

- **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



Modern CFD has revolutionized wing design. The Boeing Blended Wing Body is in design for the year 2020 or sooner.

matrix multiplications. Section 2.7 explains the basic ideas.

WEB

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 square 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.

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 i th row and j th 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 $\mathbf{a}_1, \dots, \mathbf{a}_n$, and the matrix A is written as

$$A = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_n]$$

Observe that the number a_{ij} is the i th entry (from the top) of the j th column vector \mathbf{a}_j .

The **diagonal entries** in an $m \times n$ matrix $A = [a_{ij}]$ are $a_{11}, a_{22}, a_{33}, \dots$, 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.

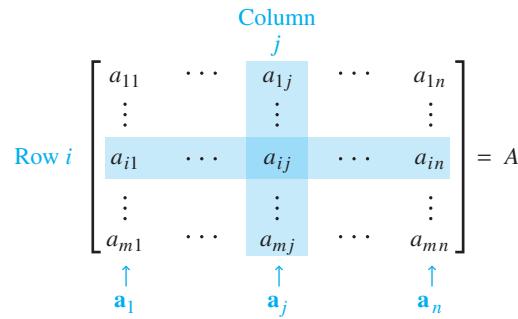


FIGURE 1 Matrix notation.

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 j th columns of A , B , and C are \mathbf{a}_j , \mathbf{b}_j , and \mathbf{c}_j , respectively, then the j th 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.

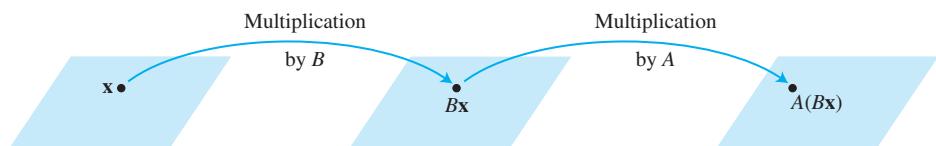


FIGURE 2 Multiplication by B and then A .

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.

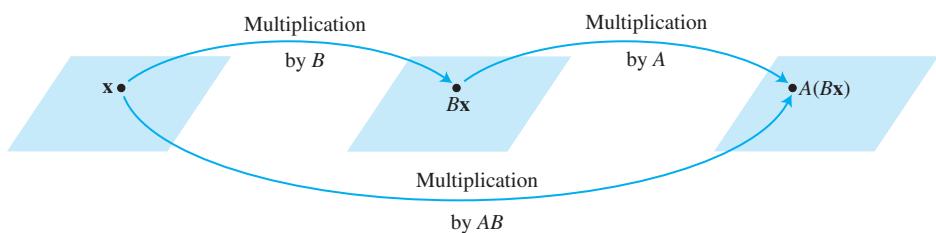


FIGURE 3 Multiplication by AB .

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 ,

$$\begin{aligned} A(B\mathbf{x}) &= A(x_1\mathbf{b}_1) + \cdots + A(x_p\mathbf{b}_p) \\ &= x_1A\mathbf{b}_1 + \cdots + x_pA\mathbf{b}_p \end{aligned}$$

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

$$A(B\mathbf{x}) = [A\mathbf{b}_1 \ A\mathbf{b}_2 \ \cdots \ A\mathbf{b}_p]\mathbf{x}$$

Thus multiplication by $[A\mathbf{b}_1 \ A\mathbf{b}_2 \ \cdots \ A\mathbf{b}_p]$ 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,

$$AB = A[\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_p] = [A\mathbf{b}_1 \ A\mathbf{b}_2 \ \cdots \ A\mathbf{b}_p]$$

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 $B = [\mathbf{b}_1 \ \mathbf{b}_2 \ \mathbf{b}_3]$, and compute:

$$\begin{aligned} A\mathbf{b}_1 &= \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix}, & A\mathbf{b}_2 &= \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix}, & A\mathbf{b}_3 &= \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 6 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} 11 \\ -1 \end{bmatrix} & &= \begin{bmatrix} 0 \\ 13 \end{bmatrix} & &= \begin{bmatrix} 21 \\ -9 \end{bmatrix} \end{aligned}$$

Then

$$AB = A[\mathbf{b}_1 \ \mathbf{b}_2 \ \mathbf{b}_3] = \begin{bmatrix} 11 & 0 & 21 \\ -1 & 13 & -9 \end{bmatrix}$$

$\uparrow \quad \uparrow \quad \uparrow$
 $A\mathbf{b}_1 \quad A\mathbf{b}_2 \quad A\mathbf{b}_3$

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:

$$\begin{array}{c} A \\ \left[\begin{array}{ccccc} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{array} \right] \\ 3 \times 5 \end{array} \quad \begin{array}{c} B \\ \left[\begin{array}{cc} * & * \\ * & * \\ * & * \\ * & * \\ * & * \end{array} \right] \\ 5 \times 2 \end{array} = \begin{array}{c} AB \\ \left[\begin{array}{cc} * & * \\ * & * \\ * & * \end{array} \right] \\ 3 \times 2 \end{array}$$

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 i th 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:

$$AB = \rightarrow \left[\begin{array}{cc} 2 & 3 \\ 1 & -5 \end{array} \right] \left[\begin{array}{ccc} 4 & 3 & 6 \\ 1 & -2 & 3 \end{array} \right] = \left[\begin{array}{ccc} \square & \square & 2(6) + 3(3) \\ \square & \square & \square \end{array} \right] = \left[\begin{array}{ccc} \square & \square & 21 \\ \square & \square & \square \end{array} \right]$$

For the entry in row 2 and column 2 of AB , use row 2 of A and column 2 of B :

$$\rightarrow \left[\begin{array}{cc} 2 & 3 \\ 1 & -5 \end{array} \right] \left[\begin{array}{ccc} 4 & 3 & 6 \\ 1 & -2 & 3 \end{array} \right] = \left[\begin{array}{ccc} \square & \square & 21 \\ \square & 1(3) + -5(-2) & \square \end{array} \right] = \left[\begin{array}{ccc} \square & \square & 21 \\ \square & 13 & \square \end{array} \right]$$

EXAMPLE 6 Find the entries in the second row of AB , where

$$A = \begin{bmatrix} 2 & -5 & 0 \\ -1 & 3 & -4 \\ 6 & -8 & -7 \\ -3 & 0 & 9 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & -6 \\ 7 & 1 \\ 3 & 2 \end{bmatrix}$$

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):

$$\begin{aligned} & \rightarrow \begin{bmatrix} 2 & -5 & 0 \\ -1 & 3 & -4 \\ 6 & -8 & -7 \\ -3 & 0 & 9 \end{bmatrix} \begin{bmatrix} 4 & -6 \\ 7 & 1 \\ 3 & 2 \end{bmatrix} \quad \downarrow \quad \downarrow \\ & = \begin{bmatrix} \square & \square \\ -4 + 21 - 12 & 6 + 3 - 8 \\ \square & \square \\ \square & \square \end{bmatrix} = \begin{bmatrix} \square & \square \\ 5 & 1 \\ \square & \square \\ \square & \square \end{bmatrix} \quad \blacksquare \end{aligned}$$

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 i th row of a matrix A . Then

$$\text{row}_i(AB) = \text{row}_i(A) \cdot B \quad (2)$$

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. $I_m A = A = AI_n$ (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

$$C = [\mathbf{c}_1 \ \cdots \ \mathbf{c}_p]$$

By the definition of matrix multiplication,

$$\begin{aligned} BC &= [B\mathbf{c}_1 \ \cdots \ B\mathbf{c}_p] \\ A(BC) &= [A(B\mathbf{c}_1) \ \cdots \ A(B\mathbf{c}_p)] \end{aligned}$$

Recall from equation (1) that the definition of AB makes $A(B\mathbf{x}) = (AB)\mathbf{x}$ for all \mathbf{x} , so

$$A(BC) = [(AB)\mathbf{c}_1 \ \cdots \ (AB)\mathbf{c}_p] = (AB)C \quad \blacksquare$$

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

$$\begin{aligned} AB &= \begin{bmatrix} 5 & 1 \\ 3 & -2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 14 & 3 \\ -2 & -6 \end{bmatrix} \\ BA &= \begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 5 & 1 \\ 3 & -2 \end{bmatrix} = \begin{bmatrix} 10 & 2 \\ 29 & -2 \end{bmatrix} \end{aligned} \quad \blacksquare$$

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

WEB

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, 4.9, 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

$$A^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}, \quad B^T = \begin{bmatrix} -5 & 1 & 0 \\ 2 & -3 & 4 \end{bmatrix}, \quad C^T = \begin{bmatrix} 1 & -3 \\ 1 & 5 \\ 1 & -2 \\ 1 & 7 \end{bmatrix} \blacksquare$$

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 33. 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.

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 $(Ax)^T$, $\mathbf{x}^T A^T$, $\mathbf{x} \mathbf{x}^T$, 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

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

$$C = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 3 & 5 \\ -1 & 4 \end{bmatrix}, \quad E = \begin{bmatrix} -5 \\ 3 \end{bmatrix}$$

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

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

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

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

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 $A = \begin{bmatrix} 2 & 5 \\ -3 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & -5 \\ 3 & k \end{bmatrix}$. What value(s) of k , if any, will make $AB = BA$?

10. Let $A = \begin{bmatrix} 2 & -3 \\ -4 & 6 \end{bmatrix}$, $B = \begin{bmatrix} 8 & 4 \\ 5 & 5 \end{bmatrix}$, and $C = \begin{bmatrix} 5 & -2 \\ 3 & 1 \end{bmatrix}$. 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 $A = \begin{bmatrix} 3 & -6 \\ -1 & 2 \end{bmatrix}$. 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 and 16 concern arbitrary matrices A , B , and C for which the indicated sums and products are defined. Mark each statement True or False. Justify each answer.
15. a. If A and B are 2×2 with columns $\mathbf{a}_1, \mathbf{a}_2$, and $\mathbf{b}_1, \mathbf{b}_2$, respectively, then $AB = [\mathbf{a}_1\mathbf{b}_1 \quad \mathbf{a}_2\mathbf{b}_2]$.
 b. Each column of AB is a linear combination of the columns of B using weights from the corresponding column of A .
 c. $AB + AC = A(B + C)$
 d. $A^T + B^T = (A + B)^T$
 e. The transpose of a product of matrices equals the product of their transposes in the same order.
16. a. If A and B are 3×3 and $B = [\mathbf{b}_1 \mathbf{b}_2 \mathbf{b}_3]$, then $AB = [A\mathbf{b}_1 + A\mathbf{b}_2 + A\mathbf{b}_3]$.
 b. The second row of AB is the second row of A multiplied on the right by B .
 c. $(AB)C = (AC)B$
 d. $(AB)^T = A^T B^T$
 e. The transpose of a sum of matrices equals the sum of their transposes.
17. If $A = \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix}$ and $AB = \begin{bmatrix} -1 & 2 & -1 \\ 6 & -9 & 3 \end{bmatrix}$, determine the first and second columns of B .
18. 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?
19. 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?
20. Suppose the second column of B is all zeros. What can you say about the second column of AB ?
21. Suppose the last column of AB is entirely zero but B itself has no column of zeros. What can you say about the columns of A ?
22. Show that if the columns of B are linearly dependent, then so are the columns of AB .
23. Suppose $CA = I_n$ (the $n \times n$ identity matrix). Show that the equation $Ax = \mathbf{0}$ has only the trivial solution. Explain why A cannot have more columns than rows.
24. Suppose $AD = I_m$ (the $m \times m$ identity matrix). Show that for any \mathbf{b} in \mathbb{R}^m , the equation $Ax = \mathbf{b}$ has a solution. [Hint: Think about the equation $AD\mathbf{b} = \mathbf{b}$.] Explain why A cannot have more rows than columns.
25. 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 .]
26. 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 27 and 28, 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.
27. 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 $\mathbf{v}\mathbf{u}^T$.
28. 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?
29. Prove Theorem 2(b) and 2(c). Use the row–column rule. The (i, j) -entry in $A(B + C)$ can be written as
- $$a_{i1}(b_{1j} + c_{1j}) + \cdots + a_{in}(b_{nj} + c_{nj}) \text{ or } \sum_{k=1}^n a_{ik}(b_{kj} + c_{kj})$$
30. Prove Theorem 2(d). [Hint: The (i, j) -entry in $(rA)B$ is $(ra_{i1})b_{1j} + \cdots + (ra_{in})b_{nj}$.]
31. 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 .
32. Show that $AI_n = A$ when A is an $m \times n$ matrix. [Hint: Use the (column) definition of AI_n .]
33. Prove Theorem 3(d). [Hint: Consider the j th row of $(AB)^T$.]
34. Give a formula for $(AB\mathbf{x})^T$, where \mathbf{x} is a vector and A and B are matrices of appropriate sizes.
35. [M] Read the documentation for your matrix program, and write the commands that will produce the following matrices (without keying in each entry of the matrix).
- A 5×6 matrix of zeros
 - 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.

- 36.** [M] 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 $-9.5 < 19(x - .5) < 9.5$.]
- 37.** [M] 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.

- 38.** [M] 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 37.) Report your conclusions. [Note: Most matrix programs use A' for A^T .]

- 39.** [M] 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 $k = 2, \dots, 6$.

- 40.** [M] Describe in words what happens when you compute A^5 , A^{10} , A^{20} , and A^{30} for

$$A = \begin{bmatrix} 1/6 & 1/2 & 1/3 \\ 1/2 & 1/4 & 1/4 \\ 1/3 & 1/4 & 5/12 \end{bmatrix}$$

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 $(A\mathbf{x})^T = [-4 \quad 2]$. Also,

$$\mathbf{x}^T A^T = [5 \quad 3] \begin{bmatrix} 1 & -2 \\ -3 & 4 \end{bmatrix} = [-4 \quad 2].$$

The quantities $(A\mathbf{x})^T$ and $\mathbf{x}^T A^T$ are equal, by Theorem 3(d). Next,

$$\begin{aligned} \mathbf{x}\mathbf{x}^T &= \begin{bmatrix} 5 \\ 3 \end{bmatrix} [5 \quad 3] = \begin{bmatrix} 25 & 15 \\ 15 & 9 \end{bmatrix} \\ \mathbf{x}^T \mathbf{x} &= [5 \quad 3] \begin{bmatrix} 5 \\ 3 \end{bmatrix} = [25 + 9] = 34 \end{aligned}$$

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

$$5^{-1} \cdot 5 = 1 \quad \text{and} \quad 5 \cdot 5^{-1} = 1$$

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**.

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 25 and 26. 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 $\det A \neq 0$.

¹One 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 23–25 in Section 2.1.

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, below.

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} .

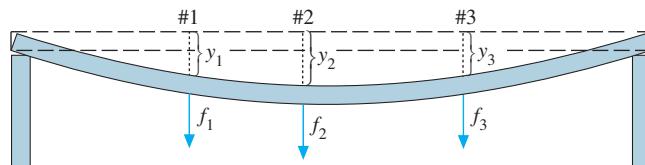


FIGURE 1 Deflection of an elastic beam.

SOLUTION Write $I_3 = [\mathbf{e}_1 \ \mathbf{e}_2 \ \mathbf{e}_3]$ and observe that

$$D = DI_3 = [D\mathbf{e}_1 \ D\mathbf{e}_2 \ D\mathbf{e}_3]$$

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

$$D^{-1} = D^{-1}I_3 = [D^{-1}\mathbf{e}_1 \ D^{-1}\mathbf{e}_2 \ D^{-1}\mathbf{e}_3]$$

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 $Ax = b$ numerically because row reduction of $[A \ b]$ 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 $Ax = b$ 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 $Ax = b$, so

$$x = A^{-1}b = \begin{bmatrix} -3 & 2 \\ 5/2 & -3/2 \end{bmatrix} \begin{bmatrix} 3 \\ 7 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$$

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, $(A^{-1})^TA^T = (AA^{-1})^T = I^T = I$. 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 27 and 28.

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 $Ax = \mathbf{b}$ 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 $A^{-1} = E_p \cdots E_1 \cdot I_n$, 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 $[A \ I]$, 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 $[A \ I]$. If A is row equivalent to I , then $[A \ I]$ is row equivalent to $[I \ A^{-1}]$. 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

$$\begin{aligned} [A \ I] &= \left[\begin{array}{ccc|ccc} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array} \right] \sim \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array} \right] \\ &\sim \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & -3 & -4 & 0 & -4 & 1 \end{array} \right] \sim \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 3 & -4 & 1 \end{array} \right] \\ &\sim \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 3/2 & -2 & 1/2 \end{array} \right] \\ &\sim \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -9/2 & 7 & -3/2 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & 3/2 & -2 & 1/2 \end{array} \right] \end{aligned}$$

Theorem 7 shows, since $A \sim I$, that A is invertible, and

$$A^{-1} = \begin{bmatrix} -9/2 & 7 & -3/2 \\ -2 & 4 & -1 \\ 3/2 & -2 & 1/2 \end{bmatrix}$$

It is a good idea to check the final answer:

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

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 $[A \ I]$ to $[I \ A^{-1}]$ 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 $[A \ \mathbf{e}_1 \ \mathbf{e}_2 \ \dots \ \mathbf{e}_n] = [A \ I]$. 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 be solved.

WEB

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

- Use determinants to determine which of the following matrices are invertible.
 - $\begin{bmatrix} 3 & -9 \\ 2 & 6 \end{bmatrix}$
 - $\begin{bmatrix} 4 & -9 \\ 0 & 5 \end{bmatrix}$
 - $\begin{bmatrix} 6 & -9 \\ -4 & 6 \end{bmatrix}$
- Find the inverse of the matrix $A = \begin{bmatrix} 1 & -2 & -1 \\ -1 & 5 & 6 \\ 5 & -4 & 5 \end{bmatrix}$, if it exists.
- 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. $\begin{bmatrix} 8 & 6 \\ 5 & 4 \end{bmatrix}$

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

3. $\begin{bmatrix} 8 & 5 \\ -7 & -5 \end{bmatrix}$

4. $\begin{bmatrix} 3 & -4 \\ 7 & -8 \end{bmatrix}$

5. Use the inverse found in Exercise 1 to solve the system

$8x_1 + 6x_2 = 2$

$5x_1 + 4x_2 = -1$

6. Use the inverse found in Exercise 3 to solve the system

$8x_1 + 5x_2 = -9$

$-7x_1 - 5x_2 = 11$

7. 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 $\mathbf{b}_4 = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$.

- Find A^{-1} , and use it to solve the four equations $A\mathbf{x} = \mathbf{b}_1$, $A\mathbf{x} = \mathbf{b}_2$, $A\mathbf{x} = \mathbf{b}_3$, $A\mathbf{x} = \mathbf{b}_4$.
- 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 $[A \ \mathbf{b}_1 \ \mathbf{b}_2 \ \mathbf{b}_3 \ \mathbf{b}_4]$.
- Use matrix algebra to show that if A is invertible and D satisfies $AD = I$, then $D = A^{-1}$.

In Exercises 9 and 10, mark each statement True or False. Justify each answer.

- In order for a matrix B to be the inverse of A , both equations $AB = I$ and $BA = I$ must be true.
- If A and B are $n \times n$ and invertible, then $A^{-1}B^{-1}$ is the inverse of AB .
- If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $ab - cd \neq 0$, then A is invertible.
- 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 .
- Each elementary matrix is invertible.
- 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.
- If A is invertible, then the inverse of A^{-1} is A itself.
- If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $ad = bc$, then A is not invertible.
- If A can be row reduced to the identity matrix, then A must be invertible.
- If A is invertible, then elementary row operations that reduce A to the identity I_n also reduce A^{-1} to I_n .
- 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 $A^{-1}B$.
- 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:

If $[A \ B] \sim \cdots \sim [I \ X]$, 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 $A^{-1}B$.

13. 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?
14. Suppose $(B - C)D = 0$, where B and C are $m \times n$ matrices and D is invertible. Show that $B = C$.
15. 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 $D(ABC) = I$.
16. 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 .]
17. Solve the equation $AB = BC$ for A , assuming that A, B , and C are square and B is invertible.

18. Suppose P is invertible and $A = PBP^{-1}$. Solve for B in terms of A .
19. 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.
20. Suppose A, B , and X are $n \times n$ matrices with A, X , and $A - AX$ invertible, and suppose

$$(A - AX)^{-1} = X^{-1}B \quad (3)$$

- a. Explain why B is invertible.
- b. Solve (3) for X . If you need to invert a matrix, explain why that matrix is invertible.

21. Explain why the columns of an $n \times n$ matrix A are linearly independent when A is invertible.
22. 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.]
23. 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 24 will be cited in Section 2.3.)
24. 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 25 and 26 prove Theorem 4 for $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

25. 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 $\mathbf{x} = \begin{bmatrix} -b \\ a \end{bmatrix}$.]

26. Show that if $ad - bc \neq 0$, the formula for A^{-1} works.

Exercises 27 and 28 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.)

27. 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.
28. Show that if row 3 of A is replaced by $\text{row}_3(A) - 4 \cdot \text{row}_1(A)$, the result is EA , where E is formed from I by replacing $\text{row}_3(I)$ by $\text{row}_3(I) - 4 \cdot \text{row}_1(I)$.

Find the inverses of the matrices in Exercises 29–32, if they exist. Use the algorithm introduced in this section.

29. $\begin{bmatrix} 1 & 2 \\ 4 & 7 \end{bmatrix}$
30. $\begin{bmatrix} 5 & 10 \\ 4 & 7 \end{bmatrix}$
31. $\begin{bmatrix} 1 & 0 & -2 \\ -3 & 1 & 4 \\ 2 & -3 & 4 \end{bmatrix}$
32. $\begin{bmatrix} 1 & -2 & 1 \\ 4 & -7 & 3 \\ -2 & 6 & -4 \end{bmatrix}$
33. 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 $BA = I$.

34. Repeat the strategy of Exercise 33 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.

35. 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.

36. [M] 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.

37. 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 $AC \neq I_3$.

38. 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?

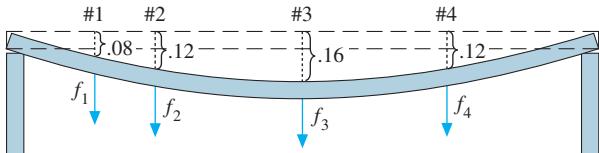
39. Let $D = \begin{bmatrix} .005 & .002 & .001 \\ .002 & .004 & .002 \\ .001 & .002 & .005 \end{bmatrix}$ be a flexibility matrix,

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.

40. [M] Compute the stiffness matrix D^{-1} for D in Exercise 39. List the forces needed to produce a deflection of .04 in. at point 3, with zero deflections at the other points.

41. [M] 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 41 and 42.

42. [M] With D as in Exercise 41, 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. $[A \ I] \sim \begin{bmatrix} 1 & -2 & -1 & 1 & 0 & 0 \\ -1 & 5 & 6 & 0 & 1 & 0 \\ 5 & -4 & 5 & 0 & 0 & 1 \end{bmatrix}$
 $\sim \begin{bmatrix} 1 & -2 & -1 & 1 & 0 & 0 \\ 0 & 3 & 5 & 1 & 1 & 0 \\ 0 & 6 & 10 & -5 & 0 & 1 \end{bmatrix}$
 $\sim \begin{bmatrix} 1 & -2 & -1 & 1 & 0 & 0 \\ 0 & 3 & 5 & 1 & 1 & 0 \\ 0 & 0 & 0 & -7 & -2 & 1 \end{bmatrix}$

So $[A \ I]$ is row equivalent to a matrix of the form $[B \ D]$, 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 $(5A)(1/5 C) = (5)(1/5)(AC) = 1I = I$, and $(1/5 C)(5A) = (1/5)(5)(CA) = 1I = I$. Thus $1/5 C$ is indeed the inverse of A , proving that A is invertible.

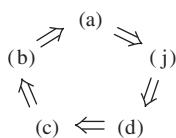
2.3 CHARACTERIZATIONS OF INVERTIBLE MATRICES

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.

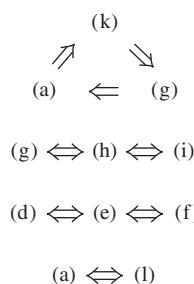
THEOREM 8**The Invertible Matrix Theorem**

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 is an invertible matrix.
- A is row equivalent to the $n \times n$ identity matrix.
- A has n pivot positions.
- The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- The columns of A form a linearly independent set.
- The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.
- The equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each \mathbf{b} in \mathbb{R}^n .
- The columns of A span \mathbb{R}^n .
- The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps \mathbb{R}^n onto \mathbb{R}^n .
- There is an $n \times n$ matrix C such that $CA = I$.
- There is an $n \times n$ matrix D such that $AD = I$.
- 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 23 in Section 2.1. (Turn back and read the exercise.) Also, $(d) \Rightarrow (c)$ by Exercise 23 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 24 in Section 2.1, and $(g) \Rightarrow (a)$ by Exercise 24 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.

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 8 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). ■

SG

Expanded Table
for the IMT 2-10

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.

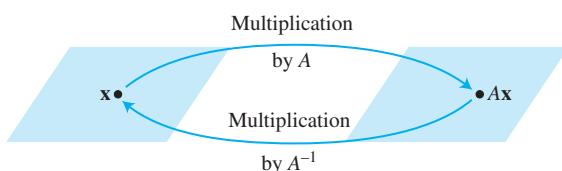


FIGURE 2 A^{-1} transforms $A\mathbf{x}$ back to \mathbf{x} .

A linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is said to be **invertible** if there exists a function $S : \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that

$$S(T(\mathbf{x})) = \mathbf{x} \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^n \quad (1)$$

$$T(S(\mathbf{x})) = \mathbf{x} \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^n \quad (2)$$

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 $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ 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 39. ■

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.

WEB

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 41–45 show that matrix computations can produce substantial error when a condition number is large.

PRACTICE PROBLEMS

- Determine if $A = \begin{bmatrix} 2 & 3 & 4 \\ 2 & 3 & 4 \\ 2 & 3 & 4 \end{bmatrix}$ is invertible.
- 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 $A\mathbf{x} = \mathbf{b}$?
- 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.
$$\begin{bmatrix} 0 & 3 & -5 \\ 1 & 0 & 2 \\ -4 & -9 & 7 \end{bmatrix}$$

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

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

8.
$$\begin{bmatrix} 1 & 3 & 7 & 4 \\ 0 & 5 & 9 & 6 \\ 0 & 0 & 2 & 8 \\ 0 & 0 & 0 & 10 \end{bmatrix}$$

9. [M]
$$\begin{bmatrix} 4 & 0 & -7 & -7 \\ -6 & 1 & 11 & 9 \\ 7 & -5 & 10 & 19 \\ -1 & 2 & 3 & -1 \end{bmatrix}$$

10. [M]
$$\begin{bmatrix} 5 & 3 & 1 & 7 & 9 \\ 6 & 4 & 2 & 8 & -8 \\ 7 & 5 & 3 & 10 & 9 \\ 9 & 6 & 4 & -9 & -5 \\ 8 & 5 & 2 & 11 & 4 \end{bmatrix}$$

In Exercises 11 and 12, 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. a. If the equation $Ax = \mathbf{0}$ has only the trivial solution, then A is row equivalent to the $n \times n$ identity matrix.
b. If the columns of A span \mathbb{R}^n , then the columns are linearly independent.
c. If A is an $n \times n$ matrix, then the equation $Ax = \mathbf{b}$ has at least one solution for each \mathbf{b} in \mathbb{R}^n .
d. If the equation $Ax = \mathbf{0}$ has a nontrivial solution, then A has fewer than n pivot positions.
e. If A^T is not invertible, then A is not invertible.
12. a. 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 $CA = I$.
b. If the columns of A are linearly independent, then the columns of A span \mathbb{R}^n .
c. If the equation $Ax = \mathbf{b}$ has at least one solution for each \mathbf{b} in \mathbb{R}^n , then the solution is unique for each \mathbf{b} .

- d. If the linear transformation $(\mathbf{x}) \mapsto A\mathbf{x}$ maps \mathbb{R}^n into \mathbb{R}^n , then A has n pivot positions.

- e. 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.

13. 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.
14. 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.
15. Can a square matrix with two identical columns be invertible? Why or why not?
16. Is it possible for a 5×5 matrix to be invertible when its columns do not span \mathbb{R}^5 ? Why or why not?
17. If A is invertible, then the columns of A^{-1} are linearly independent. Explain why.
18. 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?
19. 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?
20. If $n \times n$ matrices E and F have the property that $EF = I$, then E and F commute. Explain why.
21. 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?
22. 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?
23. If an $n \times n$ matrix K cannot be row reduced to I_n , what can you say about the columns of K ? Why?
24. 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?
25. Verify the boxed statement preceding Example 1.
26. Explain why the columns of A^2 span \mathbb{R}^n whenever the columns of A are linearly independent.
27. 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?]
28. Show that if AB is invertible, so is B .
29. 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.

30. 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.
31. 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.
32. 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 33 and 34, 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} .

33. $T(x_1, x_2) = (-5x_1 + 9x_2, 4x_1 - 7x_2)$

34. $T(x_1, x_2) = (6x_1 - 8x_2, -5x_1 + 7x_2)$

35. Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ 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.
36. 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?
37. 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?
38. Suppose a linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ 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?

39. Let $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ 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 $U(\mathbf{v})$.]
40. 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 $T(c\mathbf{x}) = cT(\mathbf{x})$.]

41. [M] Suppose an experiment leads to the following system of equations:

$$4.5x_1 + 3.1x_2 = 19.249 \quad (3)$$

$$1.6x_1 + 1.1x_2 = 6.843$$

- 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.

$$4.5x_1 + 3.1x_2 = 19.25 \quad (4)$$

$$1.6x_1 + 1.1x_2 = 6.84$$

- 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 42–44 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.

42. [M] 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} .
43. [M] Repeat Exercise 42 for the matrix in Exercise 10.
44. [M] 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 $(630, -12600, 56700, -88200, 44100)$.]

45. [M] 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.

SG

Mastering: Reviewing and Reflecting 2-13

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.

25. Without using row reduction, find the inverse of

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

26. [M] 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.

- Display the submatrix of A from rows 15 to 20 and columns 5 to 10.
- Insert a 5×10 matrix B into A , beginning at row 10 and column 20.
- 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 .]

27. [M] 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.

- Compute $A + B$.
- Compute AB .
- Solve $Ax = \mathbf{b}$ for some vector \mathbf{b} in \mathbb{R}^{50} , assuming that A can be partitioned into a 2×2 block matrix $\begin{bmatrix} A_{ij} \end{bmatrix}$, 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:

$$Ax = \mathbf{b}_1, \quad Ax = \mathbf{b}_2, \quad \dots, \quad Ax = \mathbf{b}_p \quad (1)$$

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.

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ * & 1 & 0 & 0 \\ * & * & 1 & 0 \\ * & * & * & 1 \end{bmatrix} \begin{matrix} \blacksquare & * & * & * & * \\ 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & \blacksquare & * \\ 0 & 0 & 0 & 0 & 0 \end{matrix}$$

L
 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

$$\begin{aligned} L\mathbf{y} &= \mathbf{b} \\ U\mathbf{x} &= \mathbf{y} \end{aligned} \quad (2)$$

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.

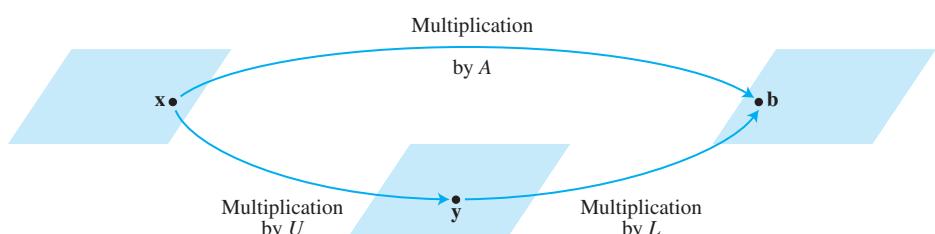


FIGURE 2 Factorization of the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

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 $Ly = \mathbf{b}$ 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.)

$$[L \quad \mathbf{b}] = \begin{bmatrix} 1 & 0 & 0 & 0 & -9 \\ -1 & 1 & 0 & 0 & 5 \\ 2 & -5 & 1 & 0 & 7 \\ -3 & 8 & 3 & 1 & 11 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 & -9 \\ 0 & 1 & 0 & 0 & -4 \\ 0 & 0 & 1 & 0 & 5 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = [I \quad \mathbf{y}]$$

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 $[U \quad \mathbf{y}]$ requires 1 division in row 4 and 3 multiplication-addition pairs to add multiples of row 4 to the rows above.)

$$[U \quad \mathbf{y}] = \begin{bmatrix} 3 & -7 & -2 & 2 & -9 \\ 0 & -2 & -1 & 2 & -4 \\ 0 & 0 & -1 & 1 & 5 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 & -6 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 3 \\ 4 \\ -6 \\ -1 \end{bmatrix}$$

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 $[A \quad \mathbf{b}]$ to $[I \quad \mathbf{x}]$ 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, \dots, 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).]

$$\begin{aligned} A &= \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ -4 & -5 & 3 & -8 & 1 \\ 2 & -5 & -4 & 1 & 8 \\ -6 & 0 & 7 & -3 & 1 \end{bmatrix} \sim \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & 3 & 1 & 2 & -3 \\ 0 & -9 & -3 & -4 & 10 \\ 0 & 12 & 4 & 12 & -5 \end{bmatrix} = A_1 \quad (5) \\ &\sim A_2 = \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & 3 & 1 & 2 & -3 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 4 & 7 \end{bmatrix} \sim \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & 3 & 1 & 2 & -3 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix} = U \end{aligned}$$

The highlighted entries on page 128 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 :

$$\begin{bmatrix} 2 \\ -4 \\ 2 \\ -6 \end{bmatrix} \begin{bmatrix} 3 \\ -9 \\ 12 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}$$

$$\begin{array}{cccc} \div 2 & \div 3 & \div 2 & \div 5 \\ \downarrow & \downarrow & \downarrow & \downarrow \end{array}$$

$$\begin{bmatrix} 1 \\ -2 \\ 1 \\ -3 \end{bmatrix}, \quad \text{and} \quad L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 1 & -3 & 1 & 0 \\ -3 & 4 & 2 & 1 \end{bmatrix}$$

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 $[L \ \mathbf{b}]$ to $[I \ \mathbf{y}]$ 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*.

SG

Permuted LU
Factorizations 2-23

WEB

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 $[A \ \mathbf{b}]$), whereas finding A^{-1} requires about $2n^3$ flops.
2. Solving $Ly = \mathbf{b}$ and $Ux = y$ 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

$$\begin{bmatrix} v_2 \\ i_2 \end{bmatrix} = A \begin{bmatrix} v_1 \\ i_1 \end{bmatrix}$$

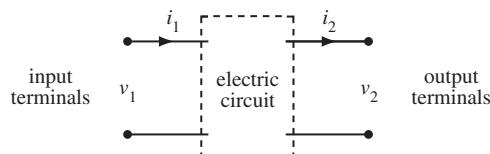


FIGURE 3 A circuit with input and output terminals.

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).

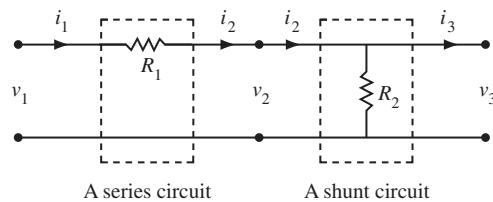


FIGURE 4 A ladder network.

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

$$\begin{bmatrix} 1 & -R_1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 \\ -1/R_2 & 1 \end{bmatrix}$$

Transfer matrix
of series circuit

Transfer matrix
of shunt circuit

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)

$$A_2A_1 = \begin{bmatrix} 1 & 0 \\ -1/R_2 & 1 \end{bmatrix} \begin{bmatrix} 1 & -R_1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -R_1 \\ -1/R_2 & 1 + R_1/R_2 \end{bmatrix} \quad (6)$$

- 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 19 and 20 in Section 2.4 and Example 2 in Section 3.3.) 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. $\begin{bmatrix} 2 & 5 \\ -3 & -4 \end{bmatrix}$

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

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

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

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

8. $\begin{bmatrix} 6 & 9 \\ 4 & 5 \end{bmatrix}$

10. $\begin{bmatrix} -5 & 3 & 4 \\ 10 & -8 & -9 \\ 15 & 1 & 2 \end{bmatrix}$

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

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

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 $0 = B \cdot 0$.

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 $Q^T Q = I$. 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?

WEB

25. (Singular Value Decomposition) Suppose $A = UDV^T$, where U and V are $n \times n$ matrices with the property that $U^T U = I$ and $V^T V = I$, and where D is a diagonal matrix with positive numbers $\sigma_1, \dots, \sigma_n$ 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

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}$$

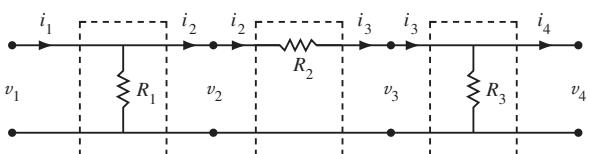
Show that this factorization is useful when computing high powers of A . Find fairly simple formulas for A^2 , A^3 , and A^k (k a positive integer), using P and the entries in D .

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.

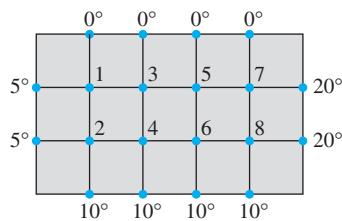
b. Let $A = \begin{bmatrix} 4/3 & -12 \\ -1/4 & 3 \end{bmatrix}$. Design a ladder network whose transfer matrix is A by finding a suitable matrix factorization of A .



30. Find a different factorization of the A in Exercise 29, and thereby design a different ladder network whose transfer matrix is A .

31. [M] 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

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

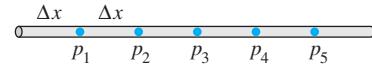
WEB

(Refer to Exercise 33 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.

32. [M] 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, \dots, 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 $k = 0, 1, 2, \dots$, 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 $A\mathbf{t}_{k+1} = \mathbf{t}_k$ ($k = 0, 1, \dots$), 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

$$\begin{aligned} A &= \begin{bmatrix} 2 & -4 & -2 & 3 \\ 6 & -9 & -5 & 8 \\ 2 & -7 & -3 & 9 \\ 4 & -2 & -2 & -1 \\ -6 & 3 & 3 & 4 \end{bmatrix} \sim \begin{bmatrix} 2 & -4 & -2 & 3 \\ 0 & 3 & 1 & -1 \\ 0 & -3 & -1 & 6 \\ 0 & 6 & 2 & -7 \\ 0 & -9 & -3 & 13 \end{bmatrix} \\ &\sim \begin{bmatrix} 2 & -4 & -2 & 3 \\ 0 & 3 & 1 & -1 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & -5 \\ 0 & 0 & 0 & 10 \end{bmatrix} \sim \begin{bmatrix} 2 & -4 & -2 & 3 \\ 0 & 3 & 1 & -1 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = U \end{aligned}$$

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.

$$\left[\begin{array}{c} 2 \\ 6 \\ 2 \\ 4 \\ -6 \end{array} \right] \left[\begin{array}{c} 3 \\ -3 \\ 6 \\ -9 \end{array} \right] \left[\begin{array}{c} 5 \\ -5 \\ 10 \end{array} \right] \xrightarrow{\begin{array}{c} \div 2 \\ \downarrow \\ \div 3 \\ \downarrow \\ \div 5 \end{array}} \left[\begin{array}{ccccc} 1 & & & & \\ 3 & 1 & & & \\ 1 & -1 & 1 & \dots & \\ 2 & 2 & -1 & & \\ -3 & -3 & 2 & & \end{array} \right], \quad L = \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 \\ 2 & 2 & -1 & 1 & 0 \\ -3 & -3 & 2 & 0 & 1 \end{array} \right]$$

2.6 THE LEONTIEF INPUT-OUTPUT MODEL

WEB

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

$$\left\{ \begin{array}{c} \text{amount} \\ \text{produced} \\ \mathbf{x} \end{array} \right\} = \left\{ \begin{array}{c} \text{intermediate} \\ \text{demand} \end{array} \right\} + \left\{ \begin{array}{c} \text{final} \\ \text{demand} \\ \mathbf{d} \end{array} \right\} \quad (1)$$

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.

SOLUTION TO PRACTICE PROBLEM

Assemble the matrices right-to-left for the three operations. Using $\mathbf{p} = (-2, 6)$, $\cos(-30^\circ) = \sqrt{3}/2$, and $\sin(-30^\circ) = -0.5$, we have

$$\begin{array}{c}
 \text{Translate} \\
 \text{back by } p \\
 \left[\begin{array}{ccc} 1 & 0 & -2 \\ 0 & 1 & 6 \\ 0 & 0 & 1 \end{array} \right]
 \end{array}
 \begin{array}{c}
 \text{Rotate around} \\
 \text{the origin} \\
 \left[\begin{array}{ccc} \sqrt{3}/2 & 1/2 & 0 \\ -1/2 & \sqrt{3}/2 & 0 \\ 0 & 0 & 1 \end{array} \right]
 \end{array}
 \begin{array}{c}
 \text{Translate} \\
 \text{by } -p \\
 \left[\begin{array}{ccc} 1 & 0 & 2 \\ 0 & 1 & -6 \\ 0 & 0 & 1 \end{array} \right]
 \end{array}
 \\
 = \left[\begin{array}{ccc} \sqrt{3}/2 & 1/2 & \sqrt{3}-5 \\ -1/2 & \sqrt{3}/2 & -3\sqrt{3}+5 \\ 0 & 0 & 1 \end{array} \right]$$

2.8 SUBSPACES OF \mathbb{R}^n

This section focuses on important sets of vectors in \mathbb{R}^n called *subspaces*. Often subspaces arise in connection with some matrix A , and they provide useful information about the equation $A\mathbf{x} = \mathbf{b}$. The concepts and terminology in this section will be used repeatedly throughout the rest of the book.¹

DEFINITION

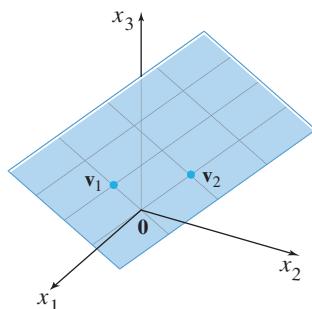


FIGURE 1

$\text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ as a plane through the origin.

A **subspace** of \mathbb{R}^n is any set H in \mathbb{R}^n that has three properties:

- a. The zero vector is in H .
- b. For each \mathbf{u} and \mathbf{v} in H , the sum $\mathbf{u} + \mathbf{v}$ is in H .
- c. For each \mathbf{u} in H and each scalar c , the vector $c\mathbf{u}$ is in H .

In words, a subspace is *closed* under addition and scalar multiplication. As you will see in the next few examples, most sets of vectors discussed in Chapter 1 are subspaces. For instance, a plane through the origin is the standard way to visualize the subspace in Example 1. See Figure 1.

EXAMPLE 1 If \mathbf{v}_1 and \mathbf{v}_2 are in \mathbb{R}^n and $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$, then H is a subspace of \mathbb{R}^n . To verify this statement, note that the zero vector is in H (because $0\mathbf{v}_1 + 0\mathbf{v}_2$ is a linear combination of \mathbf{v}_1 and \mathbf{v}_2). Now take two arbitrary vectors in H , say,

$$\mathbf{u} = s_1\mathbf{v}_1 + s_2\mathbf{v}_2 \quad \text{and} \quad \mathbf{v} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2$$

Then

$$\mathbf{u} + \mathbf{v} = (s_1 + t_1)\mathbf{v}_1 + (s_2 + t_2)\mathbf{v}_2$$

which shows that $\mathbf{u} + \mathbf{v}$ is a linear combination of \mathbf{v}_1 and \mathbf{v}_2 and hence is in H . Also, for any scalar c , the vector $c\mathbf{u}$ is in H , because $c\mathbf{u} = c(s_1\mathbf{v}_1 + s_2\mathbf{v}_2) = (cs_1)\mathbf{v}_1 + (cs_2)\mathbf{v}_2$. ■

If \mathbf{v}_1 is not zero and if \mathbf{v}_2 is a multiple of \mathbf{v}_1 , then \mathbf{v}_1 and \mathbf{v}_2 simply span a *line* through the origin. So a line through the origin is another example of a subspace.

¹ Sections 2.8 and 2.9 are included here to permit readers to postpone the study of most or all of the next two chapters and to skip directly to Chapter 5, if so desired. *Omit* these two sections if you plan to work through Chapter 4 before beginning Chapter 5.

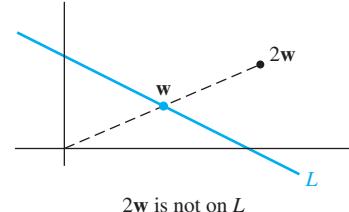
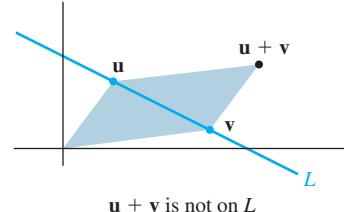
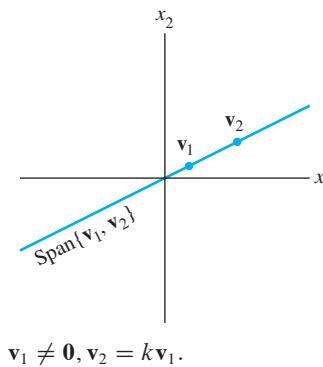


FIGURE 2

EXAMPLE 2 A line L not through the origin is *not* a subspace, because it does not contain the origin, as required. Also, Figure 2 shows that L is not closed under addition or scalar multiplication. ■

Note that \mathbb{R}^n is a subspace of itself because it has the three properties required for a subspace. Another special subspace is the set consisting of only the zero vector in \mathbb{R}^n . This set, called the **zero subspace**, also satisfies the conditions for a subspace.

Column Space and Null Space of a Matrix

Subspaces of \mathbb{R}^n usually occur in applications and theory in one of two ways. In both cases, the subspace can be related to a matrix.

DEFINITION

The **column space** of a matrix A is the set $\text{Col } A$ of all linear combinations of the columns of A .

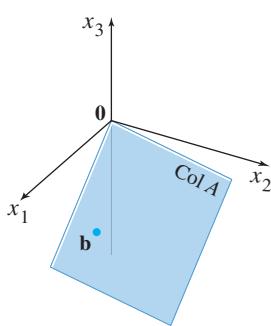
If $A = [a_1 \ \dots \ a_n]$, with the columns in \mathbb{R}^m , then $\text{Col } A$ is the same as $\text{Span}\{a_1, \dots, a_n\}$. Example 4 shows that the **column space of an $m \times n$ matrix is a subspace of \mathbb{R}^m** . Note that $\text{Col } A$ equals \mathbb{R}^m only when the columns of A span \mathbb{R}^m . Otherwise, $\text{Col } A$ is only part of \mathbb{R}^m .

EXAMPLE 4 Let $A = \begin{bmatrix} 1 & -3 & -4 \\ -4 & 6 & -2 \\ -3 & 7 & 6 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 3 \\ 3 \\ -4 \end{bmatrix}$. Determine whether \mathbf{b} is in the column space of A .

SOLUTION The vector \mathbf{b} is a linear combination of the columns of A if and only if \mathbf{b} can be written as $A\mathbf{x}$ for some \mathbf{x} , that is, if and only if the equation $A\mathbf{x} = \mathbf{b}$ has a solution. Row reducing the augmented matrix $[A \ \mathbf{b}]$,

$$\left[\begin{array}{ccc|c} 1 & -3 & -4 & 3 \\ -4 & 6 & -2 & 3 \\ -3 & 7 & 6 & -4 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & -3 & -4 & 3 \\ 0 & -6 & -18 & 15 \\ 0 & -2 & -6 & 5 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & -3 & -4 & 3 \\ 0 & -6 & -18 & 15 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

we conclude that $A\mathbf{x} = \mathbf{b}$ is consistent and \mathbf{b} is in $\text{Col } A$. ■



The solution of Example 4 shows that when a system of linear equations is written in the form $A\mathbf{x} = \mathbf{b}$, the column space of A is the set of all \mathbf{b} for which the system has a solution.

DEFINITION

The **null space** of a matrix A is the set $\text{Nul } A$ of all solutions of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

When A has n columns, the solutions of $A\mathbf{x} = \mathbf{0}$ belong to \mathbb{R}^n , and the null space of A is a subset of \mathbb{R}^n . In fact, $\text{Nul } A$ has the properties of a *subspace* of \mathbb{R}^n .

THEOREM 12

The null space of an $m \times n$ matrix A is a subspace of \mathbb{R}^n . Equivalently, the set of all solutions of a system $A\mathbf{x} = \mathbf{0}$ of m homogeneous linear equations in n unknowns is a subspace of \mathbb{R}^n .

PROOF The zero vector is in $\text{Nul } A$ (because $A\mathbf{0} = \mathbf{0}$). To show that $\text{Nul } A$ satisfies the other two properties required for a subspace, take any \mathbf{u} and \mathbf{v} in $\text{Nul } A$. That is, suppose $A\mathbf{u} = \mathbf{0}$ and $A\mathbf{v} = \mathbf{0}$. Then, by a property of matrix multiplication,

$$A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

Thus $\mathbf{u} + \mathbf{v}$ satisfies $A\mathbf{x} = \mathbf{0}$, and so $\mathbf{u} + \mathbf{v}$ is in $\text{Nul } A$. Also, for any scalar c , $A(c\mathbf{u}) = c(A\mathbf{u}) = c(\mathbf{0}) = \mathbf{0}$, which shows that $c\mathbf{u}$ is in $\text{Nul } A$. ■

To test whether a given vector \mathbf{v} is in $\text{Nul } A$, just compute $A\mathbf{v}$ to see whether $A\mathbf{v}$ is the zero vector. Because $\text{Nul } A$ is described by a condition that must be checked for each vector, we say that the null space is defined *implicitly*. In contrast, the column space is defined *explicitly*, because vectors in $\text{Col } A$ can be constructed (by linear combinations) from the columns of A . To create an explicit description of $\text{Nul } A$, solve the equation $A\mathbf{x} = \mathbf{0}$ and write the solution in parametric vector form. (See Example 6, below.)²

Basis for a Subspace

Because a subspace typically contains an infinite number of vectors, some problems involving a subspace are handled best by working with a small finite set of vectors that span the subspace. The smaller the set, the better. It can be shown that the smallest possible spanning set must be linearly independent.

DEFINITION

A **basis** for a subspace H of \mathbb{R}^n is a linearly independent set in H that spans H .

EXAMPLE 5 The columns of an invertible $n \times n$ matrix form a basis for all of \mathbb{R}^n because they are linearly independent and span \mathbb{R}^n , by the Invertible Matrix Theorem. One such matrix is the $n \times n$ identity matrix. Its columns are denoted by $\mathbf{e}_1, \dots, \mathbf{e}_n$:

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

The set $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is called the **standard basis** for \mathbb{R}^n . See Figure 3. ■

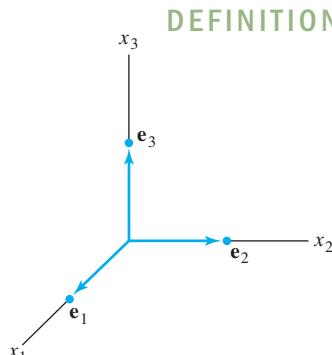


FIGURE 3

The standard basis for \mathbb{R}^3 .

²The contrast between $\text{Nul } A$ and $\text{Col } A$ is discussed further in Section 4.2.

The next example shows that the standard procedure for writing the solution set of $A\mathbf{x} = \mathbf{0}$ in parametric vector form actually identifies a basis for $\text{Nul } A$. This fact will be used throughout Chapter 5.

EXAMPLE 6 Find a basis 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 First, write the solution of $A\mathbf{x} = \mathbf{0}$ in parametric vector form:

$$[A \quad \mathbf{0}] \sim \begin{bmatrix} 1 & -2 & 0 & -1 & 3 & 0 \\ 0 & 0 & 1 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{aligned} x_1 - 2x_2 &= -x_4 + 3x_5 = 0 \\ x_3 + 2x_4 - 2x_5 &= 0 \\ 0 &= 0 \end{aligned}$$

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.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_4 + 2x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}$$

$\uparrow \quad \uparrow \quad \uparrow$

$$= x_2 \mathbf{u} + x_4 \mathbf{v} + x_5 \mathbf{w} \quad (1)$$

Equation (1) shows that $\text{Nul } A$ coincides with the set of all linear combinations of \mathbf{u} , \mathbf{v} , and \mathbf{w} . That is, $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ generates $\text{Nul } A$. In fact, this construction of \mathbf{u} , \mathbf{v} , and \mathbf{w} automatically makes them linearly independent, because equation (1) shows that $\mathbf{0} = x_2 \mathbf{u} + x_4 \mathbf{v} + x_5 \mathbf{w}$ only if the weights x_2 , x_4 , and x_5 are all zero. (Examine entries 2, 4, and 5 in the vector $x_2 \mathbf{u} + x_4 \mathbf{v} + x_5 \mathbf{w}$.) So $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a *basis* for $\text{Nul } A$. ■

Finding a basis for the column space of a matrix is actually less work than finding a basis for the null space. However, the method requires some explanation. Let's begin with a simple case.

EXAMPLE 7 Find a basis for the column space of the matrix

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

SOLUTION Denote the columns of B by $\mathbf{b}_1, \dots, \mathbf{b}_5$ and note that $\mathbf{b}_3 = -3\mathbf{b}_1 + 2\mathbf{b}_2$ and $\mathbf{b}_4 = 5\mathbf{b}_1 - \mathbf{b}_2$. The fact that \mathbf{b}_3 and \mathbf{b}_4 are combinations of the pivot columns means that any combination of $\mathbf{b}_1, \dots, \mathbf{b}_5$ is actually just a combination of \mathbf{b}_1 , \mathbf{b}_2 , and \mathbf{b}_5 . Indeed, if \mathbf{v} is any vector in $\text{Col } B$, say,

$$\mathbf{v} = c_1 \mathbf{b}_1 + c_2 \mathbf{b}_2 + c_3 \mathbf{b}_3 + c_4 \mathbf{b}_4 + c_5 \mathbf{b}_5$$

then, substituting for \mathbf{b}_3 and \mathbf{b}_4 , we can write \mathbf{v} in the form

$$\mathbf{v} = c_1 \mathbf{b}_1 + c_2 \mathbf{b}_2 + c_3(-3\mathbf{b}_1 + 2\mathbf{b}_2) + c_4(5\mathbf{b}_1 - \mathbf{b}_2) + c_5 \mathbf{b}_5$$

which is a linear combination of \mathbf{b}_1 , \mathbf{b}_2 , and \mathbf{b}_5 . So $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_5\}$ spans $\text{Col } B$. Also, \mathbf{b}_1 , \mathbf{b}_2 , and \mathbf{b}_5 are linearly independent, because they are columns from an identity matrix. So the pivot columns of B form a basis for $\text{Col } B$. ■

The matrix B in Example 7 is in reduced echelon form. To handle a general matrix A , recall that linear dependence relations among the columns of A can be expressed in the form $A\mathbf{x} = \mathbf{0}$ for some \mathbf{x} . (If some columns are not involved in a particular dependence relation, then the corresponding entries in \mathbf{x} are zero.) When A is row reduced to echelon form B , the columns are drastically changed, but the equations $A\mathbf{x} = \mathbf{0}$ and $B\mathbf{x} = \mathbf{0}$ 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 8 It can be verified that the matrix

$$A = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_5] = \begin{bmatrix} 1 & 3 & 3 & 2 & -9 \\ -2 & -2 & 2 & -8 & 2 \\ 2 & 3 & 0 & 7 & 1 \\ 3 & 4 & -1 & 11 & -8 \end{bmatrix}$$

is row equivalent to the matrix B in Example 7. Find a basis for $\text{Col } A$.

SOLUTION From Example 7, the pivot columns of A are columns 1, 2, and 5. Also, $\mathbf{b}_3 = -3\mathbf{b}_1 + 2\mathbf{b}_2$ and $\mathbf{b}_4 = 5\mathbf{b}_1 - \mathbf{b}_2$. Since row operations do not affect linear dependence relations among the columns of the matrix, we should have

$$\mathbf{a}_3 = -3\mathbf{a}_1 + 2\mathbf{a}_2 \quad \text{and} \quad \mathbf{a}_4 = 5\mathbf{a}_1 - \mathbf{a}_2$$

Check that this is true! By the argument in Example 7, \mathbf{a}_3 and \mathbf{a}_4 are not needed to generate the column space of A . Also, $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_5\}$ must be linearly independent, because any dependence relation among $\mathbf{a}_1, \mathbf{a}_2$, and \mathbf{a}_5 would imply the same dependence relation among $\mathbf{b}_1, \mathbf{b}_2$, and \mathbf{b}_5 . Since $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_5\}$ is linearly independent, $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_5\}$ is also linearly independent and hence is a basis for $\text{Col } A$. ■

The argument in Example 8 can be adapted to prove the following theorem.

THEOREM 13

The pivot columns of a matrix A form a basis for the column space of A .

Warning: Be careful to use *pivot columns of A itself* for the basis of $\text{Col } A$. The columns of an echelon form B are often not in the column space of A . (For instance, in Examples 7 and 8, the columns of B all have zeros in their last entries and cannot generate the columns of A .)

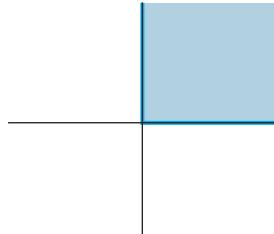
PRACTICE PROBLEMS

- Let $A = \begin{bmatrix} 1 & -1 & 5 \\ 2 & 0 & 7 \\ -3 & -5 & -3 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} -7 \\ 3 \\ 2 \end{bmatrix}$. Is \mathbf{u} in $\text{Nul } A$? Is \mathbf{u} in $\text{Col } A$? Justify each answer.
- Given $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$, find a vector in $\text{Nul } A$ and a vector in $\text{Col } A$.
- Suppose an $n \times n$ matrix A is invertible. What can you say about $\text{Col } A$? About $\text{Nul } A$?

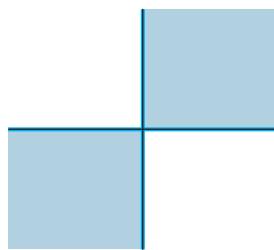
2.8 EXERCISES

Exercises 1–4 display sets in \mathbb{R}^2 . Assume the sets include the bounding lines. In each case, give a specific reason why the set H is *not* a subspace of \mathbb{R}^2 . (For instance, find two vectors in H whose sum is *not* in H , or find a vector in H with a scalar multiple that is not in H . Draw a picture.)

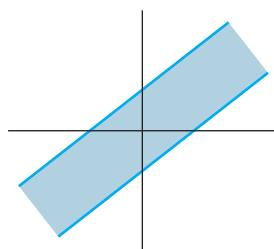
1.



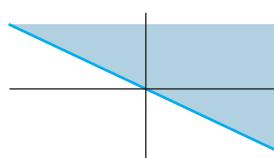
2.



3.



4.



5. Let $\mathbf{v}_1 = \begin{bmatrix} 2 \\ 3 \\ -5 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -4 \\ -5 \\ 8 \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} 8 \\ 2 \\ -9 \end{bmatrix}$. Determine if \mathbf{w} is in the subspace of \mathbb{R}^3 generated by \mathbf{v}_1 and \mathbf{v}_2 .

6. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \\ 4 \\ 3 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 4 \\ -7 \\ 9 \\ 7 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 5 \\ -8 \\ 6 \\ 5 \end{bmatrix}$, and $\mathbf{u} = \begin{bmatrix} -4 \\ 10 \\ -7 \\ -5 \end{bmatrix}$. Determine if \mathbf{u} is in the subspace of \mathbb{R}^4 generated by $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

7. Let $\mathbf{v}_1 = \begin{bmatrix} 2 \\ -8 \\ 6 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -3 \\ 8 \\ -7 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} -4 \\ 6 \\ -7 \end{bmatrix}$, $\mathbf{p} = \begin{bmatrix} 6 \\ -10 \\ 11 \end{bmatrix}$, and $A = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3]$.

- How many vectors are in $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$?
- How many vectors are in $\text{Col } A$?
- Is \mathbf{p} in $\text{Col } A$? Why or why not?

8. Let $\mathbf{v}_1 = \begin{bmatrix} -3 \\ 0 \\ 6 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -2 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 0 \\ -6 \\ 3 \end{bmatrix}$, and $\mathbf{p} = \begin{bmatrix} 1 \\ 14 \\ -9 \end{bmatrix}$. Determine if \mathbf{p} is in $\text{Col } A$, where $A = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3]$.

9. With A and \mathbf{p} as in Exercise 7, determine if \mathbf{p} is in $\text{Nul } A$.

10. With $\mathbf{u} = (-2, 3, 1)$ and A as in Exercise 8, determine if \mathbf{u} is in $\text{Nul } A$.

In Exercises 11 and 12, give integers p and q such that $\text{Nul } A$ is a subspace of \mathbb{R}^p and $\text{Col } A$ is a subspace of \mathbb{R}^q .

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

12. $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 7 \\ -5 & -1 & 0 \\ 2 & 7 & 11 \end{bmatrix}$

13. For A as in Exercise 11, find a nonzero vector in $\text{Nul } A$ and a nonzero vector in $\text{Col } A$.

14. For A as in Exercise 12, find a nonzero vector in $\text{Nul } A$ and a nonzero vector in $\text{Col } A$.

Determine which sets in Exercises 15–20 are bases for \mathbb{R}^2 or \mathbb{R}^3 . Justify each answer.

15. $\begin{bmatrix} 5 \\ -2 \end{bmatrix}, \begin{bmatrix} 10 \\ -3 \end{bmatrix}$

16. $\begin{bmatrix} -4 \\ 6 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \end{bmatrix}$

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

18. $\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \begin{bmatrix} -5 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 7 \\ 0 \\ -5 \end{bmatrix}$

19. $\begin{bmatrix} 3 \\ -8 \\ 1 \end{bmatrix}, \begin{bmatrix} 6 \\ 2 \\ -5 \end{bmatrix}$

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

In Exercises 21 and 22, mark each statement True or False. Justify each answer.

21. a. A subspace of \mathbb{R}^n is any set H such that (i) the zero vector 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 .
 - b. If $\mathbf{v}_1, \dots, \mathbf{v}_p$ are in \mathbb{R}^n , then $\text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is the same as the column space of the matrix $[\mathbf{v}_1 \ \dots \ \mathbf{v}_p]$.
 - c. The set of all solutions of a system of m homogeneous equations in n unknowns is a subspace of \mathbb{R}^m .
 - d. The columns of an invertible $n \times n$ matrix form a basis for \mathbb{R}^n .
 - e. Row operations do not affect linear dependence relations among the columns of a matrix.
-
22. a. A subset H of \mathbb{R}^n is a subspace if the zero vector is in H .
 - b. Given vectors $\mathbf{v}_1, \dots, \mathbf{v}_p$ in \mathbb{R}^n , the set of all linear combinations of these vectors is a subspace of \mathbb{R}^n .
 - c. The null space of an $m \times n$ matrix is a subspace of \mathbb{R}^n .
 - d. The column space of a matrix A is the set of solutions of $A\mathbf{x} = \mathbf{b}$.
 - e. If B is an echelon form of a matrix A , then the pivot columns of B form a basis for $\text{Col } A$.

Exercises 23–26 display a matrix A and an echelon form of A . Find a basis for $\text{Col } A$ and a basis for $\text{Nul } A$.

$$23. A = \begin{bmatrix} 4 & 5 & 9 & -2 \\ 6 & 5 & 1 & 12 \\ 3 & 4 & 8 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 6 & -5 \\ 0 & 1 & 5 & -6 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$24. A = \begin{bmatrix} -3 & 9 & -2 & -7 \\ 2 & -6 & 4 & 8 \\ 3 & -9 & -2 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 6 & 9 \\ 0 & 0 & 4 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$25. A = \begin{bmatrix} 1 & 4 & 8 & -3 & -7 \\ -1 & 2 & 7 & 3 & 4 \\ -2 & 2 & 9 & 5 & 5 \\ 3 & 6 & 9 & -5 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & 4 & 8 & 0 & 5 \\ 0 & 2 & 5 & 0 & -1 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$26. A = \begin{bmatrix} 3 & -1 & 7 & 3 & 9 \\ -2 & 2 & -2 & 7 & 5 \\ -5 & 9 & 3 & 3 & 4 \\ -2 & 6 & 6 & 3 & 7 \end{bmatrix} \sim \begin{bmatrix} 3 & -1 & 7 & 0 & 6 \\ 0 & 2 & 4 & 0 & 3 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

27. Construct a nonzero 3×3 matrix A and a nonzero vector \mathbf{b} such that \mathbf{b} is in $\text{Col } A$, but \mathbf{b} is not the same as any one of the columns of A .
28. Construct a nonzero 3×3 matrix A and a vector \mathbf{b} such that \mathbf{b} is *not* in $\text{Col } A$.
29. Construct a nonzero 3×3 matrix A and a nonzero vector \mathbf{b} such that \mathbf{b} is in $\text{Nul } A$.
30. Suppose the columns of a matrix $A = [\mathbf{a}_1 \ \dots \ \mathbf{a}_p]$ are linearly independent. Explain why $\{\mathbf{a}_1, \dots, \mathbf{a}_p\}$ is a basis for $\text{Col } A$.

In Exercises 31–36, respond as comprehensively as possible, and justify your answer.

31. Suppose F is a 5×5 matrix whose column space is not equal to \mathbb{R}^5 . What can you say about $\text{Nul } F$?
32. If R is a 6×6 matrix and $\text{Nul } R$ is *not* the zero subspace, what can you say about $\text{Col } R$?
33. If Q is a 4×4 matrix and $\text{Col } Q = \mathbb{R}^4$, what can you say about solutions of equations of the form $Q\mathbf{x} = \mathbf{b}$ for \mathbf{b} in \mathbb{R}^4 ?
34. If P is a 5×5 matrix and $\text{Nul } P$ is the zero subspace, what can you say about solutions of equations of the form $P\mathbf{x} = \mathbf{b}$ for \mathbf{b} in \mathbb{R}^5 ?
35. What can you say about $\text{Nul } B$ when B is a 5×4 matrix with linearly independent columns?
36. What can you say about the shape of an $m \times n$ matrix A when the columns of A form a basis for \mathbb{R}^m ?

[M] In Exercises 37 and 38, construct bases for the column space and the null space of the given matrix A . Justify your work.

$$37. A = \begin{bmatrix} 3 & -5 & 0 & -1 & 3 \\ -7 & 9 & -4 & 9 & -11 \\ -5 & 7 & -2 & 5 & -7 \\ 3 & -7 & -3 & 4 & 0 \end{bmatrix}$$

$$38. A = \begin{bmatrix} 5 & 2 & 0 & -8 & -8 \\ 4 & 1 & 2 & -8 & -9 \\ 5 & 1 & 3 & 5 & 19 \\ -8 & -5 & 6 & 8 & 5 \end{bmatrix}$$

WEB Column Space and Null Space

WEB A Basis for $\text{Col } A$

SOLUTIONS TO PRACTICE PROBLEMS

1. To determine whether \mathbf{u} is in $\text{Nul } A$, simply compute

$$A\mathbf{u} = \begin{bmatrix} 1 & -1 & 5 \\ 2 & 0 & 7 \\ -3 & -5 & -3 \end{bmatrix} \begin{bmatrix} -7 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

The result shows that \mathbf{u} is in $\text{Nul } A$. Deciding whether \mathbf{u} is in $\text{Col } A$ requires more work. Reduce the augmented matrix $[A \quad \mathbf{u}]$ to echelon form to determine whether the equation $A\mathbf{x} = \mathbf{u}$ is consistent:

$$\left[\begin{array}{ccc|c} 1 & -1 & 5 & -7 \\ 2 & 0 & 7 & 3 \\ -3 & -5 & -3 & 2 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & -1 & 5 & -7 \\ 0 & 2 & -3 & 17 \\ 0 & -8 & 12 & -19 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & -1 & 5 & -7 \\ 0 & 2 & -3 & 17 \\ 0 & 0 & 0 & 49 \end{array} \right]$$

The equation $A\mathbf{x} = \mathbf{u}$ has no solution, so \mathbf{u} is not in $\text{Col } A$.

2. In contrast to Practice Problem 1, finding a vector in $\text{Nul } A$ requires more work than testing whether a specified vector is in $\text{Nul } A$. However, since A is already in reduced echelon form, the equation $A\mathbf{x} = \mathbf{0}$ shows that if $\mathbf{x} = (x_1, x_2, x_3)$, then $x_2 = 0$, $x_3 = 0$, and x_1 is a free variable. Thus, a basis for $\text{Nul } A$ is $\mathbf{v} = (1, 0, 0)$. Finding just one vector in $\text{Col } A$ is trivial, since each column of A is in $\text{Col } A$. In this particular case, the same vector \mathbf{v} is in both $\text{Nul } A$ and $\text{Col } A$. For most $n \times n$ matrices, the zero vector of \mathbb{R}^n is the only vector in both $\text{Nul } A$ and $\text{Col } A$.
3. If A is invertible, then the columns of A span \mathbb{R}^n , by the Invertible Matrix Theorem. By definition, the columns of any matrix always span the column space, so in this case $\text{Col } A$ is all of \mathbb{R}^n . In symbols, $\text{Col } A = \mathbb{R}^n$. Also, since A is invertible, the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. This means that $\text{Nul } A$ is the zero subspace. In symbols, $\text{Nul } A = \{\mathbf{0}\}$.

2.9 DIMENSION AND RANK

This section continues the discussion of subspaces and bases for subspaces, beginning with the concept of a coordinate system. The definition and example below should make a useful new term, *dimension*, seem quite natural, at least for subspaces of \mathbb{R}^3 .

Coordinate Systems

The main reason for selecting a basis for a subspace H , instead of merely a spanning set, is that each vector in H can be written in only one way as a linear combination of the basis vectors. To see why, suppose $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ is a basis for H , and suppose a vector \mathbf{x} in H can be generated in two ways, say,

$$\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_p\mathbf{b}_p \quad \text{and} \quad \mathbf{x} = d_1\mathbf{b}_1 + \cdots + d_p\mathbf{b}_p \quad (1)$$

Then, subtracting gives

$$\mathbf{0} = \mathbf{x} - \mathbf{x} = (c_1 - d_1)\mathbf{b}_1 + \cdots + (c_p - d_p)\mathbf{b}_p \quad (2)$$

Since \mathcal{B} is linearly independent, the weights in (2) must all be zero. That is, $c_j = d_j$ for $1 \leq j \leq p$, which shows that the two representations in (1) are actually the same.

DEFINITION

Suppose the set $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ is a basis for a subspace H . For each \mathbf{x} in H , the **coordinates of \mathbf{x} relative to the basis \mathcal{B}** are the weights c_1, \dots, c_p such that $\mathbf{x} = c_1\mathbf{b}_1 + \dots + c_p\mathbf{b}_p$, and the vector in \mathbb{R}^p

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}$$

is called the **coordinate vector of \mathbf{x} (relative to \mathcal{B})** or the **\mathcal{B} -coordinate vector of \mathbf{x} .**¹

EXAMPLE 1 Let $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 3 \\ 12 \\ 7 \end{bmatrix}$, and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$. Then \mathcal{B} is a basis for $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ because \mathbf{v}_1 and \mathbf{v}_2 are linearly independent. 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} . Row operations show that

$$\left[\begin{array}{ccc} 3 & -1 & 3 \\ 6 & 0 & 12 \\ 2 & 1 & 7 \end{array} \right] \sim \left[\begin{array}{ccc} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{array} \right]$$

Thus $c_1 = 2$, $c_2 = 3$, and $[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. The basis \mathcal{B} determines a “coordinate system” on H , which can be visualized by the grid shown in Figure 1. ■

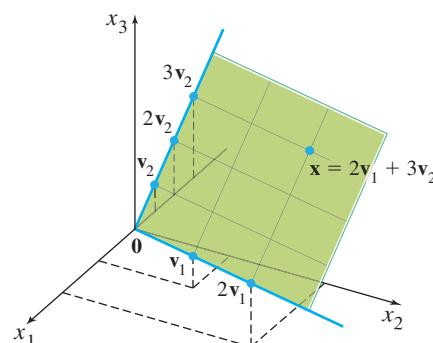


FIGURE 1 A coordinate system on a plane H in \mathbb{R}^3 .

¹ It is important that the elements of \mathcal{B} are numbered because the entries in $[\mathbf{x}]_{\mathcal{B}}$ depend on the order of the vectors in \mathcal{B} .

Notice that although points in H are also in \mathbb{R}^3 , they are completely determined by their coordinate vectors, which belong to \mathbb{R}^2 . The grid on the plane in Figure 1 makes H “look” like \mathbb{R}^2 . The correspondence $\mathbf{x} \mapsto [\mathbf{x}]_{\mathcal{B}}$ is a one-to-one correspondence between H and \mathbb{R}^2 that preserves linear combinations. We call such a correspondence an *isomorphism*, and we say that H is *isomorphic* to \mathbb{R}^2 .

In general, if $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ is a basis for H , then the mapping $\mathbf{x} \mapsto [\mathbf{x}]_{\mathcal{B}}$ is a one-to-one correspondence that makes H look and act the same as \mathbb{R}^p (even though the vectors in H themselves may have more than p entries). (Section 4.4 has more details.)

The Dimension of a Subspace

It can be shown that if a subspace H has a basis of p vectors, then every basis of H must consist of exactly p vectors. (See Exercises 27 and 28.) Thus the following definition makes sense.

DEFINITION

The **dimension** of a nonzero subspace H , denoted by $\dim H$, is the number of vectors in any basis for H . The dimension of the zero subspace $\{\mathbf{0}\}$ is defined to be zero.²

The space \mathbb{R}^n has dimension n . Every basis for \mathbb{R}^n consists of n vectors. A plane through $\mathbf{0}$ in \mathbb{R}^3 is two-dimensional, and a line through $\mathbf{0}$ is one-dimensional.

EXAMPLE 2 Recall that the null space of the matrix A in Example 6 in Section 2.8 had a basis of 3 vectors. So the dimension of $\text{Nul } A$ in this case is 3. Observe how each basis vector corresponds to a free variable in the equation $A\mathbf{x} = \mathbf{0}$. Our construction always produces a basis in this way. So, to find the dimension of $\text{Nul } A$, simply identify and count the number of free variables in $A\mathbf{x} = \mathbf{0}$. ■

DEFINITION

The **rank** of a matrix A , denoted by $\text{rank } A$, is the dimension of the column space of A .

Since the pivot columns of A form a basis for $\text{Col } A$, the rank of A is just the number of pivot columns in A .

EXAMPLE 3 Determine the rank of the matrix

$$A = \begin{bmatrix} 2 & 5 & -3 & -4 & 8 \\ 4 & 7 & -4 & -3 & 9 \\ 6 & 9 & -5 & 2 & 4 \\ 0 & -9 & 6 & 5 & -6 \end{bmatrix}$$

SOLUTION Reduce A to echelon form:

$$A \sim \left[\begin{array}{ccccc} 2 & 5 & -3 & -4 & 8 \\ 0 & -3 & 2 & 5 & -7 \\ 0 & -6 & 4 & 14 & -20 \\ 0 & -9 & 6 & 5 & -6 \end{array} \right] \sim \cdots \sim \left[\begin{array}{ccccc} 2 & 5 & -3 & -4 & 8 \\ 0 & -3 & 2 & 5 & -7 \\ 0 & 0 & 0 & 4 & -6 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

↑ ↑ ↑

Pivot columns

The matrix A has 3 pivot columns, so $\text{rank } A = 3$. ■

²The zero subspace has *no* basis (because the zero vector by itself forms a linearly dependent set).

The row reduction in Example 3 reveals that there are two free variables in $Ax = \mathbf{0}$, because two of the five columns of A are *not* pivot columns. (The nonpivot columns correspond to the free variables in $Ax = \mathbf{0}$.) Since the number of pivot columns plus the number of nonpivot columns is exactly the number of columns, the dimensions of $\text{Col } A$ and $\text{Nul } A$ have the following useful connection. (See the Rank Theorem in Section 4.6 for additional details.)

THEOREM 14

The Rank Theorem

If a matrix A has n columns, then $\text{rank } A + \dim \text{Nul } A = n$.

The following theorem is important for applications and will be needed in Chapters 5 and 6. The theorem (proved in Section 4.5) is certainly plausible, if you think of a p -dimensional subspace as isomorphic to \mathbb{R}^p . The Invertible Matrix Theorem shows that p vectors in \mathbb{R}^p are linearly independent if and only if they also span \mathbb{R}^p .

THEOREM 15

The Basis Theorem

Let H be a p -dimensional subspace of \mathbb{R}^n . Any linearly independent set of exactly p elements in H is automatically a basis for H . Also, any set of p elements of H that spans H is automatically a basis for H .

Rank and the Invertible Matrix Theorem

The various vector space concepts associated with a matrix provide several more statements for the Invertible Matrix Theorem. They are presented below to follow the statements in the original theorem in Section 2.3.

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. $\text{Col } A = \mathbb{R}^n$
- o. $\dim \text{Col } A = n$
- p. $\text{rank } A = n$
- q. $\text{Nul } A = \{\mathbf{0}\}$
- r. $\dim \text{Nul } A = 0$

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:

$$(g) \Rightarrow (n) \Rightarrow (o) \Rightarrow (p) \Rightarrow (r) \Rightarrow (q) \Rightarrow (d)$$

Statement (g), which says that the equation $Ax = \mathbf{b}$ has at least one solution for each \mathbf{b} in \mathbb{R}^n , implies statement (n), because $\text{Col } A$ is precisely the set of all \mathbf{b} such that the equation $Ax = \mathbf{b}$ is consistent. The implications $(n) \Rightarrow (o) \Rightarrow (p)$ follow from the definitions of *dimension* and *rank*. If the rank of A is n , the number of columns of A , then $\dim \text{Nul } A = 0$, by the Rank Theorem, and so $\text{Nul } A = \{\mathbf{0}\}$. Thus $(p) \Rightarrow (r) \Rightarrow (q)$.

SGExpanded Table
for the IMT 2-39

Also, statement (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. ■

WEB**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.

PRACTICE PROBLEMS

1. Determine the dimension of the subspace H of \mathbb{R}^3 spanned by the vectors \mathbf{v}_1 , \mathbf{v}_2 , and \mathbf{v}_3 . (First, find a basis for H .)

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ -8 \\ 6 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 3 \\ -7 \\ -1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -1 \\ 6 \\ -7 \end{bmatrix}$$

2. Consider the basis

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ .2 \\ .2 \end{bmatrix}, \begin{bmatrix} .2 \\ 1 \\ 1 \end{bmatrix} \right\}$$

for \mathbb{R}^3 . If $[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$, what is \mathbf{x} ?

3. Could \mathbb{R}^3 possibly contain a four-dimensional subspace? Explain.

2.9 EXERCISES

In Exercises 1 and 2, find the vector \mathbf{x} determined by the given coordinate vector $[\mathbf{x}]_{\mathcal{B}}$ and the given basis \mathcal{B} . Illustrate your answer with a figure, as in the solution of Practice Problem 2.

1. $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\}, [\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

2. $\mathcal{B} = \left\{ \begin{bmatrix} -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}, [\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$

In Exercises 3–6, the vector \mathbf{x} is in a subspace H with a basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$. Find the \mathcal{B} -coordinate vector of \mathbf{x} .

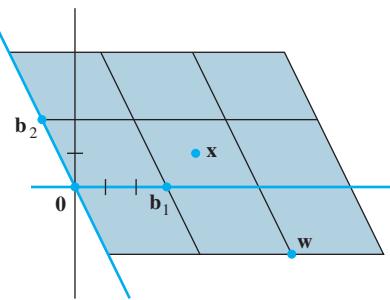
3. $\mathbf{b}_1 = \begin{bmatrix} 1 \\ -4 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} -2 \\ 7 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} -3 \\ 7 \end{bmatrix}$

4. $\mathbf{b}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} -3 \\ 5 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} -7 \\ 5 \end{bmatrix}$

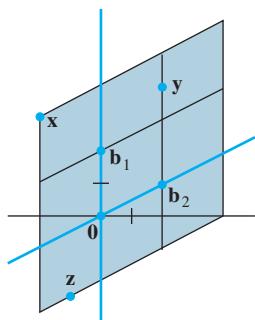
5. $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 5 \\ -3 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} -3 \\ -7 \\ 5 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 4 \\ 10 \\ -7 \end{bmatrix}$

6. $\mathbf{b}_1 = \begin{bmatrix} -3 \\ 1 \\ -4 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 7 \\ 5 \\ -6 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 11 \\ 0 \\ 7 \end{bmatrix}$

7. Let $\mathbf{b}_1 = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} 7 \\ -2 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$, and $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$. Use the figure to estimate $[\mathbf{w}]_{\mathcal{B}}$ and $[\mathbf{x}]_{\mathcal{B}}$. Confirm your estimate of $[\mathbf{x}]_{\mathcal{B}}$ by using it and $\{\mathbf{b}_1, \mathbf{b}_2\}$ to compute \mathbf{x} .



8. Let $\mathbf{b}_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$, $\mathbf{z} = \begin{bmatrix} -1 \\ -2.5 \end{bmatrix}$, and $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$. Use the figure to estimate $[\mathbf{x}]_{\mathcal{B}}$, $[\mathbf{y}]_{\mathcal{B}}$, and $[\mathbf{z}]_{\mathcal{B}}$. Confirm your estimates of $[\mathbf{y}]_{\mathcal{B}}$ and $[\mathbf{z}]_{\mathcal{B}}$ by using them and $\{\mathbf{b}_1, \mathbf{b}_2\}$ to compute \mathbf{y} and \mathbf{z} .



Exercises 9–12 display a matrix A and an echelon form of A . Find bases for $\text{Col } A$ and $\text{Nul } A$, and then state the dimensions of these subspaces.

$$9. A = \begin{bmatrix} 1 & -3 & 2 & -4 \\ -3 & 9 & -1 & 5 \\ 2 & -6 & 4 & -3 \\ -4 & 12 & 2 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 2 & -4 \\ 0 & 0 & 5 & -7 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

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

$$11. A = \begin{bmatrix} 1 & 2 & -5 & 0 & -1 \\ 2 & 5 & -8 & 4 & 3 \\ -3 & -9 & 9 & -7 & -2 \\ 3 & 10 & -7 & 11 & 7 \end{bmatrix}$$

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

$$12. A = \begin{bmatrix} 1 & 2 & -4 & 3 & 3 \\ 5 & 10 & -9 & -7 & 8 \\ 4 & 8 & -9 & -2 & 7 \\ -2 & -4 & 5 & 0 & -6 \end{bmatrix}$$

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

In Exercises 13 and 14, find a basis for the subspace spanned by the given vectors. What is the dimension of the subspace?

$$13. \begin{bmatrix} 1 \\ -3 \\ 2 \\ -4 \end{bmatrix}, \begin{bmatrix} -3 \\ 9 \\ -6 \\ 12 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 4 \\ 2 \end{bmatrix}, \begin{bmatrix} -4 \\ 5 \\ -3 \\ 7 \end{bmatrix}$$

$$14. \begin{bmatrix} 1 \\ -1 \\ -2 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ -1 \\ 6 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ -6 \\ 8 \end{bmatrix}, \begin{bmatrix} -1 \\ 4 \\ -7 \\ 7 \end{bmatrix}, \begin{bmatrix} 3 \\ -8 \\ 9 \\ -5 \end{bmatrix}$$

15. Suppose a 3×5 matrix A has three pivot columns. Is $\text{Col } A = \mathbb{R}^3$? Is $\text{Nul } A = \mathbb{R}^2$? Explain your answers.

16. Suppose a 4×7 matrix A has three pivot columns. Is $\text{Col } A = \mathbb{R}^3$? What is the dimension of $\text{Nul } A$? Explain your answers.

In Exercises 17 and 18, mark each statement True or False. Justify each answer. Here A is an $m \times n$ matrix.

17. a. If $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is a basis for a subspace H and if $\mathbf{x} = c_1\mathbf{v}_1 + \dots + c_p\mathbf{v}_p$, then c_1, \dots, c_p are the coordinates of \mathbf{x} relative to the basis \mathcal{B} .

- b. Each line in \mathbb{R}^n is a one-dimensional subspace of \mathbb{R}^n .

- c. The dimension of $\text{Col } A$ is the number of pivot columns of A .

- d. The dimensions of $\text{Col } A$ and $\text{Nul } A$ add up to the number of columns of A .

- e. If a set of p vectors spans a p -dimensional subspace H of \mathbb{R}^n , then these vectors form a basis for H .

18. a. If \mathcal{B} is a basis for a subspace H , then each vector in H can be written in only one way as a linear combination of the vectors in \mathcal{B} .

- b. If $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is a basis for a subspace H of \mathbb{R}^n , then the correspondence $\mathbf{x} \mapsto [\mathbf{x}]_{\mathcal{B}}$ makes H look and act the same as \mathbb{R}^p .

- c. The dimension of $\text{Nul } A$ is the number of variables in the equation $A\mathbf{x} = \mathbf{0}$.
- d. The dimension of the column space of A is $\text{rank } A$.
- e. If H is a p -dimensional subspace of \mathbb{R}^n , then a linearly independent set of p vectors in H is a basis for H .

In Exercises 19–24, justify each answer or construction.

19. If the subspace of all solutions of $A\mathbf{x} = \mathbf{0}$ has a basis consisting of three vectors and if A is a 5×7 matrix, what is the rank of A ?
20. What is the rank of a 4×5 matrix whose null space is three-dimensional?
21. If the rank of a 7×6 matrix A is 4, what is the dimension of the solution space of $A\mathbf{x} = \mathbf{0}$?
22. Show that a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_5\}$ in \mathbb{R}^n is linearly dependent when $\dim \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_5\} = 4$.
23. If possible, construct a 3×4 matrix A such that $\dim \text{Nul } A = 2$ and $\dim \text{Col } A = 2$.
24. Construct a 4×3 matrix with rank 1.
25. Let A be an $n \times p$ matrix whose column space is p -dimensional. Explain why the columns of A must be linearly independent.
26. Suppose columns 1, 3, 5, and 6 of a matrix A are linearly independent (but are not necessarily pivot columns) and the rank of A is 4. Explain why the four columns mentioned must be a basis for the column space of A .

27. Suppose vectors $\mathbf{b}_1, \dots, \mathbf{b}_p$ span a subspace W , and let $\{\mathbf{a}_1, \dots, \mathbf{a}_q\}$ be any set in W containing more than p vectors. Fill in the details of the following argument to show that $\{\mathbf{a}_1, \dots, \mathbf{a}_q\}$ must be linearly dependent. First, let $B = [\mathbf{b}_1 \ \dots \ \mathbf{b}_p]$ and $A = [\mathbf{a}_1 \ \dots \ \mathbf{a}_q]$.

- a. Explain why for each vector \mathbf{a}_j , there exists a vector \mathbf{c}_j in \mathbb{R}^p such that $\mathbf{a}_j = B\mathbf{c}_j$.
- b. Let $C = [\mathbf{c}_1 \ \dots \ \mathbf{c}_q]$. Explain why there is a nonzero vector \mathbf{u} such that $C\mathbf{u} = \mathbf{0}$.
- c. Use B and C to show that $A\mathbf{u} = \mathbf{0}$. This shows that the columns of A are linearly dependent.

28. Use Exercise 27 to show that if \mathcal{A} and \mathcal{B} are bases for a subspace W of \mathbb{R}^n , then \mathcal{A} cannot contain more vectors than \mathcal{B} , and, conversely, \mathcal{B} cannot contain more vectors than \mathcal{A} .

29. [M] 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} , when

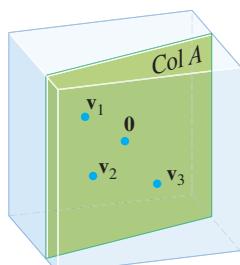
$$\mathbf{v}_1 = \begin{bmatrix} 11 \\ -5 \\ 10 \\ 7 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 14 \\ -8 \\ 13 \\ 10 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 19 \\ -13 \\ 18 \\ 15 \end{bmatrix}$$

30. [M] 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} , when

$$\mathbf{v}_1 = \begin{bmatrix} -6 \\ 4 \\ -9 \\ 4 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 8 \\ -3 \\ 7 \\ -3 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} -9 \\ 5 \\ -8 \\ 3 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 4 \\ 7 \\ -8 \\ 3 \end{bmatrix}$$

SG

Mastering: Dimension and Rank 2-41



SOLUTIONS TO PRACTICE PROBLEMS

1. Construct $A = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3]$ so that the subspace spanned by $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ is the column space of A . A basis for this space is provided by the pivot columns of A .

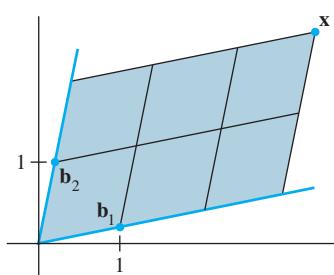
$$A = \begin{bmatrix} 2 & 3 & -1 \\ -8 & -7 & 6 \\ 6 & -1 & -7 \end{bmatrix} \sim \begin{bmatrix} 2 & 3 & -1 \\ 0 & 5 & 2 \\ 0 & -10 & -4 \end{bmatrix} \sim \begin{bmatrix} 2 & 3 & -1 \\ 0 & 5 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

The first two columns of A are pivot columns and form a basis for H . Thus $\dim H = 2$.

2. If $[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$, then \mathbf{x} is formed from a linear combination of the basis vectors using weights 3 and 2:

$$\mathbf{x} = 3\mathbf{b}_1 + 2\mathbf{b}_2 = 3 \begin{bmatrix} 1 \\ .2 \end{bmatrix} + 2 \begin{bmatrix} .2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3.4 \\ 2.6 \end{bmatrix}$$

The basis $\{\mathbf{b}_1, \mathbf{b}_2\}$ determines a *coordinate system* for \mathbb{R}^2 , illustrated by the grid in the figure. Note how \mathbf{x} is 3 units in the \mathbf{b}_1 -direction and 2 units in the \mathbf{b}_2 -direction.



3. A four-dimensional subspace would contain a basis of four linearly independent vectors. This is impossible inside \mathbb{R}^3 . Since any linearly independent set in \mathbb{R}^3 has no more than three vectors, any subspace of \mathbb{R}^3 has dimension no more than 3. The space \mathbb{R}^3 itself is the only three-dimensional subspace of \mathbb{R}^3 . Other subspaces of \mathbb{R}^3 have dimension 2, 1, or 0.

CHAPTER 2 SUPPLEMENTARY EXERCISES

1. Assume that the matrices mentioned in the statements below have appropriate sizes. Mark each statement True or False. Justify each answer.
 - a. If A and B are $m \times n$, then both AB^T and A^TB are defined.
 - b. If $AB = C$ and C has 2 columns, then A has 2 columns.
 - c. Left-multiplying a matrix B by a diagonal matrix A , with nonzero entries on the diagonal, scales the rows of B .
 - d. If $BC = BD$, then $C = D$.
 - e. If $AC = 0$, then either $A = 0$ or $C = 0$.
 - f. If A and B are $n \times n$, then $(A + B)(A - B) = A^2 - B^2$.
 - g. An elementary $n \times n$ matrix has either n or $n + 1$ nonzero entries.
 - h. The transpose of an elementary matrix is an elementary matrix.
 - i. An elementary matrix must be square.
 - j. Every square matrix is a product of elementary matrices.
 - k. If A is a 3×3 matrix with three pivot positions, there exist elementary matrices E_1, \dots, E_p such that $E_p \cdots E_1 A = I$.
 - l. If $AB = I$, then A is invertible.
 - m. If A and B are square and invertible, then AB is invertible, and $(AB)^{-1} = A^{-1}B^{-1}$.
 - n. If $AB = BA$ and if A is invertible, then $A^{-1}B = BA^{-1}$.
 - o. If A is invertible and if $r \neq 0$, then $(rA)^{-1} = rA^{-1}$.
 - p. If A is a 3×3 matrix and the equation $Ax = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ has a unique solution, then A is invertible.
2. Find the matrix C whose inverse is $C^{-1} = \begin{bmatrix} 4 & 5 \\ 6 & 7 \end{bmatrix}$.
3. Let $A = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. Show that $A^3 = 0$. Use matrix algebra to compute the product $(I - A)(I + A + A^2)$.
4. Suppose $A^n = 0$ for some $n > 1$. Find an inverse for $I - A$.
5. Suppose an $n \times n$ matrix A satisfies the equation $A^2 - 2A + I = 0$. Show that $A^3 = 3A - 2I$ and $A^4 = 4A - 3I$.
6. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. These are *Pauli spin matrices* used in the study of electron spin in quantum mechanics. Show that $A^2 = I$, $B^2 = I$, and $AB = -BA$. Matrices such that $AB = -BA$ are said to *anticommute*.
7. Let $A = \begin{bmatrix} 1 & 3 & 8 \\ 2 & 4 & 11 \\ 1 & 2 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} -3 & 5 \\ 1 & 5 \\ 3 & 4 \end{bmatrix}$. Compute $A^{-1}B$ without computing A^{-1} . [Hint: $A^{-1}B$ is the solution of the equation $AX = B$.]
8. Find a matrix A such that the transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 7 \end{bmatrix}$ into $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$, respectively. [Hint: Write a matrix equation involving A , and solve for A .]
9. Suppose $AB = \begin{bmatrix} 5 & 4 \\ -2 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 7 & 3 \\ 2 & 1 \end{bmatrix}$. Find A .
10. Suppose A is invertible. Explain why A^TA is also invertible. Then show that $A^{-1} = (A^TA)^{-1}A^T$.
11. Let x_1, \dots, x_n be fixed numbers. The matrix below, called a *Vandermonde matrix*, occurs in applications such as signal processing, error-correcting codes, and polynomial interpolation.

$$V = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix}$$

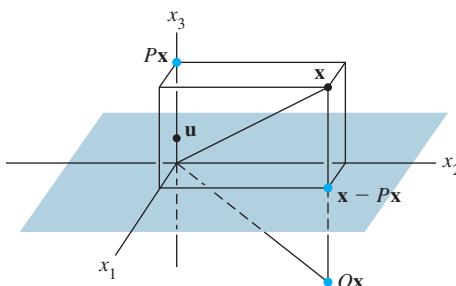
Given $\mathbf{y} = (y_1, \dots, y_n)$ in \mathbb{R}^n , suppose $\mathbf{c} = (c_0, \dots, c_{n-1})$ in \mathbb{R}^n satisfies $V\mathbf{c} = \mathbf{y}$, and define the polynomial $p(t) = c_0 + c_1t + c_2t^2 + \cdots + c_{n-1}t^{n-1}$. We call $p(t)$ an *interpolating polynomial for the points* $(x_1, y_1), \dots, (x_n, y_n)$ because the graph of $p(t)$ passes through the points.
12. Let $A = LU$, where L is an invertible lower triangular matrix and U is upper triangular. Explain why the first column

of A is a multiple of the first column of L . How is the second column of A related to the columns of L ?

13. Given \mathbf{u} in \mathbb{R}^n with $\mathbf{u}^T \mathbf{u} = 1$, let $P = \mathbf{u}\mathbf{u}^T$ (an outer product) and $Q = I - 2P$. Justify statements (a), (b), and (c).
- $P^2 = P$
 - $P^T = P$
 - $Q^2 = I$

The transformation $\mathbf{x} \mapsto P\mathbf{x}$ is called a *projection*, and $\mathbf{x} \mapsto Q\mathbf{x}$ is called a *Householder reflection*. Such reflections are used in computer programs to create multiple zeros in a vector (usually a column of a matrix).

14. Let $\mathbf{u} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} 1 \\ 5 \\ 3 \end{bmatrix}$. Determine P and Q as in Exercise 13, and compute $P\mathbf{x}$ and $Q\mathbf{x}$. The figure shows that $Q\mathbf{x}$ is the reflection of \mathbf{x} through the x_1x_2 -plane.



A Householder reflection through the plane $x_3 = 0$.

15. Suppose $C = E_3 E_2 E_1 B$, where E_1 , E_2 , and E_3 are elementary matrices. Explain why C is row equivalent to B .

16. Let A be an $n \times n$ singular matrix. Describe how to construct an $n \times n$ nonzero matrix B such that $AB = 0$.

17. Let A be a 6×4 matrix and B a 4×6 matrix. Show that the 6×6 matrix AB cannot be invertible.

18. Suppose A is a 5×3 matrix and there exists a 3×5 matrix C such that $CA = I_3$. Suppose further that for some given \mathbf{b} in \mathbb{R}^5 , the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution. Show that this solution is unique.

19. [M] Certain dynamical systems can be studied by examining powers of a matrix, such as those below. Determine what happens to A^k and B^k as k increases (for example, try $k = 2, \dots, 16$). Try to identify what is special about A and B . Investigate large powers of other matrices of this type, and make a conjecture about such matrices.

$$A = \begin{bmatrix} .4 & .2 & .3 \\ .3 & .6 & .3 \\ .3 & .2 & .4 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & .2 & .3 \\ .1 & .6 & .3 \\ .9 & .2 & .4 \end{bmatrix}$$

20. [M] Let A_n be the $n \times n$ matrix with 0's on the main diagonal and 1's elsewhere. Compute A_n^{-1} for $n = 4, 5$, and 6, and make a conjecture about the general form of A_n^{-1} for larger values of n .

5

Eigenvalues and Eigenvectors

INTRODUCTORY EXAMPLE

Dynamical Systems and Spotted Owls

In 1990, the northern spotted owl became the center of a nationwide controversy over the use and misuse of the majestic forests in the Pacific Northwest. Environmentalists convinced the federal government that the owl was threatened with extinction if logging continued in the old-growth forests (with trees more than 200 years old), where the owls prefer to live. The timber industry, anticipating the loss of 30,000 to 100,000 jobs as a result of new government restrictions on logging, argued that the owl should not be classified as a “threatened species” and cited a number of published scientific reports to support its case.¹

Caught in the crossfire of the two lobbying groups, mathematical ecologists intensified their drive to understand the population dynamics of the spotted owl. The life cycle of a spotted owl divides naturally into three stages: juvenile (up to 1 year old), subadult (1 to 2 years), and adult (older than 2 years). The owls mate for life during the subadult and adult stages, begin to breed as adults, and live for up to 20 years. Each owl pair requires about 1000 hectares (4 square miles) for its own home territory. A critical time in the life cycle is when the juveniles leave the nest. To survive and become a subadult, a juvenile must successfully find a new home range (and usually a mate).



A first step in studying the population dynamics is to model the population at yearly intervals, at times denoted by $k = 0, 1, 2, \dots$. Usually, one assumes that there is a 1:1 ratio of males to females in each life stage and counts only the females. The population at year k can be described by a vector $\mathbf{x}_k = (j_k, s_k, a_k)$, where j_k , s_k , and a_k are the numbers of females in the juvenile, subadult, and adult stages, respectively.

Using actual field data from demographic studies, R. Lamberson and co-workers considered the following *stage-matrix model*:²

$$\begin{bmatrix} j_{k+1} \\ s_{k+1} \\ a_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & .33 \\ .18 & 0 & 0 \\ 0 & .71 & .94 \end{bmatrix} \begin{bmatrix} j_k \\ s_k \\ a_k \end{bmatrix}$$

Here the number of new juvenile females in year $k + 1$ is .33 times the number of adult females in year k (based on the average birth rate per owl pair). Also, 18% of the juveniles survive to become subadults, and 71% of the subadults and 94% of the adults survive to be counted as adults.

The stage-matrix model is a difference equation of the form $\mathbf{x}_{k+1} = A\mathbf{x}_k$. Such an equation is often called a

¹ “The Great Spotted Owl War,” *Reader’s Digest*, November 1992, pp. 91–95.

² R. H. Lamberson, R. McKelvey, B. R. Noon, and C. Voss, “A Dynamic Analysis of the Viability of the Northern Spotted Owl in a Fragmented Forest Environment,” *Conservation Biology* 6 (1992), 505–512. Also, a private communication from Professor Lamberson, 1993.

dynamical system (or a **discrete linear dynamical 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 above 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.

WEB

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. Except for a brief digression in Section 5.4, all matrices in the chapter are square. The main applications described here are to discrete dynamical systems, including the spotted owls discussed above. 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} . ■

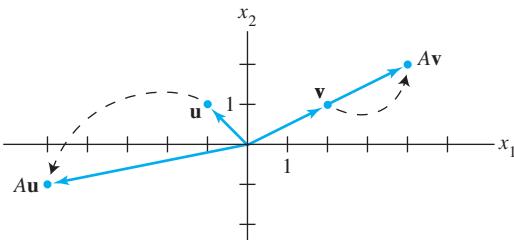


FIGURE 1 Effects of multiplication by A .

As another example, readers of Section 4.9 will recall that if A is a stochastic matrix, then the steady-state vector \mathbf{q} for A satisfies the equation $A\mathbf{x} = \mathbf{x}$. That is, $A\mathbf{q} = 1 \cdot \mathbf{q}$.

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. It is also easy to decide if a specified scalar is an eigenvalue.

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

$$\begin{aligned} A\mathbf{u} &= \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ -5 \end{bmatrix} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = -4 \begin{bmatrix} 6 \\ -5 \end{bmatrix} = -4\mathbf{u} \\ A\mathbf{v} &= \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 11 \end{bmatrix} \neq \lambda \begin{bmatrix} 3 \\ -2 \end{bmatrix} \end{aligned}$$

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

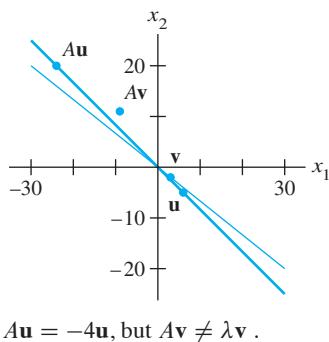
$$A - 7I = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} - \begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} -6 & 6 \\ 5 & -5 \end{bmatrix}$$

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:

$$\left[\begin{array}{ccc} -6 & 6 & 0 \\ 5 & -5 & 0 \end{array} \right] \sim \left[\begin{array}{ccc} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array} \right]$$

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.



$$A\mathbf{u} = -4\mathbf{u}, \text{ but } A\mathbf{v} \neq \lambda\mathbf{v}.$$

Warning: Although row reduction was used in Example 3 to find eigenvectors, it cannot be used to find eigenvalues. 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

$$(A - \lambda I)\mathbf{x} = \mathbf{0} \quad (3)$$

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.

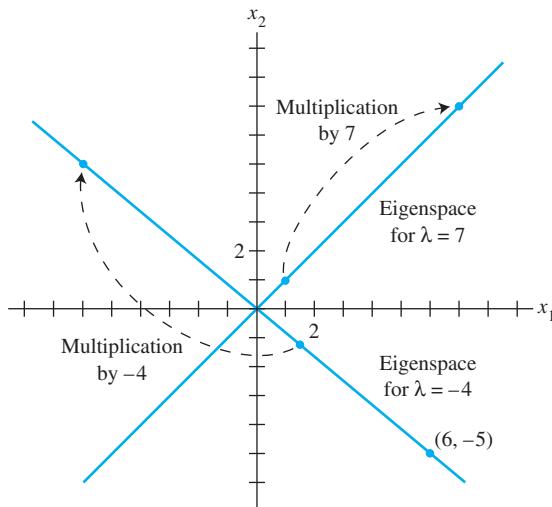


FIGURE 2 Eigenspaces for $\lambda = -4$ and $\lambda = 7$.

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 $(A - 2I)\mathbf{x} = \mathbf{0}$:

$$\left[\begin{array}{ccc|c} 2 & -1 & 6 & 0 \\ 2 & -1 & 6 & 0 \\ 2 & -1 & 6 & 0 \end{array} \right] \sim \left[\begin{array}{cccc} 2 & -1 & 6 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

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

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

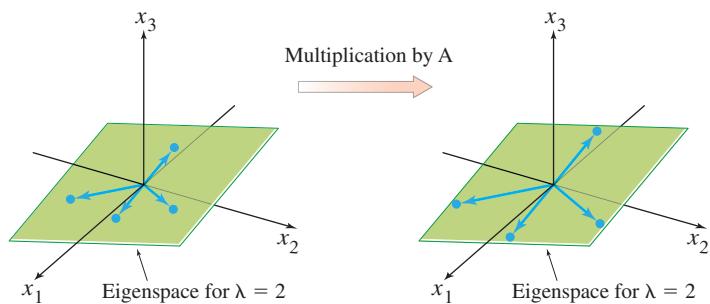


FIGURE 3 A acts as a dilation on the eigenspace.

NUMERICAL NOTE

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 28. ■

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

$$A\mathbf{x} = \mathbf{0} \quad (4)$$

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

$$c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{v}_{p+1} \quad (5)$$

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

$$\begin{aligned} c_1A\mathbf{v}_1 + \cdots + c_pA\mathbf{v}_p &= A\mathbf{v}_{p+1} \\ c_1\lambda_1\mathbf{v}_1 + \cdots + c_p\lambda_p\mathbf{v}_p &= \lambda_{p+1}\mathbf{v}_{p+1} \end{aligned} \quad (6)$$

Multiplying both sides of (5) by λ_{p+1} and subtracting the result from (6), we have

$$c_1(\lambda_1 - \lambda_{p+1})\mathbf{v}_1 + \cdots + c_p(\lambda_p - \lambda_{p+1})\mathbf{v}_p = \mathbf{0} \quad (7)$$

Since $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ 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 $i = 1, \dots, p$. But then (5) says that $\mathbf{v}_{p+1} = \mathbf{0}$, which is impossible. Hence $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ 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:

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \quad (k = 0, 1, 2, \dots) \quad (8)$$

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

$$\mathbf{x}_k = \lambda^k \mathbf{x}_0 \quad (k = 1, 2, \dots) \quad (9)$$

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 33.

PRACTICE PROBLEMS

- Is 5 an eigenvalue of $A = \begin{bmatrix} 6 & -3 & 1 \\ 3 & 0 & 5 \\ 2 & 2 & 6 \end{bmatrix}$?
- If \mathbf{x} is an eigenvector of A corresponding to λ , what is $A^3\mathbf{x}$?
- 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 $c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + (c_3\mathbf{b}_3 + c_4\mathbf{b}_4) = \mathbf{0}$.]
- 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

- Is $\lambda = 2$ an eigenvalue of $\begin{bmatrix} 3 & 2 \\ 3 & 8 \end{bmatrix}$? Why or why not?
 - Is $\lambda = -2$ an eigenvalue of $\begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix}$? Why or why not?
 - Is $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$ an eigenvector of $\begin{bmatrix} -3 & 1 \\ -3 & 8 \end{bmatrix}$? If so, find the eigenvalue.
 - Is $\begin{bmatrix} -1 + \sqrt{2} \\ 1 \end{bmatrix}$ an eigenvector of $\begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$? If so, find the eigenvalue.
 - 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.
 - Is $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ an eigenvector of $\begin{bmatrix} 3 & 6 & 7 \\ 3 & 3 & 7 \\ 5 & 6 & 5 \end{bmatrix}$? If so, find the eigenvalue.
 - 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.
 - 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. $A = \begin{bmatrix} 5 & 0 \\ 2 & 1 \end{bmatrix}, \lambda = 1, 5$

10. $A = \begin{bmatrix} 10 & -9 \\ 4 & -2 \end{bmatrix}, \lambda = 4$

11. $A = \begin{bmatrix} 4 & -2 \\ -3 & 9 \end{bmatrix}, \lambda = 10$

12. $A = \begin{bmatrix} 7 & 4 \\ -3 & -1 \end{bmatrix}, \lambda = 1, 5$

13. $A = \begin{bmatrix} 4 & 0 & 1 \\ -2 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}, \lambda = 1, 2, 3$

14. $A = \begin{bmatrix} 1 & 0 & -1 \\ 1 & -3 & 0 \\ 4 & -13 & 1 \end{bmatrix}, \lambda = -2$

15. $A = \begin{bmatrix} 4 & 2 & 3 \\ -1 & 1 & -3 \\ 2 & 4 & 9 \end{bmatrix}, \lambda = 3$

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

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 $A = \begin{bmatrix} 5 & 5 & 5 \\ 5 & 5 & 5 \\ 5 & 5 & 5 \end{bmatrix}$. Justify your answer.

In Exercises 21 and 22, A is an $n \times n$ matrix. Mark each statement True or False. Justify each answer.

21. a. If $A\mathbf{x} = \lambda\mathbf{x}$ for some vector \mathbf{x} , then λ is an eigenvalue of A .
 b. A matrix A is not invertible if and only if 0 is an eigenvalue of A .
 c. A number c is an eigenvalue of A if and only if the equation $(A - cI)\mathbf{x} = \mathbf{0}$ has a nontrivial solution.

d. Finding an eigenvector of A may be difficult, but checking whether a given vector is in fact an eigenvector is easy.

e. To find the eigenvalues of A , reduce A to echelon form.

22. a. If $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ , then \mathbf{x} is an eigenvector of A .

b. If \mathbf{v}_1 and \mathbf{v}_2 are linearly independent eigenvectors, then they correspond to distinct eigenvalues.

c. A steady-state vector for a stochastic matrix is actually an eigenvector.

d. The eigenvalues of a matrix are on its main diagonal.

e. An eigenspace of A is a null space of a certain matrix.

23. 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.

24. Construct an example of a 2×2 matrix with only one distinct eigenvalue.

25. 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 $A\mathbf{x} = \lambda\mathbf{x}$.]

26. Show that if A^2 is the zero matrix, then the only eigenvalue of A is 0.

27. 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.]

28. Use Exercise 27 to complete the proof of Theorem 1 for the case when A is lower triangular.

29. 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.]

30. 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 27 and 29.]

In Exercises 31 and 32, let A be the matrix of the linear transformation T . Without writing A , find an eigenvalue of A and describe the eigenspace.

31. T is the transformation on \mathbb{R}^2 that reflects points across some line through the origin.

32. T is the transformation on \mathbb{R}^3 that rotates points about some line through the origin.

33. 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

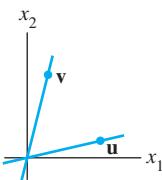
$$\mathbf{x}_k = c_1\lambda^k\mathbf{u} + c_2\mu^k\mathbf{v} \quad (k = 0, 1, 2, \dots)$$

- 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 $\mathbf{x}_{k+1} = A\mathbf{x}_k$ ($k = 0, 1, 2, \dots$).

34. Describe how you might try to build a solution of a difference equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$ ($k = 0, 1, 2, \dots$) 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 ?]

35. 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 $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ 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 $T(\mathbf{w})$.



36. Repeat Exercise 35, assuming \mathbf{u} and \mathbf{v} are eigenvectors of A that correspond to eigenvalues -1 and 3 , respectively.

[M] In Exercises 37–40, 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.

37.
$$\begin{bmatrix} 8 & -10 & -5 \\ 2 & 17 & 2 \\ -9 & -18 & 4 \end{bmatrix}$$

38.
$$\begin{bmatrix} 9 & -4 & -2 & -4 \\ -56 & 32 & -28 & 44 \\ -14 & -14 & 6 & -14 \\ 42 & -33 & 21 & -45 \end{bmatrix}$$

39.
$$\begin{bmatrix} 4 & -9 & -7 & 8 & 2 \\ -7 & -9 & 0 & 7 & 14 \\ 5 & 10 & 5 & -5 & -10 \\ -2 & 3 & 7 & 0 & 4 \\ -3 & -13 & -7 & 10 & 11 \end{bmatrix}$$

40.
$$\begin{bmatrix} -4 & -4 & 20 & -8 & -1 \\ 14 & 12 & 46 & 18 & 2 \\ 6 & 4 & -18 & 8 & 1 \\ 11 & 7 & -37 & 17 & 2 \\ 18 & 12 & -60 & 24 & 5 \end{bmatrix}$$

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:

$$\left[\begin{array}{ccc|c} 1 & -3 & 1 & 0 \\ 3 & -5 & 5 & 0 \\ 2 & 2 & 1 & 0 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & -3 & 1 & 0 \\ 0 & 4 & 2 & 0 \\ 0 & 8 & -1 & 0 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & -3 & 1 & 0 \\ 0 & 4 & 2 & 0 \\ 0 & 0 & -5 & 0 \end{array} \right]$$

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 . ■

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. However, before turning to larger matrices, we summarize the properties of determinants needed to study eigenvalues.

Determinants

Let A be an $n \times n$ matrix, let U be any echelon form obtained from A by row replacements and row interchanges (without scaling), and let r be the number of such row interchanges. Then the **determinant** of A , written as $\det A$, is $(-1)^r$ times the product of the diagonal entries u_{11}, \dots, u_{nn} in U . If A is invertible, then u_{11}, \dots, u_{nn}

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. Thus¹

$$\det A = \begin{cases} (-1)^r \cdot \left(\begin{array}{c} \text{product of} \\ \text{pivots in } U \end{array} \right), & \text{when } A \text{ is invertible} \\ 0, & \text{when } A \text{ is not invertible} \end{cases} \quad (1)$$

EXAMPLE 2 Compute $\det A$ for $A = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}$.

SOLUTION The following row reduction uses one row interchange:

$$A \sim \begin{bmatrix} 1 & 5 & 0 \\ 0 & -6 & -1 \\ 0 & -2 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & 0 \\ 0 & -2 & 0 \\ 0 & -6 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -1 \end{bmatrix} = U_1$$

So $\det A$ equals $(-1)^1(1)(-2)(-1) = -2$. The following alternative row reduction avoids the row interchange and produces a different echelon form. The last step adds $-1/3$ times row 2 to row 3:

$$A \sim \begin{bmatrix} 1 & 5 & 0 \\ 0 & -6 & -1 \\ 0 & -2 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & 0 \\ 0 & -6 & -1 \\ 0 & 0 & 1/3 \end{bmatrix} = U_2$$

This time $\det A$ is $(-1)^0(1)(-6)(1/3) = -2$, the same as before. ■

Formula (1) for the determinant shows that A is invertible if and only if $\det A$ is nonzero. This fact, and the characterization of invertibility found in Section 5.1, can be added to the Invertible Matrix Theorem.

THEOREM

The Invertible Matrix Theorem (continued)

Let A be an $n \times n$ matrix. Then A is invertible if and only if:

- s. The number 0 is *not* an eigenvalue of A .
- t. The determinant of A is *not* zero.

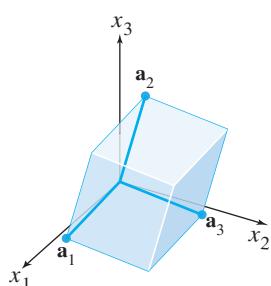


FIGURE 1

When A is a 3×3 matrix, $|\det A|$ turns out to be the *volume* of the parallelepiped determined by the columns \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 of A , as in Figure 1. (See Section 3.3 for details.) This volume is *nonzero* if and only if the vectors \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 are linearly independent, in which case the matrix A is invertible. (If the vectors are nonzero and linearly dependent, they lie in a plane or along a line.)

The next theorem lists facts needed from Sections 3.1 and 3.2. Part (a) is included here for convenient reference.

¹ Formula (1) was derived in Section 3.2. Readers who have not studied Chapter 3 may use this formula as the definition of $\det A$. It is a remarkable and nontrivial fact that any echelon form U obtained from A without scaling gives the same value for $\det A$.

THEOREM 3**Properties of Determinants**

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

- A is invertible if and only if $\det A \neq 0$.
- $\det AB = (\det A)(\det B)$.
- $\det A^T = \det A$.
- If A is triangular, then $\det A$ is the product of the entries on the main diagonal of A .
- 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.

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

$$(5 - \lambda)^2(\lambda - 3)(\lambda - 1) = 0$$

Expanding the product, we can also write

$$\lambda^4 - 14\lambda^3 + 68\lambda^2 - 130\lambda + 75 = 0$$

■

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 n th-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.

SG

Factoring a
Polynomial 5-8

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

$$\begin{aligned} \det(B - \lambda I) &= \det[P^{-1}(A - \lambda I)P] \\ &= \det(P^{-1}) \cdot \det(A - \lambda I) \cdot \det(P) \end{aligned} \tag{2}$$

Since $\det(P^{-1}) \cdot \det(P) = \det(P^{-1}P) = \det I = 1$, we see from equation (2) 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 of the dynamical system defined by $\mathbf{x}_{k+1} = A\mathbf{x}_k$ ($k = 0, 1, 2, \dots$), 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

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = [\mathbf{v}_1 \quad \mathbf{v}_2] \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \quad (3)$$

In fact,

$$\begin{aligned} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} &= [\mathbf{v}_1 \quad \mathbf{v}_2]^{-1} \mathbf{x}_0 = \begin{bmatrix} 3 & 1 \\ 5 & -1 \end{bmatrix}^{-1} \begin{bmatrix} .60 \\ .40 \end{bmatrix} \\ &= \frac{1}{-8} \begin{bmatrix} -1 & -1 \\ -5 & 3 \end{bmatrix} \begin{bmatrix} .60 \\ .40 \end{bmatrix} = \begin{bmatrix} .125 \\ .225 \end{bmatrix} \end{aligned} \quad (4)$$

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),

$$\mathbf{x}_k = .125 \begin{bmatrix} 3 \\ 5 \end{bmatrix} + .225(.92)^k \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad (k = 0, 1, 2, \dots) \quad (5)$$

This explicit formula for \mathbf{x}_k gives the solution of the difference equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$. As $k \rightarrow \infty$, $(.92)^k$ 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 4.9. Those who read that section may recognize that matrix A in Example 5 above is the same as the migration matrix M in Section 4.9, \mathbf{x}_0 is the initial population distribution between city and suburbs, and \mathbf{x}_k represents the population distribution after k years.

Theorem 18 in Section 4.9 stated that for a matrix such as A , the sequence \mathbf{x}_k tends to a steady-state vector. Now we know *why* the \mathbf{x}_k behave this way, at least for the migration matrix. The steady-state vector is $.125\mathbf{v}_1$, a multiple of the eigenvector \mathbf{v}_1 , and formula (5) for \mathbf{x}_k shows precisely why $\mathbf{x}_k \rightarrow .125\mathbf{v}_1$.

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 $n \geq 5$.
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 $\lambda_1, \dots, \lambda_n$ 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

$$A_1 = A \quad \text{and} \quad A_{k+1} = P_k^{-1} A_k P_k \quad (k = 1, 2, \dots)$$

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. $\begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$

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

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

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

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

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 either a cofactor expansion or the special formula for 3×3 determinants described

prior to Exercises 15–18 in Section 3.1. [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. $\begin{bmatrix} 4 & 0 & 0 \\ 5 & 3 & 2 \\ -2 & 0 & 2 \end{bmatrix}$

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

13. $\begin{bmatrix} 6 & -2 & 0 \\ -2 & 9 & 0 \\ 5 & 8 & 3 \end{bmatrix}$

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

For the matrices in Exercises 15–17, list the eigenvalues, repeated according to their multiplicities.

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

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 $\lambda = 5$ is two-dimensional:

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

19. Let A be an $n \times n$ matrix, and suppose A has n real eigenvalues, $\lambda_1, \dots, \lambda_n$, 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 and 22, A and B are $n \times n$ matrices. Mark each statement True or False. Justify each answer.

21. a. The determinant of A is the product of the diagonal entries in A .
 b. An elementary row operation on A does not change the determinant.
 c. $(\det A)(\det B) = \det AB$
 d. If $\lambda + 5$ is a factor of the characteristic polynomial of A , then 5 is an eigenvalue of A .

22. a. If A is 3×3 , with columns $\mathbf{a}_1, \mathbf{a}_2$, and \mathbf{a}_3 , then $\det A$ equals the volume of the parallelepiped determined by $\mathbf{a}_1, \mathbf{a}_2$ and \mathbf{a}_3 .
 b. $\det A^T = (-1) \det A$.
 c. The multiplicity of a root r of the characteristic equation of A is called the algebraic multiplicity of r as an eigenvalue of A .
 d. A row replacement operation on A does not change the 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 $A = Q_1 R_1$, where $Q_1^T = Q_1^{-1}$ and R_1 is upper triangular. The factors are interchanged to form $A_1 = R_1 Q_1$, which is again factored as $A_1 = Q_2 R_2$; then to form $A_2 = R_2 Q_2$, and so on. The similarity of A, A_1, \dots follows from the more general result in Exercise 23.

23. Show that if $A = QR$ with Q invertible, then A is similar to $A_1 = RQ$.
 24. Show that if A and B are similar, then $\det A = \det B$.
 25. Let $A = \begin{bmatrix} .6 & .3 \\ .4 & .7 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 3/7 \\ 4/7 \end{bmatrix}$, $\mathbf{x}_0 = \begin{bmatrix} .5 \\ .5 \end{bmatrix}$. [Note: A is the stochastic matrix studied in Example 5 of Section 4.9.]
 a. Find a basis for \mathbb{R}^2 consisting of \mathbf{v}_1 and another eigenvector \mathbf{v}_2 of A .
 b. Verify that \mathbf{x}_0 may be written in the form $\mathbf{x}_0 = \mathbf{v}_1 + c\mathbf{v}_2$.
 c. For $k = 1, 2, \dots$, define $\mathbf{x}_k = A^k \mathbf{x}_0$. Compute \mathbf{x}_1 and \mathbf{x}_2 , and write a formula for \mathbf{x}_k . Then show that $\mathbf{x}_k \rightarrow \mathbf{v}_1$ as k increases.

26. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Use formula (1) for a determinant (given before Example 2) to show that $\det A = ad - bc$. Consider two cases: $a \neq 0$ and $a = 0$.

27. Let $A = \begin{bmatrix} .5 & .2 & .3 \\ .3 & .8 & .3 \\ .2 & 0 & .4 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} .3 \\ .6 \\ .1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

- a. Show that $\mathbf{v}_1, \mathbf{v}_2$, and \mathbf{v}_3 are eigenvectors of A . [Note: A is the stochastic matrix studied in Example 3 of Section 4.9.]
 b. Let \mathbf{x}_0 be any vector in \mathbb{R}^3 with nonnegative entries whose sum is 1. (In Section 4.9, \mathbf{x}_0 was called a probability vector.) Explain why there are constants c_1, c_2 , and c_3 such that $\mathbf{x}_0 = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + c_3 \mathbf{v}_3$. Compute $\mathbf{w}^T \mathbf{x}_0$, and deduce that $c_1 = 1$.
 c. For $k = 1, 2, \dots$, define $\mathbf{x}_k = A^k \mathbf{x}_0$, with \mathbf{x}_0 as in part (b). Show that $\mathbf{x}_k \rightarrow \mathbf{v}_1$ as k increases.

28. [M] 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.
29. [M] Construct a random integer-valued 4×4 matrix A .
- Reduce A to echelon form U with no row scaling, and use U in formula (1) (before Example 2) to compute $\det A$. (If A happens to be singular, start over with a new random matrix.)
 - 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).

30. [M] 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 $0 \leq t \leq 3$. 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

$$\begin{aligned} 0 &= \det(A - \lambda I) = \det \begin{bmatrix} 1 - \lambda & -4 \\ 4 & 2 - \lambda \end{bmatrix} \\ &= (1 - \lambda)(2 - \lambda) - (-4)(4) = \lambda^2 - 3\lambda + 18 \end{aligned}$$

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,

$$\begin{aligned} A^2 &= (PDP^{-1})(PDP^{-1}) = PD \underbrace{(P^{-1}P)}_I DP^{-1} = PDDP^{-1} \\ &= PD^2 P^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} \end{aligned}$$

Again,

$$A^3 = (PDP^{-1})A^2 = (PDP^{-1})\underbrace{PD^2P^{-1}}_I = PDD^2P^{-1} = PD^3P^{-1}$$

In general, for $k \geq 1$,

$$\begin{aligned} A^k &= PD^k P^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 5^k & 0 \\ 0 & 3^k \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 2 \cdot 5^k - 3^k & 5^k - 3^k \\ 2 \cdot 3^k - 2 \cdot 5^k & 2 \cdot 3^k - 5^k \end{bmatrix} \quad \blacksquare \end{aligned}$$

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

$$AP = A[\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n] = [A\mathbf{v}_1 \ A\mathbf{v}_2 \ \cdots \ A\mathbf{v}_n] \quad (1)$$

while

$$PD = P \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = [\lambda_1 \mathbf{v}_1 \ \lambda_2 \mathbf{v}_2 \ \cdots \ \lambda_n \mathbf{v}_n] \quad (2)$$

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

$$[A\mathbf{v}_1 \ A\mathbf{v}_2 \ \cdots \ A\mathbf{v}_n] = [\lambda_1\mathbf{v}_1 \ \lambda_2\mathbf{v}_2 \ \cdots \ \lambda_n\mathbf{v}_n] \quad (3)$$

Equating columns, we find that

$$A\mathbf{v}_1 = \lambda_1\mathbf{v}_1, \quad A\mathbf{v}_2 = \lambda_2\mathbf{v}_2, \quad \dots, \quad A\mathbf{v}_n = \lambda_n\mathbf{v}_n \quad (4)$$

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 order of the vectors is unimportant. Using the order chosen in step 2, form

$$P = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3] = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

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

$$\begin{aligned} AP &= \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 \\ -1 & -2 & 0 \\ 1 & 0 & -2 \end{bmatrix} \\ PD &= \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 \\ -1 & -2 & 0 \\ 1 & 0 & -2 \end{bmatrix} \quad \blacksquare \end{aligned}$$

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:

$$\begin{aligned} \text{Basis for } \lambda = 1: \quad \mathbf{v}_1 &= \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \\ \text{Basis for } \lambda = -2: \quad \mathbf{v}_2 &= \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \end{aligned}$$

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 $1 \leq k \leq p$, 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.

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

¹The 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.

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.

$$\text{Basis for } \lambda = 5: \mathbf{v}_1 = \begin{bmatrix} -8 \\ 4 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} -16 \\ 4 \\ 0 \\ 1 \end{bmatrix}$$

$$\text{Basis for } \lambda = -3: \mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

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

$$P = \begin{bmatrix} -8 & -16 & 0 & 0 \\ 4 & 4 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & -3 \end{bmatrix}$$

■

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?

WEB

5.3 EXERCISES

In Exercises 1 and 2, let $A = PDP^{-1}$ and compute A^4 .

1. $P = \begin{bmatrix} 5 & 7 \\ 2 & 3 \end{bmatrix}$, $D = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$

2. $P = \begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}$, $D = \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix}$

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. $\begin{bmatrix} -2 & 12 \\ -1 & 5 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 4 \\ 1 & -3 \end{bmatrix}$

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

Diagonalize the matrices in Exercises 7–20, if possible. The eigenvalues for Exercises 11–16 are as follows: (11) $\lambda = 1, 2, 3$; (12) $\lambda = 2, 8$; (13) $\lambda = 5, 1$; (14) $\lambda = 5, 4$; (15) $\lambda = 3, 1$; (16) $\lambda = 2, 1$. For Exercise 18, one eigenvalue is $\lambda = 5$ and one eigenvector is $(-2, 1, 2)$.

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

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

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

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

15. $\begin{bmatrix} 7 & 4 & 16 \\ 2 & 5 & 8 \\ -2 & -2 & -5 \end{bmatrix}$

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

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

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

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

12. $\begin{bmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix}$

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

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

18. $\begin{bmatrix} -7 & -16 & 4 \\ 6 & 13 & -2 \\ 12 & 16 & 1 \end{bmatrix}$

20. $\begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 2 \end{bmatrix}$

In Exercises 21 and 22, A , B , P , and D are $n \times n$ matrices. Mark each statement True or False. Justify each answer. (Study Theorems 5 and 6 and the examples in this section carefully before you try these exercises.)

21. a. A is diagonalizable if $A = PDP^{-1}$ for some matrix D and some invertible matrix P .
 b. If \mathbb{R}^n has a basis of eigenvectors of A , then A is diagonalizable.
 c. A is diagonalizable if and only if A has n eigenvalues, counting multiplicities.
 d. If A is diagonalizable, then A is invertible.
22. a. A is diagonalizable if A has n eigenvectors.
 b. If A is diagonalizable, then A has n distinct eigenvalues.
 c. If $AP = PD$, with D diagonal, then the nonzero columns of P must be eigenvectors of A .
 d. If A is invertible, then A is diagonalizable.
23. 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?

24. A is a 3×3 matrix with two eigenvalues. Each eigenspace is one-dimensional. Is A diagonalizable? Why?

25. 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.

26. 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.

27. Show that if A is both diagonalizable and invertible, then so is A^{-1} .

28. Show that if A has n linearly independent eigenvectors, then so does A^T . [Hint: Use the Diagonalization Theorem.]

29. 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 $A = P_1 D_1 P_1^{-1}$.

30. With A and D as in Example 2, find an invertible P_2 unequal to the P in Example 2, such that $A = P_2 DP_2^{-1}$.

31. Construct a nonzero 2×2 matrix that is invertible but not diagonalizable.

32. Construct a nondiagonal 2×2 matrix that is diagonalizable but not invertible.

[M] Diagonalize the matrices in Exercises 33–36. Use your matrix program's eigenvalue command to find the eigenvalues, and then compute bases for the eigenspaces as in Section 5.1.

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

34. $\begin{bmatrix} 0 & 13 & 8 & 4 \\ 4 & 9 & 8 & 4 \\ 8 & 6 & 12 & 8 \\ 0 & 5 & 0 & -4 \end{bmatrix}$

35. $\begin{bmatrix} 11 & -6 & 4 & -10 & -4 \\ -3 & 5 & -2 & 4 & 1 \\ -8 & 12 & -3 & 12 & 4 \\ 1 & 6 & -2 & 3 & -1 \\ 8 & -18 & 8 & -14 & -1 \end{bmatrix}$

36. $\begin{bmatrix} 4 & 4 & 2 & 3 & -2 \\ 0 & 1 & -2 & -2 & 2 \\ 6 & 12 & 11 & 2 & -4 \\ 9 & 20 & 10 & 10 & -6 \\ 15 & 28 & 14 & 5 & -3 \end{bmatrix}$

6

Orthogonality and Least Squares

INTRODUCTORY EXAMPLE

The North American Datum and GPS Navigation

Imagine starting a massive project that you estimate will take ten years and require the efforts of scores of people to construct and solve a 1,800,000-by-900,000 system of linear equations. That is exactly what the National Geodetic Survey did in 1974, when it set out to update the North American Datum (NAD)—a network of 268,000 precisely located reference points that span the entire North American continent, together with Greenland, Hawaii, the Virgin Islands, Puerto Rico, and other Caribbean islands.

The recorded latitudes and longitudes in the NAD must be determined to within a few centimeters because they form the basis for all surveys, maps, legal property boundaries, and layouts of civil engineering projects such as highways and public utility lines. However, more than 200,000 new points had been added to the datum since the last adjustment in 1927, and errors had gradually accumulated over the years, due to imprecise measurements and shifts in the earth’s crust. Data gathering for the NAD readjustment was completed in 1983.

The system of equations for the NAD had no solution in the ordinary sense, but rather had a *least-squares solution*, which assigned latitudes and longitudes to the reference points in a way that corresponded best to the 1.8 million observations. The least-squares solution was found in 1986 by solving a related system of so-called



normal equations, which involved 928,735 equations in 928,735 variables.¹

More recently, knowledge of reference points on the ground has become crucial for accurately determining the locations of satellites in the satellite-based *Global Positioning System (GPS)*. A GPS satellite calculates its position relative to the earth by measuring the time it takes for signals to arrive from three ground transmitters. To do this, the satellites use precise atomic clocks that have been synchronized with ground stations (whose locations are known accurately because of the NAD).

The *Global Positioning System* is used both for determining the locations of new reference points on the ground and for finding a user’s position on the ground relative to established maps. When a car driver (or a mountain climber) turns on a GPS receiver, the receiver measures the relative arrival times of signals from at least three satellites. This information, together with the transmitted data about the satellites’ locations and message times, is used to adjust the GPS receiver’s time and to determine its approximate location on the earth. Given information from a fourth satellite, the GPS receiver can even establish its approximate altitude.

¹ A mathematical discussion of the solution strategy (along with details of the entire NAD project) appears in *North American Datum of 1983*, Charles R. Schwarz (ed.), National Geodetic Survey, National Oceanic and Atmospheric Administration (NOAA) Professional Paper NOS 2, 1989.

Both the NAD and GPS problems are solved by finding a vector that “approximately satisfies” an inconsistent system of equations. A careful explanation of this apparent

contradiction will require ideas developed in the first five sections of this chapter.

WEB

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, such as the system solved for the NAD report.

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

$$[u_1 \ u_2 \ \cdots \ u_n] \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

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

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} = [2 \quad -5 \quad -1] \begin{bmatrix} 3 \\ 2 \\ -3 \end{bmatrix} = (2)(3) + (-5)(2) + (-1)(-3) = -1$$

$$\mathbf{v} \cdot \mathbf{u} = \mathbf{v}^T \mathbf{u} = [3 \quad 2 \quad -3] \begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix} = (3)(2) + (2)(-5) + (-3)(-1) = -1 \quad \blacksquare$$

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 21 and 22 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

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}, \quad \text{and} \quad \|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v}$$

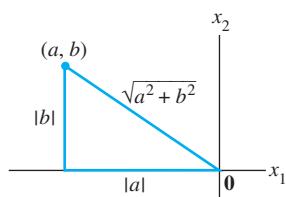


FIGURE 1

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} :

$$\begin{aligned}\|\mathbf{v}\|^2 &= \mathbf{v} \cdot \mathbf{v} = (1)^2 + (-2)^2 + (2)^2 + (0)^2 = 9 \\ \|\mathbf{v}\| &= \sqrt{9} = 3\end{aligned}$$

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 $\|\mathbf{u}\|^2 = 1$.

$$\begin{aligned}\|\mathbf{u}\|^2 &= \mathbf{u} \cdot \mathbf{u} = \left(\frac{1}{3}\right)^2 + \left(-\frac{2}{3}\right)^2 + \left(\frac{2}{3}\right)^2 + (0)^2 \\ &= \frac{1}{9} + \frac{4}{9} + \frac{4}{9} + 0 = 1\end{aligned}$$

EXAMPLE 3 Let W be the subspace of \mathbb{R}^2 spanned by $\mathbf{x} = \left(\frac{2}{3}, 1\right)$. 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 $(-2/\sqrt{13}, -3/\sqrt{13})$.

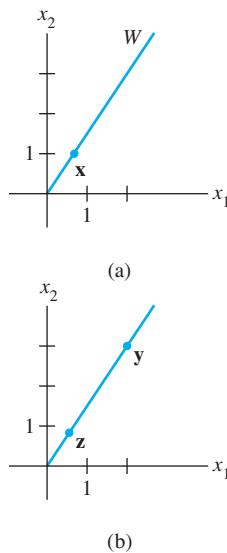


FIGURE 2

Normalizing a vector to produce a unit vector.

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 .

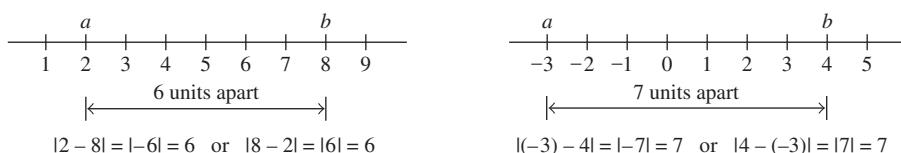


FIGURE 3 Distances in \mathbb{R} .

DEFINITION

For \mathbf{u} and \mathbf{v} in \mathbb{R}^n , the **distance between \mathbf{u} and \mathbf{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 $\mathbf{0}$. ■

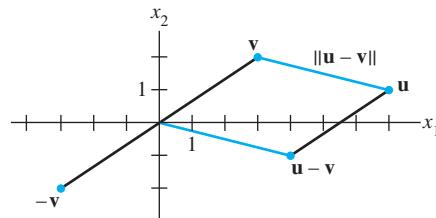


FIGURE 4 The distance between \mathbf{u} and \mathbf{v} is the length of $\mathbf{u} - \mathbf{v}$.

EXAMPLE 5 If $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$, then

$$\begin{aligned} \text{dist}(\mathbf{u}, \mathbf{v}) &= \|\mathbf{u} - \mathbf{v}\| = \sqrt{(\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v})} \\ &= \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2} \end{aligned}$$

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

$$\begin{aligned} [\text{dist}(\mathbf{u}, -\mathbf{v})]^2 &= \|\mathbf{u} - (-\mathbf{v})\|^2 = \|\mathbf{u} + \mathbf{v}\|^2 \\ &= (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) \\ &= \mathbf{u} \cdot (\mathbf{u} + \mathbf{v}) + \mathbf{v} \cdot (\mathbf{u} + \mathbf{v}) && \text{Theorem 1(b)} \\ &= \mathbf{u} \cdot \mathbf{u} + \mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v} && \text{Theorem 1(a), (b)} \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\mathbf{u} \cdot \mathbf{v} && \text{Theorem 1(a)} \end{aligned} \tag{1}$$

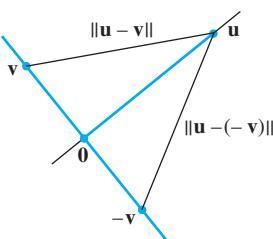


FIGURE 5

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) above 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 $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

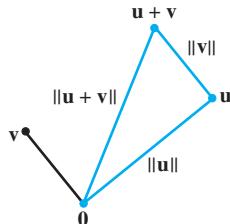


FIGURE 6

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”).

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,

$$L = W^\perp \quad \text{and} \quad W = L^\perp$$

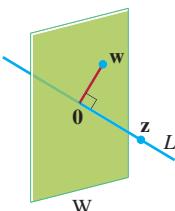


FIGURE 7

A plane and line through $\mathbf{0}$ as orthogonal complements.

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 29 and 30. Exercises 27–31 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 31 verify the claims made in Section 4.6 concerning the subspaces shown in Figure 8. (Also see Exercise 28 in Section 4.6.)

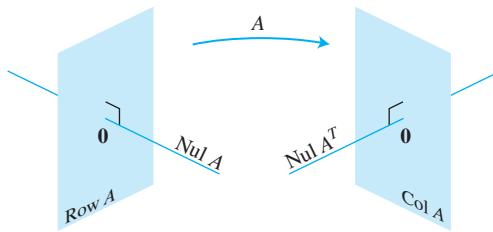


FIGURE 8 The fundamental subspaces determined by an $m \times n$ matrix A .

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 $\text{Nul } A = (\text{Row } A)^\perp$ is established by showing that $\text{Nul } A$ is a subset of $(\text{Row } A)^\perp$ and $(\text{Row } A)^\perp$ is a subset of $\text{Nul } A$. That is, an arbitrary element \mathbf{x} in $\text{Nul } A$ is shown to be in $(\text{Row } A)^\perp$, and then an arbitrary element \mathbf{x} in $(\text{Row } A)^\perp$ is shown to be in $\text{Nul } A$.

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 $\text{Nul } A$, 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 $\text{Row } A$. Conversely, if \mathbf{x} is orthogonal to $\text{Row } A$, 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 $\text{Row } A^T = \text{Col } A$. ■

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,

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\| \|\mathbf{v}\| \cos \vartheta$$

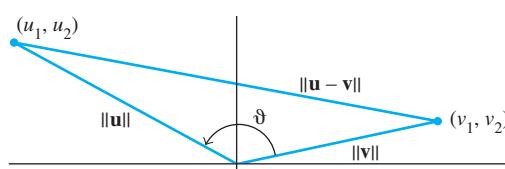


FIGURE 9 The angle between two vectors.

which can be rearranged to produce

$$\begin{aligned}\|\mathbf{u}\| \|\mathbf{v}\| \cos \vartheta &= \frac{1}{2} [\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - \|\mathbf{u} - \mathbf{v}\|^2] \\ &= \frac{1}{2} [u_1^2 + u_2^2 + v_1^2 + v_2^2 - (u_1 - v_1)^2 - (u_2 - v_2)^2] \\ &= u_1 v_1 + u_2 v_2 \\ &= \mathbf{u} \cdot \mathbf{v}\end{aligned}$$

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 \mathbf{u} and \mathbf{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 $\frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}$ and $\left(\frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}\right) \mathbf{a}$.
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 \mathbf{u} in the direction of \mathbf{c} .
 - b. Show that \mathbf{d} is orthogonal to \mathbf{c} .
 - c. Use the results of (a) and (b) to explain why \mathbf{d} must be orthogonal to the unit vector \mathbf{u} .
3. Let W be a subspace of \mathbb{R}^n . Exercise 30 establishes that W^\perp is also a subspace of \mathbb{R}^n . Prove that $\dim W + \dim W^\perp = n$.

6.1 EXERCISES

Compute the quantities in Exercises 1–8 using the vectors

$$\mathbf{u} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} 3 \\ -1 \\ -5 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 6 \\ -2 \\ 3 \end{bmatrix}$$

$$1. \mathbf{u} \cdot \mathbf{u}, \mathbf{v} \cdot \mathbf{u}, \text{ and } \frac{\mathbf{v} \cdot \mathbf{u}}{\|\mathbf{u}\| \|\mathbf{v}\|}$$

$$3. \frac{1}{\mathbf{w} \cdot \mathbf{w}} \mathbf{w}$$

$$5. \left(\frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{v}\|^2}\right) \mathbf{v}$$

$$7. \|\mathbf{w}\|$$

$$2. \mathbf{w} \cdot \mathbf{w}, \mathbf{x} \cdot \mathbf{w}, \text{ and } \frac{\mathbf{x} \cdot \mathbf{w}}{\|\mathbf{w}\| \|\mathbf{x}\|}$$

$$4. \frac{1}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u}$$

$$6. \left(\frac{\mathbf{x} \cdot \mathbf{w}}{\|\mathbf{x}\|^2}\right) \mathbf{x}$$

$$8. \|\mathbf{x}\|$$

In Exercises 9–12, find a unit vector in the direction of the given vector.

$$9. \begin{bmatrix} -30 \\ 40 \end{bmatrix}$$

$$10. \begin{bmatrix} -6 \\ 4 \\ -3 \end{bmatrix}$$

$$11. \begin{bmatrix} 7/4 \\ 1/2 \\ 1 \end{bmatrix}$$

$$12. \begin{bmatrix} 8/3 \\ 2 \end{bmatrix}$$

$$13. \text{Find the distance between } \mathbf{x} = \begin{bmatrix} 10 \\ -3 \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} -1 \\ -5 \end{bmatrix}.$$

$$14. \text{Find the distance between } \mathbf{u} = \begin{bmatrix} 0 \\ -5 \\ 2 \end{bmatrix} \text{ and } \mathbf{z} = \begin{bmatrix} -4 \\ -1 \\ 8 \end{bmatrix}.$$

Determine which pairs of vectors in Exercises 15–18 are orthogonal.

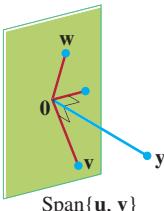
$$15. \mathbf{a} = \begin{bmatrix} 8 \\ -5 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -2 \\ -3 \end{bmatrix} \quad 16. \mathbf{u} = \begin{bmatrix} 12 \\ 3 \\ -5 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 2 \\ -3 \\ 3 \end{bmatrix}$$

$$17. \mathbf{u} = \begin{bmatrix} 3 \\ 2 \\ -5 \\ 0 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} -4 \\ 1 \\ -2 \\ 6 \end{bmatrix} \quad 18. \mathbf{y} = \begin{bmatrix} -3 \\ 7 \\ 4 \\ 0 \end{bmatrix}, \mathbf{z} = \begin{bmatrix} 1 \\ -8 \\ 15 \\ -7 \end{bmatrix}$$

In Exercises 19 and 20, all vectors are in \mathbb{R}^n . Mark each statement True or False. Justify each answer.

19. a. $\mathbf{v} \cdot \mathbf{v} = \|\mathbf{v}\|^2$.
- b. For any scalar c , $\mathbf{u} \cdot (c\mathbf{v}) = c(\mathbf{u} \cdot \mathbf{v})$.
- c. 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.
- d. For a square matrix A , vectors in $\text{Col } A$ are orthogonal to vectors in $\text{Nul } A$.

- e. 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 .
- 20.** a. $\mathbf{u} \cdot \mathbf{v} - \mathbf{v} \cdot \mathbf{u} = 0$.
b. For any scalar c , $\|\mathbf{c}\mathbf{v}\| = c\|\mathbf{v}\|$.
c. If \mathbf{x} is orthogonal to every vector in a subspace W , then \mathbf{x} is in W^\perp .
d. If $\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 = \|\mathbf{u} + \mathbf{v}\|^2$, then \mathbf{u} and \mathbf{v} are orthogonal.
e. For an $m \times n$ matrix A , vectors in the null space of A are orthogonal to vectors in the row space of A .
- 21.** Use the transpose definition of the inner product to verify parts (b) and (c) of Theorem 1. Mention the appropriate facts from Chapter 2.
- 22.** Let $\mathbf{u} = (u_1, u_2, u_3)$. Explain why $\mathbf{u} \cdot \mathbf{u} \geq 0$. When is $\mathbf{u} \cdot \mathbf{u} = 0$?
- 23.** Let $\mathbf{u} = \begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -7 \\ -4 \\ 6 \end{bmatrix}$. 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.
- 24.** Verify the *parallelogram law* for vectors \mathbf{u} and \mathbf{v} in \mathbb{R}^n :

$$\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2 = 2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2$$
- 25.** 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 $\mathbf{v} \neq \mathbf{0}$.]
- 26.** 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.
- 27.** Suppose a vector \mathbf{y} is orthogonal to vectors \mathbf{u} and \mathbf{v} . Show that \mathbf{y} is orthogonal to the vector $\mathbf{u} + \mathbf{v}$.
- 28.** Suppose \mathbf{y} is orthogonal to \mathbf{u} and \mathbf{v} . Show that \mathbf{y} is orthogonal to every \mathbf{w} in $\text{Span}\{\mathbf{u}, \mathbf{v}\}$. [Hint: An arbitrary \mathbf{w} in $\text{Span}\{\mathbf{u}, \mathbf{v}\}$ 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} .]
- 
- 29.** 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 .
- 30.** 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.
- 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 .)
 - 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?
 - Finish the proof that W^\perp is a subspace of \mathbb{R}^n .
- 31.** Show that if \mathbf{x} is in both W and W^\perp , then $\mathbf{x} = \mathbf{0}$.
- 32.** [M] Construct a pair \mathbf{u}, \mathbf{v} of random vectors in \mathbb{R}^4 , and let
- $$A = \begin{bmatrix} .5 & .5 & .5 & .5 \\ .5 & .5 & -.5 & -.5 \\ .5 & -.5 & .5 & -.5 \\ .5 & -.5 & -.5 & .5 \end{bmatrix}$$
- 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 $\mathbf{a}_3 \cdot \mathbf{a}_4$.
 - Compute and compare the lengths of \mathbf{u} , $A\mathbf{u}$, \mathbf{v} , and $A\mathbf{v}$.
 - 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 $A\mathbf{v}$.
 - Repeat parts (b) and (c) for two other pairs of random vectors. What do you conjecture about the effect of A on vectors?
- 33.** [M] 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.
- 34.** [M] 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 $\text{Nul } A$, and construct a matrix R whose rows form a basis for $\text{Row } A$ (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 $\frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{a} \cdot \mathbf{a}} = \frac{7}{5}$, and $\left(\frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{a} \cdot \mathbf{a}}\right)\mathbf{a} = \frac{7}{5}\mathbf{a} = \begin{bmatrix} -14/5 \\ 7/5 \end{bmatrix}$.

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 $\mathbf{u} = \frac{1}{\|\mathbf{y}\|}\mathbf{y} = \begin{bmatrix} 4/\sqrt{29} \\ -3/\sqrt{29} \\ 2/\sqrt{29} \end{bmatrix}$.

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 $\{\mathbf{u}_2, \mathbf{u}_3\}$.

$$\mathbf{u}_1 \cdot \mathbf{u}_2 = 3(-1) + 1(2) + 1(1) = 0$$

$$\mathbf{u}_1 \cdot \mathbf{u}_3 = 3\left(-\frac{1}{2}\right) + 1(-2) + 1\left(\frac{7}{2}\right) = 0$$

$$\mathbf{u}_2 \cdot \mathbf{u}_3 = -1\left(-\frac{1}{2}\right) + 2(-2) + 1\left(\frac{7}{2}\right) = 0$$

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 there are mutually perpendicular.

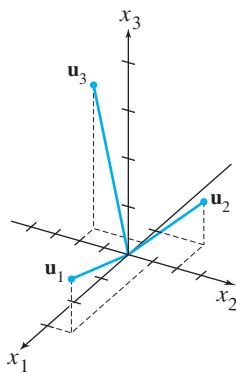


FIGURE 1

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

$$\begin{aligned} \mathbf{0} &= \mathbf{0} \cdot \mathbf{u}_1 = (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p) \cdot \mathbf{u}_1 \\ &= (c_1\mathbf{u}_1) \cdot \mathbf{u}_1 + (c_2\mathbf{u}_2) \cdot \mathbf{u}_1 + \cdots + (c_p\mathbf{u}_p) \cdot \mathbf{u}_1 \\ &= c_1(\mathbf{u}_1 \cdot \mathbf{u}_1) + c_2(\mathbf{u}_2 \cdot \mathbf{u}_1) + \cdots + c_p(\mathbf{u}_p \cdot \mathbf{u}_1) \\ &= c_1(\mathbf{u}_1 \cdot \mathbf{u}_1) \end{aligned}$$

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

$$c_j = \frac{\mathbf{y} \cdot \mathbf{u}_j}{\mathbf{u}_j \cdot \mathbf{u}_j} \quad (j = 1, \dots, p)$$

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

$$\mathbf{y} \cdot \mathbf{u}_1 = 11, \quad \mathbf{y} \cdot \mathbf{u}_2 = -12, \quad \mathbf{y} \cdot \mathbf{u}_3 = -33$$

$$\mathbf{u}_1 \cdot \mathbf{u}_1 = 11, \quad \mathbf{u}_2 \cdot \mathbf{u}_2 = 6, \quad \mathbf{u}_3 \cdot \mathbf{u}_3 = 33/2$$

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.

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

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \quad (1)$$

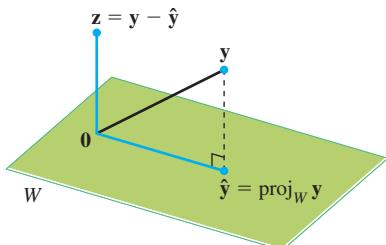


FIGURE 2

Finding α to make $\mathbf{y} - \hat{\mathbf{y}}$ orthogonal to \mathbf{u} .

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 $\alpha = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}$ and $\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}$. The vector $\hat{\mathbf{y}}$ is called the **orthogonal projection of \mathbf{y} onto \mathbf{u}** , and the vector \mathbf{z} is called the **component of \mathbf{y} orthogonal to \mathbf{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 31). 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 \mathbf{y} onto L** . That is,

$$\hat{\mathbf{y}} = \text{proj}_L \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u} \quad (2)$$

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 $\text{Span}\{\mathbf{u}\}$ and one orthogonal to \mathbf{u} .

SOLUTION Compute

$$\begin{aligned} \mathbf{y} \cdot \mathbf{u} &= \begin{bmatrix} 7 \\ 6 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 40 \\ \mathbf{u} \cdot \mathbf{u} &= \begin{bmatrix} 4 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 20 \end{aligned}$$

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,

$$\begin{bmatrix} 7 \\ 6 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

\uparrow \uparrow \uparrow
 \mathbf{y} $\hat{\mathbf{y}}$ $(\mathbf{y} - \hat{\mathbf{y}})$

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

$$\hat{\mathbf{y}} \cdot (\mathbf{y} - \hat{\mathbf{y}}) = \begin{bmatrix} 8 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 2 \end{bmatrix} = -8 + 8 = 0$$

■

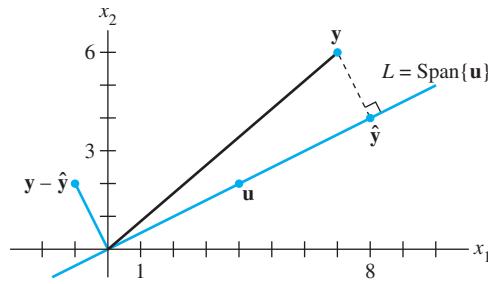


FIGURE 3 The orthogonal projection of \mathbf{y} onto a line L through the origin.

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

$$\mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 \quad (3)$$

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.

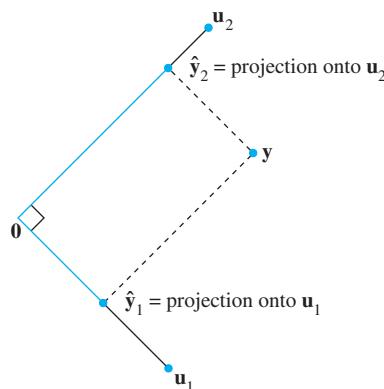


FIGURE 4 A vector decomposed into the sum of two projections.

Theorem 5 decomposes each \mathbf{y} in $\text{Span}\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ 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 made in Example 3 above.

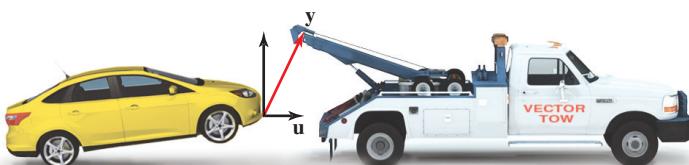


FIGURE 5

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,

$$\mathbf{v}_1 \cdot \mathbf{v}_1 = 9/11 + 1/11 + 1/11 = 1$$

$$\mathbf{v}_2 \cdot \mathbf{v}_2 = 1/6 + 4/6 + 1/6 = 1$$

$$\mathbf{v}_3 \cdot \mathbf{v}_3 = 1/66 + 16/66 + 49/66 = 1$$

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.

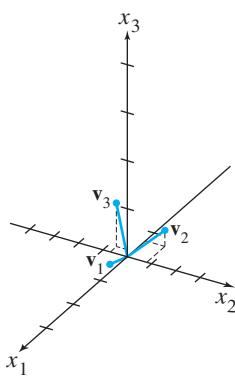


FIGURE 6

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 32. 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 $U = [\mathbf{u}_1 \ \mathbf{u}_2 \ \mathbf{u}_3]$ and compute

$$U^T U = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \mathbf{u}_3^T \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T \mathbf{u}_1 & \mathbf{u}_1^T \mathbf{u}_2 & \mathbf{u}_1^T \mathbf{u}_3 \\ \mathbf{u}_2^T \mathbf{u}_1 & \mathbf{u}_2^T \mathbf{u}_2 & \mathbf{u}_2^T \mathbf{u}_3 \\ \mathbf{u}_3^T \mathbf{u}_1 & \mathbf{u}_3^T \mathbf{u}_2 & \mathbf{u}_3^T \mathbf{u}_3 \end{bmatrix} \quad (4)$$

The entries in the matrix at the right are inner products, using transpose notation. The columns of U are orthogonal if and only if

$$\mathbf{u}_1^T \mathbf{u}_2 = \mathbf{u}_2^T \mathbf{u}_1 = 0, \quad \mathbf{u}_1^T \mathbf{u}_3 = \mathbf{u}_3^T \mathbf{u}_1 = 0, \quad \mathbf{u}_2^T \mathbf{u}_3 = \mathbf{u}_3^T \mathbf{u}_2 = 0 \quad (5)$$

The columns of U all have unit length if and only if

$$\mathbf{u}_1^T \mathbf{u}_1 = 1, \quad \mathbf{u}_2^T \mathbf{u}_2 = 1, \quad \mathbf{u}_3^T \mathbf{u}_3 = 1 \quad (6)$$

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 25 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 27 and 28. 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

- 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 .
- 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.
- Let U and \mathbf{x} be as in Example 6, and let $\mathbf{y} = \begin{bmatrix} -3\sqrt{2} \\ 6 \end{bmatrix}$. Verify that $U\mathbf{x} \cdot U\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$.
- 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.

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

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

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

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.

- $\mathbf{u}_1 = \begin{bmatrix} 2 \\ -3 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$, 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 $\mathbf{x} = \begin{bmatrix} -6 \\ 3 \end{bmatrix}$

9. $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} 8 \\ -4 \\ -3 \end{bmatrix}$

10. $\mathbf{u}_1 = \begin{bmatrix} 3 \\ -3 \\ 0 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} 5 \\ -3 \\ 1 \end{bmatrix}$

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 $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ 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 $\text{Span}\{\mathbf{u}\}$ and one orthogonal to \mathbf{u} .

14. Let $\mathbf{y} = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 7 \\ 1 \end{bmatrix}$. Write \mathbf{y} as the sum of a vector in $\text{Span}\{\mathbf{u}\}$ 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 $\mathbf{y} = \begin{bmatrix} -3 \\ 9 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. 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. $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$

19. $\begin{bmatrix} -.6 \\ .8 \end{bmatrix}, \begin{bmatrix} .8 \\ .6 \end{bmatrix}$

20. $\begin{bmatrix} -2/3 \\ 1/3 \\ 2/3 \end{bmatrix}, \begin{bmatrix} 1/3 \\ 2/3 \\ 0 \end{bmatrix}$

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 and 24, all vectors are in \mathbb{R}^n . Mark each statement True or False. Justify each answer.

23. a. Not every linearly independent set in \mathbb{R}^n is an orthogonal set.

b. 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.

c. If the vectors in an orthogonal set of nonzero vectors are normalized, then some of the new vectors may not be orthogonal.

d. A matrix with orthonormal columns is an orthogonal matrix.

e. 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 .

24. a. Not every orthogonal set in \mathbb{R}^n is linearly independent.

b. 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.

c. If the columns of an $m \times n$ matrix A are orthonormal, then the linear mapping $\mathbf{x} \mapsto A\mathbf{x}$ preserves lengths.

d. The orthogonal projection of \mathbf{y} onto \mathbf{v} is the same as the orthogonal projection of \mathbf{y} onto $c\mathbf{v}$ whenever $c \neq 0$.

e. An orthogonal matrix is invertible.

25. Prove Theorem 7. [Hint: For (a), compute $\|U\mathbf{x}\|^2$, or prove (b) first.]

26. Suppose W is a subspace of \mathbb{R}^n spanned by n nonzero orthogonal vectors. Explain why $W = \mathbb{R}^n$.

27. Let U be a square matrix with orthonormal columns. Explain why U is invertible. (Mention the theorems you use.)

28. Let U be an $n \times n$ orthogonal matrix. Show that the rows of U form an orthonormal basis of \mathbb{R}^n .

29. 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 $(UV)^T$.]

30. Let U be an orthogonal matrix, and construct V by interchanging some of the columns of U . Explain why V is an orthogonal matrix.

31. 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 $\hat{\mathbf{y}}$.

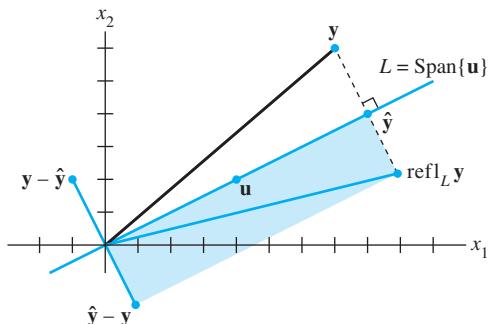
32. Let $\{\mathbf{v}_1, \mathbf{v}_2\}$ be an orthogonal set of nonzero vectors, and let c_1, c_2 be any nonzero scalars. Show that $\{c_1\mathbf{v}_1, c_2\mathbf{v}_2\}$ is also an orthogonal set. Since orthogonality of a set is defined in terms of pairs of vectors, this shows that if the vectors in an orthogonal set are normalized, the new set will still be orthogonal.

33. 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.

34. 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 \mathbf{y} in L is the point $\text{refl}_L \mathbf{y}$ defined by

$$\text{refl}_L \mathbf{y} = 2 \cdot \text{proj}_L \mathbf{y} - \mathbf{y}$$

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.

35. [M] Show that the columns of the matrix A are orthogonal by making an appropriate matrix calculation. State the calculation you use.

$$A = \begin{bmatrix} -6 & -3 & 6 & 1 \\ -1 & 2 & 1 & -6 \\ 3 & 6 & 3 & -2 \\ 6 & -3 & 6 & -1 \\ 2 & -1 & 2 & 3 \\ -3 & 6 & 3 & 2 \\ -2 & -1 & 2 & -3 \\ 1 & 2 & 1 & 6 \end{bmatrix}$$

36. [M] In parts (a)–(d), let U be the matrix formed by normalizing each column of the matrix A in Exercise 35.
- Compute $U^T U$ and $U U^T$. How do they differ?
 - Generate a random vector \mathbf{y} in \mathbb{R}^8 , and compute $\mathbf{p} = U U^T \mathbf{y}$ and $\mathbf{z} = \mathbf{y} - \mathbf{p}$. Explain why \mathbf{p} is in $\text{Col } A$. Verify that \mathbf{z} is orthogonal to \mathbf{p} .
 - Verify that \mathbf{z} is orthogonal to each column of U .
 - Notice that $\mathbf{y} = \mathbf{p} + \mathbf{z}$, with \mathbf{p} in $\text{Col } A$. 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 seem to depend on the \mathbf{u} chosen on the line. See Exercise 31.

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

$$U \mathbf{x} \cdot U \mathbf{y} = 3 + 7 + 2 = 12, \quad \text{and} \quad \mathbf{x} \cdot \mathbf{y} = -6 + 18 = 12$$

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, mentioned in the introductory example for this chapter. The full story will be told in Section 6.5.

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

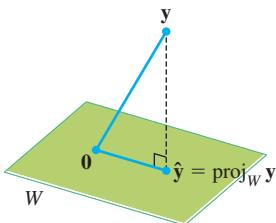


FIGURE 1

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 $\mathbf{z}_1 = c_1\mathbf{u}_1 + c_2\mathbf{u}_2$ is in $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$

and $\mathbf{z}_2 = c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5$ 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

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \quad (1)$$

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

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p} \mathbf{u}_p \quad (2)$$

and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.

The vector $\hat{\mathbf{y}}$ in (1) is called the **orthogonal projection of \mathbf{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.

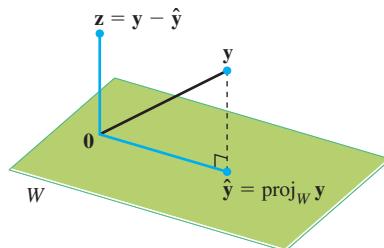


FIGURE 2 The orthogonal projection of \mathbf{y} onto W .

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 .

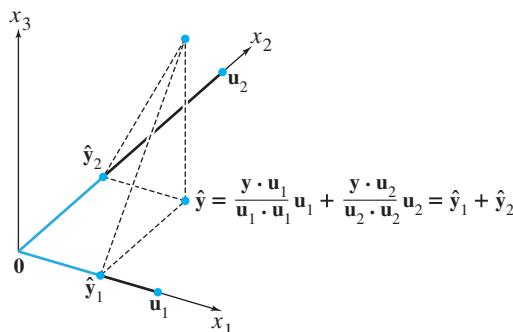


FIGURE 3 The orthogonal projection of \mathbf{y} is the sum of its projections onto one-dimensional subspaces that are mutually orthogonal.

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

$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\| \quad (3)$$

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 \mathbf{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 were 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 colored right triangle 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. ■

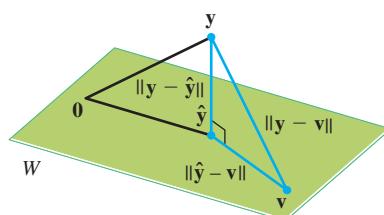


FIGURE 4 The orthogonal projection of \mathbf{y} onto W is the closest point in W to \mathbf{y} .

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

$$\text{proj}_W \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1) \mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2) \mathbf{u}_2 + \cdots + (\mathbf{y} \cdot \mathbf{u}_p) \mathbf{u}_p \quad (4)$$

If $U = [\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_p]$, then

$$\text{proj}_W \mathbf{y} = UU^T \mathbf{y} \quad \text{for all } \mathbf{y} \text{ in } \mathbb{R}^n \quad (5)$$

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). ■

WEB

Suppose U is an $n \times p$ matrix with orthonormal columns, and let W be the column space of U . Then

$$U^T U \mathbf{x} = I_p \mathbf{x} = \mathbf{x} \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^p \quad \text{Theorem 6}$$

$$UU^T \mathbf{y} = \text{proj}_W \mathbf{y} \quad \text{for all } \mathbf{y} \text{ in } \mathbb{R}^n \quad \text{Theorem 10}$$

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.

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

$\text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ and the other in $\text{Span}\{\mathbf{u}_4\}$.

2. $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 1 \\ 1 \\ -2 \\ -1 \end{bmatrix}$, $\mathbf{u}_4 = \begin{bmatrix} -1 \\ 1 \\ 1 \\ -2 \end{bmatrix}$,

$\mathbf{v} = \begin{bmatrix} 4 \\ 5 \\ -3 \\ 3 \end{bmatrix}$. Write \mathbf{v} as the sum of two vectors, one in

$\text{Span}\{\mathbf{u}_1\}$ and the other in $\text{Span}\{\mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$.

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 $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$.

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. $\mathbf{y} = \begin{bmatrix} 6 \\ 3 \\ -2 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -4 \\ 3 \\ 0 \end{bmatrix}$

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. $\mathbf{y} = \begin{bmatrix} 6 \\ 4 \\ 1 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} -4 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

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. $\mathbf{y} = \begin{bmatrix} -1 \\ 4 \\ 3 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -1 \\ 3 \\ -2 \end{bmatrix}$

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. $\mathbf{y} = \begin{bmatrix} 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 0 \\ -1 \\ 1 \\ -1 \end{bmatrix}$

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. $\mathbf{y} = \begin{bmatrix} 3 \\ -1 \\ 1 \\ 13 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \\ -1 \\ 2 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -4 \\ 1 \\ 0 \\ 3 \end{bmatrix}$

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 $\mathbf{y} = \begin{bmatrix} 5 \\ -9 \\ 5 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} -3 \\ -5 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. 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 $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$.

- Let $U = [\mathbf{u}_1 \ \mathbf{u}_2]$. Compute $U^T U$ and $U U^T$.
- Compute $\text{proj}_W \mathbf{y}$ and $(U U^T) \mathbf{y}$.

18. Let $\mathbf{y} = \begin{bmatrix} 7 \\ 9 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{bmatrix}$, and $W = \text{Span}\{\mathbf{u}_1\}$.

- Let U be the 2×1 matrix whose only column is \mathbf{u}_1 . Compute $U^T U$ and $U U^T$.
- Compute $\text{proj}_W \mathbf{y}$ and $(U U^T) \mathbf{y}$.

19. Let $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix}$, 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 $\mathbf{u}_4 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. 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 and 22, all vectors and subspaces are in \mathbb{R}^n . Mark each statement True or False. Justify each answer.

- 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 .
- For each \mathbf{y} and each subspace W , the vector $\mathbf{y} - \text{proj}_W \mathbf{y}$ is orthogonal to W .
- 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 $\hat{\mathbf{y}}$.
- If \mathbf{y} is in a subspace W , then the orthogonal projection of \mathbf{y} onto W is \mathbf{y} itself.

- If the columns of an $n \times p$ matrix U are orthonormal, then $U U^T \mathbf{y}$ is the orthogonal projection of \mathbf{y} onto the column space of U .

- 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.
 - 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 .
 - 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 .
 - The best approximation to \mathbf{y} by elements of a subspace W is given by the vector $\mathbf{y} - \text{proj}_W \mathbf{y}$.
 - If an $n \times p$ matrix U has orthonormal columns, then $U U^T \mathbf{x} = \mathbf{x}$ for all \mathbf{x} in \mathbb{R}^n .
- 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 $\text{Nul } A$. 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 $A\mathbf{p} = \mathbf{b}$.
 - Let W be a subspace of \mathbb{R}^n with an orthogonal basis $\{\mathbf{w}_1, \dots, \mathbf{w}_p\}$, and let $\{\mathbf{v}_1, \dots, \mathbf{v}_q\}$ be an orthogonal basis for W^\perp .
 - Explain why $\{\mathbf{w}_1, \dots, \mathbf{w}_p, \mathbf{v}_1, \dots, \mathbf{v}_q\}$ is an orthogonal set.
 - Explain why the set in part (a) spans \mathbb{R}^n .
 - Show that $\dim W + \dim W^\perp = n$.
 - [M] Let U be the 8×4 matrix in Exercise 36 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.
 - [M] Let U be the matrix in Exercise 25. 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.

- Using Theorem 10, let U be a matrix whose columns consist of an orthonormal basis for W . Then $\text{proj}_W \mathbf{z} = U U^T \mathbf{z} = U U^T (\mathbf{x} + \mathbf{y}) = U U^T \mathbf{x} + U U^T \mathbf{y} = \text{proj}_W \mathbf{x} + \text{proj}_W \mathbf{y} = \mathbf{u} + \mathbf{v}$.