{y}}$ is in W and $\mathbf{z}$ is in $W^\perp$. In fact, if $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is any orthogonal basis of W, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y}\cdot\mathbf{u}_p}{\mathbf{u}_p\cdot\mathbf{u}_p}\mathbf{u}_p \tag{2}$$

and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.

The vector $\hat{\mathbf{y}}$ in (2) is called the **orthogonal projection of y onto** W and often is written as $\text{proj}_W\,\mathbf{y}$. See Figure 2. When W is a one-dimensional subspace, the formula for $\hat{\mathbf{y}}$ matches the formula given in Section 6.2.

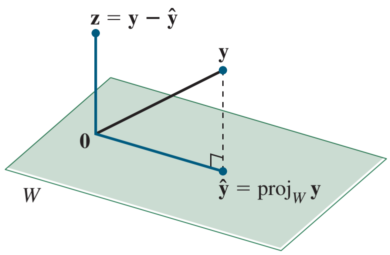


FIGURE 2 The orthogonal projection of $\mathbf{y}$ onto W.

PROOF Let $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ be any orthogonal basis for W, and define $\hat{\mathbf{y}}$ by (2).[1] Then $\hat{\mathbf{y}}$ is in W because $\hat{\mathbf{y}}$ is a linear combination of the basis $\mathbf{u}_1, \dots, \mathbf{u}_p$. Let $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$. Since $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2, \dots, \mathbf{u}_p$, it follows from (2) that

$$\begin{aligned}\mathbf{z}\cdot\mathbf{u}_1 &= (\mathbf{y} - \hat{\mathbf{y}})\cdot\mathbf{u}_1 = \mathbf{y}\cdot\mathbf{u}_1 - \left(\frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\right)\mathbf{u}_1\cdot\mathbf{u}_1 - 0 - \cdots - 0 \\ &= \mathbf{y}\cdot\mathbf{u}_1 - \mathbf{y}\cdot\mathbf{u}_1 = 0\end{aligned}$$

Thus $\mathbf{z}$ is orthogonal to $\mathbf{u}_1$. Similarly, $\mathbf{z}$ is orthogonal to each $\mathbf{u}_j$ in the basis for W. Hence $\mathbf{z}$ is orthogonal to every vector in W. That is, $\mathbf{z}$ is in $W^\perp$.

To show that the decomposition in (1) is unique, suppose $\mathbf{y}$ can also be written as $\mathbf{y} = \hat{\mathbf{y}}_1 + \mathbf{z}_1$, with $\hat{\mathbf{y}}_1$ in W and $\mathbf{z}_1$ in $W^\perp$. Then $\hat{\mathbf{y}} + \mathbf{z} = \hat{\mathbf{y}}_1 + \mathbf{z}_1$ (since both sides equal $\mathbf{y}$), and so

$$\hat{\mathbf{y}} - \hat{\mathbf{y}}_1 = \mathbf{z}_1 - \mathbf{z}$$

This equality shows that the vector $\mathbf{v} = \hat{\mathbf{y}} - \hat{\mathbf{y}}_1$ is in W and in $W^\perp$ (because $\mathbf{z}_1$ and $\mathbf{z}$ are both in $W^\perp$, and $W^\perp$ is a subspace). Hence $\mathbf{v}\cdot\mathbf{v} = 0$, which shows that $\mathbf{v} = \mathbf{0}$. This proves that $\hat{\mathbf{y}} = \hat{\mathbf{y}}_1$ and also $\mathbf{z}_1 = \mathbf{z}$. ■

The uniqueness of the decomposition (1) shows that the orthogonal projection $\hat{\mathbf{y}}$ depends only on W and not on the particular basis used in (2).

[1] We may assume that W is not the zero subspace, for otherwise $W^\perp = \mathbb{R}^n$ and (1) is simply $\mathbf{y} = \mathbf{0} + \mathbf{y}$. The next section will show that any nonzero subspace of $\mathbb{R}^n$ has an orthogonal basis.

EXAMPLE 2 Let $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Observe that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Write $\mathbf{y}$ as the sum of a vector in W and a vector orthogonal to W.

SOLUTION The orthogonal projection of $\mathbf{y}$ onto W is

$$\begin{aligned}\hat{\mathbf{y}} &= \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 \\ &= \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{15}{30}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}\end{aligned}$$

Also

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} = \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$

Theorem 8 ensures that $\mathbf{y} - \hat{\mathbf{y}}$ is in $W^{\perp}$. To check the calculations, however, it is a good idea to verify that $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to both $\mathbf{u}_1$ and $\mathbf{u}_2$ and hence to all of W. The desired decomposition of $\mathbf{y}$ is

$$\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} + \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$

■

A Geometric Interpretation of the Orthogonal Projection

When W is a one-dimensional subspace, the formula (2) for $\text{proj}_W\,\mathbf{y}$ contains just one term. Thus, when $\dim W > 1$, each term in (2) is itself an orthogonal projection of $\mathbf{y}$ onto a one-dimensional subspace spanned by one of the $\mathbf{u}$'s in the basis for W. Figure 3 illustrates this when W is a subspace of $\mathbb{R}^3$ spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$. Here $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{y}}_2$ denote the projections of $\mathbf{y}$ onto the lines spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$, respectively. The orthogonal projection $\hat{\mathbf{y}}$ of $\mathbf{y}$ onto W is the sum of the projections of $\mathbf{y}$ onto one-dimensional subspaces that are orthogonal to each other. The vector $\hat{\mathbf{y}}$ in Figure 3 corresponds to the vector $\mathbf{y}$ in Figure 4 of Section 6.2, because now it is $\hat{\mathbf{y}}$ that is in W.

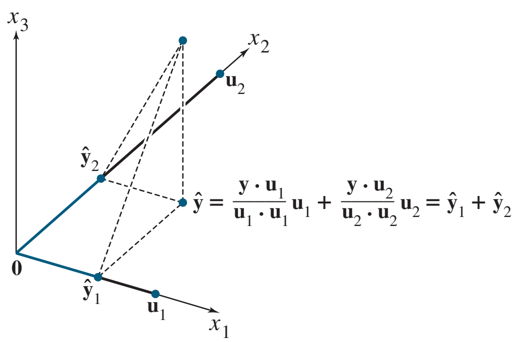


FIGURE 3 The orthogonal projection of $\mathbf{y}$ is the sum of its projections onto one-dimensional subspaces that are mutually orthogonal.

Properties of Orthogonal Projections

If $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an orthogonal basis for W and if $\mathbf{y}$ happens to be in W, then the formula for $\text{proj}_W\, \mathbf{y}$ is exactly the same as the representation of $\mathbf{y}$ given in Theorem 5 in Section 6.2. In this case, $\text{proj}_W\, \mathbf{y} = \mathbf{y}$.

If $\mathbf{y}$ is in $W = \text{Span}\,\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$, then $\text{proj}_W\, \mathbf{y} = \mathbf{y}$.

This fact also follows from the next theorem.

THEOREM 9

The Best Approximation Theorem

Let W be a subspace of $\mathbb{R}^n$, let $\mathbf{y}$ be any vector in $\mathbb{R}^n$, and let $\hat{\mathbf{y}}$ be the orthogonal projection of $\mathbf{y}$ onto W. Then $\hat{\mathbf{y}}$ is the closest point in W to $\mathbf{y}$, in the sense that

$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\| \tag{3}$$

for all $\mathbf{v}$ in W distinct from $\hat{\mathbf{y}}$.

The vector $\hat{\mathbf{y}}$ in Theorem 9 is called **the best approximation to y by elements of** W. Later sections in the text will examine problems where a given $\mathbf{y}$ must be replaced, or *approximated*, by a vector $\mathbf{v}$ in some fixed subspace W. The distance from $\mathbf{y}$ to $\mathbf{v}$, given by $\|\mathbf{y} - \mathbf{v}\|$, can be regarded as the "error" of using $\mathbf{v}$ in place of $\mathbf{y}$. Theorem 9 says that this error is minimized when $\mathbf{v} = \hat{\mathbf{y}}$.

Inequality (3) leads to a new proof that $\hat{\mathbf{y}}$ does not depend on the particular orthogonal basis used to compute it. If a different orthogonal basis for W was used to construct an orthogonal projection of $\mathbf{y}$, then this projection would also be the closest point in W to $\mathbf{y}$, namely $\hat{\mathbf{y}}$.

PROOF Take $\mathbf{v}$ in W distinct from $\hat{\mathbf{y}}$. See Figure 4. Then $\hat{\mathbf{y}} - \mathbf{v}$ is in W. By the Orthogonal Decomposition Theorem, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to W. In particular, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to $\hat{\mathbf{y}} - \mathbf{v}$ (which is in W). Since

$$\mathbf{y} - \mathbf{v} = (\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \mathbf{v})$$

the Pythagorean Theorem gives

$$\|\mathbf{y} - \mathbf{v}\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}} - \mathbf{v}\|^2$$

(See the right triangle outlined in teal in Figure 4. The length of each side is labeled.) Now $\|\hat{\mathbf{y}} - \mathbf{v}\|^2 > 0$ because $\hat{\mathbf{y}} - \mathbf{v} \neq \mathbf{0}$, and so inequality (3) follows immediately. ■

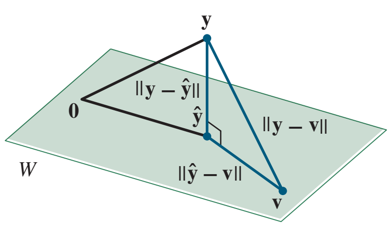


FIGURE 4 The orthogonal projection of $\mathbf{y}$ onto W is the closest point in W to $\mathbf{y}$.

EXAMPLE 3 If $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, and $W = \text{Span}\,\{\mathbf{u}_1, \mathbf{u}_2\}$, as in Example 2, then the closest point in W to $\mathbf{y}$ is

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}$$

■

EXAMPLE 4 The distance from a point $\mathbf{y}$ in $\mathbb{R}^n$ to a subspace W is defined as the distance from $\mathbf{y}$ to the nearest point in W. Find the distance from $\mathbf{y}$ to $W = \text{Span}\,\{\mathbf{u}_1, \mathbf{u}_2\}$, where

$$\mathbf{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \quad \mathbf{u}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

SOLUTION By the Best Approximation Theorem, the distance from $\mathbf{y}$ to W is $\|\mathbf{y} - \hat{\mathbf{y}}\|$, where $\hat{\mathbf{y}} = \text{proj}_W\,\mathbf{y}$. Since $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for W,

$$\hat{\mathbf{y}} = \frac{15}{30}\mathbf{u}_1 + \frac{-21}{6}\mathbf{u}_2 = \frac{1}{2}\begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} - \frac{7}{2}\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}$$

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix} - \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 6 \end{bmatrix}$$

$$\|\mathbf{y} - \hat{\mathbf{y}}\|^2 = 3^2 + 6^2 = 45$$

The distance from $\mathbf{y}$ to W is $\sqrt{45} = 3\sqrt{5}$. ■

The final theorem in this section shows how formula (2) for $\text{proj}_W\,\mathbf{y}$ is simplified when the basis for W is an orthonormal set.

THEOREM 10

If $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

$$\text{proj}_W\,\mathbf{y} = (\mathbf{y}\cdot\mathbf{u}_1)\mathbf{u}_1 + (\mathbf{y}\cdot\mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{y}\cdot\mathbf{u}_p)\mathbf{u}_p \tag{4}$$

If $U = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_p\,]$, then

$$\text{proj}_W\,\mathbf{y} = UU^T\mathbf{y} \quad \text{for all } \mathbf{y} \text{ in } \mathbb{R}^n \tag{5}$$

PROOF Formula (4) follows immediately from (2) in Theorem 8. Also, (4) shows that $\text{proj}_W\,\mathbf{y}$ is a linear combination of the columns of U using the weights $\mathbf{y}\cdot\mathbf{u}_1$, $\mathbf{y}\cdot\mathbf{u}_2, \dots, \mathbf{y}\cdot\mathbf{u}_p$. The weights can be written as $\mathbf{u}_1^T\mathbf{y}, \mathbf{u}_2^T\mathbf{y}, \dots, \mathbf{u}_p^T\mathbf{y}$, showing that they are the entries in $U^T\mathbf{y}$ and justifying (5). ■

Suppose U is an $n \times p$ matrix with orthonormal columns, and let W be the column space of U. Then

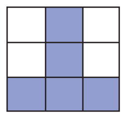
$$U^TU\mathbf{x} = I_p\mathbf{x} = \mathbf{x} \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^p \qquad \text{Theorem 6}$$

$$UU^T\mathbf{y} = \text{proj}_W\,\mathbf{y} \quad \text{for all } \mathbf{y} \text{ in } \mathbb{R}^n \qquad \text{Theorem 10}$$

If U is an $n \times n$ (square) matrix with orthonormal columns, then U is an *orthogonal* matrix, the column space W is all of $\mathbb{R}^n$, and $UU^T\mathbf{y} = I\mathbf{y} = \mathbf{y}$ for all $\mathbf{y}$ in $\mathbb{R}^n$.

Although formula (4) is important for theoretical purposes, in practice it usually involves calculations with square roots of numbers (in the entries of the $\mathbf{u}_i$). Formula (2) is recommended for hand calculations.

Example 9 of Section 2.1 illustrates how matrix multiplication and transposition are used to detect a specified pattern illustrated using blue and white squares. Now that we have more experience working with bases for W and $W^{\perp}$, we are ready to discuss how to set up the matrix M in Figure 6. Let $\mathbf{w}$ be the vector generated from a pattern of blue and white squares by turning each blue square into a 1 and each white square into a 0, and then lining up each column below the column before it. See Figure 5.



$$\rightarrow \begin{array}{|c|c|c|}\hline 0 & 1 & 0 \\ \hline 0 & 1 & 0 \\ \hline 1 & 1 & 1 \\ \hline \end{array} \rightarrow \mathbf{w} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

FIGURE 5 Creating a vector from colored squares.

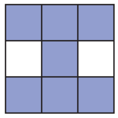
Let $W = \text{span}\,\{\mathbf{w}\}$. Choose a basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}\}$ for $W^{\perp}$. Create the matrix $B = \begin{bmatrix} \mathbf{v}_1{}^T \\ \mathbf{v}_2{}^T \\ \vdots \\ \mathbf{v}_{n-1}{}^T \end{bmatrix}$. Notice $B\mathbf{u} = \mathbf{0}$ if and only if $\mathbf{u}$ is orthogonal to a set of basis vectors for $W^{\perp}$, which happens if and only if $\mathbf{u}$ is in W. Set $M = B^T B$. Then $\mathbf{u}^T M \mathbf{u} = \mathbf{u}^T B^T B \mathbf{u} = (B\mathbf{u})^T B\mathbf{u}$. By Theorem 1, $(B\mathbf{u})^T B\mathbf{u} = 0$ if and only if $B\mathbf{u} = 0$, and hence $\mathbf{u}^T M \mathbf{u} = 0$ if and only if $\mathbf{u} \in W$. But there are only two vectors in W consisting of zeros and ones: $1\mathbf{w} = \mathbf{w}$ and $0\mathbf{w} = \mathbf{0}$. Thus we can conclude that if $\mathbf{u}^T M \mathbf{u} = 0$, but $\mathbf{u}^T \mathbf{u} \neq 0$, then $\mathbf{u} = \mathbf{w}$. See Figure 6.

EXAMPLE 5 Find a matrix M that can be used in Figure 6 to identify the perp symbol.

SOLUTION First change the symbol into a vector. Set $\mathbf{w} = [0\ \ 0\ 1\ \ 1\ \ 1\ \ 1\ \ 0\ \ 0\ \ 1]^T$. Next set $W = \text{span}\,\{\mathbf{w}\}$ and find a basis for $W^{\perp}$: solving $\mathbf{x}^T\mathbf{w} = 0$ creates the homogeneous system of equations:

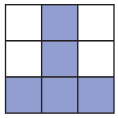
$$x_3 + x_4 + x_5 + x_6 + x_9 = 0$$

Treating x_3 as the basic variable and the remaining variables as free variables we get a basis for $W^{\perp}$. Transposing each vector in the basis and inserting it as a row of B we get



$$\mathbf{w} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \quad \Rightarrow \quad \mathbf{w}^T M \mathbf{w} = 2 \text{ and } \mathbf{w}^T\mathbf{w} = 7$$

This pattern is not the perpendicular symbol since $\mathbf{w}^T M\mathbf{w} \neq 0$.



$$\mathbf{w} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad \Rightarrow \quad \mathbf{w}^T M \mathbf{w} = 0 \text{ and } \mathbf{w}^T\mathbf{w} = 5$$

This pattern is the perpendicular symbol since $\mathbf{w}^T M\mathbf{w} = 0$, but $\mathbf{w}^T\mathbf{w} \neq 0$.

FIGURE 6 How AI detects the perp symbol.

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } M = B^T B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & -1 & -1 & -1 & 0 & 0 & -1 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Notice $\mathbf{w}^T M\mathbf{w} = 0$, but $\mathbf{w}^T\mathbf{w} \neq 0$. ■

Practice Problems

1. Let $\mathbf{u}_1 = \begin{bmatrix} -7 \\ 1 \\ 4 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -1 \\ 1 \\ -2 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} -9 \\ 1 \\ 6 \end{bmatrix}$, and $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Use the fact that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal to compute $\text{proj}_W\, \mathbf{y}$.
2. Let W be a subspace of $\mathbb{R}^n$. Let $\mathbf{x}$ and $\mathbf{y}$ be vectors in $\mathbb{R}^n$ and let $\mathbf{z} = \mathbf{x} + \mathbf{y}$. If $\mathbf{u}$ is the projection of $\mathbf{x}$ onto W and $\mathbf{v}$ is the projection of $\mathbf{y}$ onto W, show that $\mathbf{u} + \mathbf{v}$ is the projection of $\mathbf{z}$ onto W.

6.3 Exercises

In Exercises 1 and 2, you may assume that $\{\mathbf{u}_1, \dots, \mathbf{u}_4\}$ is an orthogonal basis for $\mathbb{R}^4$.

1. $\mathbf{u}_1 = \begin{bmatrix} 0 \\ 1 \\ -4 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 3 \\ 5 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ -4 \end{bmatrix}$, $\mathbf{u}_4 = \begin{bmatrix} 5 \\ -3 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 10 \\ -8 \\ 2 \\ 0 \end{bmatrix}$. Write $\mathbf{x}$ as the sum of two vectors, one in Span $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ and the other in Span $\{\mathbf{u}_4\}$.

2. $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 1 \\ 1 \\ -2 \\ -1 \end{bmatrix}$, $\mathbf{u}_4 = \begin{bmatrix} -1 \\ 1 \\ 1 \\ -2 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 4 \\ 5 \\ -2 \\ 2 \end{bmatrix}$. Write $\mathbf{v}$ as the sum of two vectors, one in Span $\{\mathbf{u}_1\}$ and the other in Span $\{\mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$.

In Exercises 3–6, verify that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal set, and then find the orthogonal projection of $\mathbf{y}$ onto Span $\{\mathbf{u}_1, \mathbf{u}_2\}$.

3. $\mathbf{y} = \begin{bmatrix} -1 \\ 4 \\ 3 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$

4. $\mathbf{y} = \begin{bmatrix} 4 \\ 3 \\ -2 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -4 \\ 3 \\ 0 \end{bmatrix}$

5. $\mathbf{y} = \begin{bmatrix} -1 \\ 2 \\ 6 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}$

6. $\mathbf{y} = \begin{bmatrix} -1 \\ 5 \\ 3 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 4 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 1 \\ -1 \\ -5 \end{bmatrix}$

In Exercises 7–10, let W be the subspace spanned by the $\mathbf{u}$'s, and write $\mathbf{y}$ as the sum of a vector in W and a vector orthogonal to W.

7. $\mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 5 \\ 1 \\ 4 \end{bmatrix}$

8. $\mathbf{y} = \begin{bmatrix} -1 \\ 6 \\ 4 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -1 \\ 4 \\ -3 \end{bmatrix}$

9. $\mathbf{y} = \begin{bmatrix} 4 \\ 3 \\ 3 \\ -1 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -1 \\ 3 \\ 1 \\ -2 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$

10. $\mathbf{y} = \begin{bmatrix} 3 \\ 4 \\ 5 \\ 4 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 0 \\ -1 \\ 1 \\ -1 \end{bmatrix}$

In Exercises 11 and 12, find the closest point to $\mathbf{y}$ in the subspace W spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$.

11. $\mathbf{y} = \begin{bmatrix} 3 \\ 1 \\ 5 \\ 1 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}$

12. $\mathbf{y} = \begin{bmatrix} 4 \\ 3 \\ 4 \\ 7 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \\ -2 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \end{bmatrix}$

In Exercises 13 and 14, find the best approximation to $\mathbf{z}$ by vectors of the form $c_1\mathbf{v}_1 + c_2\mathbf{v}_2$.

13. $\mathbf{z} = \begin{bmatrix} 3 \\ -7 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 2 \\ -1 \\ -3 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}$

14. $\mathbf{z} = \begin{bmatrix} 2 \\ 4 \\ 0 \\ -1 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 2 \\ 0 \\ -1 \\ -3 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 5 \\ -2 \\ 4 \\ 2 \end{bmatrix}$

15. Let $\mathbf{y} = \begin{bmatrix} 5 \\ -9 \\ 5 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} -3 \\ -5 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}$. Find the distance from $\mathbf{y}$ to the plane in $\mathbb{R}^3$ spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$.

16. Let $\mathbf{y}$, $\mathbf{v}_1$, and $\mathbf{v}_2$ be as in Exercise 12. Find the distance from $\mathbf{y}$ to the subspace of $\mathbb{R}^4$ spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$.

17. Let $\mathbf{y} = \begin{bmatrix} 4 \\ 8 \\ 1 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 2/3 \\ 1/3 \\ 2/3 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2/3 \\ 2/3 \\ 1/3 \end{bmatrix}$, and $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$.

a. Let $U = [\,\mathbf{u}_1 \quad \mathbf{u}_2\,]$. Compute U^TU and UU^T.

b. Compute $\text{proj}_W\,\mathbf{y}$ and $(UU^T)\mathbf{y}$.

18. Let $\mathbf{y} = \begin{bmatrix} 7 \\ 9 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{bmatrix}$, and $W = \text{Span}\{\mathbf{u}_1\}$.

a. Let U be the 2×1 matrix whose only column is $\mathbf{u}_1$. Compute U^TU and UU^T.

b. Compute $\text{proj}_W\,\mathbf{y}$ and $(UU^T)\mathbf{y}$.

19. Let $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$, and $\mathbf{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Note that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal but that $\mathbf{u}_3$ is not orthogonal to $\mathbf{u}_1$ or $\mathbf{u}_2$. It can be shown that $\mathbf{u}_3$ is not in the subspace W spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$. Use this fact to construct a nonzero vector $\mathbf{v}$ in $\mathbb{R}^3$ that is orthogonal to $\mathbf{u}_1$ and $\mathbf{u}_2$.

20. Let $\mathbf{u}_1$ and $\mathbf{u}_2$ be as in Exercise 19, and let $\mathbf{u}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. It can be shown that $\mathbf{u}_4$ is not in the subspace W spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$. Use this fact to construct a nonzero vector $\mathbf{v}$ in $\mathbb{R}^3$ that is orthogonal to $\mathbf{u}_1$ and $\mathbf{u}_2$.

In Exercises 21–30, all vectors and subspaces are in $\mathbb{R}^n$. Mark each statement True or False **(T/F)**. Justify each answer.

21. **(T/F)** If $\mathbf{z}$ is orthogonal to $\mathbf{u}_1$ and to $\mathbf{u}_2$ and if $W = \text{Span}\,\{\mathbf{u}_1, \mathbf{u}_2\}$, then $\mathbf{z}$ must be in $W^{\perp}$.

22. **(T/F)** For each $\mathbf{y}$ and each subspace W, the vector $\mathbf{y} - \text{proj}_W \mathbf{y}$ is orthogonal to W.

23. **(T/F)** The orthogonal projection $\hat{\mathbf{y}}$ of $\mathbf{y}$ onto a subspace W can sometimes depend on the orthogonal basis for W used to compute $\hat{\mathbf{y}}$.

24. **(T/F)** If $\mathbf{y}$ is in a subspace W, then the orthogonal projection of $\mathbf{y}$ onto W is $\mathbf{y}$ itself.

25. **(T/F)** The best approximation to $\mathbf{y}$ by elements of a subspace W is given by the vector $\mathbf{y} - \text{proj}_W \mathbf{y}$.

26. **(T/F)** If W is a subspace of $\mathbb{R}^n$ and if $\mathbf{v}$ is in both W and $W^{\perp}$, then $\mathbf{v}$ must be the zero vector.

27. **(T/F)** In the Orthogonal Decomposition Theorem, each term in formula (2) for $\hat{\mathbf{y}}$ is itself an orthogonal projection of $\mathbf{y}$ onto a subspace of W.

28. **(T/F)** If $\mathbf{y} = \mathbf{z}_1 + \mathbf{z}_2$, where $\mathbf{z}_1$ is in a subspace W and $\mathbf{z}_2$ is in $W^{\perp}$, then $\mathbf{z}_1$ must be the orthogonal projection of $\mathbf{y}$ onto W.

29. **(T/F)** If the columns of an $n \times p$ matrix U are orthonormal, then $UU^T\mathbf{y}$ is the orthogonal projection of $\mathbf{y}$ onto the column space of U.

30. **(T/F)** If an $n \times p$ matrix U has orthonormal columns, then $UU^T\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

31. Let A be an $m \times n$ matrix. Prove that every vector $\mathbf{x}$ in $\mathbb{R}^n$ can be written in the form $\mathbf{x} = \mathbf{p} + \mathbf{u}$, where $\mathbf{p}$ is in Row A and $\mathbf{u}$ is in Nul A. Also, show that if the equation $A\mathbf{x} = \mathbf{b}$ is consistent, then there is a unique $\mathbf{p}$ in Row A such that $A\mathbf{p} = \mathbf{b}$.

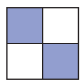
32. Let W be a subspace of $\mathbb{R}^n$ with an orthogonal basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_p\}$, and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_q\}$ be an orthogonal basis for $W^{\perp}$.

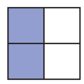
a. Explain why $\{\mathbf{w}_1, \ldots, \mathbf{w}_p, \mathbf{v}_1, \ldots, \mathbf{v}_q\}$ is an orthogonal set.

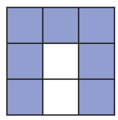
b. Explain why the set in part (a) spans $\mathbb{R}^n$.

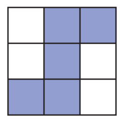
c. Show that $\dim W + \dim W^{\perp} = n$.

In Exercises 33–36, first change the given pattern into a vector $\mathbf{w}$ of zeros and ones and then use the method illustrated in Example 5 to find a matrix M so that $\mathbf{w}^T M \mathbf{w} = 0$, but $\mathbf{u}^T M \mathbf{u} \neq 0$ for all other nonzero vectors $\mathbf{u}$ of zeros and ones.

33. 

34. 

35. 

36. 

T 37. Let U be the 8×4 matrix in Exercise 43 in Section 6.2. Find the closest point to $\mathbf{y} = (1, 1, 1, 1, 1, 1, 1, 1)$ in Col U. Write the keystrokes or commands you use to solve this problem.

T 38. Let U be the matrix in Exercise 37. Find the distance from $\mathbf{b} = (1, 1, 1, 1, -1, -1, -1, -1)$ to Col U.

Solution to Practice Problems

1. Compute

$$\text{proj}_W \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 = \frac{88}{66}\mathbf{u}_1 + \frac{-2}{6}\mathbf{u}_2$$

$$= \frac{4}{3}\begin{bmatrix} -7 \\ 1 \\ 4 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} -1 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 1 \\ 6 \end{bmatrix} = \mathbf{y}$$

In this case, $\mathbf{y}$ happens to be a linear combination of $\mathbf{u}_1$ and $\mathbf{u}_2$, so $\mathbf{y}$ is in W. The closest point in W to $\mathbf{y}$ is $\mathbf{y}$ itself.

2. Using Theorem 10, let U be a matrix whose columns consist of an orthonormal basis for W. Then $\text{proj}_W\, \mathbf{z} = UU^T\, \mathbf{z} = UU^T\,(\mathbf{x} + \mathbf{y}) = UU^T\mathbf{x} + UU^T\mathbf{y} = \text{proj}_W \mathbf{x} + \text{proj}_W \mathbf{y} = \mathbf{u} + \mathbf{v}$.

6.4 The Gram–Schmidt Process

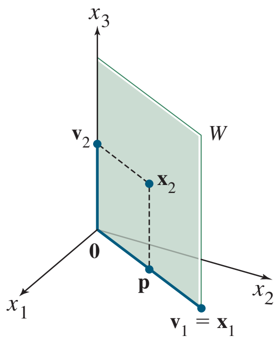


FIGURE 1
Construction of an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$.

The Gram–Schmidt process is a simple algorithm for producing an orthogonal or orthonormal basis for any nonzero subspace of $\mathbb{R}^n$. The first two examples of the process are aimed at hand calculation.

EXAMPLE 1 Let $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$, where $\mathbf{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$. Construct an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ for W.

SOLUTION The subspace W is shown in Figure 1, along with $\mathbf{x}_1$, $\mathbf{x}_2$, and the projection $\mathbf{p}$ of $\mathbf{x}_2$ onto $\mathbf{x}_1$. The component of $\mathbf{x}_2$ orthogonal to $\mathbf{x}_1$ is $\mathbf{x}_2 - \mathbf{p}$, which is in W because it is formed from $\mathbf{x}_2$ and a multiple of $\mathbf{x}_1$. Let $\mathbf{v}_1 = \mathbf{x}_1$ and

$$\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p} = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{x}_1}{\mathbf{x}_1 \cdot \mathbf{x}_1}\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \frac{15}{45}\begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

Then $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal set of nonzero vectors in W. Since $\dim W = 2$, the set $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for W. ■

The next example fully illustrates the Gram–Schmidt process. Study it carefully.

EXAMPLE 2 Let $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$. Then $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is clearly linearly independent and thus is a basis for a subspace W of $\mathbb{R}^4$. Construct an orthogonal basis for W.

SOLUTION

Step 1. Let $\mathbf{v}_1 = \mathbf{x}_1$ and $W_1 = \text{Span}\{\mathbf{x}_1\} = \text{Span}\{\mathbf{v}_1\}$.

Step 2. Let $\mathbf{v}_2$ be the vector produced by subtracting from $\mathbf{x}_2$ its projection onto the subspace W_1. That is, let

$$\begin{aligned} \mathbf{v}_2 &= \mathbf{x}_2 - \text{proj}_{W_1} \mathbf{x}_2 \\ &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \qquad \text{Since } \mathbf{v}_1 = \mathbf{x}_1 \\ &= \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} \end{aligned}$$

As in Example 1, $\mathbf{v}_2$ is the component of $\mathbf{x}_2$ orthogonal to $\mathbf{x}_1$, and $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis for the subspace W_2 spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$.

Step 2′ (optional). If appropriate, scale $\mathbf{v}_2$ to simplify later computations. Since $\mathbf{v}_2$ has fractional entries, it is convenient to scale it by a factor of 4 and replace $\{\mathbf{v}_1, \mathbf{v}_2\}$ by the orthogonal basis

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2' = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Step 3. Let $\mathbf{v}_3$ be the vector produced by subtracting from $\mathbf{x}_3$ its projection onto the subspace W_2. Use the orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2'\}$ to compute this projection onto W_2:

$$\text{proj}_{W_2}\,\mathbf{x}_3 = \underbrace{\frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1}_{\substack{\text{Projection of}\\ \mathbf{x}_3 \text{ onto } \mathbf{v}_1}} + \underbrace{\frac{\mathbf{x}_3 \cdot \mathbf{v}_2'}{\mathbf{v}_2' \cdot \mathbf{v}_2'}\mathbf{v}_2'}_{\substack{\text{Projection of}\\ \mathbf{x}_3 \text{ onto } \mathbf{v}_2'}} = \frac{2}{4}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} + \frac{2}{12}\begin{bmatrix}-3\\1\\1\\1\end{bmatrix} = \begin{bmatrix}0\\2/3\\2/3\\2/3\end{bmatrix}$$

Then $\mathbf{v}_3$ is the component of $\mathbf{x}_3$ orthogonal to W_2, namely

$$\mathbf{v}_3 = \mathbf{x}_3 - \text{proj}_{W_2}\,\mathbf{x}_3 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix} - \begin{bmatrix}0\\2/3\\2/3\\2/3\end{bmatrix} = \begin{bmatrix}0\\-2/3\\1/3\\1/3\end{bmatrix}$$

See Figure 2 for a diagram of this construction. Observe that $\mathbf{v}_3$ is in W, because $\mathbf{x}_3$ and $\text{proj}_{W_2}\mathbf{x}_3$ are both in W. Thus $\{\mathbf{v}_1, \mathbf{v}_2', \mathbf{v}_3\}$ is an orthogonal set of nonzero vectors and hence a linearly independent set in W. Note that W is three-dimensional since it was defined by a basis of three vectors. Hence, by the Basis Theorem in Section 4.5, $\{\mathbf{v}_1, \mathbf{v}_2', \mathbf{v}_3\}$ is an orthogonal basis for W. ■

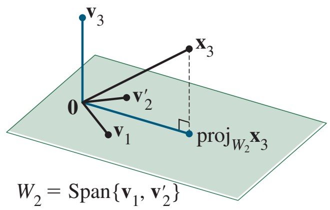


FIGURE 2 The construction of $\mathbf{v}_3$ from $\mathbf{x}_3$ and W_2.

The proof of the next theorem shows that this strategy really works. Scaling of vectors is not mentioned because that is used only to simplify hand calculations.

THEOREM 11

The Gram–Schmidt Process

Given a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\mathbf{v}_{p-1}
\end{aligned}$$

Then $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is an orthogonal basis for W. In addition

$$\text{Span}\,\{\mathbf{v}_1, \dots, \mathbf{v}_k\} = \text{Span}\,\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \qquad \text{for } 1 \le k \le p \tag{1}$$

PROOF For $1 \leq k \leq p$, let $W_k = \text{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$. Set $\mathbf{v}_1 = \mathbf{x}_1$, so that $\text{Span}\{\mathbf{v}_1\} = \text{Span}\{\mathbf{x}_1\}$. Suppose, for some $k < p$, we have constructed $\mathbf{v}_1, \ldots, \mathbf{v}_k$ so that $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthogonal basis for W_k. Define

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - \text{proj}_{W_k} \mathbf{x}_{k+1} \tag{2}$$

By the Orthogonal Decomposition Theorem, $\mathbf{v}_{k+1}$ is orthogonal to W_k. Note that $\text{proj}_{W_k} \mathbf{x}_{k+1}$ is in W_k and hence also in W_{k+1}. Since $\mathbf{x}_{k+1}$ is in W_{k+1}, so is $\mathbf{v}_{k+1}$ (because W_{k+1} is a subspace and is closed under subtraction). Furthermore, $\mathbf{v}_{k+1} \neq \mathbf{0}$ because $\mathbf{x}_{k+1}$ is not in $W_k = \text{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$. Hence $\{\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}\}$ is an orthogonal set of nonzero vectors in the $(k+1)$-dimensional space W_{k+1}. By the Basis Theorem in Section 4.5, this set is an orthogonal basis for W_{k+1}. Hence $W_{k+1} = \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}\}$. When $k + 1 = p$, the process stops. ■

Theorem 11 shows that any nonzero subspace W of $\mathbb{R}^n$ has an orthogonal basis, because an ordinary basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ is always available (by Theorem 12 in Section 4.5), and the Gram–Schmidt process depends only on the existence of orthogonal projections onto subspaces of W that already have orthogonal bases.

Orthonormal Bases

An orthonormal basis is constructed easily from an orthogonal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$: simply normalize (i.e., "scale") all the $\mathbf{v}_k$. When working problems by hand, this is easier than normalizing each $\mathbf{v}_k$ as soon as it is found (because it avoids unnecessary writing of square roots).

EXAMPLE 3 Example 1 constructed the orthogonal basis

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

An orthonormal basis is

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \frac{1}{\sqrt{45}}\begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \\ 0 \end{bmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$ ■

QR Factorization of Matrices

If an $m \times n$ matrix A has linearly independent columns $\mathbf{x}_1, \ldots, \mathbf{x}_n$, then applying the Gram–Schmidt process (with normalizations) to $\mathbf{x}_1, \ldots, \mathbf{x}_n$ amounts to *factoring* A, as described in the next theorem. This factorization is widely used in computer algorithms for various computations, such as solving equations (discussed in Section 6.5) and finding eigenvalues (mentioned in the exercises for Section 5.2).

THEOREM 12

The QR Factorization

If A is an $m \times n$ matrix with linearly independent columns, then A can be factored as $A = QR$, where Q is an $m \times n$ matrix whose columns form an orthonormal basis for Col A and R is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

PROOF The columns of A form a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ for Col A. Construct an orthonormal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ for $W = \text{Col } A$ with property (1) in Theorem 11. This basis may be constructed by the Gram–Schmidt process or some other means. Let

$$Q = [\, \mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_n \,]$$

For $k = 1, \dots, n, \mathbf{x}_k$ is in Span $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ = Span $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$. So there are constants, $r_{1k}, \dots, r_{kk}$, such that

$$\mathbf{x}_k = r_{1k}\mathbf{u}_1 + \cdots + r_{kk}\mathbf{u}_k + 0\,\mathbf{u}_{k+1} + \cdots + 0\,\mathbf{u}_n$$

We may assume that $r_{kk} \geq 0$. (If $r_{kk} < 0$, multiply both r_{kk} and $\mathbf{u}_k$ by -1.) This shows that $\mathbf{x}_k$ is a linear combination of the columns of Q using as weights the entries in the vector

$$\mathbf{r}_k = \begin{bmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

That is, $\mathbf{x}_k = Q\mathbf{r}_k$ for $k = 1, \dots, n$. Let $R = [\, \mathbf{r}_1 \quad \cdots \quad \mathbf{r}_n \,]$. Then

$$A = [\, \mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n \,] = [\, Q\mathbf{r}_1 \quad \cdots \quad Q\mathbf{r}_n \,] = QR$$

The fact that R is invertible follows easily from the fact that the columns of A are linearly independent (Exercise 23). Since R is clearly upper triangular, its nonnegative diagonal entries must be positive. ■

EXAMPLE 4 Find a QR factorization of $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.

SOLUTION The columns of A are the vectors $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ in Example 2. An orthogonal basis for Col A = Span $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ was found in that example:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2' = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

To simplify the arithmetic that follows, scale $\mathbf{v}_3$ by letting $\mathbf{v}_3' = 3\mathbf{v}_3$. Then normalize the three vectors to obtain $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$, and use these vectors as the columns of Q:

$$Q = \begin{bmatrix} 1/2 & -3/\sqrt{12} & 0 \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \end{bmatrix}$$

By construction, the first k columns of Q are an orthonormal basis of Span $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$. From the proof of Theorem 12, $A = QR$ for some R. To find R, observe that $Q^TQ = I$, because the columns of Q are orthonormal. Hence

$$Q^TA = Q^T(QR) = IR = R$$

and

$$R = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ -3/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} \\ 0 & -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 3/2 & 1 \\ 0 & 3/\sqrt{12} & 2/\sqrt{12} \\ 0 & 0 & 2/\sqrt{6} \end{bmatrix}$$

■

Numerical Notes

1. When the Gram–Schmidt process is run on a computer, roundoff error can build up as the vectors $\mathbf{u}_k$ are calculated, one by one. For j and k large but unequal, the inner products $\mathbf{u}_j^T\mathbf{u}_k$ may not be sufficiently close to zero. This loss of orthogonality can be reduced substantially by rearranging the order of the calculations.[1] However, a different computer-based QR factorization is usually preferred to this modified Gram–Schmidt method because it yields a more accurate orthonormal basis, even though the factorization requires about twice as much arithmetic.
2. To produce a QR factorization of a matrix A, a computer program usually left-multiplies A by a sequence of orthogonal matrices until A is transformed into an upper triangular matrix. This construction is analogous to the left-multiplication by elementary matrices that produces an LU factorization of A.

Practice Problems

1. Let $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$, where $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1/3 \\ 1/3 \\ -2/3 \end{bmatrix}$. Construct an orthonormal basis for W.
2. Suppose $A = QR$, where Q is an $m \times n$ matrix with orthogonal columns and R is an $n \times n$ matrix. Show that if the columns of A are linearly dependent, then R cannot be invertible.

6.4 Exercises

In Exercises 1–6, the given set is a basis for a subspace W. Use the Gram–Schmidt process to produce an orthogonal basis for W.

1. $\begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 8 \\ 5 \\ -6 \end{bmatrix}$

2. $\begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 6 \\ -7 \end{bmatrix}$

3. $\begin{bmatrix} 2 \\ -5 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ -1 \\ 2 \end{bmatrix}$

4. $\begin{bmatrix} 4 \\ -5 \\ 6 \end{bmatrix}, \begin{bmatrix} -8 \\ 17 \\ -19 \end{bmatrix}$

5. $\begin{bmatrix} 1 \\ -4 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 7 \\ -7 \\ -4 \\ 1 \end{bmatrix}$

6. $\begin{bmatrix} 1 \\ -2 \\ 1 \\ -2 \end{bmatrix}, \begin{bmatrix} 5 \\ -6 \\ 7 \\ -8 \end{bmatrix}$

[1] See *Fundamentals of Matrix Computations*, by David S. Watkins (New York: John Wiley & Sons, 1991), pp. 167–180.

7. Find an orthonormal basis of the subspace spanned by the vectors in Exercise 3.

8. Find an orthonormal basis of the subspace spanned by the vectors in Exercise 4.

Find an orthogonal basis for the column space of each matrix in Exercises 9–12.

9. $\begin{bmatrix} 3 & -5 & 1 \\ 1 & 1 & 1 \\ -1 & 5 & -2 \\ 3 & -7 & 8 \end{bmatrix}$ **10.** $\begin{bmatrix} -1 & 6 & 6 \\ 3 & -8 & 3 \\ 1 & -2 & 6 \\ 1 & -4 & -3 \end{bmatrix}$

11. $\begin{bmatrix} 1 & 2 & 5 \\ -1 & 1 & -4 \\ -1 & 4 & -3 \\ 1 & -4 & 7 \\ 1 & 2 & 1 \end{bmatrix}$ **12.** $\begin{bmatrix} 1 & 2 & 4 \\ -1 & -3 & -3 \\ 0 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 2 & 4 \end{bmatrix}$

In Exercises 13 and 14, the columns of Q were obtained by applying the Gram–Schmidt process to the columns of A. Find an upper triangular matrix R such that $A = QR$. Check your work.

13. $A = \begin{bmatrix} 5 & 9 \\ 1 & 7 \\ -3 & -5 \\ 1 & 5 \end{bmatrix}, Q = \begin{bmatrix} 5/6 & -1/6 \\ 1/6 & 5/6 \\ -3/6 & 1/6 \\ 1/6 & 3/6 \end{bmatrix}$

14. $A = \begin{bmatrix} -2 & 3 \\ 5 & 7 \\ 2 & -2 \\ 4 & 6 \end{bmatrix}, Q = \begin{bmatrix} -2/7 & 5/7 \\ 5/7 & 2/7 \\ 2/7 & -4/7 \\ 4/7 & 2/7 \end{bmatrix}$

15. Find a QR factorization of the matrix in Exercise 11.

16. Find a QR factorization of the matrix in Exercise 12.

In Exercises 17–22, all vectors and subspaces are in $\mathbb{R}^n$. Mark each statement True or False **(T/F)**. Justify each answer.

17. **(T/F)** If $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal basis for W, then multiplying $\mathbf{v}_3$ by a scalar c gives a new orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, c\mathbf{v}_3\}$.

18. **(T/F)** If $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ with $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ linearly independent, and if $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal set in W, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a basis for W.

19. **(T/F)** The Gram–Schmidt process produces from a linearly independent set $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ an orthogonal set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ with the property that for each k, the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ span the same subspace as that spanned by $\mathbf{x}_1, \dots, \mathbf{x}_k$.

20. **(T/F)** If $\mathbf{x}$ is not in a subspace W, then $\mathbf{x} - \text{proj}_W \mathbf{x}$ is not zero.

21. **(T/F)** If $A = QR$, where Q has orthonormal columns, then $R = Q^T A$.

22. **(T/F)** In a QR factorization, say $A = QR$ (when A has linearly independent columns), the columns of Q form an orthonormal basis for the column space of A.

23. Suppose $A = QR$, where Q is $m \times n$ and R is $n \times n$. Show that if the columns of A are linearly independent, then R must be invertible. [*Hint:* Study the equation $R\mathbf{x} = \mathbf{0}$ and use the fact that $A = QR$.]

24. Suppose $A = QR$, where R is an invertible matrix. Show that A and Q have the same column space. [*Hint:* Given $\mathbf{y}$ in Col A, show that $\mathbf{y} = Q\mathbf{x}$ for some $\mathbf{x}$. Also, given $\mathbf{y}$ in Col Q, show that $\mathbf{y} = A\mathbf{x}$ for some $\mathbf{x}$.]

25. Given $A = QR$ as in Theorem 12, describe how to find an orthogonal $m \times m$ (square) matrix Q_1 and an invertible $n \times n$ upper triangular matrix R such that

$$A = Q_1 \begin{bmatrix} R \\ 0 \end{bmatrix}$$

The MATLAB `qr` command supplies this "full" QR factorization when rank $A = n$.

26. Let $\mathbf{u}_1, \dots, \mathbf{u}_p$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$, and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be defined by $T(\mathbf{x}) = \text{proj}_W \mathbf{x}$. Show that T is a linear transformation.

27. Suppose $A = QR$ is a QR factorization of an $m \times n$ matrix A (with linearly independent columns). Partition A as $[A_1 \quad A_2]$, where A_1 has p columns. Show how to obtain a QR factorization of A_1, and explain why your factorization has the appropriate properties.

T **28.** Use the Gram–Schmidt process as in Example 2 to produce an orthogonal basis for the column space of

$$A = \begin{bmatrix} -10 & 13 & 7 & -11 \\ 2 & 1 & -5 & 3 \\ -6 & 3 & 13 & -3 \\ 16 & -16 & -2 & 5 \\ 2 & 1 & -5 & -7 \end{bmatrix}$$

T **29.** Use the method in this section to produce a QR factorization of the matrix in Exercise 28.

T **30.** For a matrix program, the Gram–Schmidt process works better with orthonormal vectors. Starting with $\mathbf{x}_1, \dots, \mathbf{x}_p$ as in Theorem 11, let $A = [\,\mathbf{x}_1 \ \cdots \ \mathbf{x}_p\,]$. Suppose Q is an $n \times k$ matrix whose columns form an orthonormal basis for the subspace W_k spanned by the first k columns of A. Then for $\mathbf{x}$ in $\mathbb{R}^n$, $QQ^T\mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ onto W_k (Theorem 10 in Section 6.3). If $\mathbf{x}_{k+1}$ is the next column of A, then equation (2) in the proof of Theorem 11 becomes

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - Q(Q^T\mathbf{x}_{k+1})$$

(The parentheses above reduce the number of arithmetic operations.) Let $\mathbf{u}_{k+1} = \mathbf{v}_{k+1}/\|\mathbf{v}_{k+1}\|$. The new Q for the next step is $[\,Q \quad \mathbf{u}_{k+1}\,]$. Use this procedure to compute the QR factorization of the matrix in Exercise 28. Write the keystrokes or commands you use.

Solution to Practice Problems

1. Let $\mathbf{v}_1 = \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \mathbf{x}_2 - \dfrac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \mathbf{x}_2 - 0\mathbf{v}_1 = \mathbf{x}_2$. So $\{\mathbf{x}_1, \mathbf{x}_2\}$ is already orthogonal. All that is needed is to normalize the vectors. Let

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

Instead of normalizing $\mathbf{v}_2$ directly, normalize $\mathbf{v}_2' = 3\mathbf{v}_2$ instead:

$$\mathbf{u}_2 = \frac{1}{\|\mathbf{v}_2'\|}\mathbf{v}_2' = \frac{1}{\sqrt{1^2 + 1^2 + (-2)^2}}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{bmatrix}$$

Then $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthonormal basis for W.

2. Since the columns of A are linearly dependent, there is a nontrivial vector $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$. But then $QR\mathbf{x} = \mathbf{0}$. Applying Theorem 7 from Section 6.2 results in $\|R\mathbf{x}\| = \|QR\mathbf{x}\| = \|\mathbf{0}\| = 0$. But $\|R\mathbf{x}\| = 0$ implies $R\mathbf{x} = \mathbf{0}$, by Theorem 1 from Section 6.1. Thus there is a nontrivial vector $\mathbf{x}$ such that $R\mathbf{x} = \mathbf{0}$ and hence, by the Invertible Matrix Theorem, R cannot be invertible.

6.5 Least-Squares Problems

Inconsistent systems arise often in applications. When a solution is demanded and none exists, the best one can do is to find an $\mathbf{x}$ that makes $A\mathbf{x}$ as close as possible to $\mathbf{b}$.

Think of $A\mathbf{x}$ as an *approximation* to $\mathbf{b}$. The smaller the distance between $\mathbf{b}$ and $A\mathbf{x}$, given by $\|\mathbf{b} - A\mathbf{x}\|$, the better the approximation. The **general least-squares problem** is to find an $\mathbf{x}$ that makes $\|\mathbf{b} - A\mathbf{x}\|$ as small as possible. The adjective "least-squares" arises from the fact that $\|\mathbf{b} - A\mathbf{x}\|$ is the square root of a sum of squares.

DEFINITION

If A is $m \times n$ and $\mathbf{b}$ is in $\mathbb{R}^m$, a **least-squares solution** of $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \le \|\mathbf{b} - A\mathbf{x}\|$$

for all $\mathbf{x}$ in $\mathbb{R}^n$.

The most important aspect of the least-squares problem is that no matter what $\mathbf{x}$ we select, the vector $A\mathbf{x}$ will necessarily be in the column space, Col A. So we seek an $\mathbf{x}$ that makes $A\mathbf{x}$ the closest point in Col A to $\mathbf{b}$. See Figure 1. (Of course, if $\mathbf{b}$ happens to be in Col A, then $\mathbf{b}$ *is* $A\mathbf{x}$ for some $\mathbf{x}$, and such an $\mathbf{x}$ is a "least-squares solution.")

Solution of the General Least-Squares Problem

Given A and $\mathbf{b}$ as above, apply the Best Approximation Theorem in Section 6.3 to the subspace Col A. Let

$$\hat{\mathbf{b}} = \text{proj}_{\text{Col}\,A}\,\mathbf{b}$$

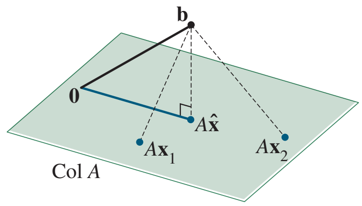


FIGURE 1 The vector $\mathbf{b}$ is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other $\mathbf{x}$.

Because $\hat{\mathbf{b}}$ is in the column space of A, the equation $A\mathbf{x} = \hat{\mathbf{b}}$ *is* consistent, and there is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$A\hat{\mathbf{x}} = \hat{\mathbf{b}} \tag{1}$$

Since $\hat{\mathbf{b}}$ is the closest point in Col A to $\mathbf{b}$, a vector $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if $\hat{\mathbf{x}}$ satisfies (1). Such an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ is a list of weights that will build $\hat{\mathbf{b}}$ out of the columns of A. See Figure 2. [There are many solutions of (1) if the equation has free variables.]

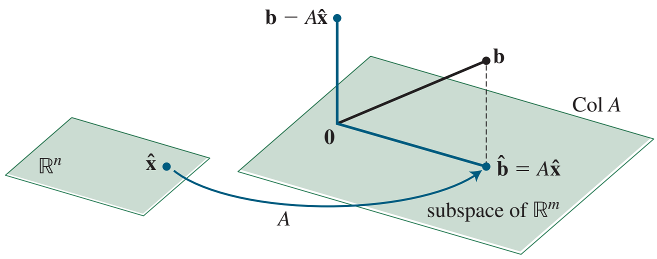


FIGURE 2 The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

Suppose $\hat{\mathbf{x}}$ satisfies $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. By the Orthogonal Decomposition Theorem in Section 6.3, the projection $\hat{\mathbf{b}}$ has the property that $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to Col A, so $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to each column of A. If $\mathbf{a}_j$ is any column of A, then $\mathbf{a}_j \cdot (\mathbf{b} - A\hat{\mathbf{x}}) = 0$, and $\mathbf{a}_j^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0$. Since each $\mathbf{a}_j^T$ is a row of A^T,

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0} \tag{2}$$

(This equation also follows from Theorem 3 in Section 6.1.) Thus

$$\begin{aligned} A^T\mathbf{b} - A^TA\hat{\mathbf{x}} &= \mathbf{0} \\ A^TA\hat{\mathbf{x}} &= A^T\mathbf{b} \end{aligned}$$

These calculations show that each least-squares solution of $A\mathbf{x} = \mathbf{b}$ satisfies the equation

$$A^TA\mathbf{x} = A^T\mathbf{b} \tag{3}$$

The matrix equation (3) represents a system of equations called the **normal equations** for $A\mathbf{x} = \mathbf{b}$. A solution of (3) is often denoted by $\hat{\mathbf{x}}$.

THEOREM 13

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$.

PROOF As shown, the set of least-squares solutions is nonempty and each least-squares solution $\hat{\mathbf{x}}$ satisfies the normal equations. Conversely, suppose $\hat{\mathbf{x}}$ satisfies $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$. Then $\hat{\mathbf{x}}$ satisfies (2), which shows that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the rows of A^T and hence is orthogonal to the columns of A. Since the columns of A span Col A, the vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to all of Col A. Hence the equation

$$\mathbf{b} = A\hat{\mathbf{x}} + (\mathbf{b} - A\hat{\mathbf{x}})$$

is a decomposition of $\mathbf{b}$ into the sum of a vector in Col A and a vector orthogonal to Col A. By the uniqueness of the orthogonal decomposition, $A\hat{\mathbf{x}}$ must be the orthogonal projection of $\mathbf{b}$ onto Col A. That is, $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, and $\hat{\mathbf{x}}$ is a least-squares solution. ∎

EXAMPLE 1 Find a least-squares solution of the inconsistent system $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

SOLUTION To use normal equations (3), compute:

$$A^TA = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$A^T\mathbf{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

Then the equation $A^TA\mathbf{x} = A^T\mathbf{b}$ becomes

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

Row operations can be used to solve this system, but since A^TA is invertible and 2×2, it is probably faster to compute

$$(A^TA)^{-1} = \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$$

and then to solve $A^TA\mathbf{x} = A^T\mathbf{b}$ as

$$\begin{aligned} \hat{\mathbf{x}} &= (A^TA)^{-1}A^T\mathbf{b} \\ &= \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix} \begin{bmatrix} 19 \\ 11 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 84 \\ 168 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \end{aligned}$$

∎

In many calculations, A^TA is invertible, but this is not always the case. The next example involves a matrix of the sort that appears in what are called *analysis of variance* problems in statistics.

EXAMPLE 2 Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}$$

SOLUTION Compute

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}$$

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}$$

The augmented matrix for $A^TA\mathbf{x} = A^T\mathbf{b}$ is

$$\begin{bmatrix} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The general solution is $x_1 = 3 - x_4$, $x_2 = -5 + x_4$, $x_3 = -2 + x_4$, and x_4 is free. So the general least-squares solution of $A\mathbf{x} = \mathbf{b}$ has the form

$$\hat{\mathbf{x}} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

■

The next theorem gives useful criteria for determining when there is only one least-squares solution of $A\mathbf{x} = \mathbf{b}$. (Of course, the orthogonal projection $\hat{\mathbf{b}}$ is always unique.)

THEOREM 14

Let A be an $m \times n$ matrix. The following statements are logically equivalent:

a. The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$.

b. The columns of A are linearly independent.

c. The matrix A^TA is invertible.

When these statements are true, the least-squares solution $\hat{\mathbf{x}}$ is given by

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} \tag{4}$$

The main elements of a proof of Theorem 14 are outlined in Exercises 27–29, which also review concepts from Chapter 4. Formula (4) for $\hat{\mathbf{x}}$ is useful mainly for theoretical purposes and for hand calculations when A^TA is a 2×2 invertible matrix.

When a least-squares solution $\hat{\mathbf{x}}$ is used to produce $A\hat{\mathbf{x}}$ as an approximation to $\mathbf{b}$, the distance from $\mathbf{b}$ to $A\hat{\mathbf{x}}$ is called the **least-squares error** of this approximation.

EXAMPLE 3 Given A and $\mathbf{b}$ as in Example 1, determine the least-squares error in the least-squares solution of $A\mathbf{x} = \mathbf{b}$.

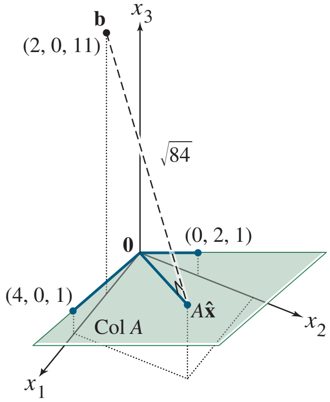


FIGURE 3

SOLUTION From Example 1,

$$\mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} \quad \text{and} \quad A\hat{\mathbf{x}} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$$

Hence

$$\mathbf{b} - A\hat{\mathbf{x}} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ -4 \\ 8 \end{bmatrix}$$

and

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| = \sqrt{(-2)^2 + (-4)^2 + 8^2} = \sqrt{84}$$

The least-squares error is $\sqrt{84}$. For any $\mathbf{x}$ in $\mathbb{R}^2$, the distance between $\mathbf{b}$ and the vector $A\mathbf{x}$ is at least $\sqrt{84}$. See Figure 3. Note that the least-squares solution $\hat{\mathbf{x}}$ itself does not appear in the figure. ■

Alternative Calculations of Least-Squares Solutions

The next example shows how to find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ when the columns of A are orthogonal. Such matrices often appear in linear regression problems, discussed in the next section.

EXAMPLE 4 Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

SOLUTION Because the columns $\mathbf{a}_1$ and $\mathbf{a}_2$ of A are orthogonal, the orthogonal projection of $\mathbf{b}$ onto Col A is given by

$$\hat{\mathbf{b}} = \frac{\mathbf{b} \cdot \mathbf{a}_1}{\mathbf{a}_1 \cdot \mathbf{a}_1}\mathbf{a}_1 + \frac{\mathbf{b} \cdot \mathbf{a}_2}{\mathbf{a}_2 \cdot \mathbf{a}_2}\mathbf{a}_2 = \frac{8}{4}\mathbf{a}_1 + \frac{45}{90}\mathbf{a}_2 \tag{5}$$

$$= \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} + \begin{bmatrix} -3 \\ -1 \\ 1/2 \\ 7/2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 5/2 \\ 11/2 \end{bmatrix}$$

Now that $\hat{\mathbf{b}}$ is known, we can solve $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. But this is trivial, since we already know what weights to place on the columns of A to produce $\hat{\mathbf{b}}$. It is clear from (5) that

$$\hat{\mathbf{x}} = \begin{bmatrix} 8/4 \\ 45/90 \end{bmatrix} = \begin{bmatrix} 2 \\ 1/2 \end{bmatrix}$$ ■

In some cases, the normal equations for a least-squares problem can be *ill-conditioned*; that is, small errors in the calculations of the entries of A^TA can sometimes cause relatively large errors in the solution $\hat{\mathbf{x}}$. If the columns of A are linearly independent, the least-squares solution can often be computed more reliably through a QR factorization of A (described in Section 6.4).[1]

[1] The QR method is compared with the standard normal equation method in G. Golub and C. Van Loan, *Matrix Computations*, 3rd ed. (Baltimore: Johns Hopkins Press, 1996), pp. 230–231.

THEOREM 15

Given an $m \times n$ matrix A with linearly independent columns, let $A = QR$ be a QR factorization of A as in Theorem 12. Then, for each $\mathbf{b}$ in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution, given by

$$\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b} \tag{6}$$

PROOF Let $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$. Then

$$A\hat{\mathbf{x}} = QR\hat{\mathbf{x}} = QRR^{-1}Q^T\mathbf{b} = QQ^T\mathbf{b}$$

By Theorem 12, the columns of Q form an orthonormal basis for Col A. Hence, by Theorem 10, $QQ^T\mathbf{b}$ is the orthogonal projection $\hat{\mathbf{b}}$ of $\mathbf{b}$ onto Col A. Then $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, which shows that $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$. The uniqueness of $\hat{\mathbf{x}}$ follows from Theorem 14. ■

Numerical Notes

Since R in Theorem 15 is upper triangular, $\hat{\mathbf{x}}$ should be calculated as the exact solution of the equation

$$R\mathbf{x} = Q^T\mathbf{b} \tag{7}$$

It is much faster to solve (7) by back-substitution or row operations than to compute R^{-1} and use (6).

EXAMPLE 5 Find the least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix}$$

SOLUTION The QR factorization of A can be obtained as in Section 6.4.

$$A = QR = \begin{bmatrix} 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix}$$

Then

$$Q^T\mathbf{b} = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix} = \begin{bmatrix} 6 \\ -6 \\ 4 \end{bmatrix}$$

The least-squares solution $\hat{\mathbf{x}}$ satisfies $R\mathbf{x} = Q^T\mathbf{b}$; that is,

$$\begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ -6 \\ 4 \end{bmatrix}$$

This equation is solved easily and yields $\hat{\mathbf{x}} = \begin{bmatrix} 10 \\ -6 \\ 2 \end{bmatrix}$. ■

Practice Problems

1. Let $A = \begin{bmatrix} 1 & -3 & -3 \\ 1 & 5 & 1 \\ 1 & 7 & 2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 5 \\ -3 \\ -5 \end{bmatrix}$. Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$, and compute the associated least-squares error.
2. What can you say about the least-squares solution of $A\mathbf{x} = \mathbf{b}$ when $\mathbf{b}$ is orthogonal to the columns of A?

6.5 Exercises

In Exercises 1–4, find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ by (a) constructing the normal equations for $\hat{\mathbf{x}}$ and (b) solving for $\hat{\mathbf{x}}$.

1. $A = \begin{bmatrix} -1 & 2 \\ 2 & -3 \\ -1 & 3 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}$

2. $A = \begin{bmatrix} 2 & 1 \\ -2 & 0 \\ 2 & 3 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -5 \\ 8 \\ 1 \end{bmatrix}$

3. $A = \begin{bmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 3 \\ 2 & 5 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ 1 \\ -4 \\ 2 \end{bmatrix}$

4. $A = \begin{bmatrix} 1 & 1 \\ 1 & -4 \\ 1 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 9 \\ 2 \\ 5 \end{bmatrix}$

In Exercises 5 and 6, describe all least-squares solutions of the equation $A\mathbf{x} = \mathbf{b}$.

5. $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ 3 \\ 8 \\ 2 \end{bmatrix}$

6. $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 7 \\ 2 \\ 3 \\ 6 \\ 5 \\ 4 \end{bmatrix}$

7. Compute the least-squares error associated with the least-squares solution found in Exercise 3.

8. Compute the least-squares error associated with the least-squares solution found in Exercise 4.

In Exercises 9–12, find (a) the orthogonal projection of $\mathbf{b}$ onto Col A and (b) a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

9. $A = \begin{bmatrix} 1 & 5 \\ 3 & 1 \\ -2 & 4 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 4 \\ -2 \\ -3 \end{bmatrix}$

10. $A = \begin{bmatrix} 1 & 2 \\ -1 & 4 \\ 1 & 2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}$

11. $A = \begin{bmatrix} 1 & -1 & -4 \\ 1 & -4 & 1 \\ 3 & 0 & 1 \\ 5 & 1 & 0 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ -2 \\ -4 \\ 7 \end{bmatrix}$

12. $A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & 2 & -1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ 9 \\ 9 \\ 3 \end{bmatrix}$

13. Let $A = \begin{bmatrix} 3 & 4 \\ -2 & 1 \\ 3 & 4 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 11 \\ -9 \\ 5 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 5 \\ -1 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$. Compute $A\mathbf{u}$ and $A\mathbf{v}$, and compare them with $\mathbf{b}$. Could $\mathbf{u}$ possibly be a least-squares solution of $A\mathbf{x} = \mathbf{b}$? (Answer this without computing a least-squares solution.)

14. Let $A = \begin{bmatrix} 2 & 1 \\ -3 & -4 \\ 3 & 2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 5 \\ 4 \\ 4 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 4 \\ -5 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$. Compute $A\mathbf{u}$ and $A\mathbf{v}$, and compare them with $\mathbf{b}$. Is it possible that at least one of $\mathbf{u}$ or $\mathbf{v}$ could be a least-squares solution of $A\mathbf{x} = \mathbf{b}$? (Answer this without computing a least-squares solution.)

In Exercises 15 and 16, use the factorization $A = QR$ to find the least-squares solution of $A\mathbf{x} = \mathbf{b}$.

15. $A = \begin{bmatrix} 2 & 3 \\ 2 & 4 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2/3 & -1/3 \\ 2/3 & 2/3 \\ 1/3 & -2/3 \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 0 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 7 \\ 3 \\ 1 \end{bmatrix}$

16. $A = \begin{bmatrix} 3 & 5 \\ 3 & 0 \\ 3 & 0 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \\ 1/2 & -1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 6 & 5 \\ 0 & 5 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 9 \\ -8 \\ 5 \\ -3 \end{bmatrix}$

In Exercises 17–26, A is an $m \times n$ matrix and $\mathbf{b}$ is in $\mathbb{R}^m$. Mark each statement True or False **(T/F)**. Justify each answer.

17. **(T/F)** The general least-squares problem is to find an $\mathbf{x}$ that makes $A\mathbf{x}$ as close as possible to $\mathbf{b}$.

18. **(T/F)** If $\mathbf{b}$ is in the column space of A, then every solution of $A\mathbf{x} = \mathbf{b}$ is a least-squares solution.

19. **(T/F)** A least-squares solution of $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}}$ that satisfies $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, where $\hat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{b}$ onto Col A.

20. **(T/F)** A least-squares solution of $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}}$ such that $\|\mathbf{b} - A\mathbf{x}\| \leq \|\mathbf{b} - A\hat{\mathbf{x}}\|$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

21. **(T/F)** Any solution of $A^TA\mathbf{x} = A^T\mathbf{b}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

22. **(T/F)** If the columns of A are linearly independent, then the equation $A\mathbf{x} = \mathbf{b}$ has exactly one least-squares solution.

23. **(T/F)** The least-squares solution of $A\mathbf{x} = \mathbf{b}$ is the point in the column space of A closest to $\mathbf{b}$.

24. **(T/F)** A least-squares solution of $A\mathbf{x} = \mathbf{b}$ is a list of weights that, when applied to the columns of A, produces the orthogonal projection of $\mathbf{b}$ onto Col A.

25. **(T/F)** The normal equations always provide a reliable method for computing least-squares solutions.

26. **(T/F)** If A has a QR factorization, say $A = QR$, then the best way to find the least-squares solution of $A\mathbf{x} = \mathbf{b}$ is to compute $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$.

27. Let A be an $m \times n$ matrix. Use the steps below to show that a vector $\mathbf{x}$ in $\mathbb{R}^n$ satisfies $A\mathbf{x} = \mathbf{0}$ if and only if $A^TA\mathbf{x} = \mathbf{0}$. This will show that Nul A = Nul A^TA.

a. Show that if $A\mathbf{x} = \mathbf{0}$, then $A^TA\mathbf{x} = \mathbf{0}$.

b. Suppose $A^TA\mathbf{x} = \mathbf{0}$. Explain why $\mathbf{x}^TA^TA\mathbf{x} = \mathbf{0}$, and use this to show that $A\mathbf{x} = \mathbf{0}$.

28. Let A be an $m \times n$ matrix such that A^TA is invertible. Show that the columns of A are linearly independent. [*Careful:* You may not assume that A is invertible; it may not even be square.]

29. Let A be an $m \times n$ matrix whose columns are linearly independent. [*Careful:* A need not be square.]

a. Use Exercise 27 to show that A^TA is an invertible matrix.

b. Explain why A must have at least as many rows as columns.

c. Determine the rank of A.

30. Use Exercise 27 to show that rank A^TA = rank A. [*Hint:* How many columns does A^TA have? How is this connected with the rank of A^TA?]

31. Suppose A is $m \times n$ with linearly independent columns and $\mathbf{b}$ is in $\mathbb{R}^m$. Use the normal equations to produce a formula for $\hat{\mathbf{b}}$, the projection of $\mathbf{b}$ onto Col A. [*Hint:* Find $\hat{\mathbf{x}}$ first. The formula does not require an orthogonal basis for Col A.]

32. Find a formula for the least-squares solution of $A\mathbf{x} = \mathbf{b}$ when the columns of A are orthonormal.

33. Describe all least-squares solutions of the system

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

T **34.** Example 2 in Section 4.8 displayed a low-pass linear filter that changed a signal $\{y_k\}$ into $\{y_{k+1}\}$ and changed a higher-frequency signal $\{w_k\}$ into the zero signal, where $y_k = \cos(\pi k/4)$ and $w_k = \cos(3\pi k/4)$. The following calculations will design a filter with approximately those properties. The filter equation is

$$a_0 y_{k+2} + a_1 y_{k+1} + a_2 y_k = z_k \qquad \text{for all } k \tag{8}$$

Because the signals are periodic, with period 8, it suffices to study equation (8) for $k = 0, \ldots, 7$. The action on the two signals described above translates into two sets of eight equations, shown below:

$$\begin{array}{c} \\ k=0 \\ k=1 \\ \vdots \\ \\ \\ \\ \\ k=7 \end{array}
\begin{array}{c} \begin{array}{ccc} y_{k+2} & y_{k+1} & y_k \end{array} \\ \begin{bmatrix} 0 & .7 & 1 \\ -.7 & 0 & .7 \\ -1 & -.7 & 0 \\ -.7 & -1 & -.7 \\ 0 & -.7 & -1 \\ .7 & 0 & -.7 \\ 1 & .7 & 0 \\ .7 & 1 & .7 \end{bmatrix} \end{array}
\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} =
\begin{array}{c} y_{k+1} \\ \begin{bmatrix} .7 \\ 0 \\ -.7 \\ -1 \\ -.7 \\ 0 \\ .7 \\ 1 \end{bmatrix} \end{array}$$

$$\begin{array}{c} \\ k=0 \\ k=1 \\ \vdots \\ \\ \\ \\ \\ k=7 \end{array}
\begin{array}{c} \begin{array}{ccc} w_{k+2} & w_{k+1} & w_k \end{array} \\ \begin{bmatrix} 0 & -.7 & 1 \\ .7 & 0 & -.7 \\ -1 & .7 & 0 \\ .7 & -1 & .7 \\ 0 & .7 & -1 \\ -.7 & 0 & .7 \\ 1 & -.7 & 0 \\ -.7 & 1 & -.7 \end{bmatrix} \end{array}
\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} =
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Write an equation $A\mathbf{x} = \mathbf{b}$, where A is a 16×3 matrix formed from the two coefficient matrices above and where $\mathbf{b}$ in $\mathbb{R}^{16}$ is formed from the two right sides of the equations. Find a_0, a_1, and a_2 given by the least-squares solution of $A\mathbf{x} = \mathbf{b}$. (The .7 in the data above was used as an approximation for $\sqrt{2}/2$, to illustrate how a typical computation in an applied problem might proceed. If .707 were used instead, the resulting filter coefficients would agree to at least seven decimal places with $\sqrt{2}/4$, $1/2$, and $\sqrt{2}/4$, the values produced by exact arithmetic calculations.)

Solutions to Practice Problems

1. First, compute

$$A^TA = \begin{bmatrix} 1 & 1 & 1 \\ -3 & 5 & 7 \\ -3 & 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & -3 & -3 \\ 1 & 5 & 1 \\ 1 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 9 & 0 \\ 9 & 83 & 28 \\ 0 & 28 & 14 \end{bmatrix}$$

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 1 & 1 \\ -3 & 5 & 7 \\ -3 & 1 & 2 \end{bmatrix}\begin{bmatrix} 5 \\ -3 \\ -5 \end{bmatrix} = \begin{bmatrix} -3 \\ -65 \\ -28 \end{bmatrix}$$

Next, row reduce the augmented matrix for the normal equations, $A^TA\mathbf{x} = A^T\mathbf{b}$:

$$\begin{bmatrix} 3 & 9 & 0 & -3 \\ 9 & 83 & 28 & -65 \\ 0 & 28 & 14 & -28 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 0 & -1 \\ 0 & 56 & 28 & -56 \\ 0 & 28 & 14 & -28 \end{bmatrix} \sim \cdots \sim \begin{bmatrix} 1 & 0 & -3/2 & 2 \\ 0 & 1 & 1/2 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The general least-squares solution is $x_1 = 2 + \frac{3}{2}x_3$, $x_2 = -1 - \frac{1}{2}x_3$, with x_3 free. For one specific solution, take $x_3 = 0$ (for example), and get

$$\hat{\mathbf{x}} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$$

To find the least-squares error, compute

$$\hat{\mathbf{b}} = A\hat{\mathbf{x}} = \begin{bmatrix} 1 & -3 & -3 \\ 1 & 5 & 1 \\ 1 & 7 & 2 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \\ -5 \end{bmatrix}$$

It turns out that $\hat{\mathbf{b}} = \mathbf{b}$, so $\|\mathbf{b} - \hat{\mathbf{b}}\| = 0$. The least-squares error is zero because $\mathbf{b}$ happens to be in Col A.

2. If $\mathbf{b}$ is orthogonal to the columns of A, then the projection of $\mathbf{b}$ onto the column space of A is $\mathbf{0}$. In this case, a least-squares solution $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ satisfies $A\hat{\mathbf{x}} = \mathbf{0}$.

6.6 Machine Learning and Linear Models

Machine Learning

Machine learning uses linear models in situations where the machine is being *trained* to predict the outcome (dependent variables) based on the values of the inputs (independent variables). The machine is given a set of training data where the values of the independent and dependent variables are known. The machine then *learns* the relationship between the independent variables and the dependent variables. One type of learning is to fit a curve, such as a least-squares line or parabola, to the data. Once the machine has learned the pattern from the training data, it can then estimate the value of the output based on a given value for the input.

Least-Squares Lines

A common task in science and engineering is to analyze and understand relationships among several quantities that vary. This section describes a variety of situations in which data are used to build or verify a formula that predicts the value of one variable as a function of other variables. In each case, the problem will amount to solving a least-squares problem.

For easy application of the discussion to real problems that you may encounter later in your career, we choose notation that is commonly used in the statistical analysis of scientific and engineering data. Instead of $A\mathbf{x} = \mathbf{b}$, we write $X\boldsymbol{\beta} = \mathbf{y}$ and refer to X as the **design matrix**, $\boldsymbol{\beta}$ as the **parameter vector**, and $\mathbf{y}$ as the **observation vector**.

The simplest relation between two variables x and y is the linear equation $y = \beta_0 + \beta_1 x$.[1] Experimental data often produce points $(x_1, y_1), \dots, (x_n, y_n)$ that, when graphed, seem to lie close to a line. We want to determine the parameters β_0 and β_1 that make the line as "close" to the points as possible.

Suppose β_0 and β_1 are fixed, and consider the line $y = \beta_0 + \beta_1 x$ in Figure 1. Corresponding to each data point (x_j, y_j) there is a point $(x_j, \beta_0 + \beta_1 x_j)$ on the line with the same x-coordinate. We call y_j the *observed* value of y and $\beta_0 + \beta_1 x_j$ the *predicted* y-value (determined by the line). The difference between an observed y-value and a predicted y-value is called a *residual*.

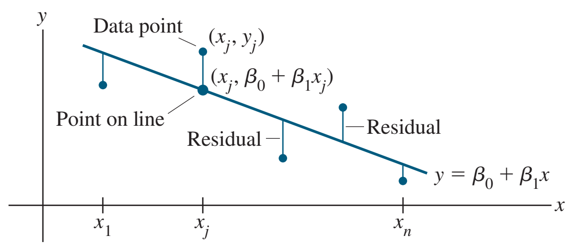


FIGURE 1 Fitting a line to experimental data.

There are several ways to measure how "close" the line is to the data. The usual choice (primarily because the mathematical calculations are simple) is to add the squares of the residuals. The **least-squares line** is the line $y = \beta_0 + \beta_1 x$ that minimizes the sum of the squares of the residuals. This line is also called a **line of regression of y on x**, because any errors in the data are assumed to be only in the y-coordinates. The coefficients β_0, β_1 of the line are called (linear) **regression coefficients**.[2]

If the data points were on the line, the parameters β_0 and β_1 would satisfy the equations

Predicted y-value		**Observed y-value**
$\beta_0 + \beta_1 x_1$	=	y_1
$\beta_0 + \beta_1 x_2$	=	y_2
$\vdots$		$\vdots$
$\beta_0 + \beta_1 x_n$	=	y_n

[1] This notation is commonly used for least-squares lines instead of $y = mx + b$.

[2] If the measurement errors are in x instead of y, simply interchange the coordinates of the data (x_j, y_j) before plotting the points and computing the regression line. If both coordinates are subject to possible error, then you might choose the line that minimizes the sum of the squares of the *orthogonal* (perpendicular) distances from the points to the line.

We can write this system as

$$X\boldsymbol{\beta} = \mathbf{y}, \quad \text{where } X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{1}$$

Of course, if the data points don't lie on a line, then there are no parameters β_0, β_1 for which the predicted y-values in $X\boldsymbol{\beta}$ equal the observed y-values in $\mathbf{y}$, and $X\boldsymbol{\beta} = \mathbf{y}$ has no solution. This is a least-squares problem, $A\mathbf{x} = \mathbf{b}$, with different notation!

The square of the distance between the vectors $X\boldsymbol{\beta}$ and $\mathbf{y}$ is precisely the sum of the squares of the residuals. The $\boldsymbol{\beta}$ that minimizes this sum also minimizes the distance between $X\boldsymbol{\beta}$ and $\mathbf{y}$. *Computing the least-squares solution of $X\boldsymbol{\beta} = \mathbf{y}$ is equivalent to finding the $\boldsymbol{\beta}$ that determines the least-squares line in Figure 1.*

EXAMPLE 1 Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points $(2, 1)$, $(5, 2)$, $(7, 3)$, and $(8, 3)$.

SOLUTION Use the x-coordinates of the data to build the design matrix X in (1) and the y-coordinates to build the observation vector $\mathbf{y}$:

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$$

For the least-squares solution of $X\boldsymbol{\beta} = \mathbf{y}$, obtain the normal equations (with the new notation):

$$X^T X\boldsymbol{\beta} = X^T\mathbf{y}$$

That is, compute

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

The normal equations are

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

Hence

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 142 & -22 \\ -22 & 4 \end{bmatrix} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 24 \\ 30 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$$

Thus the least-squares line has the equation

$$y = \frac{2}{7} + \frac{5}{14}x$$

See Figure 2. ■

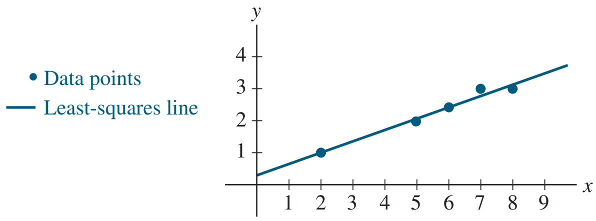


FIGURE 2 The least-squares line $y = \frac{2}{7} + \frac{5}{14}x$.

EXAMPLE 2 If a machine learns the data from Example 1 by creating a least-squares line, what outcome will it predict for the inputs 4 and 6?

SOLUTION The machine would perform the same calculations as in Example 1 to arrive at the least-squares line

$$y = \frac{2}{7} + \frac{5}{14}x$$

as a reasonable pattern to use to predict the outcomes.

For the value $x = 4$, the machine will predict an output of $y = \frac{2}{7} + \frac{5}{14}(4) = \frac{12}{7}$.

For the value $x = 6$, the machine will predict an output of $y = \frac{2}{7} + \frac{5}{14}(6) = \frac{17}{7}$.

See Figure 3.

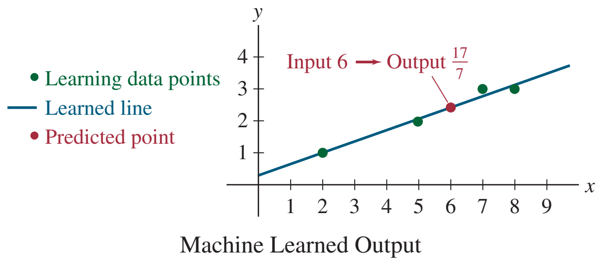


FIGURE 3 Machine-learned output. ■

A common practice before computing a least-squares line is to compute the average $\overline{x}$ of the original x-values and form a new variable $x^* = x - \overline{x}$. The new x-data are said to be in **mean-deviation form**. In this case, the two columns of the design matrix will be orthogonal. Solution of the normal equations is simplified, just as in Example 4 in Section 6.5. See Exercises 23 and 24.

The General Linear Model

In some applications, it is necessary to fit data points with something other than a straight line. In the examples that follow, the matrix equation is still $X\boldsymbol{\beta} = \mathbf{y}$, but the specific form of X changes from one problem to the next. Statisticians usually introduce a **residual vector** $\boldsymbol{\epsilon}$, defined by $\boldsymbol{\epsilon} = \mathbf{y} - X\boldsymbol{\beta}$, and write

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Any equation of this form is referred to as a **linear model**. Once X and $\mathbf{y}$ are determined, the goal is to minimize the length of $\boldsymbol{\epsilon}$, which amounts to finding a least-squares solution of $X\boldsymbol{\beta} = \mathbf{y}$. In each case, the least-squares solution $\hat{\boldsymbol{\beta}}$ is a solution of the normal equations

$$X^T X\boldsymbol{\beta} = X^T\mathbf{y}$$

Least-Squares Fitting of Other Curves

When data points $(x_1, y_1), \dots, (x_n, y_n)$ on a scatter plot do not lie close to any line, it may be appropriate to postulate some other functional relationship between x and y.

The next two examples show how to fit data by curves that have the general form

$$y = \beta_0 f_0(x) + \beta_1 f_1(x) + \cdots + \beta_k f_k(x) \tag{2}$$

where $f_0, \dots, f_k$ are known functions and $\beta_0, \dots, \beta_k$ are parameters that must be determined. As we will see, equation (2) describes a linear model because it is linear in the unknown parameters.

For a particular value of x, (2) gives a predicted, or "fitted," value of y. The difference between the observed value and the predicted value is the residual. The parameters $\beta_0, \dots, \beta_k$ must be determined so as to minimize the sum of the squares of the residuals.

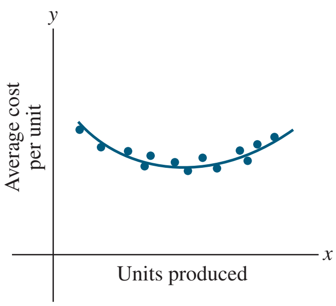


FIGURE 4
Average cost curve.

EXAMPLE 3 Suppose data points $(x_1, y_1), \dots, (x_n, y_n)$ appear to lie along some sort of parabola instead of a straight line. For instance, if the x-coordinate denotes the production level for a company, and y denotes the average cost per unit of operating at a level of x units per day, then a typical average cost curve looks like a parabola that opens upward (Figure 4). In ecology, a parabolic curve that opens downward is used to model the net primary production of nutrients in a plant, as a function of the surface area of the foliage (Figure 5). Suppose we wish to approximate the data by an equation of the form

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 \tag{3}$$

Describe the linear model that produces a "least-squares fit" of the data by equation (3).

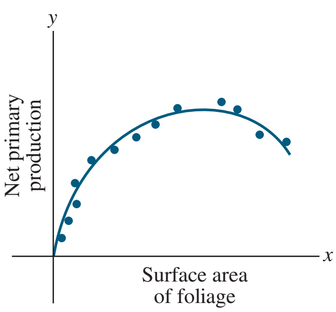


FIGURE 5
Production of nutrients.

SOLUTION Equation (3) describes the ideal relationship. Suppose the actual values of the parameters are $\beta_0, \beta_1, \beta_2$. Then the coordinates of the first data point (x_1, y_1) satisfy an equation of the form

$$y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \epsilon_1$$

where ϵ_1 is the residual error between the observed value y_1 and the predicted y-value $\beta_0 + \beta_1 x_1 + \beta_2 x_1^2$. Each data point determines a similar equation:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \epsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_2 + \beta_2 x_2^2 + \epsilon_2 \\ &\vdots \qquad\qquad\quad \vdots \\ y_n &= \beta_0 + \beta_1 x_n + \beta_2 x_n^2 + \epsilon_n \end{aligned}$$

It is a simple matter to write this system of equations in the form $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$. To find X, inspect the first few rows of the system and look for the pattern.

$$\begin{array}{ccccccc} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} & = & \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} & \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} & + & \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \\ \mathbf{y} & = & X & \boldsymbol{\beta} & + & \boldsymbol{\epsilon} \end{array}$$

■

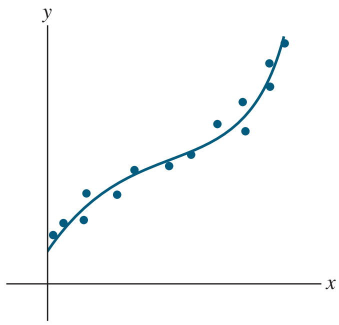


FIGURE 6
Data points along a cubic curve.

EXAMPLE 4 If data points tend to follow a pattern such as in Figure 6, then an appropriate model might be an equation of the form

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

Such data, for instance, could come from a company's total costs, as a function of the level of production. Describe the linear model that gives a least-squares fit of this type to data $(x_1, y_1), \ldots, (x_n, y_n)$.

SOLUTION By an analysis similar to that in Example 2, we obtain

$$\begin{array}{cccc} \text{Observation vector} & \text{Design matrix} & \text{Parameter vector} & \text{Residual vector} \\ \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, & X = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}, & \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, & \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \end{array}$$

■

Multiple Regression

Suppose an experiment involves two independent variables—say, u and v—and one dependent variable, y. A simple equation for predicting y from u and v has the form

$$y = \beta_0 + \beta_1 u + \beta_2 v \tag{4}$$

A more general prediction equation might have the form

$$y = \beta_0 + \beta_1 u + \beta_2 v + \beta_3 u^2 + \beta_4 uv + \beta_5 v^2 \tag{5}$$

This equation is used in geology, for instance, to model erosion surfaces, glacial cirques, soil pH, and other quantities. In such cases, the least-squares fit is called a *trend surface*.

Equations (4) and (5) both lead to a linear model because they are linear in the unknown parameters (even though u and v are multiplied). In general, a linear model will arise whenever y is to be predicted by an equation of the form

$$y = \beta_0 f_0(u, v) + \beta_1 f_1(u, v) + \cdots + \beta_k f_k(u, v)$$

with $f_0, \ldots, f_k$ any sort of known functions and $\beta_0, \ldots, \beta_k$ unknown weights.

EXAMPLE 5 In geography, local models of terrain are constructed from data $(u_1, v_1, y_1), \dots, (u_n, v_n, y_n)$, where u_j, v_j, and y_j are latitude, longitude, and altitude, respectively. Describe the linear model based on (4) that gives a least-squares fit to such data. The solution is called the *least-squares plane*. See Figure 7.

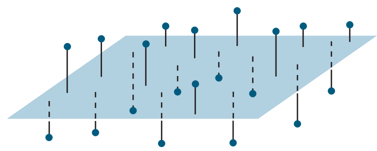


FIGURE 7 A least-squares plane.

SOLUTION We expect the data to satisfy the following equations:

$$
\begin{aligned}
y_1 &= \beta_0 + \beta_1 u_1 + \beta_2 v_1 + \epsilon_1 \\
y_2 &= \beta_0 + \beta_1 u_2 + \beta_2 v_2 + \epsilon_2 \\
&\ \ \vdots \qquad\qquad\quad \vdots \\
y_n &= \beta_0 + \beta_1 u_n + \beta_2 v_n + \epsilon_n
\end{aligned}
$$

This system has the matrix form $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where

$$
\begin{array}{cccc}
\text{Observation vector} & \text{Design matrix} & \text{Parameter vector} & \text{Residual vector} \\
\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, &
X = \begin{bmatrix} 1 & u_1 & v_1 \\ 1 & u_2 & v_2 \\ \vdots & \vdots & \vdots \\ 1 & u_n & v_n \end{bmatrix}, &
\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, &
\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}
\end{array}
$$

■

Example 5 shows that the linear model for multiple regression has the same abstract form as the model for the simple regression in the earlier examples. Linear algebra gives us the power to understand the general principle behind all the linear models. Once X is defined properly, the normal equations for $\boldsymbol{\beta}$ have the same matrix form, no matter how many variables are involved. Thus, for any linear model where X^TX is invertible, the least-squares $\hat{\boldsymbol{\beta}}$ is given by $(X^TX)^{-1}X^T\mathbf{y}$.

STUDY GUIDE offers additional resources for understanding the geometry of a linear model.

Practice Problem

When the monthly sales of a product are subject to seasonal fluctuations, a curve that approximates the sales data might have the form

$$y = \beta_0 + \beta_1 x + \beta_2 \sin(2\pi x/12)$$

where x is the time in months. The term $\beta_0 + \beta_1 x$ gives the basic sales trend, and the sine term reflects the seasonal changes in sales. Give the design matrix and the parameter vector for the linear model that leads to a least-squares fit of the equation above. Assume the data are $(x_1, y_1), \dots, (x_n, y_n)$.

6.6 Exercises

In Exercises 1–4, find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the given data points.

1. $(0, 1), (1, 1), (2, 2), (3, 2)$
2. $(1, 0), (2, 2), (3, 7), (4, 9)$
3. $(-1, 0), (0, 1), (1, 2), (2, 4)$
4. $(2, 3), (3, 2), (5, 1), (6, 0)$
5. If a machine learns the least-squares line that best fits the data in Exercise 1, what will the machine pick for the value of y when $x = 4$?
6. If a machine learns the least-squares line that best fits the data in Exercise 2, what will the machine pick for the value of y when $x = 3$?
7. If a machine learns the least-squares line that best fits the data in Exercise 1, what will the machine pick for the value of y when $x = 3$? How closely does this match the data point at $x = 3$ fed into the machine?
8. If a machine learns the least-squares line that best fits the data in Exercise 2, what will the machine pick for the value of y when $x = 4$? How closely does this match the data point at $x = 4$ fed into the machine?
9. If you enter the data from Exercise 1 into a machine and it returns a y value of 20 when $x = 2.5$, should you trust the machine? Justify your answer.
10. If you enter the data from Exercise 2 into a machine and it returns a y value of -1 when $x = 1.5$, should you trust the machine? Justify your answer.
11. Let X be the design matrix used to find the least-squares line to fit data $(x_1, y_1), \ldots, (x_n, y_n)$. Use a theorem in Section 6.5 to show that the normal equations have a unique solution if and only if the data include at least two data points with different x-coordinates.
12. Let X be the design matrix in Example 2 corresponding to a least-squares fit of a parabola to data $(x_1, y_1), \ldots, (x_n, y_n)$. Suppose x_1, x_2, and x_3 are distinct. Explain why there is only one parabola that fits the data best, in a least-squares sense. (See Exercise 11.)
13. A certain experiment produces the data $(1, 2.5)$, $(2, 4.3)$, $(3, 5.5)$, $(4, 6.1)$, $(5, 6.1)$. Describe the model that produces a least-squares fit of these points by a function of the form

$$y = \beta_1 x + \beta_2 x^2$$

Such a function might arise, for example, as the revenue from the sale of x units of a product, when the amount offered for sale affects the price to be set for the product.

a. Give the design matrix, the observation vector, and the unknown parameter vector.

T b. Find the associated least-squares curve for the data.

c. If a machine learned the curve you found in (b), what output would it provide for an input of $x = 6$?

14. A simple curve that often makes a good model for the variable costs of a company, as a function of the sales level x, has the form $y = \beta_1 x + \beta_2 x^2 + \beta_3 x^3$. There is no constant term because fixed costs are not included.

a. Give the design matrix and the parameter vector for the linear model that leads to a least-squares fit of the equation above, with data $(x_1, y_1), \ldots, (x_n, y_n)$.

T b. Find the least-squares curve of the form above to fit the data $(4, 1.58)$, $(6, 2.08)$, $(8, 2.5)$, $(10, 2.8)$, $(12, 3.1)$, $(14, 3.4)$, $(16, 3.8)$, and $(18, 4.32)$, with values in thousands. If possible, produce a graph that shows the data points and the graph of the cubic approximation.

c. If a machine learned the curve you found in (b), what output would it provide for an input of $x = 9$?

15. A certain experiment produces the data $(1, 7.9)$, $(2, 5.4)$, and $(3, -.9)$. Describe the model that produces a least-squares fit of these points by a function of the form

$$y = A \cos x + B \sin x$$

16. Suppose radioactive substances A and B have decay constants of .02 and .07, respectively. If a mixture of these two substances at time $t = 0$ contains M_A grams of A and M_B grams of B, then a model for the total amount y of the mixture present at time t is

$$y = M_A e^{-.02t} + M_B e^{-.07t} \tag{6}$$

Suppose the initial amounts M_A and M_B are unknown, but a scientist is able to measure the total amounts present at several times and records the following points (t_i, y_i): $(10, 21.34)$, $(11, 20.68)$, $(12, 20.05)$, $(14, 18.87)$, and $(15, 18.30)$.

a. Describe a linear model that can be used to estimate M_A and M_B.

T b. Find the least-squares curve based on (6).

Halley's Comet last appeared in 1986 and will reappear in 2061.

T 17. According to Kepler's first law, a comet should have an elliptic, parabolic, or hyperbolic orbit (with gravitational attractions from the planets ignored). In suitable polar coordinates, the position (r, ϑ) of a comet satisfies an equation of the form

$$r = \beta + e(r \cdot \cos \vartheta)$$

where β is a constant and e is the *eccentricity* of the orbit, with $0 \le e < 1$ for an ellipse, $e = 1$ for a parabola, and $e > 1$ for a hyperbola. Suppose observations of a newly discovered comet provide the data below. Determine the type of orbit, and predict where the comet will be when $\vartheta = 4.6$ (radians).[3]

ϑ	.88	1.10	1.42	1.77	2.14
r	3.00	2.30	1.65	1.25	1.01

T 18. A healthy child's systolic blood pressure p (in millimeters of mercury) and weight w (in pounds) are approximately related by the equation

$$\beta_0 + \beta_1 \ln w = p$$

Use the following experimental data to estimate the systolic blood pressure of a healthy child weighing 100 pounds.

w	44	61	81	113	131
$\ln w$	3.78	4.11	4.39	4.73	4.88
p	91	98	103	110	112

T 19. To measure the takeoff performance of an airplane, the horizontal position of the plane was measured every second, from $t = 0$ to $t = 12$. The positions (in feet) were: 0, 8.8, 29.9, 62.0, 104.7, 159.1, 222.0, 294.5, 380.4, 471.1, 571.7, 686.8, and 809.2.

a. Find the least-squares cubic curve $y = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$ for these data.

b. If a machine learned the curve given in part (a), what would it estimate the velocity of the plane to be when $t = 4.5$ seconds?

20. Let $\overline{x} = \frac{1}{n}(x_1 + \cdots + x_n)$ and $\overline{y} = \frac{1}{n}(y_1 + \cdots + y_n)$. Show that the least-squares line for the data $(x_1, y_1), \ldots, (x_n, y_n)$ must pass through $(\overline{x}, \overline{y})$. That is, show that $\overline{x}$ and $\overline{y}$ satisfy the linear equation $\overline{y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{x}$. [*Hint:* Derive this equation from the vector equation $\mathbf{y} = X\hat{\boldsymbol{\beta}} + \boldsymbol{\epsilon}$. Denote the first column of X by $\mathbf{1}$. Use the fact that the residual vector $\boldsymbol{\epsilon}$ is orthogonal to the column space of X and hence is orthogonal to $\mathbf{1}$.]

Given data for a least-squares problem, $(x_1, y_1), \ldots, (x_n, y_n)$, the following abbreviations are helpful:

$$\sum x = \sum_{i=1}^{n} x_i, \quad \sum x^2 = \sum_{i=1}^{n} x_i^2,$$
$$\sum y = \sum_{i=1}^{n} y_i, \quad \sum xy = \sum_{i=1}^{n} x_i y_i$$

The normal equations for a least-squares line $y = \hat{\beta}_0 + \hat{\beta}_1 x$ may be written in the form

$$\begin{aligned} n\hat{\beta}_0 + \hat{\beta}_1 \textstyle\sum x &= \textstyle\sum y \\ \hat{\beta}_0 \textstyle\sum x + \hat{\beta}_1 \textstyle\sum x^2 &= \textstyle\sum xy \end{aligned} \tag{7}$$

21. Derive the normal equations (7) from the matrix form given in this section.

22. Use a matrix inverse to solve the system of equations in (7) and thereby obtain formulas for $\hat{\beta}_0$ and $\hat{\beta}_1$ that appear in many statistics texts.

23. a. Rewrite the data in Example 1 with new x-coordinates in mean deviation form. Let X be the associated design matrix. Why are the columns of X orthogonal?

b. Write the normal equations for the data in part (a), and solve them to find the least-squares line, $y = \beta_0 + \beta_1 x^*$, where $x^* = x - 5.5$.

24. Suppose the x-coordinates of the data $(x_1, y_1), \ldots, (x_n, y_n)$ are in mean deviation form, so that $\sum x_i = 0$. Show that if X is the design matrix for the least-squares line in this case, then $X^T X$ is a diagonal matrix.

Exercises 25 and 26 involve a design matrix X with two or more columns and a least-squares solution $\hat{\boldsymbol{\beta}}$ of $\mathbf{y} = X\boldsymbol{\beta}$. Consider the following numbers.

(i) $\|X\hat{\boldsymbol{\beta}}\|^2$—the sum of the squares of the "regression term." Denote this number by SS(R).

(ii) $\|\mathbf{y} - X\hat{\boldsymbol{\beta}}\|^2$—the sum of the squares for the error term. Denote this number by SS(E).

(iii) $\|\mathbf{y}\|^2$—the "total" sum of the squares of the y-values. Denote this number by SS(T).

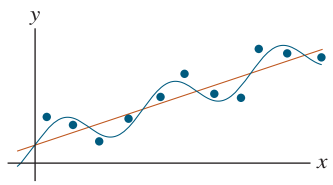
Every statistics text that discusses regression and the linear model $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ introduces these numbers, though terminology and notation vary somewhat. To simplify matters, assume that the mean of the y-values is zero. In this case, SS(T) is proportional to what is called the *variance* of the set of y-values.

25. Justify the equation SS(T) = SS(R) + SS(E). [*Hint:* Use a theorem, and explain why the hypotheses of the theorem are satisfied.] This equation is extremely important in statistics, both in regression theory and in the analysis of variance.

26. Show that $\|X\hat{\boldsymbol{\beta}}\|^2 = \hat{\boldsymbol{\beta}}^T X^T \mathbf{y}$. [*Hint:* Rewrite the left side and use the fact that $\hat{\boldsymbol{\beta}}$ satisfies the normal equations.] This formula for SS(R) is used in statistics. From this and from Exercise 25, obtain the standard formula for SS(E):

$$\text{SS(E)} = \mathbf{y}^T\mathbf{y} - \hat{\boldsymbol{\beta}}^T X^T \mathbf{y}$$

[3] The basic idea of least-squares fitting of data is due to K. F. Gauss (and, independently, to A. Legendre), whose initial rise to fame occurred in 1801 when he used the method to determine the path of the asteroid *Ceres*. Forty days after the asteroid was discovered, it disappeared behind the sun. Gauss predicted it would appear ten months later and gave its location. The accuracy of the prediction astonished the European scientific community.



Sales trend with seasonal fluctuations.

Solution to Practice Problem

Construct X and $\boldsymbol{\beta}$ so that the kth row of $X\boldsymbol{\beta}$ is the predicted y-value that corresponds to the data point (x_k, y_k), namely

$$\beta_0 + \beta_1 x_k + \beta_2 \sin(2\pi x_k/12)$$

It should be clear that

$$X = \begin{bmatrix} 1 & x_1 & \sin(2\pi x_1/12) \\ \vdots & \vdots & \vdots \\ 1 & x_n & \sin(2\pi x_n/12) \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

6.7 Inner Product Spaces

Notions of length, distance, and orthogonality are often important in applications involving a vector space. For $\mathbb{R}^n$, these concepts were based on the properties of the inner product listed in Theorem 1 of Section 6.1. For other spaces, we need analogues of the inner product with the same properties. The conclusions of Theorem 1 now become *axioms* in the following definition.

DEFINITION

An **inner product** on a vector space V is a function that, to each pair of vectors $\mathbf{u}$ and $\mathbf{v}$ in V, associates a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ and satisfies the following axioms, for all $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V and all scalars c:

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
3. $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$
4. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$

A vector space with an inner product is called an **inner product space**.

The vector space $\mathbb{R}^n$ with the standard inner product is an inner product space, and nearly everything discussed in this chapter for $\mathbb{R}^n$ carries over to inner product spaces. The examples in this section and the next lay the foundation for a variety of applications treated in courses in engineering, physics, mathematics, and statistics.

EXAMPLE 1 Fix any two positive numbers—say, 4 and 5—and for vectors $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$ in $\mathbb{R}^2$, set

$$\langle \mathbf{u}, \mathbf{v} \rangle = 4u_1v_1 + 5u_2v_2 \tag{1}$$

Show that equation (1) defines an inner product.

SOLUTION Certainly Axiom 1 is satisfied, because $\langle \mathbf{u}, \mathbf{v} \rangle = 4u_1v_1 + 5u_2v_2 = 4v_1u_1 + 5v_2u_2 = \langle \mathbf{v}, \mathbf{u} \rangle$. If $\mathbf{w} = (w_1, w_2)$, then

$$\begin{aligned} \langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle &= 4(u_1 + v_1)w_1 + 5(u_2 + v_2)w_2 \\ &= 4u_1w_1 + 5u_2w_2 + 4v_1w_1 + 5v_2w_2 \\ &= \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle \end{aligned}$$

This verifies Axiom 2. For Axiom 3, compute

$$\langle c\mathbf{u}, \mathbf{v}\rangle = 4(cu_1)v_1 + 5(cu_2)v_2 = c(4u_1v_1 + 5u_2v_2) = c\langle \mathbf{u}, \mathbf{v}\rangle$$

For Axiom 4, note that $\langle \mathbf{u}, \mathbf{u}\rangle = 4u_1^2 + 5u_2^2 \geq 0$, and $4u_1^2 + 5u_2^2 = 0$ only if $u_1 = u_2 = 0$, that is, if $\mathbf{u} = \mathbf{0}$. Also, $\langle \mathbf{0}, \mathbf{0}\rangle = 0$. So (1) defines an inner product on $\mathbb{R}^2$. ■

Inner products similar to (1) can be defined on $\mathbb{R}^n$. They arise naturally in connection with "weighted least-squares" problems, in which weights are assigned to the various entries in the sum for the inner product in such a way that more importance is given to the more reliable measurements.

From now on, when an inner product space involves polynomials or other functions, we will write the functions in the familiar way, rather than use the boldface type for vectors. Nevertheless, it is important to remember that each function *is* a vector when it is treated as an element of a vector space.

EXAMPLE 2 Let $t_0, \dots, t_n$ be distinct real numbers. For p and q in $\mathbb{P}_n$, define

$$\langle p, q\rangle = p(t_0)q(t_0) + p(t_1)q(t_1) + \cdots + p(t_n)q(t_n) \tag{2}$$

Inner product Axioms 1–3 are readily checked. For Axiom 4, note that

$$\langle p, p\rangle = [p(t_0)]^2 + [p(t_1)]^2 + \cdots + [p(t_n)]^2 \geq 0$$

Also, $\langle \mathbf{0}, \mathbf{0}\rangle = 0$. (The boldface zero here denotes the zero polynomial, the zero vector in $\mathbb{P}_n$.) If $\langle p, p\rangle = 0$, then p must vanish at $n+1$ points: $t_0, \dots, t_n$. This is possible only if p is the zero polynomial, because the degree of p is less than $n+1$. Thus (2) defines an inner product on $\mathbb{P}_n$. ■

EXAMPLE 3 Let V be $\mathbb{P}_2$, with the inner product from Example 2, where $t_0 = 0$, $t_1 = \frac{1}{2}$, and $t_2 = 1$. Let $p(t) = 12t^2$ and $q(t) = 2t - 1$. Compute $\langle p, q\rangle$ and $\langle q, q\rangle$.

SOLUTION

$$\begin{aligned}
\langle p, q\rangle &= p(0)q(0) + p\left(\tfrac{1}{2}\right)q\left(\tfrac{1}{2}\right) + p(1)q(1) \\
&= (0)(-1) + (3)(0) + (12)(1) = 12 \\
\langle q, q\rangle &= [q(0)]^2 + [q\left(\tfrac{1}{2}\right)]^2 + [q(1)]^2 \\
&= (-1)^2 + (0)^2 + (1)^2 = 2
\end{aligned}$$

■

Lengths, Distances, and Orthogonality

Let V be an inner product space, with the inner product denoted by $\langle \mathbf{u}, \mathbf{v}\rangle$. Just as in $\mathbb{R}^n$, we define the **length**, or **norm**, of a vector $\mathbf{v}$ to be the scalar

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}$$

Equivalently, $\|\mathbf{v}\|^2 = \langle \mathbf{v}, \mathbf{v}\rangle$. (This definition makes sense because $\langle \mathbf{v}, \mathbf{v}\rangle \geq 0$, but the definition *does not* say that $\langle \mathbf{v}, \mathbf{v}\rangle$ is a "sum of squares," because $\mathbf{v}$ need not be an element of $\mathbb{R}^n$.)

A **unit vector** is one whose length is 1. The **distance between u and v** is $\|\mathbf{u} - \mathbf{v}\|$. Vectors $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if $\langle \mathbf{u}, \mathbf{v}\rangle = 0$.

EXAMPLE 4 Let $\mathbb{P}_2$ have the inner product (2) of Example 3. Compute the lengths of the vectors $p(t) = 12t^2$ and $q(t) = 2t - 1$.

SOLUTION

$$\begin{aligned}\|p\|^2 &= \langle p, p\rangle = [p(0)]^2 + \left[p\left(\tfrac{1}{2}\right)\right]^2 + [p(1)]^2 \\ &= 0 + [3]^2 + [12]^2 = 153 \\ \|p\| &= \sqrt{153}\end{aligned}$$

From Example 3, $\langle q, q\rangle = 2$. Hence $\|q\| = \sqrt{2}$. ■

The Gram–Schmidt Process

The existence of orthogonal bases for finite-dimensional subspaces of an inner product space can be established by the Gram–Schmidt process, just as in $\mathbb{R}^n$. Certain orthogonal bases that arise frequently in applications can be constructed by this process.

The orthogonal projection of a vector onto a subspace W with an orthogonal basis can be constructed as usual. The projection does not depend on the choice of orthogonal basis, and it has the properties described in the Orthogonal Decomposition Theorem and the Best Approximation Theorem.

EXAMPLE 5 Let V be $\mathbb{P}_4$ with the inner product in Example 2, involving evaluation of polynomials at -2, -1, 0, 1, and 2, and view $\mathbb{P}_2$ as a subspace of V. Produce an orthogonal basis for $\mathbb{P}_2$ by applying the Gram–Schmidt process to the polynomials 1, t, and t^2.

SOLUTION The inner product depends only on the values of a polynomial at $-2, \dots, 2$, so we list the values of each polynomial as a vector in $\mathbb{R}^5$, underneath the name of the polynomial:[1]

$$\begin{array}{rccc} \text{Polynomial:} & 1 & t & t^2 \\ \text{Vector of values:} & \begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}, & \begin{bmatrix}-2\\-1\\0\\1\\2\end{bmatrix}, & \begin{bmatrix}4\\1\\0\\1\\4\end{bmatrix} \end{array}$$

The inner product of two polynomials in V equals the (standard) inner product of their corresponding vectors in $\mathbb{R}^5$. Observe that t is orthogonal to the constant function 1. So take $p_0(t) = 1$ and $p_1(t) = t$. For p_2, use the vectors in $\mathbb{R}^5$ to compute the projection of t^2 onto Span$\{p_0, p_1\}$:

$$\begin{aligned}\langle t^2, p_0\rangle &= \langle t^2, 1\rangle = 4 + 1 + 0 + 1 + 4 = 10 \\ \langle p_0, p_0\rangle &= 5 \\ \langle t^2, p_1\rangle &= \langle t^2, t\rangle = -8 + (-1) + 0 + 1 + 8 = 0\end{aligned}$$

The orthogonal projection of t^2 onto Span$\{1, t\}$ is $\frac{10}{5}p_0 + 0p_1$. Thus

$$p_2(t) = t^2 - 2p_0(t) = t^2 - 2$$

[1] Each polynomial in $\mathbb{P}_4$ is uniquely determined by its value at the five numbers $-2, \dots, 2$. In fact, the correspondence between p and its vector of values is an isomorphism, that is, a one-to-one mapping onto $\mathbb{R}^5$ that preserves linear combinations.

An orthogonal basis for the subspace $\mathbb{P}_2$ of V is

$$\begin{array}{lccc} \text{Polynomial} & p_0 & p_1 & p_2 \\ \text{Vector of values} & \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, & \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, & \begin{bmatrix} 2 \\ -1 \\ -2 \\ -1 \\ 2 \end{bmatrix} \end{array} \tag{3}$$

■

Best Approximation in Inner Product Spaces

A common problem in applied mathematics involves a vector space V whose elements are functions. The problem is to approximate a function f in V by a function g from a specified subspace W of V. The "closeness" of the approximation of f depends on the way $\|f - g\|$ is defined. We will consider only the case in which the distance between f and g is determined by an inner product. In this case, the *best approximation to f by functions in W* is the orthogonal projection of f onto the subspace W.

EXAMPLE 6 Let V be $\mathbb{P}_4$ with the inner product in Example 5, and let p_0, p_1, and p_2 be the orthogonal basis found in Example 5 for the subspace $\mathbb{P}_2$. Find the best approximation to $p(t) = 5 - \frac{1}{2}t^4$ by polynomials in $\mathbb{P}_2$.

SOLUTION The values of p_0, p_1, and p_2 at the numbers $-2, -1, 0, 1$, and 2 are listed in $\mathbb{R}^5$ vectors in (3) above. The corresponding values for p are -3, 9/2, 5, 9/2, and -3. Compute

$$\begin{array}{lll} \langle p, p_0\rangle = 8, & \langle p, p_1\rangle = 0, & \langle p, p_2\rangle = -31 \\ \langle p_0, p_0\rangle = 5, & & \langle p_2, p_2\rangle = 14 \end{array}$$

Then the best approximation in V to p by polynomials in $\mathbb{P}_2$ is

$$\begin{aligned} \hat{p} = \text{proj}_{\mathbb{P}_2} p &= \frac{\langle p, p_0\rangle}{\langle p_0, p_0\rangle} p_0 + \frac{\langle p, p_1\rangle}{\langle p_1, p_1\rangle} p_1 + \frac{\langle p, p_2\rangle}{\langle p_2, p_2\rangle} p_2 \\ &= \tfrac{8}{5} p_0 + \tfrac{-31}{14} p_2 = \tfrac{8}{5} - \tfrac{31}{14}(t^2 - 2). \end{aligned}$$

This polynomial is the closest to p of all polynomials in $\mathbb{P}_2$, when the distance between polynomials is measured only at $-2, -1, 0, 1$, and 2. See Figure 1. ■

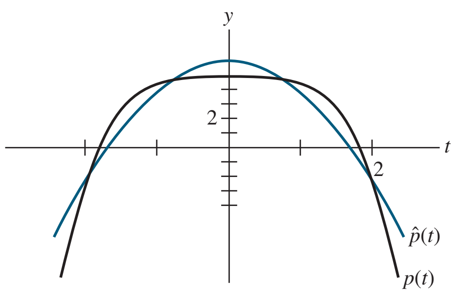


FIGURE 1

The polynomials p_0, p_1, and p_2 in Examples 5 and 6 belong to a class of polynomials that are referred to in statistics as *orthogonal polynomials*.[2] The orthogonality refers to the type of inner product described in Example 2.

Two Inequalities

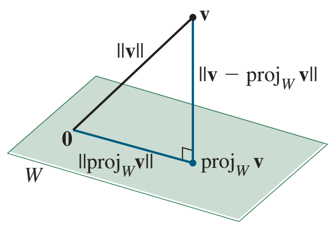


FIGURE 2
The hypotenuse is the longest side.

Given a vector $\mathbf{v}$ in an inner product space V and given a finite-dimensional subspace W, we may apply the Pythagorean Theorem to the orthogonal decomposition of $\mathbf{v}$ with respect to W and obtain

$$\|\mathbf{v}\|^2 = \|\operatorname{proj}_W \mathbf{v}\|^2 + \|\mathbf{v} - \operatorname{proj}_W \mathbf{v}\|^2$$

See Figure 2. In particular, this shows that the norm of the projection of $\mathbf{v}$ onto W does not exceed the norm of $\mathbf{v}$ itself. This simple observation leads to the following important inequality.

THEOREM 16

The Cauchy–Schwarz Inequality

For all $\mathbf{u}$, $\mathbf{v}$ in V,

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \le \|\mathbf{u}\|\,\|\mathbf{v}\| \tag{4}$$

PROOF If $\mathbf{u} = \mathbf{0}$, then both sides of (4) are zero, and hence the inequality is true in this case. (See Practice Problem 1.) If $\mathbf{u} \neq \mathbf{0}$, let W be the subspace spanned by $\mathbf{u}$. Recall that $\|c\mathbf{u}\| = |c|\,\|\mathbf{u}\|$ for any scalar c. Thus

$$\|\operatorname{proj}_W \mathbf{v}\| = \left\| \frac{\langle \mathbf{v}, \mathbf{u}\rangle}{\langle \mathbf{u}, \mathbf{u}\rangle}\mathbf{u} \right\| = \frac{|\langle \mathbf{v}, \mathbf{u}\rangle|}{|\langle \mathbf{u}, \mathbf{u}\rangle|}\|\mathbf{u}\| = \frac{|\langle \mathbf{v}, \mathbf{u}\rangle|}{\|\mathbf{u}\|^2}\|\mathbf{u}\| = \frac{|\langle \mathbf{u}, \mathbf{v}\rangle|}{\|\mathbf{u}\|}$$

Since $\|\operatorname{proj}_W \mathbf{v}\| \le \|\mathbf{v}\|$, we have $\dfrac{|\langle \mathbf{u}, \mathbf{v}\rangle|}{\|\mathbf{u}\|} \le \|\mathbf{v}\|$, which gives (4). ■

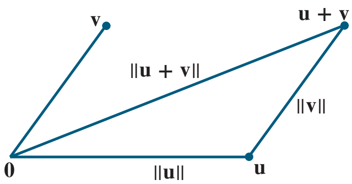


FIGURE 3
The lengths of the sides of a triangle.

The Cauchy–Schwarz inequality is useful in many branches of mathematics. A few simple applications are presented in the exercises. Our main need for this inequality here is to prove another fundamental inequality involving norms of vectors. See Figure 3.

THEOREM 17

The Triangle Inequality

For all $\mathbf{u}$, $\mathbf{v}$ in V,

$$\|\mathbf{u} + \mathbf{v}\| \le \|\mathbf{u}\| + \|\mathbf{v}\|$$

PROOF

$$\begin{aligned}
\|\mathbf{u} + \mathbf{v}\|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{u}\rangle + 2\langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle \\
&\le \|\mathbf{u}\|^2 + 2|\langle \mathbf{u}, \mathbf{v}\rangle| + \|\mathbf{v}\|^2 \\
&\le \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\,\|\mathbf{v}\| + \|\mathbf{v}\|^2 \qquad \text{Cauchy–Schwarz} \\
&= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2
\end{aligned}$$

The triangle inequality follows immediately by taking square roots of both sides. ■

[2] See *Statistics and Experimental Design in Engineering and the Physical Sciences*, 2nd ed., by Norman L. Johnson and Fred C. Leone (New York: John Wiley & Sons, 1977). Tables there list "Orthogonal Polynomials," which are simply the values of the polynomial at numbers such as −2, −1, 0, 1, and 2.

An Inner Product for $C[a, b]$ (Calculus required)

Probably the most widely used inner product space for applications is the vector space $C[a, b]$ of all continuous functions on an interval $a \le t \le b$, with an inner product that we will describe.

We begin by considering a polynomial p and any integer n larger than or equal to the degree of p. Then p is in $\mathbb{P}_n$, and we may compute a "length" for p using the inner product of Example 2 involving evaluation at $n + 1$ points in $[a, b]$. However, this length of p captures the behavior at only those $n + 1$ points. Since p is in $\mathbb{P}_n$ for all large n, we could use a much larger n, with many more points for the "evaluation" inner product. See Figure 4.

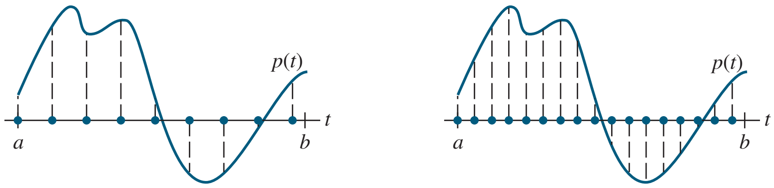
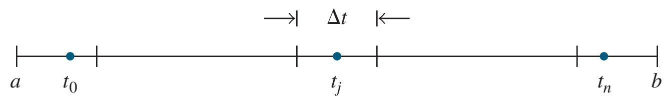


FIGURE 4 Using different numbers of evaluation points in $[a, b]$ to compute $\|p\|^2$.

Let us partition $[a, b]$ into $n + 1$ subintervals of length $\Delta t = (b - a)/(n + 1)$, and let $t_0, \dots, t_n$ be arbitrary points in these subintervals.



If n is large, the inner product on $\mathbb{P}_n$ determined by $t_0, \dots, t_n$ will tend to give a large value to $\langle p, p \rangle$, so we scale it down and divide by $n + 1$. Observe that $1/(n + 1) = \Delta t/(b - a)$, and define

$$\langle p, q \rangle = \frac{1}{n+1} \sum_{j=0}^{n} p(t_j)q(t_j) = \frac{1}{b-a} \left[\sum_{j=0}^{n} p(t_j)q(t_j)\Delta t \right]$$

Now, let n increase without bound. Since polynomials p and q are continuous functions, the expression in brackets is a Riemann sum that approaches a definite integral, and we are led to consider the *average value of* $p(t)q(t)$ on the interval $[a, b]$:

$$\frac{1}{b-a} \int_a^b p(t)q(t)\,dt$$

This quantity is defined for polynomials of any degree (in fact, for all continuous functions), and it has all the properties of an inner product, as the next example shows. The scale factor $1/(b - a)$ is inessential and is often omitted for simplicity.

EXAMPLE 7 For f, g in $C[a, b]$, set

$$\langle f, g \rangle = \int_a^b f(t)g(t)\,dt \tag{5}$$

Show that (5) defines an inner product on $C[a, b]$.

SOLUTION Inner product Axioms 1–3 follow from elementary properties of definite integrals. For Axiom 4, observe that

$$\langle f, f \rangle = \int_a^b [f(t)]^2 \, dt \geq 0$$

The function $[f(t)]^2$ is continuous and nonnegative on $[a, b]$. If the definite integral of $[f(t)]^2$ is zero, then $[f(t)]^2$ must be identically zero on $[a, b]$, by a theorem in advanced calculus, in which case f is the zero function. Thus $\langle f, f \rangle = 0$ implies that f is the zero function on $[a, b]$. So (5) defines an inner product on $C[a, b]$. ■

EXAMPLE 8 Let V be the space $C[0, 1]$ with the inner product of Example 7, and let W be the subspace spanned by the polynomials $p_1(t) = 1$, $p_2(t) = 2t - 1$, and $p_3(t) = 12t^2$. Use the Gram–Schmidt process to find an orthogonal basis for W.

SOLUTION Let $q_1 = p_1$, and compute

$$\langle p_2, q_1 \rangle = \int_0^1 (2t - 1)(1) \, dt = (t^2 - t)\Big|_0^1 = 0$$

So p_2 is already orthogonal to q_1, and we can take $q_2 = p_2$. For the projection of p_3 onto $W_2 = \text{Span}\{q_1, q_2\}$, compute

$$\langle p_3, q_1 \rangle = \int_0^1 12t^2 \cdot 1 \, dt = 4t^3\Big|_0^1 = 4$$

$$\langle q_1, q_1 \rangle = \int_0^1 1 \cdot 1 \, dt = t\Big|_0^1 = 1$$

$$\langle p_3, q_2 \rangle = \int_0^1 12t^2(2t - 1) \, dt = \int_0^1 (24t^3 - 12t^2) \, dt = 2$$

$$\langle q_2, q_2 \rangle = \int_0^1 (2t - 1)^2 \, dt = \frac{1}{6}(2t - 1)^3\Big|_0^1 = \frac{1}{3}$$

Then

$$\text{proj}_{W_2} p_3 = \frac{\langle p_3, q_1 \rangle}{\langle q_1, q_1 \rangle} q_1 + \frac{\langle p_3, q_2 \rangle}{\langle q_2, q_2 \rangle} q_2 = \frac{4}{1} q_1 + \frac{2}{1/3} q_2 = 4q_1 + 6q_2$$

and

$$q_3 = p_3 - \text{proj}_{W_2} p_3 = p_3 - 4q_1 - 6q_2$$

As a function, $q_3(t) = 12t^2 - 4 - 6(2t - 1) = 12t^2 - 12t + 2$. The orthogonal basis for the subspace W is $\{q_1, q_2, q_3\}$. ■

Practice Problems

Use the inner product axioms to verify the following statements.

1. $\langle \mathbf{v}, \mathbf{0} \rangle = \langle \mathbf{0}, \mathbf{v} \rangle = 0$.
2. $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$.

6.7 Exercises

1. Let $\mathbb{R}^2$ have the inner product of Example 1, and let $\mathbf{x} = (1, 1)$ and $\mathbf{y} = (5, -1)$.

a. Find $\|\mathbf{x}\|$, $\|\mathbf{y}\|$, and $|\langle \mathbf{x}, \mathbf{y}\rangle|^2$.

b. Describe all vectors (z_1, z_2) that are orthogonal to $\mathbf{y}$.

2. Let $\mathbb{R}^2$ have the inner product of Example 1. Show that the Cauchy–Schwarz inequality holds for $\mathbf{x} = (3, -4)$ and $\mathbf{y} = (-4, 3)$. [*Suggestion:* Study $|\langle \mathbf{x}, \mathbf{y}\rangle|^2$.]

Exercises 3–8 refer to $\mathbb{P}_2$ with the inner product given by evaluation at -1, 0, and 1. (See Example 2.)

3. Compute $\langle p, q\rangle$, where $p(t) = 4 + t$, $q(t) = 5 - 4t^2$.

4. Compute $\langle p, q\rangle$, where $p(t) = 4t - 3t^2$, $q(t) = 1 + 9t^2$.

5. Compute $\|p\|$ and $\|q\|$, for p and q in Exercise 3.

6. Compute $\|p\|$ and $\|q\|$, for p and q in Exercise 4.

7. Compute the orthogonal projection of q onto the subspace spanned by p, for p and q in Exercise 3.

8. Compute the orthogonal projection of q onto the subspace spanned by p, for p and q in Exercise 4.

9. Let $\mathbb{P}_3$ have the inner product given by evaluation at -3, -1, 1, and 3. Let $p_0(t) = 1$, $p_1(t) = t$, and $p_2(t) = t^2$.

a. Compute the orthogonal projection of p_2 onto the subspace spanned by p_0 and p_1.

b. Find a polynomial q that is orthogonal to p_0 and p_1, such that $\{p_0, p_1, q\}$ is an orthogonal basis for Span $\{p_0, p_1, p_2\}$. Scale the polynomial q so that its vector of values at $(-3, -1, 1, 3)$ is $(1, -1, -1, 1)$.

10. Let $\mathbb{P}_3$ have the inner product as in Exercise 9, with p_0, p_1, and q the polynomials described there. Find the best approximation to $p(t) = t^3$ by polynomials in Span $\{p_0, p_1, q\}$.

11. Let p_0, p_1, and p_2 be the orthogonal polynomials described in Example 5, where the inner product on $\mathbb{P}_4$ is given by evaluation at -2, -1, 0, 1, and 2. Find the orthogonal projection of t^3 onto Span $\{p_0, p_1, p_2\}$.

12. Find a polynomial p_3 such that $\{p_0, p_1, p_2, p_3\}$ (see Exercise 11) is an orthogonal basis for the subspace $\mathbb{P}_3$ of $\mathbb{P}_4$. Scale the polynomial p_3 so that its vector of values is $(-1, 2, 0, -2, 1)$.

13. Let A be any invertible $n \times n$ matrix. Show that for $\mathbf{u}$, $\mathbf{v}$ in $\mathbb{R}^n$, the formula $\langle \mathbf{u}, \mathbf{v}\rangle = (A\mathbf{u})\cdot(A\mathbf{v}) = (A\mathbf{u})^T(A\mathbf{v})$ defines an inner product on $\mathbb{R}^n$.

14. Let T be a one-to-one linear transformation from a vector space V into $\mathbb{R}^n$. Show that for $\mathbf{u}$, $\mathbf{v}$ in V, the formula $\langle \mathbf{u}, \mathbf{v}\rangle = T(\mathbf{u})\cdot T(\mathbf{v})$ defines an inner product on V.

Use the inner product axioms and other results of this section to verify the statements in Exercises 15–18.

15. $\langle \mathbf{u}, c\mathbf{v}\rangle = c\langle \mathbf{u}, \mathbf{v}\rangle$ for all scalars c.

16. If $\{\mathbf{u}, \mathbf{v}\}$ is an orthonormal set in V, then $\|\mathbf{u} - \mathbf{v}\| = \sqrt{2}$.

17. $\langle \mathbf{u}, \mathbf{v}\rangle = \frac{1}{4}\|\mathbf{u} + \mathbf{v}\|^2 - \frac{1}{4}\|\mathbf{u} - \mathbf{v}\|^2$.

18. $\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2 = 2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2$.

In Exercises 19–24, $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors. Mark each statement True or False **(T/F)**. Justify each answer.

19. (T/F) If $\langle \mathbf{u}, \mathbf{u}\rangle = 0$, then $\mathbf{u} = \mathbf{0}$.

20. (T/F) If $\langle \mathbf{u}, \mathbf{v}\rangle = 0$, then either $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$.

21. (T/F) $\langle \mathbf{u} + \mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{w}, \mathbf{u}\rangle + \langle \mathbf{w}, \mathbf{v}\rangle$.

22. (T/F) $\langle c\mathbf{u}, c\mathbf{v}\rangle = c\langle \mathbf{u}, \mathbf{v}\rangle$.

23. (T/F) $|\langle \mathbf{u}, \mathbf{u}\rangle| = \langle \mathbf{u}, \mathbf{u}\rangle$.

24. (T/F) $|\langle \mathbf{u}, \mathbf{v}\rangle| \le \|\mathbf{u}\|\,\|\mathbf{v}\|$.

25. Given $a \ge 0$ and $b \ge 0$, let $\mathbf{u} = \begin{bmatrix} \sqrt{a} \\ \sqrt{b} \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} \sqrt{b} \\ \sqrt{a} \end{bmatrix}$. Use the Cauchy–Schwarz inequality to compare the geometric mean $\sqrt{ab}$ with the arithmetic mean $(a + b)/2$.

26. Let $\mathbf{u} = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Use the Cauchy–Schwarz inequality to show that

$$\left(\frac{a + b}{2}\right)^2 \le \frac{a^2 + b^2}{2}$$

Exercises 27–30 refer to $V = C[0, 1]$, with the inner product given by an integral, as in Example 7.

27. Compute $\langle f, g\rangle$, where $f(t) = 1 - 3t^2$ and $g(t) = t - t^3$.

28. Compute $\langle f, g\rangle$, where $f(t) = 5t - 2$ and $g(t) = 7t^3 - 6t^2$.

29. Compute $\|f\|$ for f in Exercise 27.

30. Compute $\|g\|$ for g in Exercise 28.

31. Let V be the space $C[-1, 1]$ with the inner product of Example 7. Find an orthogonal basis for the subspace spanned by the polynomials 1, t, and t^2. The polynomials in this basis are called *Legendre polynomials*.

32. Let V be the space $C[-2, 2]$ with the inner product of Example 7. Find an orthogonal basis for the subspace spanned by the polynomials 1, t, and t^2.

T 33. Let $\mathbb{P}_4$ have the inner product as in Example 5, and let p_0, p_1, p_2 be the orthogonal polynomials from that example. Using your matrix program, apply the Gram–Schmidt process to the set $\{p_0, p_1, p_2, t^3, t^4\}$ to create an orthogonal basis for $\mathbb{P}_4$.

T 34. Let V be the space $C[0, 2\pi]$ with the inner product of Example 7. Use the Gram–Schmidt process to create an orthogonal basis for the subspace spanned by $\{1, \cos t, \cos^2 t, \cos^3 t\}$. Use a matrix program or computational program to compute the appropriate definite integrals.

Solutions to Practice Problems

1. By Axiom 1, $\langle \mathbf{v}, \mathbf{0} \rangle = \langle \mathbf{0}, \mathbf{v} \rangle$. Then $\langle \mathbf{0}, \mathbf{v} \rangle = \langle 0\mathbf{v}, \mathbf{v} \rangle = 0\langle \mathbf{v}, \mathbf{v} \rangle$, by Axiom 3, so $\langle \mathbf{0}, \mathbf{v} \rangle = 0$.
2. By Axioms 1, 2, and then 1 again, $\langle \mathbf{u}, \mathbf{v}+\mathbf{w} \rangle = \langle \mathbf{v}+\mathbf{w}, \mathbf{u} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{w}, \mathbf{u} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$.

6.8 Applications of Inner Product Spaces

The examples in this section suggest how the inner product spaces defined in Section 6.7 arise in practical problems. Like in Section 6.6, important components of machine learning are analyzed.

Weighted Least-Squares

Let $\mathbf{y}$ be a vector of n observations, $y_1, \dots, y_n$, and suppose we wish to approximate $\mathbf{y}$ by a vector $\hat{\mathbf{y}}$ that belongs to some specified subspace of $\mathbb{R}^n$. (In Section 6.5, $\hat{\mathbf{y}}$ was written as $A\mathbf{x}$ so that $\hat{\mathbf{y}}$ was in the column space of A.) Denote the entries in $\hat{\mathbf{y}}$ by $\hat{y}_1, \dots, \hat{y}_n$. Then the *sum of the squares for error*, or SS(E), in approximating $\mathbf{y}$ by $\hat{\mathbf{y}}$ is

$$\text{SS(E)} = (y_1 - \hat{y}_1)^2 + \cdots + (y_n - \hat{y}_n)^2 \tag{1}$$

This is simply $\|\mathbf{y} - \hat{\mathbf{y}}\|^2$, using the standard length in $\mathbb{R}^n$.

Now suppose the measurements that produced the entries in $\mathbf{y}$ are not equally reliable. The entries in $\mathbf{y}$ might be computed from various samples of measurements, with unequal sample sizes. Then it becomes appropriate to weight the squared errors in (1) in such a way that more importance is assigned to the more reliable measurements.[1] If the weights are denoted by $w_1^2, \dots, w_n^2$, then the weighted sum of the squares for error is

$$\text{Weighted SS(E)} = w_1^2(y_1 - \hat{y}_1)^2 + \cdots + w_n^2(y_n - \hat{y}_n)^2 \tag{2}$$

This is the square of the length of $\mathbf{y} - \hat{\mathbf{y}}$, where the length is derived from an inner product analogous to that in Example 1 in Section 6.7, namely

$$\langle \mathbf{x}, \mathbf{y} \rangle = w_1^2 x_1 y_1 + \cdots + w_n^2 x_n y_n$$

It is sometimes convenient to transform a weighted least-squares problem into an equivalent ordinary least-squares problem. Let W be the diagonal matrix with (positive) $w_1, \dots, w_n$ on its diagonal, so that

$$W\mathbf{y} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & & \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & w_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} w_1 y_1 \\ w_2 y_2 \\ \vdots \\ w_n y_n \end{bmatrix}$$

with a similar expression for $W\hat{\mathbf{y}}$. Observe that the jth term in (2) can be written as

$$w_j^2(y_j - \hat{y}_j)^2 = (w_j y_j - w_j \hat{y}_j)^2$$

[1] Note for readers with a background in statistics: Suppose the errors in measuring the y_i are independent random variables with means equal to zero and variances of $\sigma_1^2, \dots, \sigma_n^2$. Then the appropriate weights in (2) are $w_i^2 = 1/\sigma_i^2$. The larger the variance of the error, the smaller the weight.

7.1 Diagonalization of Symmetric Matrices

A **symmetric** matrix is a matrix A such that $A^T = A$. Such a matrix is necessarily square. Its main diagonal entries are arbitrary, but its other entries occur in pairs—on opposite sides of the main diagonal.

EXAMPLE 1 Of the following matrices, only the first three are symmetric:

$$\text{Symmetric:} \quad \begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix}, \quad \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & 8 \\ 0 & 8 & -7 \end{bmatrix}, \quad \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}$$

$$\text{Nonsymmetric:} \quad \begin{bmatrix} 1 & -3 \\ 3 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & -4 & 0 \\ -6 & 1 & -4 \\ 0 & -6 & 1 \end{bmatrix}, \quad \begin{bmatrix} 5 & 4 & 3 & 2 \\ 4 & 3 & 2 & 1 \\ 3 & 2 & 1 & 0 \end{bmatrix}$$

■

To begin the study of symmetric matrices, it is helpful to review the diagonalization process of Section 5.3.

EXAMPLE 2 If possible, diagonalize the matrix $A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}$.

SOLUTION The characteristic equation of A is

$$0 = -\lambda^3 + 17\lambda^2 - 90\lambda + 144 = -(\lambda - 8)(\lambda - 6)(\lambda - 3)$$

Standard calculations produce a basis for each eigenspace:

$$\lambda = 8: \ \mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}; \qquad \lambda = 6: \ \mathbf{v}_2 = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}; \qquad \lambda = 3: \ \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

These three vectors form a basis for $\mathbb{R}^3$. In fact, it is easy to check that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an *orthogonal* basis for $\mathbb{R}^3$. Experience from Chapter 6 suggests that an *orthonormal* basis might be useful for calculations, so here are the normalized (unit) eigenvectors.

$$\mathbf{u}_1 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

Let

$$P = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}, \quad D = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Then $A = PDP^{-1}$, as usual. But this time, since P is square and has orthonormal columns, P is an *orthogonal* matrix, and P^{-1} is simply P^T. (See Section 6.2.) ■

Theorem 1 explains why the eigenvectors in Example 2 are orthogonal—they correspond to distinct eigenvalues.

THEOREM 1

> If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.

PROOF Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be eigenvectors that correspond to distinct eigenvalues, say, λ_1 and λ_2. To show that $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$, compute

$$\begin{aligned} \lambda_1 \mathbf{v}_1 \cdot \mathbf{v}_2 &= (\lambda_1 \mathbf{v}_1)^T \mathbf{v}_2 = (A\mathbf{v}_1)^T \mathbf{v}_2 && \text{Since } \mathbf{v}_1 \text{ is an eigenvector} \\ &= (\mathbf{v}_1^T A^T)\mathbf{v}_2 = \mathbf{v}_1^T (A\mathbf{v}_2) && \text{Since } A^T = A \\ &= \mathbf{v}_1^T (\lambda_2 \mathbf{v}_2) && \text{Since } \mathbf{v}_2 \text{ is an eigenvector} \\ &= \lambda_2 \mathbf{v}_1^T \mathbf{v}_2 = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2 \end{aligned}$$

Hence $(\lambda_1 - \lambda_2)\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. But $\lambda_1 - \lambda_2 \neq 0$, so $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. ■

The special type of diagonalization in Example 2 is crucial for the theory of symmetric matrices. An $n \times n$ matrix A is said to be **orthogonally diagonalizable** if there are an orthogonal matrix P (with $P^{-1} = P^T$) and a diagonal matrix D such that

$$A = PDP^T = PDP^{-1} \tag{1}$$

Such a diagonalization requires n linearly independent and orthonormal eigenvectors. When is this possible? If A is orthogonally diagonalizable as in (1), then

$$A^T = (PDP^T)^T = P^{TT} D^T P^T = PDP^T = A$$

Thus A is symmetric! Theorem 2 below shows that, conversely, every symmetric matrix is orthogonally diagonalizable. The proof is much harder and is omitted; the main idea for a proof will be given after Theorem 3.

THEOREM 2

> An $n \times n$ matrix A is orthogonally diagonalizable if and only if A is a symmetric matrix.

This theorem is rather amazing, because the work in Chapter 5 would suggest that it is usually impossible to tell when a matrix is diagonalizable. But this is not the case for symmetric matrices.

The next example treats a matrix whose eigenvalues are not all distinct.

EXAMPLE 3 Orthogonally diagonalize the matrix $A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$, whose characteristic equation is

$$0 = -\lambda^3 + 12\lambda^2 - 21\lambda - 98 = -(\lambda - 7)^2(\lambda + 2)$$

SOLUTION The usual calculations produce bases for the eigenspaces:

$$\lambda = 7: \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix}; \qquad \lambda = -2: \mathbf{v}_3 = \begin{bmatrix} -1 \\ -1/2 \\ 1 \end{bmatrix}$$

Although $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent, they are not orthogonal. Recall from Section 6.2 that the projection of $\mathbf{v}_2$ onto $\mathbf{v}_1$ is $\frac{\mathbf{v}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1$, and the component of $\mathbf{v}_2$ orthogonal to $\mathbf{v}_1$ is

$$\mathbf{z}_2 = \mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix} - \frac{-1/2}{2}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1/4 \\ 1 \\ 1/4 \end{bmatrix}$$

Then $\{\mathbf{v}_1, \mathbf{z}_2\}$ is an orthogonal set in the eigenspace for $\lambda = 7$. (Note that $\mathbf{z}_2$ is a linear combination of the eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$, so $\mathbf{z}_2$ is in the eigenspace. This construction of $\mathbf{z}_2$ is just the Gram–Schmidt process of Section 6.4.) Since the eigenspace is two-dimensional (with basis $\mathbf{v}_1, \mathbf{v}_2$), the orthogonal set $\{\mathbf{v}_1, \mathbf{z}_2\}$ is an *orthogonal basis* for the eigenspace, by the Basis Theorem. (See Section 2.9 or 4.5.)

Normalize $\mathbf{v}_1$ and $\mathbf{z}_2$ to obtain the following orthonormal basis for the eigenspace for $\lambda = 7$:

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1/\sqrt{18} \\ 4/\sqrt{18} \\ 1/\sqrt{18} \end{bmatrix}$$

An orthonormal basis for the eigenspace for $\lambda = -2$ is

$$\mathbf{u}_3 = \frac{1}{\|2\mathbf{v}_3\|}2\mathbf{v}_3 = \frac{1}{3}\begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}$$

By Theorem 1, $\mathbf{u}_3$ is orthogonal to the other eigenvectors $\mathbf{u}_1$ and $\mathbf{u}_2$. Hence $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal set. Let

$$P = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3\,] = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{18} & -2/3 \\ 0 & 4/\sqrt{18} & -1/3 \\ 1/\sqrt{2} & 1/\sqrt{18} & 2/3 \end{bmatrix}, \quad D = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

Then P orthogonally diagonalizes A, and $A = PDP^{-1}$. ■

In Example 3, the eigenvalue 7 has multiplicity two and the eigenspace is two-dimensional. This fact is not accidental, as the next theorem shows.

The Spectral Theorem

The set of eigenvalues of a matrix A is sometimes called the *spectrum* of A, and the following description of the eigenvalues is called a *spectral theorem.*

THEOREM 3

The Spectral Theorem for Symmetric Matrices

An $n \times n$ symmetric matrix A has the following properties:

a. A has n real eigenvalues, counting multiplicities.

b. The dimension of the eigenspace for each eigenvalue λ equals the multiplicity of λ as a root of the characteristic equation.

c. The eigenspaces are mutually orthogonal, in the sense that eigenvectors corresponding to different eigenvalues are orthogonal.

d. A is orthogonally diagonalizable.

Part (a) follows from Exercise 28 in Section 5.5. Part (b) follows easily from part (d). (See Exercise 37.) Part (c) is Theorem 1. Because of (a), a proof of (d) can be given using Exercise 38 and the Schur factorization discussed in Supplementary Exercise 34 in Chapter 6. The details are omitted.

Spectral Decomposition

Suppose $A = PDP^{-1}$, where the columns of P are orthonormal eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_n$ of A and the corresponding eigenvalues $\lambda_1, \dots, \lambda_n$ are in the diagonal matrix D. Then, since $P^{-1} = P^T$,

$$\begin{aligned} A = PDP^T &= \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\mathbf{u}_1 & \cdots & \lambda_n\mathbf{u}_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \end{aligned}$$

Using the column–row expansion of a product (Theorem 10 in Section 2.4), we can write

$$A = \lambda_1\mathbf{u}_1\mathbf{u}_1^T + \lambda_2\mathbf{u}_2\mathbf{u}_2^T + \cdots + \lambda_n\mathbf{u}_n\mathbf{u}_n^T \tag{2}$$

This representation of A is called a **spectral decomposition** of A because it breaks up A into pieces determined by the spectrum (eigenvalues) of A. Each term in (2) is an $n \times n$ matrix of rank 1. For example, every column of $\lambda_1\mathbf{u}_1\mathbf{u}_1^T$ is a multiple of $\mathbf{u}_1$. Furthermore, each matrix $\mathbf{u}_j\mathbf{u}_j^T$ is a **projection matrix** in the sense that for each $\mathbf{x}$ in $\mathbb{R}^n$, the vector $(\mathbf{u}_j\mathbf{u}_j^T)\mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ onto the subspace spanned by $\mathbf{u}_j$. (See Exercise 41.)

EXAMPLE 4 Construct a spectral decomposition of the matrix A that has the orthogonal diagonalization

$$A = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 8 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$$

SOLUTION Denote the columns of P by $\mathbf{u}_1$ and $\mathbf{u}_2$. Then

$$A = 8\mathbf{u}_1\mathbf{u}_1^T + 3\mathbf{u}_2\mathbf{u}_2^T$$

To verify this decomposition of A, compute

$$\mathbf{u}_1\mathbf{u}_1^T = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 4/5 & 2/5 \\ 2/5 & 1/5 \end{bmatrix}$$

$$\mathbf{u}_2\mathbf{u}_2^T = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 1/5 & -2/5 \\ -2/5 & 4/5 \end{bmatrix}$$

and

$$8\mathbf{u}_1\mathbf{u}_1^T + 3\mathbf{u}_2\mathbf{u}_2^T = \begin{bmatrix} 32/5 & 16/5 \\ 16/5 & 8/5 \end{bmatrix} + \begin{bmatrix} 3/5 & -6/5 \\ -6/5 & 12/5 \end{bmatrix} = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix} = A \quad \blacksquare$$

Numerical Notes

When A is symmetric and not too large, modern high-performance computer algorithms calculate eigenvalues and eigenvectors with great precision. They apply a sequence of similarity transformations to A involving orthogonal matrices. The diagonal entries of the transformed matrices converge rapidly to the eigenvalues of A. (See the Numerical Notes in Section 5.2.) Using orthogonal matrices generally prevents numerical errors from accumulating during the process. When A is symmetric, the sequence of orthogonal matrices combines to form an orthogonal matrix whose columns are eigenvectors of A.

A nonsymmetric matrix cannot have a full set of orthogonal eigenvectors, but the algorithm still produces fairly accurate eigenvalues. After that, nonorthogonal techniques are needed to calculate eigenvectors.

Practice Problems

1. Show that if A is a symmetric matrix, then A^2 is symmetric.
2. Show that if A is orthogonally diagonalizable, then so is A^2.

7.1 Exercises

Determine which of the matrices in Exercises 1–6 are symmetric.

1. $\begin{bmatrix} 4 & 3 \\ 3 & -8 \end{bmatrix}$

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

3. $\begin{bmatrix} 3 & 5 \\ 3 & 7 \end{bmatrix}$

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

5. $\begin{bmatrix} -2 & 4 & 5 \\ 4 & -2 & 3 \\ 5 & 3 & -2 \end{bmatrix}$

6. $\begin{bmatrix} 2 & 1 & 1 & 2 \\ 3 & 3 & 3 & 2 \\ 1 & 1 & 2 & 1 \end{bmatrix}$

Determine which of the matrices in Exercises 7–12 are orthogonal. If orthogonal, find the inverse.

7. $\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

8. $\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}$

9. $\begin{bmatrix} -3/5 & 4/5 \\ 4/5 & 3/5 \end{bmatrix}$

10. $\begin{bmatrix} 2/3 & 1/3 & -2/3 \\ -2/3 & 2/3 & -1/3 \\ 1/3 & 2/3 & 2/3 \end{bmatrix}$

11. $\begin{bmatrix} -2/3 & 1/3 & 2/3 \\ 0 & 2/3 & -1/3 \\ 5/3 & 2/3 & 4/3 \end{bmatrix}$

12. $\begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & -3/\sqrt{12} \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} & 0 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \end{bmatrix}$

Orthogonally diagonalize the matrices in Exercises 13–22, giving an orthogonal matrix P and a diagonal matrix D. To save you time, the eigenvalues in Exercises 17–22 are the following: (17) $-5, 5, 8$; (18) $1, 2, 5$; (19) $8, -1$; (20) $-3, 15$; (21) $3, 5, 9$; (22) $4, 6$.

13. $\begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$

14. $\begin{bmatrix} 2 & -3 \\ -3 & 2 \end{bmatrix}$

15. $\begin{bmatrix} 5 & 6 \\ 6 & 10 \end{bmatrix}$

16. $\begin{bmatrix} 5 & -4 \\ -4 & 11 \end{bmatrix}$

17. $\begin{bmatrix} 1 & 1 & 6 \\ 1 & 6 & 1 \\ 6 & 1 & 1 \end{bmatrix}$

18. $\begin{bmatrix} 2 & -1 & 1 \\ -1 & 4 & -1 \\ 1 & -1 & 2 \end{bmatrix}$

19. $\begin{bmatrix} 4 & -2 & 4 \\ -2 & 7 & 2 \\ 4 & 2 & 4 \end{bmatrix}$

20. $\begin{bmatrix} 5 & -8 & 4 \\ -8 & 5 & -4 \\ 4 & -4 & -1 \end{bmatrix}$

21. $\begin{bmatrix} 5 & 4 & 1 & 1 \\ 4 & 5 & 1 & 1 \\ 1 & 1 & 5 & 4 \\ 1 & 1 & 4 & 5 \end{bmatrix}$

22. $\begin{bmatrix} 5 & 0 & 1 & 0 \\ 0 & 5 & 0 & 1 \\ 1 & 0 & 5 & 0 \\ 0 & 1 & 0 & 5 \end{bmatrix}$

23. Let $A = \begin{bmatrix} 5 & -1 & -1 \\ -1 & 5 & -1 \\ -1 & -1 & 5 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Verify that 3 is an eigenvalue of A and $\mathbf{v}$ is an eigenvector. Then orthogonally diagonalize A.

24. Let $A = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$. Verify that $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors of A. Then orthogonally diagonalize A.

In Exercises 25–32, mark each statement True or False **(T/F)**. Justify each answer.

25. **(T/F)** An $n \times n$ matrix that is orthogonally diagonalizable must be symmetric.

26. **(T/F)** There are symmetric matrices that are not orthogonally diagonalizable.

27. **(T/F)** An orthogonal matrix is orthogonally diagonalizable.

28. **(T/F)** If $B = PDP^T$, where $P^T = P^{-1}$ and D is a diagonal matrix, then B is a symmetric matrix.

29. **(T/F)** For a nonzero $\mathbf{v}$ in $\mathbb{R}^n$, the matrix $\mathbf{v}\mathbf{v}^T$ is called a projection matrix.

30. **(T/F)** If $A^T = A$ and if vectors $\mathbf{u}$ and $\mathbf{v}$ satisfy $A\mathbf{u} = 3\mathbf{u}$ and $A\mathbf{v} = 4\mathbf{v}$, then $\mathbf{u} \cdot \mathbf{v} = 0$.

31. **(T/F)** An $n \times n$ symmetric matrix has n distinct real eigenvalues.

32. **(T/F)** The dimension of an eigenspace of a symmetric matrix is sometimes less than the multiplicity of the corresponding eigenvalue.

33. Show that if A is an $n \times n$ symmetric matrix, then $(A\mathbf{x})\cdot\mathbf{y} = \mathbf{x}\cdot(A\mathbf{y})$ for all $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^n$.

34. Suppose A is a symmetric $n \times n$ matrix and B is any $n \times m$ matrix. Show that B^TAB, B^TB, and BB^T are symmetric matrices.

35. Suppose A is invertible and orthogonally diagonalizable. Explain why A^{-1} is also orthogonally diagonalizable.

36. Suppose A and B are both orthogonally diagonalizable and $AB = BA$. Explain why AB is also orthogonally diagonalizable.

37. Let $A = PDP^{-1}$, where P is orthogonal and D is diagonal, and let λ be an eigenvalue of A of multiplicity k. Then λ appears k times on the diagonal of D. Explain why the dimension of the eigenspace for λ is k.

38. Suppose $A = PRP^{-1}$, where P is orthogonal and R is upper triangular. Show that if A is symmetric, then R is symmetric and hence is actually a diagonal matrix.

39. Construct a spectral decomposition of A from Example 2.

40. Construct a spectral decomposition of A from Example 3.

41. Let $\mathbf{u}$ be a unit vector in $\mathbb{R}^n$, and let $B = \mathbf{u}\mathbf{u}^T$.

a. Given any $\mathbf{x}$ in $\mathbb{R}^n$, compute $B\mathbf{x}$ and show that $B\mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ onto $\mathbf{u}$, as described in Section 6.2.

b. Show that B is a symmetric matrix and $B^2 = B$.

c. Show that $\mathbf{u}$ is an eigenvector of B. What is the corresponding eigenvalue?

42. Let B be an $n \times n$ symmetric matrix such that $B^2 = B$. Any such matrix is called a **projection matrix** (or an **orthogonal projection matrix**). Given any $\mathbf{y}$ in $\mathbb{R}^n$, let $\hat{\mathbf{y}} = B\mathbf{y}$ and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.

a. Show that $\mathbf{z}$ is orthogonal to $\hat{\mathbf{y}}$.

b. Let W be the column space of B. Show that $\mathbf{y}$ is the sum of a vector in W and a vector in $W^\perp$. Why does this prove that $B\mathbf{y}$ is the orthogonal projection of $\mathbf{y}$ onto the column space of B?

Orthogonally diagonalize the matrices in Exercises 43–46. To practice the methods of this section, do not use an eigenvector routine from your matrix program. Instead, use the program to find the eigenvalues, and, for each eigenvalue λ, find an orthonormal basis for $\text{Nul}(A - \lambda I)$, as in Examples 2 and 3.

T **43.** $\begin{bmatrix} 6 & 2 & 9 & -6 \\ 2 & 6 & -6 & 9 \\ 9 & -6 & 6 & 2 \\ -6 & 9 & 2 & 6 \end{bmatrix}$

T **44.** $\begin{bmatrix} .63 & -.18 & -.06 & -.04 \\ -.18 & .84 & -.04 & .12 \\ -.06 & -.04 & .72 & -.12 \\ -.04 & .12 & -.12 & .66 \end{bmatrix}$

T **45.** $\begin{bmatrix} .31 & .58 & .08 & .44 \\ .58 & -.56 & .44 & -.58 \\ .08 & .44 & .19 & -.08 \\ .44 & -.58 & -.08 & .31 \end{bmatrix}$

T **46.** $\begin{bmatrix} 8 & 2 & 2 & -6 & 9 \\ 2 & 8 & 2 & -6 & 9 \\ 2 & 2 & 8 & -6 & 9 \\ -6 & -6 & -6 & 24 & 9 \\ 9 & 9 & 9 & 9 & -21 \end{bmatrix}$

Solutions to Practice Problems

1. $(A^2)^T = (AA)^T = A^TA^T$, by a property of transposes. By hypothesis, $A^T = A$. So $(A^2)^T = AA = A^2$, which shows that A^2 is symmetric.

2. If A is orthogonally diagonalizable, then A is symmetric, by Theorem 2. By Practice Problem 1, A^2 is symmetric and hence is orthogonally diagonalizable (Theorem 2).

7.2 Quadratic Forms

Until now, our attention in this text has focused on linear equations, except for the sums of squares encountered in Chapter 6 when computing $\mathbf{x}^T\mathbf{x}$. Such sums and more general expressions, called *quadratic forms*, occur frequently in applications of linear algebra to engineering (in design criteria and optimization) and signal processing (as output noise power). They also arise, for example, in physics (as potential and kinetic energy), differential geometry (as normal curvature of surfaces), economics (as utility functions), and statistics (in confidence ellipsoids). Some of the mathematical background for such applications flows easily from our work on symmetric matrices.

A **quadratic form** on $\mathbb{R}^n$ is a function Q defined on $\mathbb{R}^n$ whose value at a vector $\mathbf{x}$ in $\mathbb{R}^n$ can be computed by an expression of the form $Q(\mathbf{x}) = \mathbf{x}^TA\mathbf{x}$, where A is an $n \times n$ symmetric matrix. The matrix A is called the **matrix of the quadratic form**.

The simplest example of a nonzero quadratic form is $Q(\mathbf{x}) = \mathbf{x}^TI\mathbf{x} = \|\mathbf{x}\|^2$. Examples 1 and 2 show the connection between any symmetric matrix A and the quadratic form $\mathbf{x}^TA\mathbf{x}$.

EXAMPLE 1 Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Compute $\mathbf{x}^TA\mathbf{x}$ for the following matrices:

a. $A = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}$ b. $A = \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix}$

SOLUTION

a. $\mathbf{x}^TA\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 4x_1 \\ 3x_2 \end{bmatrix} = 4x_1^2 + 3x_2^2.$

b. There are two -2 entries in A. Watch how they enter the calculations. The $(1, 2)$-entry in A is in boldface type.

$$\begin{aligned}
\mathbf{x}^TA\mathbf{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 3 & \mathbf{-2} \\ -2 & 7 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 3x_1 - \mathbf{2}x_2 \\ -2x_1 + 7x_2 \end{bmatrix} \\
&= x_1(3x_1 - \mathbf{2}x_2) + x_2(-2x_1 + 7x_2) \\
&= 3x_1^2 - \mathbf{2}x_1x_2 - 2x_2x_1 + 7x_2^2 \\
&= 3x_1^2 - 4x_1x_2 + 7x_2^2
\end{aligned}$$ ■

The presence of $-4x_1x_2$ in the quadratic form in Example 1(b) is due to the -2 entries off the diagonal in the matrix A. In contrast, the quadratic form associated with the diagonal matrix A in Example 1(a) has no x_1x_2 *cross-product* term.

EXAMPLE 2 For $\mathbf{x}$ in $\mathbb{R}^3$, let $Q(\mathbf{x}) = 5x_1^2 + 3x_2^2 + 2x_3^2 - x_1x_2 + 8x_2x_3$. Write this quadratic form as $\mathbf{x}^TA\mathbf{x}$.

SOLUTION The coefficients of x_1^2, x_2^2, x_3^2 go on the diagonal of A. To make A symmetric, the coefficient of x_ix_j for $i \neq j$ must be split evenly between the (i, j)- and (j, i)-entries in A. The coefficient of x_1x_3 is 0. It is readily checked that

$$Q(\mathbf{x}) = \mathbf{x}^TA\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}\begin{bmatrix} 5 & -1/2 & 0 \\ -1/2 & 3 & 4 \\ 0 & 4 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$ ■

EXAMPLE 3 Let $Q(\mathbf{x}) = x_1^2 - 8x_1x_2 - 5x_2^2$. Compute the value of $Q(\mathbf{x})$ for $\mathbf{x} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \end{bmatrix}$, and $\begin{bmatrix} 1 \\ -3 \end{bmatrix}$.

SOLUTION

$$\begin{aligned} Q(-3, 1) &= (-3)^2 - 8(-3)(1) - 5(1)^2 = 28 \\ Q(2, -2) &= (2)^2 - 8(2)(-2) - 5(-2)^2 = 16 \\ Q(1, -3) &= (1)^2 - 8(1)(-3) - 5(-3)^2 = -20 \end{aligned}$$

In some cases, quadratic forms are easier to use when they have no cross-product terms—that is, when the matrix of the quadratic form is a diagonal matrix. Fortunately, the cross-product term can be eliminated by making a suitable change of variable.

Change of Variable in a Quadratic Form

If $\mathbf{x}$ represents a variable vector in $\mathbb{R}^n$, then a **change of variable** is an equation of the form

$$\mathbf{x} = P\mathbf{y}, \qquad \text{or equivalently,} \qquad \mathbf{y} = P^{-1}\mathbf{x} \tag{1}$$

where P is an invertible matrix and $\mathbf{y}$ is a new variable vector in $\mathbb{R}^n$. Here $\mathbf{y}$ is the coordinate vector of $\mathbf{x}$ relative to the basis of $\mathbb{R}^n$ determined by the columns of P. (See Section 4.4.)

If the change of variable (1) is made in a quadratic form $\mathbf{x}^TA\mathbf{x}$, then

$$\mathbf{x}^TA\mathbf{x} = (P\mathbf{y})^TA(P\mathbf{y}) = \mathbf{y}^TP^TAP\mathbf{y} = \mathbf{y}^T(P^TAP)\mathbf{y} \tag{2}$$

and the new matrix of the quadratic form is P^TAP. Since A is symmetric, Theorem 2 guarantees that there is an *orthogonal* matrix P such that P^TAP is a diagonal matrix D, and the quadratic form in (2) becomes $\mathbf{y}^TD\mathbf{y}$. This is the strategy of the next example.

EXAMPLE 4 Make a change of variable that transforms the quadratic form in Example 3 into a quadratic form with no cross-product term.

SOLUTION The matrix of the quadratic form in Example 3 is

$$A = \begin{bmatrix} 1 & -4 \\ -4 & -5 \end{bmatrix}$$

The first step is to orthogonally diagonalize A. Its eigenvalues turn out to be $\lambda = 3$ and $\lambda = -7$. Associated unit eigenvectors are

$$\lambda = 3: \begin{bmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{bmatrix}; \qquad \lambda = -7: \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

These vectors are automatically orthogonal (because they correspond to distinct eigenvalues) and so provide an orthonormal basis for $\mathbb{R}^2$. Let

$$P = \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}, \qquad D = \begin{bmatrix} 3 & 0 \\ 0 & -7 \end{bmatrix}$$

Then $A = PDP^{-1}$ and $D = P^{-1}AP = P^TAP$, as pointed out earlier. A suitable change of variable is

$$\mathbf{x} = P\mathbf{y}, \qquad \text{where } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

Then

$$\begin{aligned} x_1^2 - 8x_1x_2 - 5x_2^2 &= \mathbf{x}^T A\mathbf{x} = (P\mathbf{y})^T A(P\mathbf{y}) \\ &= \mathbf{y}^T P^T A P\mathbf{y} = \mathbf{y}^T D\mathbf{y} \\ &= 3y_1^2 - 7y_2^2 \end{aligned}$$

To illustrate the meaning of the equality of quadratic forms in Example 4, we can compute $Q(\mathbf{x})$ for $\mathbf{x} = (2, -2)$ using the new quadratic form. First, since $\mathbf{x} = P\mathbf{y}$,

$$\mathbf{y} = P^{-1}\mathbf{x} = P^T\mathbf{x}$$

so

$$\mathbf{y} = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 2 \\ -2 \end{bmatrix} = \begin{bmatrix} 6/\sqrt{5} \\ -2/\sqrt{5} \end{bmatrix}$$

Hence

$$\begin{aligned} 3y_1^2 - 7y_2^2 &= 3(6/\sqrt{5})^2 - 7(-2/\sqrt{5})^2 = 3(36/5) - 7(4/5) \\ &= 80/5 = 16 \end{aligned}$$

This is the value of $Q(\mathbf{x})$ in Example 3 when $\mathbf{x} = (2, -2)$. See Figure 1.

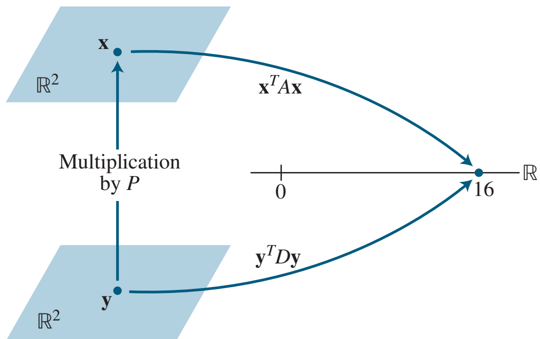


FIGURE 1 Change of variable in $\mathbf{x}^T A\mathbf{x}$.

Example 4 illustrates the following theorem. The proof of the theorem was essentially given before Example 4.

THEOREM 4

The Principal Axes Theorem

Let A be an $n \times n$ symmetric matrix. Then there is an orthogonal change of variable, $\mathbf{x} = P\mathbf{y}$, that transforms the quadratic form $\mathbf{x}^T A\mathbf{x}$ into a quadratic form $\mathbf{y}^T D\mathbf{y}$ with no cross-product term.

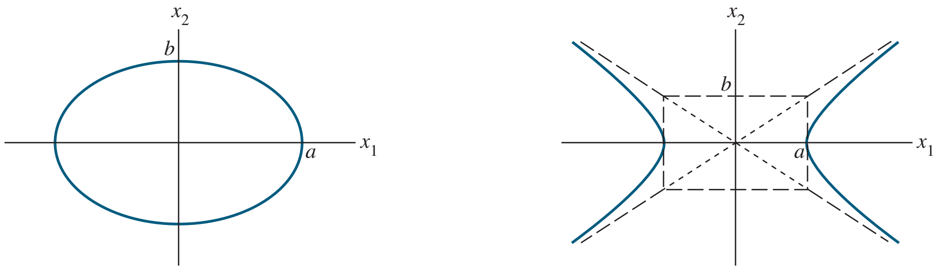
The columns of P in the theorem are called the **principal axes** of the quadratic form $\mathbf{x}^T A\mathbf{x}$. The vector $\mathbf{y}$ is the coordinate vector of $\mathbf{x}$ relative to the orthonormal basis of $\mathbb{R}^n$ given by these principal axes.

A Geometric View of Principal Axes

Suppose $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, where A is an invertible 2×2 symmetric matrix, and let c be a constant. It can be shown that the set of all $\mathbf{x}$ in $\mathbb{R}^2$ that satisfy

$$\mathbf{x}^T A\mathbf{x} = c \tag{3}$$

either corresponds to an ellipse (or circle), a hyperbola, two intersecting lines, or a single point, or contains no points at all. If A is a diagonal matrix, the graph is in *standard position*, such as in Figure 2. If A is not a diagonal matrix, the graph of equation (3) is



$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1, \; a > b > 0$$

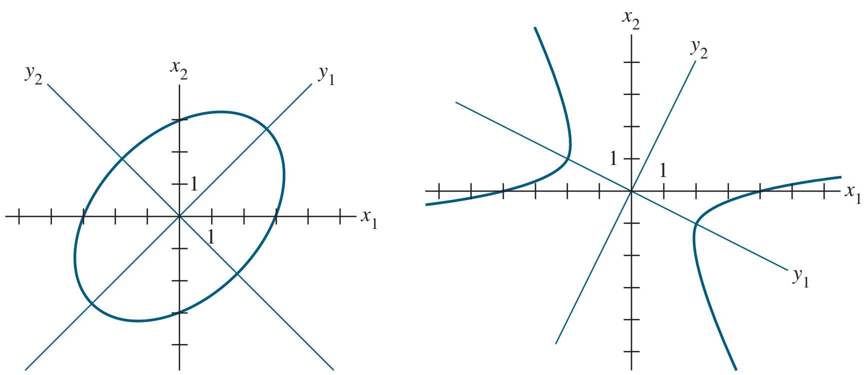
Ellipse

$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1, \; a > b > 0$$

Hyperbola

FIGURE 2 An ellipse and a hyperbola in standard position.

rotated out of standard position, as in Figure 3. Finding the *principal axes* (determined by the eigenvectors of A) amounts to finding a new coordinate system with respect to which the graph is in standard position.



(a) $5x_1^2 - 4x_1x_2 + 5x_2^2 = 48$ (b) $x_1^2 - 8x_1x_2 - 5x_2^2 = 16$

FIGURE 3 An ellipse and a hyperbola *not* in standard position.

The hyperbola in Figure 3(b) is the graph of the equation $\mathbf{x}^T A\mathbf{x} = 16$, where A is the matrix in Example 4. The positive y_1-axis in Figure 3(b) is in the direction of the first column of the matrix P in Example 4, and the positive y_2-axis is in the direction of the second column of P.

EXAMPLE 5 The ellipse in Figure 3(a) is the graph of the equation $5x_1^2 - 4x_1x_2 + 5x_2^2 = 48$. Find a change of variable that removes the cross-product term from the equation.

SOLUTION The matrix of the quadratic form is $A = \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$. The eigenvalues of A turn out to be 3 and 7, with corresponding unit eigenvectors

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

Let $P = [\,\mathbf{u}_1 \quad \mathbf{u}_2\,] = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$. Then P orthogonally diagonalizes A, so the change of variable $\mathbf{x} = P\mathbf{y}$ produces the quadratic form $\mathbf{y}^T D\mathbf{y} = 3y_1^2 + 7y_2^2$. The new axes for this change of variable are shown in Figure 3(a). ■

Classifying Quadratic Forms

When A is an $n \times n$ matrix, the quadratic form $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is a real-valued function with domain $\mathbb{R}^n$. Figure 4 displays the graphs of four quadratic forms with domain $\mathbb{R}^2$. For each point $\mathbf{x} = (x_1, x_2)$ in the domain of a quadratic form Q, the graph displays the point (x_1, x_2, z) where $z = Q(\mathbf{x})$. Notice that except at $\mathbf{x} = \mathbf{0}$, the values of $Q(\mathbf{x})$ are all positive in Figure 4(a) and all negative in Figure 4(d). The horizontal cross-sections of the graphs are ellipses in Figures 4(a) and 4(d) and hyperbolas in Figure 4(c).

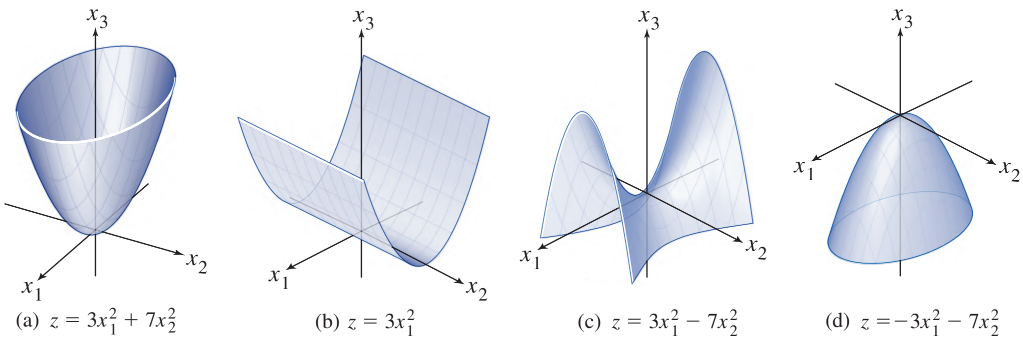


FIGURE 4 Graphs of quadratic forms.

The simple 2×2 examples in Figure 4 illustrate the following definitions.

DEFINITION

A quadratic form Q is

a. **positive definite** if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$,

b. **negative definite** if $Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$,

c. **indefinite** if $Q(\mathbf{x})$ assumes both positive and negative values.

Also, Q is said to be **positive semidefinite** if $Q(\mathbf{x}) \geq 0$ for all $\mathbf{x}$, and to be **negative semidefinite** if $Q(\mathbf{x}) \leq 0$ for all $\mathbf{x}$. The quadratic forms in parts (a) and (b) of Figure 4 are both positive semidefinite, but the form in (a) is better described as positive definite.

Theorem 5 characterizes some quadratic forms in terms of eigenvalues.

THEOREM 5

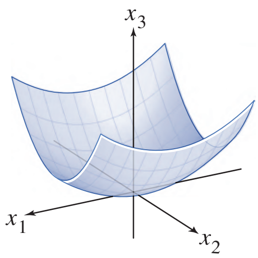
Quadratic Forms and Eigenvalues

Let A be an $n \times n$ symmetric matrix. Then a quadratic form $\mathbf{x}^T A\mathbf{x}$ is

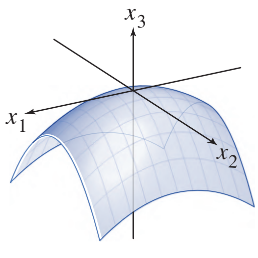
a. positive definite if and only if the eigenvalues of A are all positive,

b. negative definite if and only if the eigenvalues of A are all negative, or

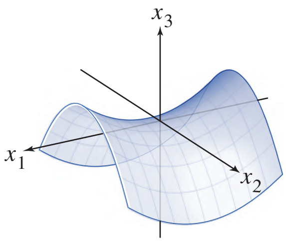
c. indefinite if and only if A has both positive and negative eigenvalues.



Positive definite



Negative definite



Indefinite

PROOF By the Principal Axes Theorem, there exists an orthogonal change of variable $\mathbf{x} = P\mathbf{y}$ such that

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \tag{4}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. Since P is invertible, there is a one-to-one correspondence between all nonzero $\mathbf{x}$ and all nonzero $\mathbf{y}$. Thus the values of $Q(\mathbf{x})$ for $\mathbf{x} \neq \mathbf{0}$ coincide with the values of the expression on the right side of (4), which is obviously controlled by the signs of the eigenvalues $\lambda_1, \ldots, \lambda_n$, in the three ways described in the theorem. ■

EXAMPLE 6 Is $Q(\mathbf{x}) = 3x_1^2 + 2x_2^2 + x_3^2 + 4x_1x_2 + 4x_2x_3$ positive definite?

SOLUTION Because of all the plus signs, this form "looks" positive definite. But the matrix of the form is

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

and the eigenvalues of A turn out to be 5, 2, and -1. So Q is an indefinite quadratic form, not positive definite. ■

The classification of a quadratic form is often carried over to the matrix of the form. Thus a **positive definite matrix** A is a *symmetric* matrix for which the quadratic form $\mathbf{x}^T A\mathbf{x}$ is positive definite. Other terms, such as **positive semidefinite matrix**, are defined analogously.

Numerical Notes

A fast way to determine whether a symmetric matrix A is positive definite is to attempt to factor A in the form $A = R^T R$, where R is upper triangular with positive diagonal entries. (A slightly modified algorithm for an LU factorization is one approach.) Such a *Cholesky factorization* is possible if and only if A is positive definite. See Supplementary Exercise 23 at the end of Chapter 7.

Practice Problem

Describe a positive semidefinite matrix A in terms of its eigenvalues.

7.2 Exercises

1. Compute the quadratic form $\mathbf{x}^T A\mathbf{x}$, when $A = \begin{bmatrix} 3 & 1/4 \\ 1/4 & 1 \end{bmatrix}$ and

a. $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ b. $\mathbf{x} = \begin{bmatrix} 8 \\ 1 \end{bmatrix}$ c. $\mathbf{x} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$

2. Compute the quadratic form $\mathbf{x}^T A\mathbf{x}$, for $A = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 1 & 3 \\ 0 & 3 & 0 \end{bmatrix}$ and

a. $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ b. $\mathbf{x} = \begin{bmatrix} -3 \\ -1 \\ 4 \end{bmatrix}$ c. $\mathbf{x} = \begin{bmatrix} 1/\sqrt{5} \\ 1/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$

3. Find the matrix of the quadratic form. Assume $\mathbf{x}$ is in $\mathbb{R}^2$.

a. $4x_1^2 - 6x_1x_2 + 5x_2^2$ b. $5x_1^2 + 4x_1x_2$

4. Find the matrix of the quadratic form. Assume $\mathbf{x}$ is in $\mathbb{R}^2$.

a. $7x_1^2 + 18x_1x_2 - 7x_2^2$ b. $8x_1x_2$

5. Find the matrix of the quadratic form. Assume $\mathbf{x}$ is in $\mathbb{R}^3$.
   a. $5x_1^2 + 3x_2^2 - 7x_3^2 - 4x_1x_2 + 6x_1x_3 - 2x_2x_3$
   b. $8x_1x_2 + 10x_1x_3 - 6x_2x_3$

6. Find the matrix of the quadratic form. Assume $\mathbf{x}$ is in $\mathbb{R}^3$.
   a. $5x_1^2 - 3x_2^2 + 7x_3^2 + 8x_1x_2 - 4x_1x_3$
   b. $6x_3^2 - 4x_1x_2 - 2x_2x_3$

7. Make the change of variable, $\mathbf{x} = P\mathbf{y}$, that transforms the quadratic form $x_1^2 + 12x_1x_2 + x_2^2$ into a quadratic form with no cross-product terms. Give P and the new quadratic form.

8. Let A be the matrix of the quadratic form

$$7x_1^2 + 5x_2^2 + 9x_3^2 - 8x_1x_2 + 8x_1x_3$$

It can be shown that the eigenvalues of A are 1, 7, and 13. Find an orthogonal matrix P such that the change of variable $\mathbf{x} = P\mathbf{y}$ transforms $\mathbf{x}^TA\mathbf{x}$ into a quadratic form with no cross-product term. Give P and the new quadratic form.

Classify the quadratic forms in Exercises 9–18. Then make a change of variable, $\mathbf{x} = P\mathbf{y}$ that transform the quadratic form into one with no cross-product terms. Write the new quadratic form. Construct P using the methods of Section 7.1.

9. $6x_1^2 - 4x_1x_2 + 3x_2^2$

10. $3x_1^2 + 8x_1x_2 - 3x_2^2$

11. $4x_1^2 - 8x_1x_2 - 2x_2^2$

12. $-2x_1^2 - 4x_1x_2 - 2x_2^2$

13. $x_1^2 - 4x_1x_2 + 4x_2^2$

14. $5x_1^2 + 12x_1x_2$

T 15. $-3x_1^2 - 7x_2^2 - 10x_3^2 - 10x_4^2 + 4x_1x_2 + 4x_1x_3 + 4x_1x_4 + 6x_3x_4$

T 16. $4x_1^2 + 4x_2^2 + 4x_3^2 + 4x_4^2 + 8x_1x_2 + 8x_3x_4 - 6x_1x_4 + 6x_2x_3$

T 17. $11x_1^2 + 11x_2^2 + 11x_3^2 + 11x_4^2 + 16x_1x_2 - 12x_1x_4 + 12x_2x_3 + 16x_3x_4$

T 18. $2x_1^2 + 2x_2^2 - 6x_1x_2 - 6x_1x_3 - 6x_1x_4 - 6x_2x_3 - 6x_2x_4 - 2x_3x_4$

19. What is the largest possible value of the quadratic form $4x_1^2 + 9x_2^2$ if $\mathbf{x} = (x_1, x_2)$ and $\mathbf{x}^T\mathbf{x} = 1$, that is, if $x_1^2 + x_2^2 = 1$? (Try some examples of $\mathbf{x}$)

20. What is the largest possible value of the quadratic form $7x_1^2 - 5x_2^2$ if $\mathbf{x}^T\mathbf{x} = 1$?

In Exercises 21–30, matrices are $n \times n$ and vectors are in $\mathbb{R}^n$. Mark each statement True or False **(T/F)**. Justify each answer.

21. **(T/F)** The matrix of a quadratic form is a symmetric matrix.

22. **(T/F)** The expression $\|\mathbf{x}\|^2$ is not a quadratic form.

23. **(T/F)** A quadratic form has no cross-product terms if and only if the matrix of the quadratic form is a diagonal matrix.

24. **(T/F)** If A is symmetric and P is an orthogonal matrix, then the change of variable $\mathbf{x} = P\mathbf{y}$ transforms $\mathbf{x}^TA\mathbf{x}$ into a quadratic form with no cross-product term.

25. **(T/F)** The principal axes of a quadratic form $\mathbf{x}^TA\mathbf{x}$ are eigenvectors of A.

26. **(T/F)** If the eigenvalues of a symmetric matrix A are all positive, then the quadratic form $\mathbf{x}^TA\mathbf{x}$ is positive definite.

27. **(T/F)** A positive definite quadratic form Q satisfies $Q(\mathbf{x}) > 0$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

28. **(T/F)** An indefinite quadratic form is neither positive semidefinite nor negative semidefinite.

29. **(T/F)** A Cholesky factorization of a symmetric matrix A has the form $A = R^TR$, for an upper triangular matrix R with positive diagonal entries.

30. **(T/F)** If A is symmetric and the quadratic form $\mathbf{x}^TA\mathbf{x}$ has only negative values for $\mathbf{x} \neq \mathbf{0}$, then the eigenvalues of A are all positive.

Exercises 31 and 32 show how to classify a quadratic form $Q(\mathbf{x}) = \mathbf{x}^TA\mathbf{x}$, when $A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$ and $\det A \neq 0$, without finding the eigenvalues of A.

31. If λ_1 and λ_2 are the eigenvalues of A, then the characteristic polynomial of A can be written in two ways: $\det(A - \lambda I)$ and $(\lambda - \lambda_1)(\lambda - \lambda_2)$. Use this fact to show that $\lambda_1 + \lambda_2 = a + d$ (the diagonal entries of A) and $\lambda_1\lambda_2 = \det A$.

32. Verify the following statements:
   a. Q is positive definite if $\det A > 0$ and $a > 0$.
   b. Q is negative definite if $\det A > 0$ and $a < 0$.
   c. Q is indefinite if $\det A < 0$.

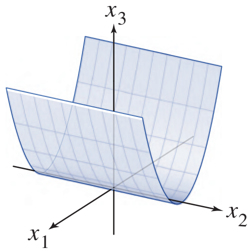
33. Show that if B is $m \times n$, then B^TB is positive semidefinite; and if B is $n \times n$ and invertible, then B^TB is positive definite.

34. Show that if an $n \times n$ matrix A is positive definite, then there exists a positive definite matrix B such that $A = B^TB$. [*Hint:* Write $A = PDP^T$, with $P^T = P^{-1}$. Produce a diagonal matrix C such that $D = C^TC$, and let $B = PCP^T$. Show that B works.]

35. Let A and B be symmetric $n \times n$ matrices whose eigenvalues are all positive. Show that the eigenvalues of $A + B$ are all positive. [*Hint:* Consider quadratic forms.]

36. Let A be an $n \times n$ invertible symmetric matrix. Show that if the quadratic form $\mathbf{x}^TA\mathbf{x}$ is positive definite, then so is the quadratic form $\mathbf{x}^TA^{-1}\mathbf{x}$. [*Hint:* Consider eigenvalues.]

STUDY GUIDE offers additional resources on diagonalization and quadratic forms.



Positive semidefinite

Solution to Practice Problem

Make an orthogonal change of variable $\mathbf{x} = P\mathbf{y}$, and write

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

as in equation (4). If an eigenvalue—say, λ_i—were negative, then $\mathbf{x}^T A\mathbf{x}$ would be negative for the $\mathbf{x}$ corresponding to $\mathbf{y} = \mathbf{e}_i$ (the ith column of I_n). So the eigenvalues of a positive semidefinite quadratic form must all be nonnegative. Conversely, if the eigenvalues are nonnegative, the expansion above shows that $\mathbf{x}^T A\mathbf{x}$ must be positive semidefinite.

7.3 Constrained Optimization

Engineers, economists, scientists, and mathematicians often need to find the maximum or minimum value of a quadratic form $Q(\mathbf{x})$ for $\mathbf{x}$ in some specified set. Typically, the problem can be arranged so that $\mathbf{x}$ varies over the set of unit vectors. This *constrained optimization problem* has an interesting and elegant solution. Example 6 and the discussion in Section 7.5 will illustrate how such problems arise in practice.

The requirement that a vector $\mathbf{x}$ in $\mathbb{R}^n$ be a unit vector can be stated in several equivalent ways:

$$\|\mathbf{x}\| = 1, \qquad \|\mathbf{x}\|^2 = 1, \qquad \mathbf{x}^T\mathbf{x} = 1$$

and

$$x_1^2 + x_2^2 + \cdots + x_n^2 = 1 \tag{1}$$

The expanded version (1) of $\mathbf{x}^T\mathbf{x} = 1$ is commonly used in applications.

When a quadratic form Q has no cross-product terms, it is easy to find the maximum and minimum of $Q(\mathbf{x})$ for $\mathbf{x}^T\mathbf{x} = 1$.

EXAMPLE 1 Find the maximum and minimum values of $Q(\mathbf{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$ subject to the constraint $\mathbf{x}^T\mathbf{x} = 1$.

SOLUTION Since x_2^2 and x_3^2 are nonnegative, note that

$$4x_2^2 \le 9x_2^2 \qquad \text{and} \qquad 3x_3^2 \le 9x_3^2$$

and hence

$$\begin{aligned} Q(\mathbf{x}) &= 9x_1^2 + 4x_2^2 + 3x_3^2 \\ &\le 9x_1^2 + 9x_2^2 + 9x_3^2 \\ &= 9(x_1^2 + x_2^2 + x_3^2) \\ &= 9 \end{aligned}$$

whenever $x_1^2 + x_2^2 + x_3^2 = 1$. So the maximum value of $Q(\mathbf{x})$ cannot exceed 9 when $\mathbf{x}$ is a unit vector. Furthermore, $Q(\mathbf{x}) = 9$ when $\mathbf{x} = (1, 0, 0)$. Thus 9 is the maximum value of $Q(\mathbf{x})$ for $\mathbf{x}^T\mathbf{x} = 1$.

To find the minimum value of $Q(\mathbf{x})$, observe that

$$9x_1^2 \ge 3x_1^2, \qquad 4x_2^2 \ge 3x_2^2$$

and hence

$$Q(\mathbf{x}) \ge 3x_1^2 + 3x_2^2 + 3x_3^2 = 3(x_1^2 + x_2^2 + x_3^2) = 3$$

whenever $x_1^2 + x_2^2 + x_3^2 = 1$. Also, $Q(\mathbf{x}) = 3$ when $x_1 = 0$, $x_2 = 0$, and $x_3 = 1$. So 3 is the minimum value of $Q(\mathbf{x})$ when $\mathbf{x}^T\mathbf{x} = 1$. ■