CHAPTER

1 MATRICES, VECTORS, AND SYSTEMS OF LINEAR EQUATIONS

The most common use of linear algebra is to solve systems of linear equations, which arise in applications to such diverse disciplines as physics, biology, economics, engineering, and sociology. In this chapter, we describe the most efficient algorithm for solving systems of linear equations, *Gaussian elimination*. This algorithm, or some variation of it, is used by most mathematics software (such as MATLAB).

We can write systems of linear equations compactly, using arrays called *matrices* and *vectors*. More importantly, the arithmetic properties of these arrays enable us to compute solutions of such systems or to determine if no solutions exist. This chapter begins by developing the basic properties of matrices and vectors. In Sections 1.3 and 1.4, we begin our study of systems of linear equations. In Sections 1.6 and 1.7, we introduce two other important concepts of vectors, namely, generating sets and linear independence, which provide information about the existence and uniqueness of solutions of a system of linear equations.

1.1 MATRICES AND VECTORS

Many types of numerical data are best displayed in two-dimensional arrays, such as tables.

For example, suppose that a company owns two bookstores, each of which sells newspapers, magazines, and books. Assume that the sales (in hundreds of dollars) of the two bookstores for the months of July and August are represented by the following tables:

	July	
Store	1	2
Newspapers	6	8
Magazines	15	20
Books	45	64

and

	August	
Store	1	2
Newspapers	7	9
Magazines	18	31
Books	52	68

The first column of the July table shows that store 1 sold \$1500 worth of magazines and \$4500 worth of books during July. We can represent the information on July sales more simply as

$$\begin{bmatrix} 6 & 8 \\ 15 & 20 \\ 45 & 64 \end{bmatrix}.$$

Such a rectangular array of real numbers is called a *matrix*.[1] It is customary to refer to real numbers as **scalars** (originally from the word *scale*) when working with a matrix. We denote the set of real numbers by $\mathcal{R}$.

tal

Definitions A **matrix** (*plural, matrices*) is a rectangular array of scalars. If the matrix has m rows and n columns, we say that the **size** of the matrix is $\boldsymbol{m}$ **by** $\boldsymbol{n}$, written $m \times n$. The matrix is **square** if $m = n$. The scalar in the ith row and jth column is called the **(i, j)-entry** of the matrix.

If A is a matrix, we denote its (i,j)-entry by a_{ij}. We say that two matrices A and B are **equal** if they have the same size and have equal corresponding entries; that is, $a_{ij} = b_{ij}$ for all i and j. Symbolically, we write $A = B$.

In our bookstore example, the July and August sales are contained in the matrices

$$B = \begin{bmatrix} 6 & 8 \\ 15 & 20 \\ 45 & 64 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 7 & 9 \\ 18 & 31 \\ 52 & 68 \end{bmatrix}.$$

Note that $b_{12} = 8$ and $c_{12} = 9$, so $B \neq C$. Both B and C are 3×2 matrices. Because of the context in which these matrices arise, they are called *inventory matrices*.

Other examples of matrices are

$$\begin{bmatrix} \frac{2}{3} & -4 & 0 \\ \pi & 1 & 6 \end{bmatrix}, \quad \begin{bmatrix} 3 \\ 8 \\ 4 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} -2 & 0 & 1 & 1 \end{bmatrix}.$$

The first matrix has size 2×3, the second has size 3×1, and the third has size 1×4.

Practice Problem 1 ▶ Let $A = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$.

(a) What is the $(1,2)$-entry of A?

(b) What is a_{22}? ◀

Sometimes we are interested in only a part of the information contained in a matrix. For example, suppose that we are interested in only magazine and book sales in July. Then the relevant information is contained in the last two rows of B; that is, in the matrix E defined by

$$E = \begin{bmatrix} 15 & 20 \\ 45 & 64 \end{bmatrix}.$$

E is called a *submatrix* of B. In general, a **submatrix** of a matrix M is obtained by deleting from M entire rows, entire columns, or both. It is permissible, when forming a submatrix of M, to delete none of the rows or none of the columns of M. As another example, if we delete the first row and the second column of B, we obtain the submatrix

$$\begin{bmatrix} 15 \\ 45 \end{bmatrix}.$$

[1] James Joseph Sylvester (1814–1897) coined the term *matrix* in the 1850s.

MATRIX SUMS AND SCALAR MULTIPLICATION

entry = indgang

Matrices are more than convenient devices for storing information. Their usefulness lies in their *arithmetic*. As an example, suppose that we want to know the total numbers of newspapers, magazines, and books sold by both stores during July and August. It is natural to form one matrix whose entries are the sum of the corresponding entries of the matrices B and C, namely,

$$\begin{array}{r c} \text{Store} & \begin{array}{cc} 1 & \;\;2 \end{array} \\ \begin{array}{r} \text{Newspapers} \\ \text{Magazines} \\ \text{Books} \end{array} & \begin{bmatrix} 13 & 17 \\ 33 & 51 \\ 97 & 132 \end{bmatrix} \end{array}.$$

If A and B are $m \times n$ matrices, the **sum** of A and B, denoted by $A + B$, is the $m \times n$ matrix obtained by adding the corresponding entries of A and B; that is, $A + B$ is the $m \times n$ matrix whose (i,j)-entry is $a_{ij} + b_{ij}$. Notice that the matrices A and B must have the same size for their sum to be defined.

Suppose that in our bookstore example, July sales were to double in all categories. Then the new matrix of July sales would be

$$\begin{bmatrix} 12 & 16 \\ 30 & 40 \\ 90 & 128 \end{bmatrix}.$$

We denote this matrix by $2B$.

Let A be an $m \times n$ matrix and c be a scalar. The **scalar multiple** cA is the $m \times n$ matrix whose entries are c times the corresponding entries of A; that is, cA is the $m \times n$ matrix whose (i,j)-entry is ca_{ij}. Note that $1A = A$. We denote the matrix $(-1)A$ by $-A$ and the matrix $0A$ by O. We call the $m \times n$ matrix O in which each entry is 0 the $m \times n$ **zero matrix**.

Example 1 Compute the matrices $A + B$, $3A$, $-A$, and $3A + 4B$, where

$$A = \begin{bmatrix} 3 & 4 & 2 \\ 2 & -3 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -4 & 1 & 0 \\ 5 & -6 & 1 \end{bmatrix}.$$

Solution We have

$$A + B = \begin{bmatrix} -1 & 5 & 2 \\ 7 & -9 & 1 \end{bmatrix}, \quad 3A = \begin{bmatrix} 9 & 12 & 6 \\ 6 & -9 & 0 \end{bmatrix}, \quad -A = \begin{bmatrix} -3 & -4 & -2 \\ -2 & 3 & 0 \end{bmatrix},$$

and

$$3A + 4B = \begin{bmatrix} 9 & 12 & 6 \\ 6 & -9 & 0 \end{bmatrix} + \begin{bmatrix} -16 & 4 & 0 \\ 20 & -24 & 4 \end{bmatrix} = \begin{bmatrix} -7 & 16 & 6 \\ 26 & -33 & 4 \end{bmatrix}.$$

Just as we have defined addition of matrices, we can also define **subtraction**. For any matrices A and B of the same size, we define $A - B$ to be the matrix obtained by subtracting each entry of B from the corresponding entry of A. Thus the (i,j)-entry of $A - B$ is $a_{ij} - b_{ij}$. Notice that $A - A = O$ for all matrices A.

If, as in Example 1, we have

$$A = \begin{bmatrix} 3 & 4 & 2 \\ 2 & -3 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} -4 & 1 & 0 \\ 5 & -6 & 1 \end{bmatrix}, \quad \text{and} \quad O = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

then

$$-B = \begin{bmatrix} 4 & -1 & 0 \\ -5 & 6 & -1 \end{bmatrix}, \quad A - B = \begin{bmatrix} 7 & 3 & 2 \\ -3 & 3 & -1 \end{bmatrix}, \quad \text{and} \quad A - O = \begin{bmatrix} 3 & 4 & 2 \\ 2 & -3 & 0 \end{bmatrix}.$$

Practice Problem 2 ▶ Let $A = \begin{bmatrix} 2 & -1 & 1 \\ 3 & 0 & -2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 3 & 0 \\ 2 & -1 & 4 \end{bmatrix}$. Compute the following matrices:

(a) $A - B$

(b) $2A$

(c) $A + 3B$ ◀

We have now defined the operations of matrix addition and scalar multiplication. The power of linear algebra lies in the natural relations between these operations, which are described in our first theorem.

THEOREM 1.1

(Properties of Matrix Addition and Scalar Multiplication) Let A, B, and C be $m \times n$ matrices, and let s and t be any scalars. Then

(a) $A + B = B + A$. (commutative law of matrix addition)

(b) $(A + B) + C = A + (B + C)$. (associative law of matrix addition)

(c) $A + O = A$.

(d) $A + (-A) = O$.

(e) $(st)A = s(tA)$.

(f) $s(A + B) = sA + sB$.

(g) $(s + t)A = sA + tA$.

PROOF We prove parts (b) and (f). The rest are left as exercises.

(b) The matrices on each side of the equation are $m \times n$ matrices. We must show that each entry of $(A + B) + C$ is the same as the corresponding entry of $A + (B + C)$. Consider the (i,j)-entries. Because of the definition of matrix addition, the (i,j)-entry of $(A + B) + C$ is the sum of the (i,j)-entry of $A + B$, which is $a_{ij} + b_{ij}$, and the (i,j)-entry of C, which is c_{ij}. Therefore this sum equals $(a_{ij} + b_{ij}) + c_{ij}$. Similarly, the (i,j)-entry of $A + (B + C)$ is $a_{ij} + (b_{ij} + c_{ij})$. Because the associative law holds for addition of scalars, $(a_{ij} + b_{ij}) + c_{ij} = a_{ij} + (b_{ij} + c_{ij})$. Therefore the (i,j)-entry of $(A + B) + C$ equals the (i,j)-entry of $A + (B + C)$, proving (b).

(f) The matrices on each side of the equation are $m \times n$ matrices. As in the proof of (b), we consider the (i,j)-entries of each matrix. The (i,j)-entry of $s(A + B)$ is defined to be the product of s and the (i,j)-entry of $A + B$, which is $a_{ij} + b_{ij}$. This product equals $s(a_{ij} + b_{ij})$. The (i,j)-entry of $sA + sB$ is the sum of the (i,j)-entry of sA, which is sa_{ij}, and the (i,j)-entry of sB, which is sb_{ij}. This sum is $sa_{ij} + sb_{ij}$. Since $s(a_{ij} + b_{ij}) = sa_{ij} + sb_{ij}$, (f) is proved. ■

Because of the associative law of matrix addition, sums of three or more matrices can be written unambiguously without parentheses. Thus we may write $A + B + C$ instead of either $(A + B) + C$ or $A + (B + C)$.

MATRIX TRANSPOSES

In the bookstore example, we could have recorded the information about July sales in the following form:

Store	Newspapers	Magazines	Books
1	6	15	45
2	8	20	64

This representation produces the matrix

$$\begin{bmatrix} 6 & 15 & 45 \\ 8 & 20 & 64 \end{bmatrix}.$$

Compare this with

$$B = \begin{bmatrix} 6 & 8 \\ 15 & 20 \\ 45 & 64 \end{bmatrix}.$$

The rows of the first matrix are the columns of B, and the columns of the first matrix are the rows of B. This new matrix is called the *transpose* of B. In general, the **transpose** of an $m \times n$ matrix A is the $n \times m$ matrix denoted by A^T whose (i,j)-entry is the (j,i)-entry of A.

The matrix C in our bookstore example and its transpose are

$$C = \begin{bmatrix} 7 & 9 \\ 18 & 31 \\ 52 & 68 \end{bmatrix} \quad \text{and} \quad C^T = \begin{bmatrix} 7 & 18 & 52 \\ 9 & 31 & 68 \end{bmatrix}.$$

Practice Problem 3 ▶ Let $A = \begin{bmatrix} 2 & -1 & 1 \\ 3 & 0 & -2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 3 & 0 \\ 2 & -1 & 4 \end{bmatrix}$. Compute the following matrices:

(a) A^T
(b) $(3B)^T$
(c) $(A+B)^T$ ◀

The following theorem shows that the transpose preserves the operations of matrix addition and scalar multiplication:

THEOREM 1.2

(Properties of the Transpose) Let A and B be $m \times n$ matrices, and let s be any scalar. Then

(a) $(A+B)^T = A^T + B^T$.
(b) $(sA)^T = sA^T$.
(c) $(A^T)^T = A$.

PROOF We prove part (a). The rest are left as exercises.

(a) The matrices on each side of the equation are $n \times m$ matrices. So we show that the (i,j)-entry of $(A+B)^T$ equals the (i,j)-entry of $A^T + B^T$. By the definition of transpose, the (i,j)-entry of $(A+B)^T$ equals the (j,i)-entry of $A+B$, which is $a_{ji} + b_{ji}$. On the other hand, the (i,j)-entry of $A^T + B^T$ equals the sum of the (i,j)-entry of A^T and the (i,j)-entry of B^T, that is, $a_{ji} + b_{ji}$. Because the (i,j)-entries of $(A+B)^T$ and $A^T + B^T$ are equal, (a) is proved. ■

In the following exercise, use either a calculator with matrix capabilities or computer software such as MATLAB to solve the problem:

84. Consider the matrices

$$A = \begin{bmatrix} 1.3 & 2.1 & -3.3 & 6.0 \\ 5.2 & 2.3 & -1.1 & 3.4 \\ 3.2 & -2.6 & 1.1 & -4.0 \\ 0.8 & -1.3 & -12.1 & 5.7 \\ -1.4 & 3.2 & 0.7 & 4.4 \end{bmatrix}$$

and

$$B = \begin{bmatrix} 2.6 & -1.3 & 0.7 & -4.4 \\ 2.2 & -2.6 & 1.3 & -3.2 \\ 7.1 & 1.5 & -8.3 & 4.6 \\ -0.9 & -1.2 & 2.4 & 5.9 \\ 3.3 & -0.9 & 1.4 & 6.2 \end{bmatrix}.$$

(a) Compute $A + 2B$.
(b) Compute $A - B$.
(c) Compute $A^T + B^T$.

SOLUTIONS TO THE PRACTICE PROBLEMS

1. (a) The $(1, 2)$-entry of A is 2.
   (b) The $(2, 2)$-entry of A is 3.

2. (a) $A - B = \begin{bmatrix} 2 & -1 & 1 \\ 3 & 0 & -2 \end{bmatrix} - \begin{bmatrix} 1 & 3 & 0 \\ 2 & -1 & 4 \end{bmatrix}$

$$= \begin{bmatrix} 1 & 4 & -1 \\ 1 & 1 & -6 \end{bmatrix}$$

   (b) $2A = 2\begin{bmatrix} 2 & -1 & 1 \\ 3 & 0 & -2 \end{bmatrix} = \begin{bmatrix} 4 & -2 & 2 \\ 6 & 0 & -4 \end{bmatrix}$

   (c) $A + 3B = \begin{bmatrix} 2 & -1 & 1 \\ 3 & 0 & -2 \end{bmatrix} + 3\begin{bmatrix} 1 & 3 & 0 \\ 2 & -1 & 4 \end{bmatrix}$

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

$$= \begin{bmatrix} 5 & 8 & 1 \\ 9 & -3 & 10 \end{bmatrix}$$

3. (a) $A^T = \begin{bmatrix} 2 & 3 \\ -1 & 0 \\ 1 & -2 \end{bmatrix}$

   (b) $(3B)^T = \begin{bmatrix} 3 & 9 & 0 \\ 6 & -3 & 12 \end{bmatrix}^T = \begin{bmatrix} 3 & 6 \\ 9 & -3 \\ 0 & 12 \end{bmatrix}$

   (c) $(A + B)^T = \begin{bmatrix} 3 & 2 & 1 \\ 5 & -1 & 2 \end{bmatrix}^T = \begin{bmatrix} 3 & 5 \\ 2 & -1 \\ 1 & 2 \end{bmatrix}$

1.2 LINEAR COMBINATIONS, MATRIX-VECTOR PRODUCTS, AND SPECIAL MATRICES

In this section, we explore some applications involving matrix operations and introduce the product of a matrix and a vector.

Suppose that 20 students are enrolled in a linear algebra course, in which two tests, a quiz, and a final exam are given. Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{20} \end{bmatrix}$, where u_i denotes the score of the ith student on the first test. Likewise, define vectors $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{z}$ similarly for the second test, quiz, and final exam, respectively. Assume that the instructor computes a student's course average by counting each test score twice as much as a quiz score, and the final exam score three times as much as a test score. Thus the *weights* for the tests, quiz, and final exam score are, respectively, 2/11, 2/11, 1/11, 6/11 (the weights must sum to one). Now consider the vector

$$\mathbf{y} = \frac{2}{11}\mathbf{u} + \frac{2}{11}\mathbf{v} + \frac{1}{11}\mathbf{w} + \frac{6}{11}\mathbf{z}.$$

The first component y_1 represents the first student's course average, the second component y_2 represents the second student's course average, and so on. Notice that $\mathbf{y}$ is a sum of scalar multiples of $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{z}$. This form of vector sum is so important that it merits its own definition.

Definitions A **linear combination** of vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ is a vector of the form

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k,$$

where $c_1, c_2, \ldots, c_k$ are scalars. These scalars are called the **coefficients** of the linear combination.

Note that a linear combination of one vector is simply a scalar multiple of that vector.

In the previous example, the vector $\mathbf{y}$ of the students' course averages is a linear combination of the vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{z}$. The coefficients are the weights. Indeed, any weighted average produces a linear combination of the scores.

Notice that

$$\begin{bmatrix} 2 \\ 8 \end{bmatrix} = (-3)\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 4\begin{bmatrix} 1 \\ 3 \end{bmatrix} + 1\begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Thus $\begin{bmatrix} 2 \\ 8 \end{bmatrix}$ is a linear combination of $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$, and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$, with coefficients -3, 4, and 1. We can also write

$$\begin{bmatrix} 2 \\ 8 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 3 \end{bmatrix} - 1\begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

This equation also expresses $\begin{bmatrix} 2 \\ 8 \end{bmatrix}$ as a linear combination of $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$, and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$, but now the coefficients are 1, 2, and -1. So the set of coefficients that express one vector as a linear combination of the others need not be unique.

Example 1

(a) Determine whether $\begin{bmatrix} 4 \\ -1 \end{bmatrix}$ is a linear combination of $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$.

(b) Determine whether $\begin{bmatrix} -4 \\ -2 \end{bmatrix}$ is a linear combination of $\begin{bmatrix} 6 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

(c) Determine whether $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ is a linear combination of $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 6 \\ 4 \end{bmatrix}$.

Solution (a) We seek scalars x_1 and x_2 such that

$$\begin{bmatrix} 4 \\ -1 \end{bmatrix} = x_1\begin{bmatrix} 2 \\ 3 \end{bmatrix} + x_2\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 3x_1 \end{bmatrix} + \begin{bmatrix} 3x_2 \\ 1x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 + 3x_2 \\ 3x_1 + x_2 \end{bmatrix}.$$

That is, we seek a solution of the system of equations

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

Because these equations represent nonparallel lines in the plane, there is exactly one solution, namely, $x_1 = -1$ and $x_2 = 2$. Therefore $\begin{bmatrix} 4 \\ -1 \end{bmatrix}$ is a (unique) linear combination of the vectors $\begin{bmatrix}2\\3\end{bmatrix}$ and $\begin{bmatrix}3\\1\end{bmatrix}$, namely,

$$\begin{bmatrix}4\\-1\end{bmatrix} = (-1)\begin{bmatrix}2\\3\end{bmatrix} + 2\begin{bmatrix}3\\1\end{bmatrix}.$$

(See Figure 1.8.)

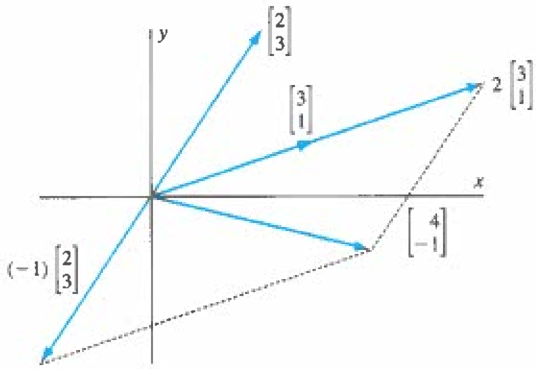


Figure 1.8 The vector $\begin{bmatrix}4\\-1\end{bmatrix}$ is a linear combination of $\begin{bmatrix}2\\3\end{bmatrix}$ and $\begin{bmatrix}3\\1\end{bmatrix}$.

(b) To determine whether $\begin{bmatrix}-4\\-2\end{bmatrix}$ is a linear combination of $\begin{bmatrix}6\\3\end{bmatrix}$ and $\begin{bmatrix}2\\1\end{bmatrix}$, we perform a similar computation and produce the set of equations

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

Since the first equation is twice the second, we need only solve $3x_1 + x_2 = -2$. This equation represents a line in the plane, and the coordinates of any point on the line give a solution. For example, we can let $x_1 = -2$ and $x_2 = 4$. In this case, we have

$$\begin{bmatrix}-4\\-2\end{bmatrix} = (-2)\begin{bmatrix}6\\3\end{bmatrix} + 4\begin{bmatrix}2\\1\end{bmatrix}.$$

There are infinitely many solutions. (See Figure 1.9.)

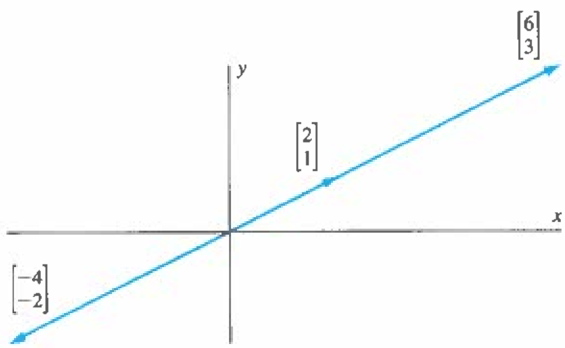


Figure 1.9 The vector $\begin{bmatrix}-4\\-2\end{bmatrix}$ is a linear combination of $\begin{bmatrix}6\\3\end{bmatrix}$ and $\begin{bmatrix}2\\1\end{bmatrix}$.

(c) To determine if $\begin{bmatrix}3\\4\end{bmatrix}$ is a linear combination of $\begin{bmatrix}3\\2\end{bmatrix}$ and $\begin{bmatrix}6\\4\end{bmatrix}$, we must solve the system of equations

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

If we add $-\frac{2}{3}$ times the first equation to the second, we obtain $0 = 2$, an equation with no solutions. Indeed, the two original equations represent parallel lines in the plane, so the original system has no solutions. We conclude that $\begin{bmatrix}3\\4\end{bmatrix}$ is not a linear combination of $\begin{bmatrix}3\\2\end{bmatrix}$ and $\begin{bmatrix}6\\4\end{bmatrix}$. (See Figure 1.10.)

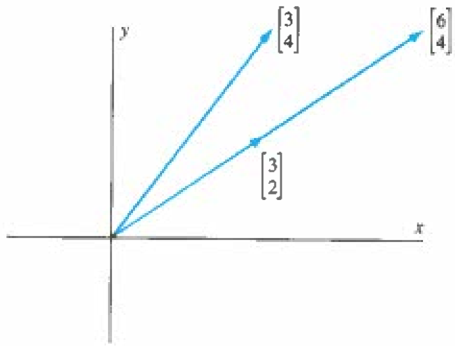


Figure 1.10 The vector $\begin{bmatrix}3\\4\end{bmatrix}$ is *not* a linear combination of $\begin{bmatrix}3\\2\end{bmatrix}$ and $\begin{bmatrix}6\\4\end{bmatrix}$.

Example 2 Given vectors $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$, show that the sum of any two linear combinations of these vectors is also a linear combination of these vectors.

Solution Suppose that $\mathbf{w}$ and $\mathbf{z}$ are linear combinations of $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$. Then we may write

$$\mathbf{w} = a\mathbf{u}_1 + b\mathbf{u}_2 + c\mathbf{u}_3 \quad \text{and} \quad \mathbf{z} = a'\mathbf{u}_1 + b'\mathbf{u}_2 + c'\mathbf{u}_3,$$

where a, b, c, a', b', c' are scalars. So

$$\mathbf{w} + \mathbf{z} = (a + a')\mathbf{u}_1 + (b + b')\mathbf{u}_2 + (c + c')\mathbf{u}_3,$$

which is also a linear combination of $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$.

STANDARD VECTORS

We can write any vector $\begin{bmatrix}a\\b\end{bmatrix}$ in $\mathcal{R}^2$ as a linear combination of the two vectors $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$ as follows:

$$\begin{bmatrix}a\\b\end{bmatrix} = a\begin{bmatrix}1\\0\end{bmatrix} + b\begin{bmatrix}0\\1\end{bmatrix}$$

The vectors $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$ are called the *standard vectors* of $\mathcal{R}^2$. Similarly, we can write any vector $\begin{bmatrix}a\\b\\c\end{bmatrix}$ in $\mathcal{R}^3$ as a linear combination of the vectors $\begin{bmatrix}1\\0\\0\end{bmatrix}$, $\begin{bmatrix}0\\1\\0\end{bmatrix}$, and $\begin{bmatrix}0\\0\\1\end{bmatrix}$ as follows:

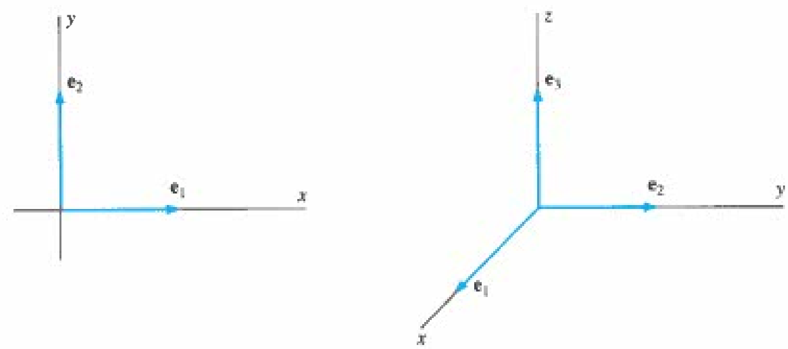
$$\begin{bmatrix}a\\b\\c\end{bmatrix} = a\begin{bmatrix}1\\0\\0\end{bmatrix} + b\begin{bmatrix}0\\1\\0\end{bmatrix} + c\begin{bmatrix}0\\0\\1\end{bmatrix}$$

The vectors $\begin{bmatrix}1\\0\\0\end{bmatrix}$, $\begin{bmatrix}0\\1\\0\end{bmatrix}$, and $\begin{bmatrix}0\\0\\1\end{bmatrix}$ are called the *standard vectors* of $\mathcal{R}^3$.

In general, we define the **standard vectors** of $\mathcal{R}^n$ by

$$\mathbf{e}_1 = \begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix}0\\1\\\vdots\\0\end{bmatrix}, \quad \ldots, \quad \mathbf{e}_n = \begin{bmatrix}0\\0\\\vdots\\1\end{bmatrix}.$$

(See Figure 1.11.)



The standard vectors of $\mathcal{R}^2$ The standard vectors of $\mathcal{R}^3$

Figure 1.11

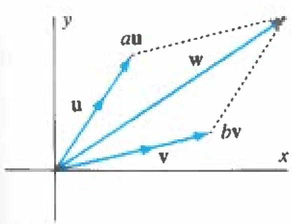


Figure 1.12 The vector $\mathbf{w}$ is a linear combination of the nonparallel vectors $\mathbf{u}$ and $\mathbf{v}$.

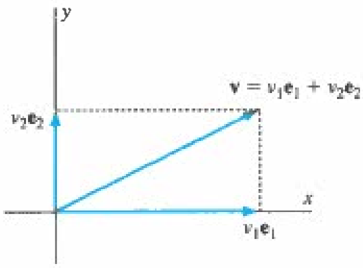
From the preceding equations, it is easy to see that every vector in $\mathcal{R}^n$ is a linear combination of the standard vectors of $\mathcal{R}^n$. In fact, for any vector $\mathbf{v}$ in $\mathcal{R}^n$,

$$\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \cdots + v_n\mathbf{e}_n.$$

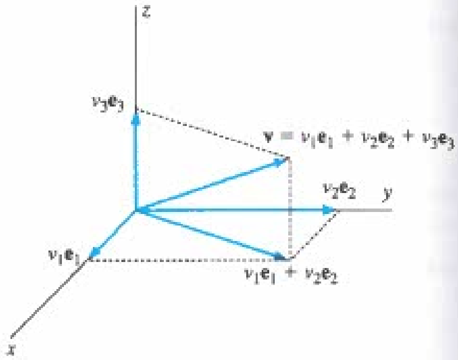
(See Figure 1.13.)

Now let $\mathbf{u}$ and $\mathbf{v}$ be nonparallel vectors, and let $\mathbf{w}$ be any vector in $\mathcal{R}^2$. Begin with the endpoint of $\mathbf{w}$ and create a parallelogram with sides $a\mathbf{u}$ and $b\mathbf{v}$, so that $\mathbf{w}$ is its diagonal. It follows that $\mathbf{w} = a\mathbf{u} + b\mathbf{v}$; that is, $\mathbf{w}$ is a linear combination of the vectors $\mathbf{u}$ and $\mathbf{v}$. (See Figure 1.12.) More generally, the following statement is true:

> If $\mathbf{u}$ and $\mathbf{v}$ are any nonparallel vectors in $\mathcal{R}^2$, then every vector in $\mathcal{R}^2$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$.



The vector **v** is a linear combination of standard vectors in $\mathcal{R}^2$.



The vector **v** is a linear combination of standard vectors in $\mathcal{R}^3$.

Figure 1.13

Practice Problem 1 ▶ Let $\mathbf{w} = \begin{bmatrix} -1 \\ 10 \end{bmatrix}$ and $\mathcal{S} = \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \end{bmatrix} \right\}$.

(a) Without doing any calculations, explain why **w** can be written as a linear combination of the vectors in $\mathcal{S}$.

(b) Express **w** as a linear combination of the vectors in $\mathcal{S}$. ◀

Suppose that a garden supply store sells three mixtures of grass seed. The deluxe mixture is 80% bluegrass and 20% rye, the standard mixture is 60% bluegrass and 40% rye, and the economy mixture is 40% bluegrass and 60% rye. One way to record this information is with the following 2×3 matrix:

$$\begin{array}{cccc} & \text{deluxe} & \text{standard} & \text{economy} & \\ B = & \left[\begin{array}{c} .80 \\ .20 \end{array}\right. & \begin{array}{c} .60 \\ .40 \end{array} & \left.\begin{array}{c} .40 \\ .60 \end{array}\right] & \begin{array}{c} \text{bluegrass} \\ \text{rye} \end{array} \end{array}$$

A customer wants to purchase a blend of grass seed containing 5 lb of bluegrass and 3 lb of rye. There are two natural questions that arise:

1. Is it possible to combine the three mixtures of seed into a blend that has exactly the desired amounts of bluegrass and rye, with no surplus of either?
2. If so, how much of each mixture should the store clerk add to the blend?

Let x_1, x_2, and x_3 denote the number of pounds of deluxe, standard, and economy mixtures, respectively, to be used in the blend. Then we have

$$\begin{aligned} .80x_1 + .60x_2 + .40x_3 &= 5 \\ .20x_1 + .40x_2 + .60x_3 &= 3. \end{aligned}$$

This is a *system of two linear equations in three unknowns.* Finding a solution of this system is equivalent to answering our second question. The technique for solving general systems is explored in great detail in Sections 1.3 and 1.4.

Using matrix notation, we may rewrite these equations in the form

$$\begin{bmatrix} .80x_1 + .60x_2 + .40x_3 \\ .20x_1 + .40x_2 + .60x_3 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}.$$

Now we use matrix operations to rewrite this matrix equation, using the columns of B, as

$$x_1\begin{bmatrix}.80\\.20\end{bmatrix}+x_2\begin{bmatrix}.60\\.40\end{bmatrix}+x_3\begin{bmatrix}.40\\.60\end{bmatrix}=\begin{bmatrix}5\\3\end{bmatrix}.$$

Thus we can rephrase the first question as follows: Is $\begin{bmatrix}5\\3\end{bmatrix}$ a linear combination of the columns $\begin{bmatrix}.80\\.20\end{bmatrix}$, $\begin{bmatrix}.60\\.40\end{bmatrix}$, and $\begin{bmatrix}.40\\.60\end{bmatrix}$ of B? The result in the box on page 17 provides an affirmative answer. Because no two of the three vectors are parallel, $\begin{bmatrix}5\\3\end{bmatrix}$ is a linear combination of any pair of these vectors.

MATRIX–VECTOR PRODUCTS

A convenient way to represent systems of linear equations is by *matrix–vector products.* For the preceding example, we represent the variables by the vector $\mathbf{x}=\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$ and define the *matrix–vector product* $B\mathbf{x}$ to be the linear combination

$$B\mathbf{x}=\begin{bmatrix}.80&.60&.40\\.20&.40&.60\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=x_1\begin{bmatrix}.80\\.20\end{bmatrix}+x_2\begin{bmatrix}.60\\.40\end{bmatrix}+x_3\begin{bmatrix}.40\\.60\end{bmatrix}.$$

This definition provides another way to state the first question in the preceding example: Does the vector $\begin{bmatrix}5\\3\end{bmatrix}$ equal $B\mathbf{x}$ for some vector $\mathbf{x}$? Notice that for the matrix–vector product to make sense, the number of columns of B must equal the number of components in $\mathbf{x}$. The general definition of a matrix–vector product is given next.

Definition Let A be an $m\times n$ matrix and $\mathbf{v}$ be an $n\times 1$ vector. We define the **matrix–vector product** of A and $\mathbf{v}$, denoted by $A\mathbf{v}$, to be the linear combination of the columns of A whose coefficients are the corresponding components of $\mathbf{v}$. That is,

$$A\mathbf{v}=v_1\mathbf{a}_1+v_2\mathbf{a}_2+\cdots+v_n\mathbf{a}_n.$$

As we have noted, for $A\mathbf{v}$ to exist, the number of columns of A must equal the number of components of $\mathbf{v}$. For example, suppose that

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

Notice that A has two columns and $\mathbf{v}$ has two components. Then

$$A\mathbf{v}=\begin{bmatrix}1&2\\3&4\\5&6\end{bmatrix}\begin{bmatrix}7\\8\end{bmatrix}=7\begin{bmatrix}1\\3\\5\end{bmatrix}+8\begin{bmatrix}2\\4\\6\end{bmatrix}=\begin{bmatrix}7\\21\\35\end{bmatrix}+\begin{bmatrix}16\\32\\48\end{bmatrix}=\begin{bmatrix}23\\53\\83\end{bmatrix}.$$

Returning to the preceding garden supply store example, suppose that the store has 140 lb of seed in stock: 60 lb of the deluxe mixture, 50 lb of the standard mixture, and 30 lb of the economy mixture. We let $\mathbf{v} = \begin{bmatrix} 60 \\ 50 \\ 30 \end{bmatrix}$ represent this information. Now the matrix–vector product

$$B\mathbf{v} = \begin{bmatrix} .80 & .60 & .40 \\ .20 & .40 & .60 \end{bmatrix} \begin{bmatrix} 60 \\ 50 \\ 30 \end{bmatrix}$$

$$= 60 \begin{bmatrix} .80 \\ .20 \end{bmatrix} + 50 \begin{bmatrix} .60 \\ .40 \end{bmatrix} + 30 \begin{bmatrix} .40 \\ .60 \end{bmatrix}$$

$$= \begin{array}{cc} \text{seed (lb)} & \\ \begin{bmatrix} 90 \\ 50 \end{bmatrix} & \begin{array}{c} \text{bluegrass} \\ \text{rye} \end{array} \end{array}$$

gives the number of pounds of each type of seed contained in the 140 pounds of seed that the garden supply store has in stock. For example, there are 90 pounds of bluegrass because $90 = .80(60) + .60(50) + .40(30)$.

There is another approach to computing the matrix–vector product that relies more on the entries of A than on its columns. Consider the following example:

$$A\mathbf{v} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

$$= v_1 \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} + v_2 \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} + v_3 \begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}v_1 + a_{12}v_2 + a_{13}v_3 \\ a_{21}v_1 + a_{22}v_2 + a_{23}v_3 \end{bmatrix}$$

Notice that the first component of the vector $A\mathbf{v}$ is the sum of products of the corresponding entries of the first row of A and the components of $\mathbf{v}$. Likewise, the second component of $A\mathbf{v}$ is the sum of products of the corresponding entries of the second row of A and the components of $\mathbf{v}$. With this approach to computing a matrix–vector product, we can omit the intermediate step in the preceding illustration. For example, suppose

$$A = \begin{bmatrix} 2 & 3 & 1 \\ 1 & -2 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix}.$$

Then

$$A\mathbf{v} = \begin{bmatrix} 2 & 3 & 1 \\ 1 & -2 & 3 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} (2)(-1) + (3)(1) + (1)(3) \\ (1)(-1) + (-2)(1) + (3)(3) \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}.$$

In general, you can use this technique to compute $A\mathbf{v}$ when A is an $m \times n$ matrix and $\mathbf{v}$ is a vector in $\mathcal{R}^n$. In this case, the ith component of $A\mathbf{v}$ is

$$[a_{i1} \;\; a_{i2} \;\; \dots \;\; a_{in}]\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = a_{i1}v_1 + a_{i2}v_2 + \cdots + a_{in}v_n,$$

which is the matrix–vector product of the ith row of A and $\mathbf{v}$. The computation of all the components of the matrix–vector product $A\mathbf{v}$ is given by

$$A\mathbf{v} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} a_{11}v_1 + a_{12}v_2 + \cdots + a_{1n}v_n \\ a_{21}v_1 + a_{22}v_2 + \cdots + a_{2n}v_n \\ \vdots \\ a_{m1}v_1 + a_{m2}v_2 + \cdots + a_{mn}v_n \end{bmatrix}.$$

Practice Problem 2 ▶ Let $A = \begin{bmatrix} 2 & -1 & 1 \\ 3 & 0 & -2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}$. Compute the following vectors:

(a) $A\mathbf{v}$

(b) $(A\mathbf{v})^T$ ◀

Example 3 A sociologist is interested in studying the population changes within a metropolitan area as people move between the city and suburbs. From empirical evidence, she has discovered that in any given year, 15% of those living in the city will move to the suburbs and 3% of those living in the suburbs will move to the city. For simplicity, we assume that the metropolitan population remains stable. This information may be represented by the following matrix:

$$\begin{array}{ccc} & & \text{From} \\ & & \begin{array}{cc}\text{City} & \text{Suburbs}\end{array} \\ \text{To} & \begin{array}{c}\text{City} \\ \text{Suburbs}\end{array} & \begin{bmatrix} .85 & .03 \\ .15 & .97 \end{bmatrix} = A \end{array}$$

Notice that the entries of A are nonnegative and that the entries of each column sum to 1. Such a matrix is called a **stochastic matrix**. Suppose that there are now 500 thousand people living in the city and 700 thousand people living in the suburbs. The sociologist would like to know how many people will be living in each of the two areas next year. Figure 1.14 describes the changes of population from one year to the next. It follows that the number of people (in thousands) who will be living in the city next year is $(.85)(500) + (.03)(700) = 446$ thousand, and the number of people living in the suburbs is $(.15)(500) + (.97)(700) = 754$ thousand.

If we let $\mathbf{p}$ represent the vector of current populations of the city and suburbs, we have

$$\mathbf{p} = \begin{bmatrix} 500 \\ 700 \end{bmatrix}.$$

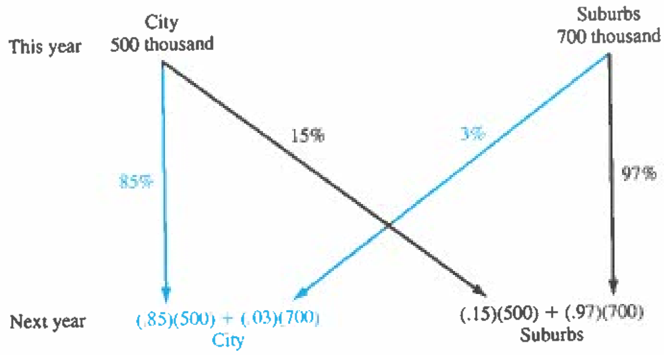


Figure 1.14 Movement between the city and suburbs

We can find the populations in the next year by computing the matrix–vector product:

$$A\mathbf{p} = \begin{bmatrix} .85 & .03 \\ .15 & .97 \end{bmatrix} \begin{bmatrix} 500 \\ 700 \end{bmatrix} = \begin{bmatrix} (.85)(500) + (.03)(700) \\ (.15)(500) + (.97)(700) \end{bmatrix} = \begin{bmatrix} 446 \\ 754 \end{bmatrix}$$

In other words, $A\mathbf{p}$ is the vector of populations in the next year. If we want to determine the populations in two years, we can repeat this procedure by multiplying A by the vector $A\mathbf{p}$. That is, in two years, the vector of populations is $A(A\mathbf{p})$.

IDENTITY MATRICES

Suppose we let $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\mathbf{v}$ be any vector in $\mathcal{R}^2$. Then

$$I_2\mathbf{v} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \mathbf{v}.$$

So multiplication by I_2 leaves every vector $\mathbf{v}$ in $\mathcal{R}^2$ unchanged. The same property holds in a more general context.

Definition For each positive integer n, the $n \times n$ **identity matrix** I_n is the $n \times n$ matrix whose respective columns are the standard vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ in $\mathcal{R}^n$.

For example,

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Because the columns of I_n are the standard vectors of $\mathcal{R}^n$, it follows easily that $I_n\mathbf{v} = \mathbf{v}$ for any $\mathbf{v}$ in $\mathcal{R}^n$.

ROTATION MATRICES

Consider a point $P_0 = (x_0, y_0)$ in $\mathcal{R}^2$ with polar coordinates (r, α), where $r \geq 0$ and α is the angle between the segment $\overline{OP_0}$ and the positive x-axis. (See Figure 1.15.) Then $x_0 = r\cos\alpha$ and $y_0 = r\sin\alpha$. Suppose that $\overline{OP_0}$ is rotated by an angle θ to the

3. (a) Clearly, every entry of A is either 0 or 1, so A is a $(0, 1)$-matrix.

(b) The $(2, 3)$-entry of A^2 is

$$a_{21}a_{13} + a_{22}a_{23} + a_{23}a_{33} + a_{24}a_{43}.$$

A typical term has the form $a_{2k}a_{k3}$, which equals 1 or 0. This term equals 1 if and only if $a_{2k} = 1$ and $a_{k3} = 1$. Consequently, this term equals 1 if and only if there is a nonstop flight between city 2 and city k, as well as a nonstop flight between city k and city 3. That is, $a_{2k}a_{k3} = 1$ means that there is a flight with one *layover* (the plane stops at city k) from city 2 to city 3. Therefore we may interpret the $(2, 3)$-entry of A^2 as the number of flights with one layover from city 2 to city 3.

(c) $A^2 = \begin{bmatrix} 2 & 0 & 0 & 1 \\ 0 & 2 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$.

(d) Because the $(2, 3)$-entry of A^2 is 1, there is one flight with one layover from city 2 to city 3.

(e) We compute A^3 to find the number of flights with two layovers from city 1 to city 2. We have

$$A^3 = \begin{bmatrix} 0 & 3 & 2 & 0 \\ 3 & 0 & 0 & 2 \\ 2 & 0 & 0 & 1 \\ 0 & 2 & 1 & 0 \end{bmatrix}.$$

Because the $(1, 2)$-entry of A^3 is 3, we see that there are three flights with two layovers from city 1 to city 2.

(f) From the entries of A, we see that there are nonstop flights between cities 1 and 2, cities 1 and 3, and cities 2 and 4. From A^2, we see that there are flights between cities 1 and 4, as well as between cities 2 and 3. Finally, from A^3, we discover that there is a flight between cities 3 and 4. We conclude that there are flights between all pairs of cities.

2.3 INVERTIBILITY AND ELEMENTARY MATRICES

In this section, we introduce the concept of *invertible matrix* and examine special invertible matrices that are intimately associated with elementary row operations, the *elementary matrices*.

For any real number $a \neq 0$, there is a unique real number b, called the *multiplicative inverse* of a, with the property that $ab = ba = 1$. For example, if $a = 2$, then $b = 1/2$. In the context of matrices, the identity matrix I_n is a multiplicative identity; so it is natural to ask for what matrices A does there exist a matrix B such that $AB = BA = I_n$. Notice that this last equation is possible only if both A and B are $n \times n$ matrices. This discussion motivates the following definitions:

Definitions An $n \times n$ matrix A is called **invertible** if there exists an $n \times n$ matrix B such that $AB = BA = I_n$. In this case, B is called an **inverse** of A.

If A is an invertible matrix, then its inverse is unique. For if both B and C are inverses of A, then $AB = BA = I_n$ and $AC = CA = I_n$. Hence

$$B = BI_n = B(AC) = (BA)C = I_nC = C.$$

When A is invertible, we denote the unique inverse of A by A^{-1}, so that $AA^{-1} = A^{-1}A = I_n$. Notice the similarity of this statement and $2 \cdot 2^{-1} = 2^{-1} \cdot 2 = 1$, where 2^{-1} is the multiplicative inverse of the real number 2.

Example 1 Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix}$. Then

$$AB = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}\begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2,$$

and

$$BA = \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2.$$

So A is invertible, and B is the inverse of A; that is, $A^{-1} = B$.

Practice Problem 1 ▶ If $A = \begin{bmatrix} -1 & 0 & 1 \\ 1 & 2 & -2 \\ 2 & -1 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 1 & 2 \\ 3 & 1 & 1 \\ 5 & 1 & 2 \end{bmatrix}$, is $B = A^{-1}$? ◀

Because the roles of the matrices A and B are the same in the preceding definition, it follows that if B is the inverse of A, then A is also the inverse of B. Thus, in Example 1, we also have

$$B^{-1} = A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}.$$

Just as the real number 0 has no multiplicative inverse, the $n \times n$ zero matrix O has no inverse because $OB = O \neq I_n$ for any $n \times n$ matrix B. But there are also other square matrices that are not invertible; for example, $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$. For if $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is any 2×2 matrix, then

$$AB = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a+c & b+d \\ 2a+2c & 2b+2d \end{bmatrix}.$$

Since the second row of the matrix on the right equals twice its first row, it cannot be the identity matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. So B cannot be the inverse of A, and hence A is not invertible.

In the next section, we learn which matrices are invertible and how to compute their inverses. In this section, we discuss some elementary properties of invertible matrices.

The inverse of a real number can be used to solve certain equations. For example, the equation $2x = 14$ can be solved by multiplying both sides of the equation by the inverse of 2:

$$\begin{aligned} 2^{-1}(2x) &= 2^{-1}(14) \\ (2^{-1}2)x &= 7 \\ 1x &= 7 \\ x &= 7 \end{aligned}$$

In a similar manner, if A is an invertible $n \times n$ matrix, then we can use A^{-1} to solve matrix equations in which an unknown matrix is multiplied by A. For example, if A is invertible, then we can solve the matrix equation $A\mathbf{x} = \mathbf{b}$ as follows:[2]

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ A^{-1}(A\mathbf{x}) &= A^{-1}\mathbf{b} \end{aligned}$$

---

[2] Although matrix inverses can be used to solve systems whose coefficient matrices are invertible, the method of solving systems presented in Chapter 1 is far more efficient.

$$(A^{-1}A)\mathbf{x} = A^{-1}\mathbf{b}$$
$$I_n\mathbf{x} = A^{-1}\mathbf{b}$$
$$\mathbf{x} = A^{-1}\mathbf{b}$$

If A is an invertible $n \times n$ matrix, then for every $\mathbf{b}$ in $\mathcal{R}^n$, $A\mathbf{x} = \mathbf{b}$ has the unique solution $A^{-1}\mathbf{b}$.

In solving a system of linear equations by using the inverse of a matrix A, we observe that A^{-1} "reverses" the action of A; that is, if A is an invertible $n \times n$ matrix and $\mathbf{u}$ is a vector in $\mathcal{R}^n$, then $A^{-1}(A\mathbf{u}) = \mathbf{u}$. (See Figure 2.7.)

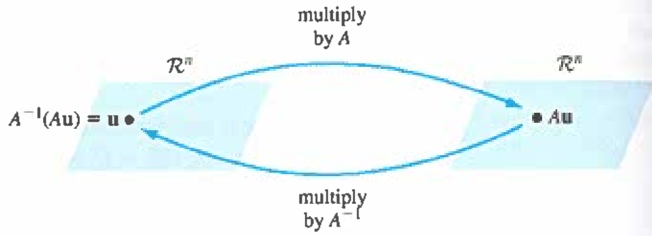


Figure 2.7 Multiplication by a matrix and its inverse

Example 2 Use a matrix inverse to solve the system of linear equations

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

Solution This system is the same as the matrix equation $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 4 \\ 7 \end{bmatrix}.$$

We saw in Example 1 that A is invertible. Hence we can solve this equation for $\mathbf{x}$ by multiplying both sides of the equation on the left by

$$A^{-1} = \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix}$$

as follows:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{x} = A^{-1}\mathbf{b} = \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix}\begin{bmatrix} 4 \\ 7 \end{bmatrix} = \begin{bmatrix} -6 \\ 5 \end{bmatrix}$$

Therefore $x_1 = -6$ and $x_2 = 5$ is the unique solution of the system.

Practice Problem 2 ▶ Use the answer to Practice Problem 1 to solve the following system of linear equations:

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

Example 3 Recall the rotation matrix

$$A_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

considered in Section 1.2 and Example 4 of Section 2.1. Notice that for $\theta = 0^\circ$, $A_\theta = I_2$. Furthermore, for any angle α,

$$A_\alpha A_{-\alpha} = A_{\alpha+(-\alpha)} = A_{0^\circ} = I_2.$$

Similarly, $A_{-\alpha}A_\alpha = I_2$. Hence A_α satisfies the definition of an invertible matrix with inverse $A_{-\alpha}$. Therefore $(A_\alpha)^{-1} = A_{-\alpha}$.

Another way of viewing $A_{-\alpha}A_\alpha$ is that it represents a rotation by α, followed by a rotation by $-\alpha$, which results in a net rotation of 0°. This is the same as multiplying by the identity matrix.

The following theorem states some useful properties of matrix inverses:

THEOREM 2.2

Let A and B be $n \times n$ matrices.

(a) If A is invertible, then A^{-1} is invertible and $(A^{-1})^{-1} = A$.

(b) If A and B are invertible, then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.

(c) If A is invertible, then A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.

PROOF The proof of (a) is a simple consequence of the definition of matrix inverse.

(b) Suppose that A and B are invertible. Then

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1} = AA^{-1} = I_n.$$

Similarly, $(B^{-1}A^{-1})(AB) = I_n$. Hence AB satisfies the definition of an invertible matrix with inverse $B^{-1}A^{-1}$; that is, $(AB)^{-1} = B^{-1}A^{-1}$.

(c) Suppose that A is invertible. Then $A^{-1}A = I_n$. Using Theorem 2.1(g), we obtain

$$A^T(A^{-1})^T = (A^{-1}A)^T = I_n^T = I_n.$$

Similarly, $(A^{-1})^TA^T = I_n$. Hence A^T satisfies the definition of an invertible matrix with the inverse $(A^{-1})^T$; that is, $(A^T)^{-1} = (A^{-1})^T$. ■

Part (b) of Theorem 2.2 can be easily extended to products of more than two matrices.

Let $A_1, A_2, \ldots, A_k$ be $n \times n$ invertible matrices. Then the product $A_1A_2\cdots A_k$ is invertible, and

$$(A_1A_2\cdots A_k)^{-1} = (A_k)^{-1}(A_{k-1})^{-1}\cdots(A_1)^{-1}.$$

Thus

$$\left\{\begin{bmatrix}-1\\2\\1\\0\end{bmatrix}, \begin{bmatrix}2\\-3\\0\\1\end{bmatrix}\right\}$$

is a basis for $W^{\perp}$.

Next, we apply the methods used in Practice Problem 1 to obtain an orthonormal basis $\{\mathbf{w}_1, \mathbf{w}_2\}$ for W, where

$$\mathbf{w}_1 = \frac{1}{2}\begin{bmatrix}1\\1\\-1\\1\end{bmatrix} \quad \text{and} \quad \mathbf{w}_2 = \frac{1}{2\sqrt{5}}\begin{bmatrix}3\\1\\1\\-3\end{bmatrix}.$$

Thus

$$\begin{aligned}\mathbf{w} &= (\mathbf{u}\cdot\mathbf{w}_1)\mathbf{w}_1 + (\mathbf{u}\cdot\mathbf{w}_2)\mathbf{w}_2\\ &= (5)\cdot\frac{1}{2}\begin{bmatrix}1\\1\\-1\\1\end{bmatrix} + (-\sqrt{5})\cdot\frac{1}{2\sqrt{5}}\begin{bmatrix}3\\1\\1\\-3\end{bmatrix}\\ &= \begin{bmatrix}1\\2\\-3\\4\end{bmatrix}.\end{aligned}$$

Finally,

$$\mathbf{z} = \mathbf{u} - \mathbf{w} = \begin{bmatrix}0\\7\\4\\7\end{bmatrix} - \begin{bmatrix}1\\2\\-3\\4\end{bmatrix} = \begin{bmatrix}-1\\5\\7\\3\end{bmatrix}.$$

(b) Let

$$C = \begin{bmatrix}1 & 3\\1 & 2\\-1 & -1\\1 & 0\end{bmatrix}.$$

Then

$$\begin{aligned}P_W &= C(C^TC)^{-1}C^T\\ &= \begin{bmatrix}1 & 3\\1 & 2\\-1 & -1\\1 & 0\end{bmatrix}\begin{bmatrix}0.7 & -0.3\\-0.3 & 0.2\end{bmatrix}\begin{bmatrix}1 & 1 & -1 & 1\\3 & 2 & -1 & 0\end{bmatrix}\\ &= \begin{bmatrix}0.7 & 0.4 & -0.1 & -0.2\\0.4 & 0.3 & -0.2 & 0.1\\-0.1 & -0.2 & 0.3 & -0.4\\-0.2 & 0.1 & -0.4 & 0.7\end{bmatrix}.\end{aligned}$$

Observe that the product $P_W\mathbf{u}$ gives the same result as obtained in (a).

(c) The distance from $\mathbf{u}$ to W is the distance between $\mathbf{u}$ and the orthogonal projection of $\mathbf{u}$ on W, which is

$$\|\mathbf{z}\| = \left\|\begin{bmatrix}-1\\5\\7\\3\end{bmatrix}\right\| = \sqrt{84}.$$

6.4 LEAST-SQUARES APPROXIMATION AND ORTHOGONAL PROJECTION MATRICES

In almost all areas of empirical research, there is an interest in finding simple mathematical relationships between variables. In economics, the variables might be the gross domestic product, the unemployment rate, and the annual deficit. In the life sciences, the variables of interest might be the incidence of smoking and heart disease. In sociology, it might be birth order and frequency of juvenile delinquency.

Many relationships in science are *deterministic*; that is, information about one variable completely determines the value of another variable. For example, the relationship between force f and acceleration a of an object of mass m is given by the equation $f = ma$ (Newton's second law). Another example is the height of a freely falling object and the time that it has been falling. On the other hand, the relationship between the height and the weight of an individual is not deterministic. There are many people with the same height, but different weights. Yet, in hospitals, there exist charts that give the recommended weights for given heights. Relationships that are not deterministic are often called *probabilistic* or *stochastic*.

We can apply what we know about orthogonal projections to identify relationships between variables. We begin with a given set of data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

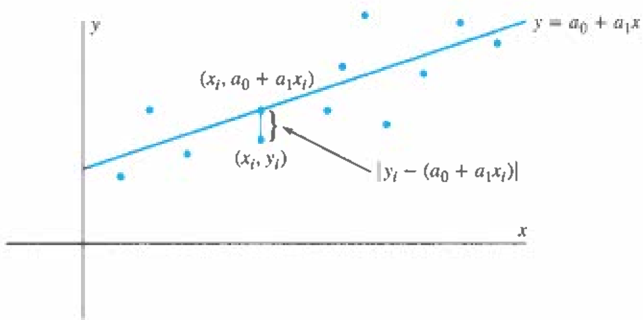


Figure 6.16 A plot of the data

obtained by empirical measurements. For example, we might have a randomly selected sample of n people, where x_i represents the number of years of education and y_i represents the annual income of the ith person. The data are plotted as in Figure 6.16. Notice that there is an approximately linear (straight line) relationship between x and y. To obtain this relationship, we would like to find the line $y = a_0 + a_1x$ that *best fits* the data. The usual criterion that statisticians use for defining the line of best fit is that the sum of the squared vertical distances of the data from it is smaller than from any other line. From Figure 6.16, we see that we must find a_0 and a_1 so that the quantity

$$E = [y_1 - (a_0 + a_1x_1)]^2 + [y_2 - (a_0 + a_1x_2)]^2 + \cdots + [y_n - (a_0 + a_1x_n)]^2 \tag{3}$$

is minimized. The technique to find this line is called the **method of least squares,**[10] E is called the **error sum of squares**, and the line for which E is minimized is called the **least-squares line.**

To find the least-squares line, we let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \text{and} \quad C = [\mathbf{v}_1 \;\; \mathbf{v}_2].$$

With this notation, equation (3) can be rewritten in the notation of vectors as

$$E = \|\mathbf{y} - (a_0\mathbf{v}_1 + a_1\mathbf{v}_2)\|^2. \tag{4}$$

(See Exercise 33.) Notice that $\sqrt{E} = \|\mathbf{y} - (a_0\mathbf{v}_1 + a_1\mathbf{v}_2)\|$ is the distance between $\mathbf{y}$ and the vector $a_0\mathbf{v}_1 + a_1\mathbf{v}_2$, which lies in $W = \text{Span}\,\{\mathbf{v}_1, \mathbf{v}_2\}$. So to minimize E, we need only choose the vector in W that is nearest to $\mathbf{y}$. But from the closest vector property, this vector is the orthogonal projection of $\mathbf{y}$ on W. Thus we want

$$a_0\mathbf{v}_1 + a_1\mathbf{v}_2 = C\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = P_W\mathbf{y},$$

the orthogonal projection of $\mathbf{y}$ on W.

For any reasonable set of data, the x_i's are not all equal, and hence $\mathbf{v}_1$ and $\mathbf{v}_2$ are not multiples of one another. Thus the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent,

[10] The method of least squares first appeared in a paper by Adrien Marie Legendre (1752–1833), entitled *Nouvelles Méthodes pour la détermination des orbites des comètes.*

and so $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for W. Since the columns of C form a basis for W, we may apply Theorem 6.8 to obtain

$$C\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = C(C^TC)^{-1}C^T\mathbf{y}.$$

Multiplying on the left by C^T gives

$$C^TC\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = C^TC(C^TC)^{-1}C^T\mathbf{y} = C^T\mathbf{y}.$$

The matrix equation $C^TC\mathbf{x} = C^T\mathbf{y}$ corresponds to a system of linear equations called the **normal equations**. Thus the line of best fit occurs when $\begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$ is the solution of the normal equations. Since C^TC is invertible by the lemma preceding Theorem 6.8, we see that the least-squares line has the equation $y = a_0 + a_1x$, where

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = (C^TC)^{-1}C^T\mathbf{y}.$$

Example 1

In the manufacture of refrigerators, it is necessary to finish connecting rods. If the weight of the finished rod is above a certain amount, the rod must be discarded. As the finishing process is expensive, it would be of considerable value to the manufacturer to be able to estimate the relationship between the finished weight and the initial rough weight. Then, those rods whose rough weights are too high could be discarded before they are finished. From past experience, the manufacturer knows that this relationship is approximately linear.

From a sample of five rods, we let x_i and y_i denote the rough weight and the finished weight, respectively, of the ith rod. The data are given in the following table:

Rough weight x_i (in pounds)	Finished weight y_i (in pounds)
2.60	2.00
2.72	2.10
2.75	2.10
2.67	2.03
2.68	2.04

From this information, we let

$$C = \begin{bmatrix} 1 & 2.60 \\ 1 & 2.72 \\ 1 & 2.75 \\ 1 & 2.67 \\ 1 & 2.68 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} 2.00 \\ 2.10 \\ 2.10 \\ 2.03 \\ 2.04 \end{bmatrix}.$$

Then

$$C^TC = \begin{bmatrix} 5.0000 & 13.4200 \\ 13.4200 & 36.0322 \end{bmatrix} \quad \text{and} \quad C^T\mathbf{y} = \begin{bmatrix} 10.2700 \\ 27.5743 \end{bmatrix},$$

and the solution of the normal equations is

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} \approx \begin{bmatrix} 0.056 \\ 0.745 \end{bmatrix}.$$

Thus the approximate relationship between the finished weight y and the rough weight x is given by the equation of the least-squares line

$$y = 0.056 + 0.745x.$$

For example, if the rough weight of a rod is 2.65 pounds, then the finished weight is approximately

$$0.056 + 0.745(2.65) \approx 2.030 \text{ pounds.}$$

Practice Problem 1 ▶ Find the equation of the least-squares line for the data $(1, 62)$, $(3, 54)$, $(4, 50)$, $(5, 48)$, and $(7, 40)$. ◀

The method we have developed for finding the best fit to data points (x_1, y_1), (x_2, y_2), ..., (x_n, y_n) by a linear polynomial $a_0 + a_1x$ can be modified to find the best fit by a quadratic polynomial $y = a_0 + a_1x + a_2x^2$. The only change in the method is that the new error sum of squares is

$$E = [y_1 - (a_0 + a_1x_1 + a_2x_1^2)]^2 + \cdots + [y_n - (a_0 + a_1x_n + a_2x_n^2)]^2.$$

In this case, let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_n^2 \end{bmatrix}, \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Assuming that the x_i's are distinct and $n \geq 3$, which in practice is always the case, the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are linearly independent (see Exercise 34), and hence they form a basis for a 3-dimensional subspace W of $\mathcal{R}^n$. So we let C be the $n \times 3$ matrix $C = [\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3]$. As in the linear case, we can obtain the normal equations

$$C^TC \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = C^T\mathbf{y},$$

whose solution is

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = (C^TC)^{-1}C^T\mathbf{y}.$$

Example 2 It is known from physics that if a ball is thrown upward at a velocity of v_0 feet per second from a building of height s_0 feet, then the height of the ball after t seconds is given by $s = s_0 + v_0t + \frac{1}{2}gt^2$, where g represents the acceleration due to gravity. To provide an empirical estimate of g, a ball is thrown upward from a building 100 feet high at a velocity of 30 feet per second. The height of the ball is observed at the times given in the following table:

Time (in seconds)	Height (in feet)
0	100
1	118
2	92
3	48
3.5	7

For these data, we let

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 3.5 & 12.25 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} 100 \\ 118 \\ 92 \\ 48 \\ 7 \end{bmatrix}.$$

Thus the quadratic polynomial $y = a_0 + a_1x + a_2x^2$ of best fit satisfies

$$\begin{bmatrix} s_0 \\ v_0 \\ \frac{1}{2}g \end{bmatrix} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = (C^TC)^{-1}C^T\mathbf{y} \approx \begin{bmatrix} 101.00 \\ 29.77 \\ -16.11 \end{bmatrix}.$$

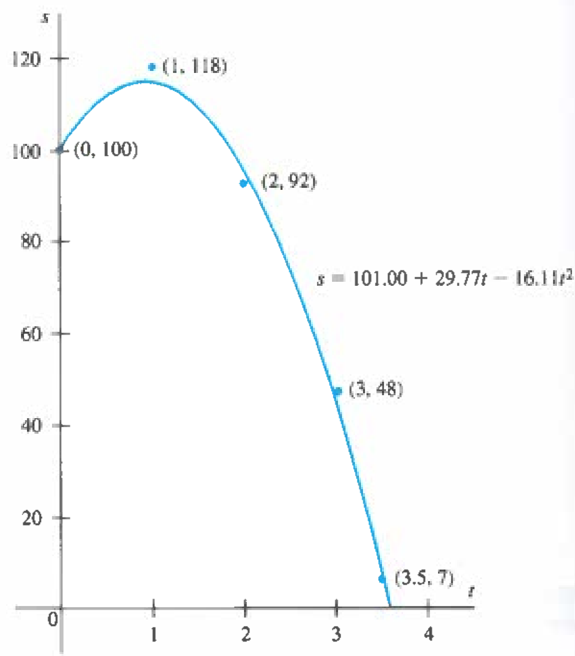


Figure 6.17 The quadratic polynomial of best fit for the data in Example 2

This yields the approximate relationship

$$s = 101.00 + 29.77t - 16.11t^2.$$

(See Figure 6.17.) Setting $\frac{1}{2}g = -16.11$, we obtain -32.22 feet per second per second as the estimate for g.

It should be pointed out that the same method may be extended to find the best-fitting polynomial[11] of any desired maximum degree, provided that the data set is sufficiently large. Furthermore, by using the appropriate change of variable, many more complicated relationships may be estimated by the same type of matrix computations.

The material treated in the rest of this section will be revisited from a different perspective in Section 6.7.

INCONSISTENT SYSTEMS OF LINEAR EQUATIONS*

The preceding examples are special cases of inconsistent systems of linear equations for which it is desirable to obtain approximate solutions. In general, a system of linear equations $A\mathbf{x} = \mathbf{b}$ arising from the application of a theoretical model to real data may be inconsistent because the entries of A and $\mathbf{b}$ that are obtained from empirical measurements are not precise or because the model only approximates reality. In these circumstances, we are interested in obtaining a vector $\mathbf{z}$ for which $\|A\mathbf{z} - \mathbf{b}\|$ is a minimum. Let W denote the set of all vectors of the form $A\mathbf{u}$. Then W is the column space of A. By the closest vector property, the vector in W that is closest to $\mathbf{b}$ is the orthogonal projection of $\mathbf{b}$ on W, which can be computed as $P_W\mathbf{b}$. Thus a vector $\mathbf{z}$ minimizes $\|A\mathbf{z} - \mathbf{b}\|$ if and only if it is a solution of the system of linear equations

$$A\mathbf{x} = P_W\mathbf{b},$$

which is guaranteed to be consistent. (See Figure 6.18.)

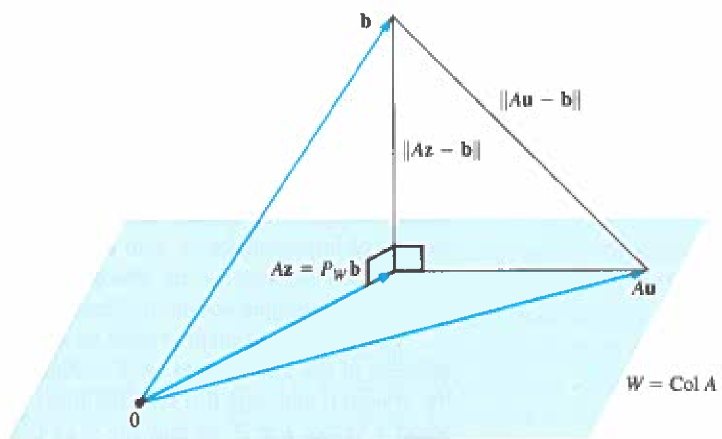


Figure 6.18 The vector $\mathbf{z}$ minimizes $\|A\mathbf{z} - \mathbf{b}\|$ if and only if it is a solution of the system of linear equations $A\mathbf{x} = P_W\mathbf{b}$.

[11] Caution! The MATLAB function `polyfit` returns the coefficients of the polynomial of best fit with terms written in *descending* order (rather than in ascending order, as in this book).

* The remainder of this section may be omitted without loss of continuity.

Example 3 Given the inconsistent system of linear equations $A\mathbf{x} = \mathbf{b}$, with

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

use the method of least squares to describe the vectors $\mathbf{z}$ for which $\|A\mathbf{z} - \mathbf{b}\|$ is a minimum.

Solution By computing the reduced row echelon form of A, we see that the rank of A is 2 and that the first two columns of A are linearly independent. Thus the first two columns of A form a basis for $W = \operatorname{Col} A$. Let C be the 4×2 matrix with these two vectors as its columns. Then

$$P_W\mathbf{b} = C(C^TC)^{-1}C^T\mathbf{b} = \frac{1}{3}\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 1 \\ 1 & -1 & 2 & 0 \\ 1 & 1 & 0 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 7 \\ -4 \\ 8 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 5 \\ 19 \\ -14 \\ 24 \end{bmatrix}.$$

As noted, the vectors that minimize $\|A\mathbf{z} - \mathbf{b}\|$ are the solutions to $A\mathbf{x} = P_W\mathbf{b}$. The general solution of this system is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 14 \\ -9 \\ 0 \end{bmatrix} + x_3\begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}.$$

So these are the vectors that minimize $\|A\mathbf{z} - \mathbf{b}\|$. Note that, for each of these vectors, we have

$$\|A\mathbf{z} - \mathbf{b}\| = \|P_W\mathbf{b} - \mathbf{b}\| = \left\| \frac{1}{3}\begin{bmatrix} 5 \\ 19 \\ -14 \\ 24 \end{bmatrix} - \begin{bmatrix} 1 \\ 7 \\ -4 \\ 8 \end{bmatrix} \right\| = \frac{2}{\sqrt{3}}.$$

SOLUTIONS OF LEAST NORM

In solving the problem posed in Example 3, we obtained an infinite set of solutions of a nonhomogeneous system of linear equations. In general, given a nonhomogeneous system of linear equations with an infinite set of solutions, it is often useful to select the solution of least norm. We show, using orthogonal projections, that any such system has a unique solution of least norm.

Consider a consistent system $A\mathbf{x} = \mathbf{c}$ of linear equations with $\mathbf{c} \neq \mathbf{0}$. Let $\mathbf{v}_0$ be any solution of the system, and let $Z = \operatorname{Null} A$. By Exercise 35, a vector $\mathbf{v}$ is a solution of the system if and only if it is of the form $\mathbf{v} = \mathbf{v}_0 + \mathbf{z}$, where $\mathbf{z}$ is in Z. Here, we wish to select a vector $\mathbf{z}$ in Z so that $\|\mathbf{v}_0 + \mathbf{z}\|$ is a minimum. Since $\|\mathbf{v}_0 + \mathbf{z}\| = \|-\mathbf{v}_0 - \mathbf{z}\|$, which is the distance between $-\mathbf{v}_0$ and $\mathbf{z}$, the vector in Z that minimizes this distance is, of course, the orthogonal projection of $-\mathbf{v}_0$ on Z; that is, $\mathbf{z} = P_Z(-\mathbf{v}_0) = -P_Z\mathbf{v}_0$. Thus $\mathbf{v}_0 + \mathbf{z} = \mathbf{v}_0 - P_Z\mathbf{v}_0$ is the unique solution of the system of least norm.

Example 4 Find the solution of least norm to the equation $A\mathbf{x} = P_W\mathbf{b}$ in Example 3.

Solution Based on the vector form of the solution given in Example 3, a vector $\mathbf{v}$ is a solution if and only if $\mathbf{v} = \mathbf{v}_0 + \mathbf{z}$, for $\mathbf{z}$ in Z, where

$$\mathbf{v}_0 = \frac{1}{3}\begin{bmatrix} 14 \\ -9 \\ 0 \end{bmatrix} \quad \text{and} \quad Z = \text{Null}\, A = \text{Span}\left\{\begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}\right\}.$$

Setting $C = \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}$, we compute the orthogonal projection matrix

$$P_Z = C(C^TC)^{-1}C^T = \frac{1}{14}\begin{bmatrix} 9 & -6 & 3 \\ -6 & 4 & 2 \\ -3 & 2 & 1 \end{bmatrix}.$$

Thus

$$\mathbf{v}_0 - P_Z\mathbf{v}_0 = (I_3 - P_Z)\mathbf{v}_0 = \frac{1}{21}\begin{bmatrix} 8 \\ -3 \\ 30 \end{bmatrix}$$

is the solution of least norm.

EXERCISES

In Exercises 1–8, find the equation of the least-squares line for the given data.

1. $(1, 14), (3, 17), (5, 19), (7, 20)$
2. $(1, 30), (2, 27), (4, 21), (7, 14)$
3. $(1, 5), (2, 6), (3, 8), (4, 10), (5, 11)$
4. $(1, 2), (2, 4), (3, 7), (4, 8), (5, 10)$
5. $(1, 40), (3, 36), (7, 23), (8, 21), (10, 13)$
6. $(1, 19), (2, 17), (3, 16), (4, 14), (5, 12)$
7. $(1, 4), (4, 24), (5, 30), (8, 32), (12, 36)$
8. $(1, 21), (3, 32), (9, 38), (12, 41), (15, 51)$
9. Suppose that a spring whose natural length is L inches is attached to a wall. A force y is applied to the free end of the spring, stretching the spring s inches beyond its natural length. *Hooke's law* states (within certain limits) that $y = ks$, where k is a constant called the *spring constant.* Now suppose that after the force y is applied, the new length of the spring is x. Then $s = x - L$, and Hooke's law yields

$$y = ks = k(x - L) = a + kx,$$

where $a = -kL$. Apply the method of least squares to the following data to estimate k and L:

Length x in inches	Force y in pounds
3.5	1.0
4.0	2.2
4.5	2.8
5.0	4.3

In Exercises 10–15, use the method of least squares to find the polynomial of degree at most n that best fits the given data.

10. $n = 2$ with data $(0, 2), (1, 2), (2, 4), (3, 8)$
11. $n = 2$ with data $(0, 3), (1, 3), (2, 5), (3, 9)$
12. $n = 2$ with data $(0, 1), (1, 2), (2, 3), (3, 4)$
13. $n = 2$ with data $(0, 2), (1, 3), (2, 5), (3, 8)$
14. $n = 3$ with data $(-2, -5), (-1, -1), (0, -1), (1, 1), (2, 11)$
15. $n = 3$ with data $(-2, -4), (-1, -5), (0, 5), (1, -3), (2, 12)$