5 Orthogonality

. . . that sprightly Scot of Scots, Douglas, that runs a-horseback up a hill perpendicular—

—William Shakespeare
Henry IV, Part I
Act II, Scene IV

5.0 Introduction: Shadows on a Wall

In this chapter, we will extend the notion of orthogonal projection that we encountered first in Chapter 1 and then again in Chapter 3. Until now, we have discussed only projection onto a single vector (or, equivalently, the one-dimensional subspace spanned by that vector). In this section, we will see if we can find the analogous formulas for projection onto a plane in $\mathbb{R}^3$. Figure 5.1 shows what happens, for example, when parallel light rays create a shadow on a wall. A similar process occurs when a three-dimensional object is displayed on a two-dimensional screen, such as a computer monitor. Later in this chapter, we will consider these ideas in full generality.

To begin, let's take another look at what we already know about projections. In Section 3.6, we showed that, in $\mathbb{R}^2$, the standard matrix of a projection onto the line through the origin with direction vector $\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$ is

$$P = \frac{1}{d_1^2 + d_2^2}\begin{bmatrix} d_1^2 & d_1 d_2 \\ d_1 d_2 & d_2^2 \end{bmatrix} = \begin{bmatrix} d_1^2/(d_1^2 + d_2^2) & d_1 d_2/(d_1^2 + d_2^2) \\ d_1 d_2/(d_1^2 + d_2^2) & d_2^2/(d_1^2 + d_2^2) \end{bmatrix}$$

Hence, the projection of the vector $\mathbf{v}$ onto this line is just $P\mathbf{v}$.

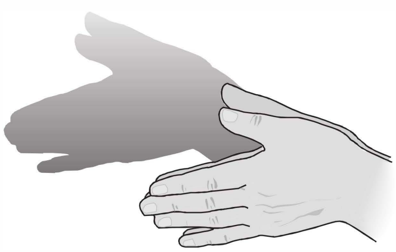


Figure 5.1
Shadows on a wall are projections

Problem 1 Show that P can be written in the equivalent form

$$P = \begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{bmatrix}$$

(What does θ represent here?)

Problem 2 Show that P can also be written in the form $P = \mathbf{u}\mathbf{u}^T$, where $\mathbf{u}$ is a *unit* vector in the direction of $\mathbf{d}$.

Problem 3 Using Problem 2, find P and then find the projection of $\mathbf{v} = \begin{bmatrix} 3 \\ -4 \end{bmatrix}$ onto the lines with the following unit direction vectors:

$$\text{(a) } \mathbf{u} = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \quad \text{(b) } \mathbf{u} = \begin{bmatrix} \frac{4}{5} \\ \frac{3}{5} \end{bmatrix} \quad \text{(c) } \mathbf{u} = \begin{bmatrix} -\frac{3}{5} \\ \frac{4}{5} \end{bmatrix}$$

Problem 4 Using the form $P = \mathbf{u}\mathbf{u}^T$, show that (a) $P^T = P$ (i.e., P is symmetric) and (b) $P^2 = P$ (i.e., P is idempotent).

Problem 5 Explain why, if P is a 2×2 projection matrix, the line onto which it projects vectors is the column space of P.

Now we will move into $\mathbb{R}^3$ and consider projections onto planes through the origin. We will explore several approaches.

Figure 5.2 shows one way to proceed. If $\mathcal{P}$ is a plane through the origin in $\mathbb{R}^3$ with normal vector $\mathbf{n}$ and if $\mathbf{v}$ is a vector in $\mathbb{R}^3$, then $\mathbf{p} = \text{proj}_{\mathcal{P}}(\mathbf{v})$ is a vector in $\mathcal{P}$ such that $\mathbf{v} - c\mathbf{n} = \mathbf{p}$ for some scalar c.

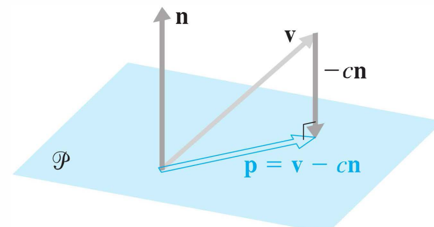


Figure 5.2
Projection onto a plane

Problem 6 Using the fact that $\mathbf{n}$ is orthogonal to every vector in $\mathcal{P}$, solve $\mathbf{v} - c\mathbf{n} = \mathbf{p}$ for c to find an expression for $\mathbf{p}$ in terms of $\mathbf{v}$ and $\mathbf{n}$.

Problem 7 Use the method of Problem 6 to find the projection of

$$\mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}$$

onto the planes with the following equations:

(a) $x + y + z = 0$ (b) $x - 2z = 0$ (c) $2x - 3y + z = 0$

Another approach to the problem of finding the projection of a vector onto a plane is suggested by Figure 5.3. We can decompose the projection of $\mathbf{v}$ onto $\mathcal{P}$ into the *sum* of its projections onto the direction vectors for $\mathcal{P}$. This works only if the direction vectors are orthogonal unit vectors. Accordingly, let $\mathbf{u}_1$ and $\mathbf{u}_2$ be direction vectors for $\mathcal{P}$ with the property that

$$\|\mathbf{u}_1\| = \|\mathbf{u}_2\| = 1 \quad \text{and} \quad \mathbf{u}_1 \cdot \mathbf{u}_2 = 0$$

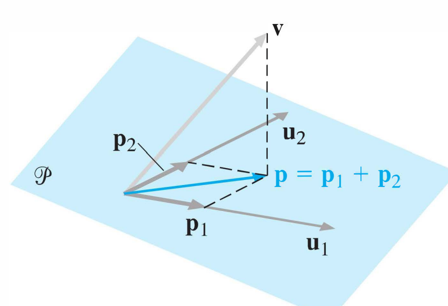


Figure 5.3

By Problem 2, the projections of $\mathbf{v}$ onto $\mathbf{u}_1$ and $\mathbf{u}_2$ are

$$\mathbf{p}_1 = \mathbf{u}_1\mathbf{u}_1^T\mathbf{v} \quad \text{and} \quad \mathbf{p}_2 = \mathbf{u}_2\mathbf{u}_2^T\mathbf{v}$$

respectively. To show that $\mathbf{p}_1 + \mathbf{p}_2$ gives the projection of $\mathbf{v}$ onto $\mathcal{P}$, we need to show that $\mathbf{v} - (\mathbf{p}_1 + \mathbf{p}_2)$ is orthogonal to $\mathcal{P}$. It is enough to show that $\mathbf{v} - (\mathbf{p}_1 + \mathbf{p}_2)$ is orthogonal to both $\mathbf{u}_1$ and $\mathbf{u}_2$. (Why?)

Problem 8 Show that $\mathbf{u}_1 \cdot (\mathbf{v} - (\mathbf{p}_1 + \mathbf{p}_2)) = 0$ and $\mathbf{u}_2 \cdot (\mathbf{v} - (\mathbf{p}_1 + \mathbf{p}_2)) = 0$. [*Hint:* Use the alternative form of the dot product, $\mathbf{x}^T\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$, together with the fact that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal unit vectors.]

It follows from Problem 8 and the comments preceding it that the matrix of the projection onto the subspace $\mathcal{P}$ of $\mathbb{R}^3$ spanned by orthogonal unit vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ is

$$P = \mathbf{u}_1\mathbf{u}_1^T + \mathbf{u}_2\mathbf{u}_2^T \tag{1}$$

Problem 9 Repeat Problem 7, using the formula for P given by Equation (1). Use the same $\mathbf{v}$ and use $\mathbf{u}_1$ and $\mathbf{u}_2$, as indicated below. (First, verify that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal unit vectors in the given plane.)

(a) $x + y + z = 0$ with $\mathbf{u}_1 = \begin{bmatrix} -2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$

(b) $x - 2z = 0$ with $\mathbf{u}_1 = \begin{bmatrix} 2/\sqrt{5} \\ 0 \\ 1/\sqrt{5} \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

(c) $2x - 3y + z = 0$ with $\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 2/\sqrt{6} \\ 1/\sqrt{6} \\ -1/\sqrt{6} \end{bmatrix}$

Problem 10 Show that a projection matrix given by Equation (1) satisfies properties (a) and (b) of Problem 4.

Problem 11 Show that the matrix P of a projection onto a plane in $\mathbb{R}^3$ can be expressed as

$$P = AA^T$$

for some 3×2 matrix A. [*Hint:* Show that Equation (1) is an outer product expansion.]

Problem 12 Show that if P is the matrix of a projection onto a plane in $\mathbb{R}^3$, then $\operatorname{rank}(P) = 2$.

In this chapter, we will look at the concepts of orthogonality and orthogonal projection in greater detail. We will see that the ideas introduced in this section can be generalized and that they have many important applications.

5.1 Orthogonality in $\mathbb{R}^n$

In this section, we will generalize the notion of orthogonality of vectors in $\mathbb{R}^n$ from two vectors to sets of vectors. In doing so, we will see that two properties make the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ of $\mathbb{R}^n$ easy to work with: First, any two distinct vectors in the set are orthogonal. Second, each vector in the set is a unit vector. These two properties lead us to the notion of orthogonal bases and orthonormal bases—concepts that we will be able to fruitfully apply to a variety of applications.

Orthogonal and Orthonormal Sets of Vectors

Definition A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ is called an ***orthogonal set*** if all pairs of distinct vectors in the set are orthogonal—that is, if

$$\mathbf{v}_i \cdot \mathbf{v}_j = 0 \quad \text{whenever} \quad i \neq j \quad \text{for } i, j = 1, 2, \ldots, k$$

The standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ of $\mathbb{R}^n$ is an orthogonal set, as is any subset of it. As the first example illustrates, there are many other possibilities.

Example 5.1

Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal set in $\mathbb{R}^3$ if

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

Solution We must show that every pair of vectors from this set is orthogonal. This is true, since

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = 2(0) + 1(1) + (-1)(1) = 0$$
$$\mathbf{v}_2 \cdot \mathbf{v}_3 = 0(1) + 1(-1) + (1)(1) = 0$$
$$\mathbf{v}_1 \cdot \mathbf{v}_3 = 2(1) + 1(-1) + (-1)(1) = 0$$

Geometrically, the vectors in Example 5.1 are mutually perpendicular, as Figure 5.4 shows.

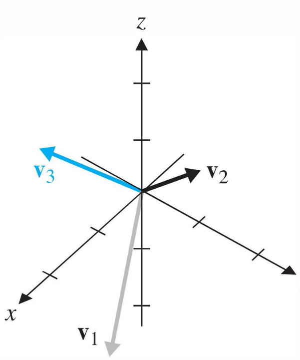


Figure 5.4
An orthogonal set of vectors

One of the main advantages of working with orthogonal sets of vectors is that they are necessarily linearly independent, as Theorem 5.1 shows.

Theorem 5.1 If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then these vectors are linearly independent.

Proof If $c_1, \ldots, c_k$ are scalars such that $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$, then

$$(c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) \cdot \mathbf{v}_i = \mathbf{0} \cdot \mathbf{v}_i = 0$$

or, equivalently,

$$c_1(\mathbf{v}_1 \cdot \mathbf{v}_i) + \cdots + c_i(\mathbf{v}_i \cdot \mathbf{v}_i) + \cdots + c_k(\mathbf{v}_k \cdot \mathbf{v}_i) = 0 \quad (1)$$

Since $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is an orthogonal set, all of the dot products in Equation (1) are zero, except $\mathbf{v}_i \cdot \mathbf{v}_i$. Thus, Equation (1) reduces to

$$c_i(\mathbf{v}_i \cdot \mathbf{v}_i) = 0$$

Now, $\mathbf{v}_i \cdot \mathbf{v}_i \neq 0$ because $\mathbf{v}_i \neq \mathbf{0}$ by hypothesis. So we must have $c_i = 0$. The fact that this is true for all $i = 1, \ldots, k$ implies that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is a linearly independent set.

Remark Thanks to Theorem 5.1, we know that if a set of vectors is orthogonal, it is automatically linearly independent. For example, we can immediately deduce that the three vectors in Example 5.1 are linearly independent. Contrast this approach with the work needed to establish their linear independence directly!

Definition An ***orthogonal basis*** for a subspace W of $\mathbb{R}^n$ is a basis of W that is an orthogonal set.

Example 5.2

The vectors

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

from Example 5.1 are orthogonal and, hence, linearly independent. Since any three linearly independent vectors in $\mathbb{R}^3$ form a basis for $\mathbb{R}^3$, by the Fundamental Theorem of Invertible Matrices, it follows that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal basis for $\mathbb{R}^3$.

Remark In Example 5.2, suppose only the orthogonal vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ were given and you were asked to find a third vector $\mathbf{v}_3$ to make $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ an orthogonal basis for $\mathbb{R}^3$. One way to do this is to remember that in $\mathbb{R}^3$, the cross product of two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ is orthogonal to each of them. (See Exploration: The Cross Product in Chapter 1.) Hence we may take

$$\mathbf{v}_3 = \mathbf{v}_1 \times \mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 2 \end{bmatrix}$$

Note that the resulting vector is a multiple of the vector $\mathbf{v}_3$ in Example 5.2, as it must be.

Example 5.3

Find an orthogonal basis for the subspace W of $\mathbb{R}^3$ given by

$$W = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : x - y + 2z = 0 \right\}$$

Solution Section 5.3 gives a general procedure for problems of this sort. For now, we will find the orthogonal basis by brute force. The subspace W is a plane through the origin in $\mathbb{R}^3$. From the equation of the plane, we have $x = y - 2z$, so W consists of vectors of the form

$$\begin{bmatrix} y - 2z \\ y \\ z \end{bmatrix} = y \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

It follows that $\mathbf{u} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$ are a basis for W, but they are *not* orthogonal. It suffices to find another nonzero vector in W that is orthogonal to either one of these.

Suppose $\mathbf{w} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ is a vector in W that is orthogonal to $\mathbf{u}$. Then $x - y + 2z = 0$, since $\mathbf{w}$ is in the plane W. Since $\mathbf{u} \cdot \mathbf{w} = 0$, we also have $x + y = 0$. Solving the linear system

$$\begin{aligned} x - y + 2z &= 0 \\ x + y \qquad &= 0 \end{aligned}$$

we find that $x = -z$ and $y = z$. (Check this.) Thus, any nonzero vector $\mathbf{w}$ of the form

$$\mathbf{w} = \begin{bmatrix} -z \\ z \\ z \end{bmatrix}$$

will do. To be specific, we could take $\mathbf{w} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$. It is easy to check that $\{\mathbf{u}, \mathbf{w}\}$ is an orthogonal set in W and, hence, an orthogonal basis for W, since $\dim W = 2$.

Another advantage of working with an orthogonal basis is that the coordinates of a vector with respect to such a basis are easy to compute. Indeed, there is a formula for these coordinates, as the following theorem establishes.

Theorem 5.2 Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$ and let $\mathbf{w}$ be any vector in W. Then the unique scalars $c_1, \ldots, c_k$ such that

$$\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$$

are given by

$$c_i = \frac{\mathbf{w} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i} \quad \text{for } i = 1, \ldots, k$$

Proof Since $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is a basis for W, we know that there are unique scalars $c_1, \ldots, c_k$ such that $\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$ (from Theorem 3.29). To establish the formula for c_i, we take the dot product of this linear combination with $\mathbf{v}_i$ to obtain

$$\begin{aligned} \mathbf{w} \cdot \mathbf{v}_i &= (c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) \cdot \mathbf{v}_i \\ &= c_1(\mathbf{v}_1 \cdot \mathbf{v}_i) + \cdots + c_i(\mathbf{v}_i \cdot \mathbf{v}_i) + \cdots + c_k(\mathbf{v}_k \cdot \mathbf{v}_i) \\ &= c_i(\mathbf{v}_i \cdot \mathbf{v}_i) \end{aligned}$$

since $\mathbf{v}_j \cdot \mathbf{v}_i = 0$ for $j \neq i$. Since $\mathbf{v}_i \neq \mathbf{0}$, $\mathbf{v}_i \cdot \mathbf{v}_i \neq 0$. Dividing by $\mathbf{v}_i \cdot \mathbf{v}_i$, we obtain the desired result.

Example 5.4

Find the coordinates of $\mathbf{w} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ with respect to the orthogonal basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ of Examples 5.1 and 5.2.

Solution Using Theorem 5.2, we compute

$$c_1 = \frac{\mathbf{w} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} = \frac{2 + 2 - 3}{4 + 1 + 1} = \frac{1}{6}$$

$$c_2 = \frac{\mathbf{w} \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2} = \frac{0 + 2 + 3}{0 + 1 + 1} = \frac{5}{2}$$

$$c_3 = \frac{\mathbf{w} \cdot \mathbf{v}_3}{\mathbf{v}_3 \cdot \mathbf{v}_3} = \frac{1 - 2 + 3}{1 + 1 + 1} = \frac{2}{3}$$

Thus,

$$\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = \tfrac{1}{6}\mathbf{v}_1 + \tfrac{5}{2}\mathbf{v}_2 + \tfrac{2}{3}\mathbf{v}_3$$

(Check this.) With the notation introduced in Section 3.5, we can also write the above equation as

$$[\mathbf{w}]_{\mathcal{B}} = \begin{bmatrix} \frac{1}{6} \\ \frac{5}{2} \\ \frac{2}{3} \end{bmatrix}$$

Compare the procedure in Example 5.4 with the work required to find these coordinates directly and you should start to appreciate the value of orthogonal bases.

As noted at the beginning of this section, the other property of the standard basis in $\mathbb{R}^n$ is that each standard basis vector is a unit vector. Combining this property with orthogonality, we have the following definition.

Definition A set of vectors in $\mathbb{R}^n$ is an ***orthonormal set*** if it is an orthogonal set of unit vectors. An ***orthonormal basis*** for a subspace W of $\mathbb{R}^n$ is a basis of W that is an orthonormal set.

Remark If $S = \{\mathbf{q}_1, \ldots, \mathbf{q}_k\}$ is an orthonormal set of vectors, then $\mathbf{q}_i \cdot \mathbf{q}_j = 0$ for $i \neq j$ and $\|\mathbf{q}_i\| = 1$. The fact that each $\mathbf{q}_i$ is a unit vector is equivalent to $\mathbf{q}_i \cdot \mathbf{q}_i = 1$. It follows that we can summarize the statement that S is orthonormal as

$$\mathbf{q}_i \cdot \mathbf{q}_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

Example 5.5

Show that $S = \{\mathbf{q}_1, \mathbf{q}_2\}$ is an orthonormal set in $\mathbb{R}^3$ if

$$\mathbf{q}_1 = \begin{bmatrix} 1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix} \quad \text{and} \quad \mathbf{q}_2 = \begin{bmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$$

Solution We check that

$$\begin{aligned} \mathbf{q}_1 \cdot \mathbf{q}_2 &= 1/\sqrt{18} - 2/\sqrt{18} + 1/\sqrt{18} = 0 \\ \mathbf{q}_1 \cdot \mathbf{q}_1 &= 1/3 + 1/3 + 1/3 = 1 \\ \mathbf{q}_2 \cdot \mathbf{q}_2 &= 1/6 + 4/6 + 1/6 = 1 \end{aligned}$$

If we have an orthogonal set, we can easily obtain an orthonormal set from it: We simply normalize each vector.

Example 5.6

Construct an orthonormal basis for $\mathbb{R}^3$ from the vectors in Example 5.1.

Solution Since we already know that $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are an orthogonal basis, we normalize them to get

$$\mathbf{q}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \frac{1}{\sqrt{6}}\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{6} \\ 1/\sqrt{6} \\ -1/\sqrt{6} \end{bmatrix}$$

$$\mathbf{q}_2 = \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$$\mathbf{q}_3 = \frac{1}{\|\mathbf{v}_3\|}\mathbf{v}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

Then $\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$ is an orthonormal basis for $\mathbb{R}^3$.

Since any orthonormal set of vectors is, in particular, orthogonal, it is linearly independent, by Theorem 5.1. If we have an orthonormal basis, Theorem 5.2 becomes even simpler.

Theorem 5.3

Let $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_k\}$ be an orthonormal basis for a subspace W of $\mathbb{R}^n$ and let $\mathbf{w}$ be any vector in W. Then

$$\mathbf{w} = (\mathbf{w} \cdot \mathbf{q}_1)\mathbf{q}_1 + (\mathbf{w} \cdot \mathbf{q}_2)\mathbf{q}_2 + \cdots + (\mathbf{w} \cdot \mathbf{q}_k)\mathbf{q}_k$$

and this representation is unique.

Proof Apply Theorem 5.2 and use the fact that $\mathbf{q}_i \cdot \mathbf{q}_i = 1$ for $i = 1, \ldots, k$.

Orthogonal Matrices

Matrices whose columns form an orthonormal set arise frequently in applications, as you will see in Section 5.5. Such matrices have several attractive properties, which we now examine.

Theorem 5.4 The columns of an $m \times n$ matrix Q form an orthonormal set if and only if $Q^TQ = I_n$.

Proof We need to show that

$$(Q^TQ)_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

Let $\mathbf{q}_i$ denote the ith column of Q (and, hence, the ith row of Q^T). Since the (i, j) entry of Q^TQ is the dot product of the ith row of Q^T and the jth column of Q, it follows that

$$(Q^TQ)_{ij} = \mathbf{q}_i \cdot \mathbf{q}_j \tag{2}$$

by the definition of matrix multiplication.

Now the columns Q form an orthonormal set if and only if

$$\mathbf{q}_i \cdot \mathbf{q}_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

which, by Equation (2), holds if and only if

$$(Q^TQ)_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

This completes the proof.

If the matrix Q in Theorem 5.4 is a *square* matrix, it has a special name.

Orthogonal matrix is an unfortunate bit of terminology. "Orthonormal matrix" would clearly be a better term, but it is not standard. Moreover, there is no term for a nonsquare matrix with orthonormal columns.

Definition An $n \times n$ matrix Q whose columns form an orthonormal set is called an ***orthogonal matrix.***

The most important fact about orthogonal matrices is given by the next theorem.

Theorem 5.5 A square matrix Q is orthogonal if and only if $Q^{-1} = Q^T$.

Proof By Theorem 5.4, Q is orthogonal if and only if $Q^TQ = I$. This is true if and only if Q is invertible and $Q^{-1} = Q^T$, by Theorem 3.13.

Example 5.7

Show that the following matrices are orthogonal and find their inverses:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

Solution The columns of A are just the standard basis vectors for $\mathbb{R}^3$, which are clearly orthonormal. Hence, A is orthogonal and

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

For B, we check directly that

$$B^TB = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

$$= \begin{bmatrix} \cos^2\theta + \sin^2\theta & -\cos\theta\sin\theta + \sin\theta\cos\theta \\ -\sin\theta\cos\theta + \cos\theta\sin\theta & \sin^2\theta + \cos^2\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

Therefore, B is orthogonal, by Theorem 5.5, and

$$B^{-1} = B^T = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

Remark Matrix A in Example 5.7 is an example of a permutation matrix, a matrix obtained by permuting the columns of an identity matrix. In general, any $n \times n$ permutation matrix is orthogonal (see Exercise 25). Matrix B is the matrix of a rotation through the angle θ in $\mathbb{R}^2$. Any rotation has the property that it is a *length-preserving* transformation (known as an ***isometry*** in geometry). The next theorem shows that every orthogonal matrix transformation is an isometry. Orthogonal matrices also preserve dot products. In fact, orthogonal matrices are characterized by either one of these properties.

The word *isometry* literally means "length preserving," since it is derived from the Greek roots *isos* ("equal") and *metron* ("measure").

Theorem 5.6 Let Q be an $n \times n$ matrix. The following statements are equivalent:

a. Q is orthogonal.
b. $\|Q\mathbf{x}\| = \|\mathbf{x}\|$ for every $\mathbf{x}$ in $\mathbb{R}^n$.
c. $Q\mathbf{x}\cdot Q\mathbf{y} = \mathbf{x}\cdot\mathbf{y}$ for every $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$.

Proof We will prove that (a) $\Rightarrow$ (c) $\Rightarrow$ (b) $\Rightarrow$ (a). To do so, we will need to make use of the fact that if $\mathbf{x}$ and $\mathbf{y}$ are (column) vectors in $\mathbb{R}^n$, then $\mathbf{x}\cdot\mathbf{y} = \mathbf{x}^T\mathbf{y}$.

(a) $\Rightarrow$ (c) Assume that Q is orthogonal. Then $Q^TQ = I$, and we have

$$Q\mathbf{x}\cdot Q\mathbf{y} = (Q\mathbf{x})^TQ\mathbf{y} = \mathbf{x}^TQ^TQ\mathbf{y} = \mathbf{x}^TI\mathbf{y} = \mathbf{x}^T\mathbf{y} = \mathbf{x}\cdot\mathbf{y}$$

(c) $\Rightarrow$ (b) Assume that $Q\mathbf{x}\cdot Q\mathbf{y} = \mathbf{x}\cdot\mathbf{y}$ for every $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$. Then, taking $\mathbf{y} = \mathbf{x}$, we have $Q\mathbf{x}\cdot Q\mathbf{x} = \mathbf{x}\cdot\mathbf{x}$, so $\|Q\mathbf{x}\| = \sqrt{Q\mathbf{x}\cdot Q\mathbf{x}} = \sqrt{\mathbf{x}\cdot\mathbf{x}} = \|\mathbf{x}\|$.

(b) $\Rightarrow$ (a) Assume that property (b) holds and let $\mathbf{q}_i$ denote the ith column of Q. Using Exercise 63 in Section 1.2 and property (b), we have

$$\begin{aligned}\mathbf{x}\cdot\mathbf{y} &= \tfrac{1}{4}(\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2)\\ &= \tfrac{1}{4}(\|Q(\mathbf{x}+\mathbf{y})\|^2 - \|Q(\mathbf{x}-\mathbf{y})\|^2)\\ &= \tfrac{1}{4}(\|Q\mathbf{x}+Q\mathbf{y}\|^2 - \|Q\mathbf{x}-Q\mathbf{y}\|^2)\\ &= Q\mathbf{x}\cdot Q\mathbf{y}\end{aligned}$$

for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$. [This shows that (b) $\Rightarrow$ (c).]

Now if $\mathbf{e}_i$ is the ith standard basis vector, then $\mathbf{q}_i = Q\mathbf{e}_i$. Consequently,

$$\mathbf{q}_i\cdot\mathbf{q}_j = Q\mathbf{e}_i\cdot Q\mathbf{e}_j = \mathbf{e}_i\cdot\mathbf{e}_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

Thus, the columns of Q form an orthonormal set, so Q is an orthogonal matrix.

Looking at the orthogonal matrices A and B in Example 5.7, you may notice that not only do their columns form orthonormal sets—so do their *rows*. In fact, every orthogonal matrix has this property, as the next theorem shows.

Theorem 5.7 If Q is an orthogonal matrix, then its rows form an orthonormal set.

Proof From Theorem 5.5, we know that $Q^{-1} = Q^T$. Therefore,

$$(Q^T)^{-1} = (Q^{-1})^{-1} = Q = (Q^T)^T$$

so Q^T is an orthogonal matrix. Thus, the columns of Q^T—which are just the rows of Q—form an orthonormal set.

The final theorem in this section lists some other properties of orthogonal matrices.

Theorem 5.8 Let Q be an orthogonal matrix.

a. Q^{-1} is orthogonal.
b. $\det Q = \pm 1$
c. If λ is an eigenvalue of Q, then $|\lambda| = 1$.
d. If Q_1 and Q_2 are orthogonal $n \times n$ matrices, then so is Q_1Q_2.

Proof We will prove property (c) and leave the proofs of the remaining properties as exercises.

(c) Let λ be an eigenvalue of Q with corresponding eigenvector $\mathbf{v}$. Then $Q\mathbf{v} = \lambda\mathbf{v}$, and, using Theorem 5.6(b), we have

$$\|\mathbf{v}\| = \|Q\mathbf{v}\| = \|\lambda\mathbf{v}\| = |\lambda|\,\|\mathbf{v}\|$$

Since $\|\mathbf{v}\| \neq 0$, this implies that $|\lambda| = 1$.

a + bi

Remark Property (c) holds even for complex eigenvalues. The matrix $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ is orthogonal with eigenvalues i and $-i$, both of which have absolute value 1.

Exercises 5.1

In Exercises 1–6, determine which sets of vectors are orthogonal.

1. $\begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$

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

3. $\begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 4 \end{bmatrix}$

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

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

6. $\begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \\ 1 \end{bmatrix}$

In Exercises 7–10, show that the given vectors form an orthogonal basis for $\mathbb{R}^2$ or $\mathbb{R}^3$. Then use Theorem 5.2 to express $\mathbf{w}$ *as a linear combination of these basis vectors. Give the coordinate vector* $[\mathbf{w}]_{\mathcal{B}}$ *of* $\mathbf{w}$ *with respect to the basis* $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$ *of* $\mathbb{R}^2$ *or* $\mathcal{B} = \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ *of* $\mathbb{R}^3$.

7. $\mathbf{v}_1 = \begin{bmatrix} 4 \\ -2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}; \mathbf{w} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$

8. $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -6 \\ 2 \end{bmatrix}; \mathbf{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

9. $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}; \mathbf{w} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

10. $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}; \mathbf{w} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

In Exercises 11–15, determine whether the given orthogonal set of vectors is orthonormal. If it is not, normalize the vectors to form an orthonormal set.

11. $\begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix}, \begin{bmatrix} -\frac{4}{5} \\ \frac{3}{5} \end{bmatrix}$

12. $\begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}, \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$

13. $\begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}, \begin{bmatrix} \frac{2}{3} \\ -\frac{1}{3} \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -\frac{5}{2} \end{bmatrix}$

14. $\begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix}, \begin{bmatrix} 0 \\ \frac{1}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}, \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{6} \\ \frac{1}{6} \\ -\frac{1}{6} \end{bmatrix}$

15. $\begin{bmatrix} 1/2 \\ 1/2 \\ -1/2 \\ 1/2 \end{bmatrix}, \begin{bmatrix} 0 \\ \sqrt{6}/3 \\ 1/\sqrt{6} \\ -1/\sqrt{6} \end{bmatrix}, \begin{bmatrix} \sqrt{3}/2 \\ -\sqrt{3}/6 \\ \sqrt{3}/6 \\ -\sqrt{3}/6 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$

In Exercises 16–21, determine whether the given matrix is orthogonal. If it is, find its inverse.

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

17. $\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

18. $\begin{bmatrix} \frac{1}{3} & \frac{1}{2} & \frac{1}{5} \\ \frac{1}{3} & -\frac{1}{2} & \frac{1}{5} \\ -\frac{1}{3} & 0 & \frac{2}{5} \end{bmatrix}$

19. $\begin{bmatrix} \cos\theta\sin\theta & -\cos\theta & -\sin^2\theta \\ \cos^2\theta & \sin\theta & -\cos\theta\sin\theta \\ \sin\theta & 0 & \cos\theta \end{bmatrix}$

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

21. $\begin{bmatrix} 1 & 0 & 0 & 1/\sqrt{6} \\ 0 & 2/3 & 1/\sqrt{2} & 1/\sqrt{6} \\ 0 & -2/3 & 1/\sqrt{2} & -1/\sqrt{6} \\ 0 & 1/3 & 0 & 1/\sqrt{2} \end{bmatrix}$

22. Prove Theorem 5.8(a).

23. Prove Theorem 5.8(b).

24. Prove Theorem 5.8(d).

25. Prove that every permutation matrix is orthogonal.

26. If Q is an orthogonal matrix, prove that any matrix obtained by rearranging the rows of Q is also orthogonal.

27. Let Q be an orthogonal 2×2 matrix and let $\mathbf{x}$ and $\mathbf{y}$ be vectors in $\mathbb{R}^2$. If θ is the angle between $\mathbf{x}$ and $\mathbf{y}$, prove that the angle between $Q\mathbf{x}$ and $Q\mathbf{y}$ is also θ. (This proves that the linear transformations defined by orthogonal matrices are *angle-preserving* in $\mathbb{R}^2$, a fact that is true in general.)

28. **(a)** Prove that an orthogonal 2×2 matrix must have the form

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} a & b \\ b & -a \end{bmatrix}$$

where $\begin{bmatrix} a \\ b \end{bmatrix}$ is a unit vector.

(b) Using part (a), show that every orthogonal 2×2 matrix is of the form

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

where $0 \le \theta < 2\pi$.

(c) Show that every orthogonal 2×2 matrix corresponds to either a rotation or a reflection in $\mathbb{R}^2$.

(d) Show that an orthogonal 2×2 matrix Q corresponds to a rotation in $\mathbb{R}^2$ if $\det Q = 1$ and a reflection in $\mathbb{R}^2$ if $\det Q = -1$.

In Exercises 29–32, use Exercise 28 to determine whether the given orthogonal matrix represents a rotation or a reflection. If it is a rotation, give the angle of rotation; if it is a reflection, give the line of reflection.

29. $\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

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

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

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

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

(a) Prove that $A(A^T + B^T)B = A + B$.

(b) Use part (a) to prove that, if $\det A + \det B = 0$, then $A + B$ is not invertible.

34. Let $\mathbf{x}$ be a unit vector in $\mathbb{R}^n$. Partition $\mathbf{x}$ as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \hdashline x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 \\ \mathbf{y} \end{bmatrix}$$

Let

$$Q = \left[\begin{array}{c:c} x_1 & \mathbf{y}^T \\ \hdashline \mathbf{y} & I - \left(\dfrac{1}{1 - x_1}\right)\mathbf{y}\mathbf{y}^T \end{array}\right]$$

Prove that Q is orthogonal. (This procedure gives a quick method for finding an orthonormal basis for $\mathbb{R}^n$ with a prescribed first vector $\mathbf{x}$, a construction that is frequently useful in applications.)

35. Prove that if an upper triangular matrix is orthogonal, then it must be a diagonal matrix.

36. Prove that if $n > m$, then there is no $m \times n$ matrix A such that $\|A\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

37. Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an orthonormal basis for $\mathbb{R}^n$.

(a) Prove that, for any $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$,

$$\mathbf{x} \cdot \mathbf{y} = (\mathbf{x} \cdot \mathbf{v}_1)(\mathbf{y} \cdot \mathbf{v}_1) + (\mathbf{x} \cdot \mathbf{v}_2)(\mathbf{y} \cdot \mathbf{v}_2) + \cdots + (\mathbf{x} \cdot \mathbf{v}_n)(\mathbf{y} \cdot \mathbf{v}_n)$$

(This identity is called ***Parseval's Identity.***)

(b) What does Parseval's Identity imply about the relationship between the dot products $\mathbf{x} \cdot \mathbf{y}$ and $[\mathbf{x}]_{\mathcal{B}} \cdot [\mathbf{y}]_{\mathcal{B}}$?

5.2 Orthogonal Complements and Orthogonal Projections

In this section, we generalize two concepts that we encountered in Chapter 1. The notion of a normal vector to a plane will be extended to orthogonal complements, and the projection of one vector onto another will give rise to the concept of orthogonal projection onto a subspace.

Orthogonal Complements

$W^{\perp}$ is pronounced "W perp."

A normal vector $\mathbf{n}$ to a plane is orthogonal to every vector in that plane. If the plane passes through the origin, then it is a subspace W of $\mathbb{R}^3$, as is span($\mathbf{n}$). Hence, we have two subspaces of $\mathbb{R}^3$ with the property that every vector of one is orthogonal to every vector of the other. This is the idea behind the following definition.

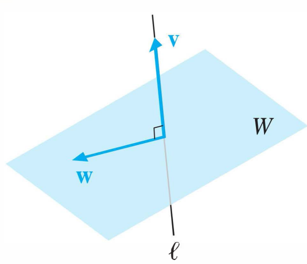


Figure 5.5
$\ell = W^{\perp}$ and $W = \ell^{\perp}$

Definition Let W be a subspace of $\mathbb{R}^n$. We say that a vector $\mathbf{v}$ in $\mathbb{R}^n$ is ***orthogonal to*** W if $\mathbf{v}$ is orthogonal to every vector in W. The set of all vectors that are orthogonal to W is called the ***orthogonal complement of*** W, denoted $W^{\perp}$. That is,

$$W^{\perp} = \{\mathbf{v} \text{ in } \mathbb{R}^n : \mathbf{v} \cdot \mathbf{w} = 0 \quad \text{for all } \mathbf{w} \text{ in } W\}$$

Example 5.8

If W is a plane through the origin in $\mathbb{R}^3$ and ℓ is the line through the origin perpendicular to W (i.e., parallel to the normal vector to W), then every vector $\mathbf{v}$ on ℓ is orthogonal to every vector $\mathbf{w}$ in W; hence, $\ell = W^{\perp}$. Moreover, W consists *precisely* of those vectors $\mathbf{w}$ that are orthogonal to every $\mathbf{v}$ on ℓ; hence, we also have $W = \ell^{\perp}$. Figure 5.5 illustrates this situation.

In Example 5.8, the orthogonal complement of a subspace turned out to be another subspace. Also, the complement of the complement of a subspace was the original subspace. These properties are true in general and are proved as properties (a) and (b) of Theorem 5.9. Properties (c) and (d) will also be useful. (Recall that the *intersection* $A \cap B$ of sets A and B consists of their common elements. See Appendix A.)

Theorem 5.9 Let W be a subspace of $\mathbb{R}^n$.

a. $W^{\perp}$ is a subspace of $\mathbb{R}^n$.
b. $(W^{\perp})^{\perp} = W$
c. $W \cap W^{\perp} = \{\mathbf{0}\}$
d. If $W = \text{span}(\mathbf{w}_1, \ldots, \mathbf{w}_k)$, then $\mathbf{v}$ is in $W^{\perp}$ if and only if $\mathbf{v} \cdot \mathbf{w}_i = 0$ for all $i = 1, \ldots, k$.

Proof (a) Since $\mathbf{0} \cdot \mathbf{w} = 0$ for all $\mathbf{w}$ in W, $\mathbf{0}$ is in $W^{\perp}$. Let $\mathbf{u}$ and $\mathbf{v}$ be in $W^{\perp}$ and let c be a scalar. Then

$$\mathbf{u} \cdot \mathbf{w} = \mathbf{v} \cdot \mathbf{w} = 0 \quad \text{for all } \mathbf{w} \text{ in } W$$

Therefore,

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

so $\mathbf{u} + \mathbf{v}$ is in $W^{\perp}$.

We also have

$$(c\mathbf{u}) \cdot \mathbf{w} = c(\mathbf{u} \cdot \mathbf{w}) = c(0) = 0$$

from which we see that $c\mathbf{u}$ is in $W^{\perp}$. It follows that $W^{\perp}$ is a subspace of $\mathbb{R}^n$.

(b) We will prove this property as Corollary 5.12.

(c) You are asked to prove this property in Exercise 23.

(d) You are asked to prove this property in Exercise 24.

We can now express some fundamental relationships involving the subspaces associated with an $m \times n$ matrix.

Theorem 5.10 Let A be an $m \times n$ matrix. Then 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:

$$(\text{row}(A))^{\perp} = \text{null}(A) \quad \text{and} \quad (\text{col}(A))^{\perp} = \text{null}(A^T)$$

Proof If $\mathbf{x}$ is a vector in $\mathbb{R}^n$, then $\mathbf{x}$ is in $(\text{row}(A))^{\perp}$ if and only if $\mathbf{x}$ is orthogonal to every row of A. But this is true if and only if $A\mathbf{x} = \mathbf{0}$, which is equivalent to $\mathbf{x}$ being in $\text{null}(A)$, so we have established the first identity. To prove the second identity, we simply replace A by A^T and use the fact that $\text{row}(A^T) = \text{col}(A)$.

Thus, an $m \times n$ matrix has four subspaces: $\text{row}(A)$, $\text{null}(A)$, $\text{col}(A)$, and $\text{null}(A^T)$. The first two are orthogonal complements in $\mathbb{R}^n$, and the last two are orthogonal

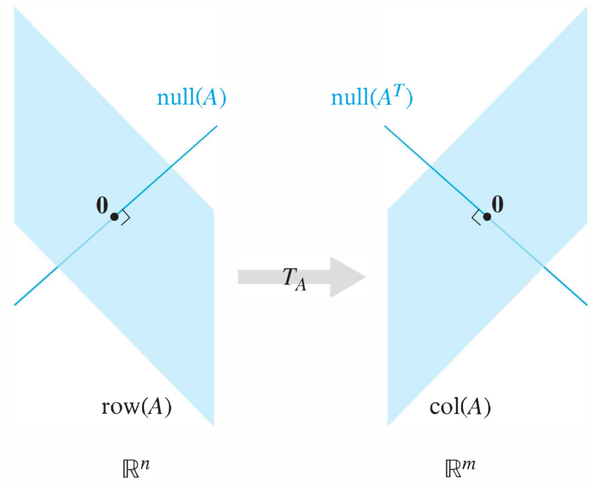


Figure 5.6
The four fundamental subspaces

complements in $\mathbb{R}^m$. The $m \times n$ matrix A defines a linear transformation from $\mathbb{R}^n$ into $\mathbb{R}^m$ whose range is col(A). Moreover, this transformation sends null(A) to $\mathbf{0}$ in $\mathbb{R}^m$. Figure 5.6 illustrates these ideas schematically. These four subspaces are called the ***fundamental subspaces*** of the $m \times n$ matrix A.

Example 5.9

Find bases for the four fundamental subspaces of

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

and verify Theorem 5.10.

Solution In Examples 3.45, 3.47, and 3.48, we computed bases for the row space, column space, and null space of A. We found that row(A) = span($\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$), where

$$\mathbf{u}_1 = [1 \quad 0 \quad 1 \quad 0 \quad -1], \quad \mathbf{u}_2 = [0 \quad 1 \quad 2 \quad 0 \quad 3], \quad \mathbf{u}_3 = [0 \quad 0 \quad 0 \quad 1 \quad 4]$$

Also, null(A) = span($\mathbf{x}_1, \mathbf{x}_2$), where

$$\mathbf{x}_1 = \begin{bmatrix} -1 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ -3 \\ 0 \\ -4 \\ 1 \end{bmatrix}$$

To show that $(\text{row}(A))^\perp = \text{null}(A)$, it is enough to show that every $\mathbf{u}_i$ is orthogonal to each $\mathbf{x}_j$, which is an easy exercise. (Why is this sufficient?)

The column space of A is $\text{col}(A) = \text{span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$, where

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 4 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 1 \\ -1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 1 \\ 1 \\ -2 \\ 1 \end{bmatrix}$$

We still need to compute the null space of A^T. Row reduction produces

$$[A^T \mid \mathbf{0}] = \left[\begin{array}{cccc|c} 1 & 2 & -3 & 4 & 0 \\ 1 & -1 & 2 & 1 & 0 \\ 3 & 0 & 1 & 6 & 0 \\ 1 & 1 & -2 & 1 & 0 \\ 6 & -1 & 1 & 3 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 6 & 0 \\ 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

So, if $\mathbf{y}$ is in the null space of A^T, then $y_1 = -y_4$, $y_2 = -6y_4$, and $y_3 = -3y_4$. It follows that

$$\text{null}(A^T) = \left\{ \begin{bmatrix} -y_4 \\ -6y_4 \\ -3y_4 \\ y_4 \end{bmatrix} \right\} = \text{span}\left(\begin{bmatrix} -1 \\ -6 \\ -3 \\ 1 \end{bmatrix} \right)$$

and it is easy to check that this vector is orthogonal to $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$.

The method of Example 5.9 is easily adapted to other situations.

Example 5.10

Let W be the subspace of $\mathbb{R}^5$ spanned by

$$\mathbf{w}_1 = \begin{bmatrix} 1 \\ -3 \\ 5 \\ 0 \\ 5 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} -1 \\ 1 \\ 2 \\ -2 \\ 3 \end{bmatrix}, \quad \mathbf{w}_3 = \begin{bmatrix} 0 \\ -1 \\ 4 \\ -1 \\ 5 \end{bmatrix}$$

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

Solution The subspace W spanned by $\mathbf{w}_1$, $\mathbf{w}_2$, and $\mathbf{w}_3$ is the same as the column space of

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

Therefore, by Theorem 5.10, $W^\perp = (\text{col}(A))^\perp = \text{null}(A^T)$, and we may proceed as in the previous example. We compute

$$[A^T \mid \mathbf{0}] = \left[\begin{array}{rrrrr|r} 1 & -3 & 5 & 0 & 5 & 0 \\ -1 & 1 & 2 & -2 & 3 & 0 \\ 0 & -1 & 4 & -1 & 5 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{rrrrr|r} 1 & 0 & 0 & 3 & 4 & 0 \\ 0 & 1 & 0 & 1 & 3 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 \end{array}\right]$$

Hence, $\mathbf{y}$ is in $W^\perp$ if and only if $y_1 = -3y_4 - 4y_5$, $y_2 = -y_4 - 3y_5$, and $y_3 = -2y_5$. It follows that

$$W^\perp = \left\{ \begin{bmatrix} -3y_4 - 4y_5 \\ -y_4 - 3y_5 \\ -2y_5 \\ y_4 \\ y_5 \end{bmatrix} \right\} = \text{span}\left(\begin{bmatrix} -3 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -4 \\ -3 \\ -2 \\ 0 \\ 1 \end{bmatrix} \right)$$

and these two vectors form a basis for $W^\perp$.

Orthogonal Projections

Recall that, in $\mathbb{R}^2$, the projection of a vector $\mathbf{v}$ onto a nonzero vector $\mathbf{u}$ is given by

$$\text{proj}_{\mathbf{u}}(\mathbf{v}) = \left(\frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}}\right)\mathbf{u}$$

Furthermore, the vector $\text{perp}_{\mathbf{u}}(\mathbf{v}) = \mathbf{v} - \text{proj}_{\mathbf{u}}(\mathbf{v})$ is orthogonal to $\text{proj}_{\mathbf{u}}(\mathbf{v})$, and we can decompose $\mathbf{v}$ as

$$\mathbf{v} = \text{proj}_{\mathbf{u}}(\mathbf{v}) + \text{perp}_{\mathbf{u}}(\mathbf{v})$$

as shown in Figure 5.7.

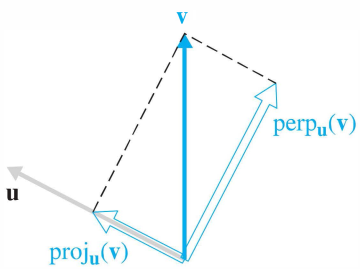


Figure 5.7
$\mathbf{v} = \text{proj}_{\mathbf{u}}(\mathbf{v}) + \text{perp}_{\mathbf{u}}(\mathbf{v})$

If we let $W = \text{span}(\mathbf{u})$, then $\mathbf{w} = \text{proj}_{\mathbf{u}}(\mathbf{v})$ is in W and $\mathbf{w}^\perp = \text{perp}_{\mathbf{u}}(\mathbf{v})$ is in $W^\perp$. We therefore have a way of "decomposing" $\mathbf{v}$ into the sum of two vectors, one from W and the other orthogonal to W—namely, $\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$. We now generalize this idea to $\mathbb{R}^n$.

Definition Let W be a subspace of $\mathbb{R}^n$ and let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be an orthogonal basis for W. For any vector $\mathbf{v}$ in $\mathbb{R}^n$, the ***orthogonal projection of* v *onto W*** is defined as

$$\text{proj}_W(\mathbf{v}) = \left(\frac{\mathbf{u}_1 \cdot \mathbf{v}}{\mathbf{u}_1 \cdot \mathbf{u}_1}\right)\mathbf{u}_1 + \cdots + \left(\frac{\mathbf{u}_k \cdot \mathbf{v}}{\mathbf{u}_k \cdot \mathbf{u}_k}\right)\mathbf{u}_k$$

The ***component of* v *orthogonal to W*** is the vector

$$\text{perp}_W(\mathbf{v}) = \mathbf{v} - \text{proj}_W(\mathbf{v})$$

Each summand in the definition of $\text{proj}_W(\mathbf{v})$ is also a projection onto a single vector (or, equivalently, the one-dimensional subspace spanned by it—in our previous sense). Therefore, with the notation of the preceding definition, we can write

$$\text{proj}_W(\mathbf{v}) = \text{proj}_{\mathbf{u}_1}(\mathbf{v}) + \cdots + \text{proj}_{\mathbf{u}_k}(\mathbf{v})$$

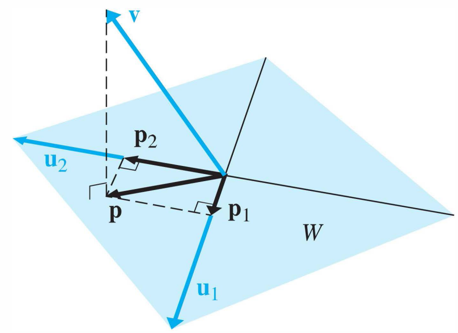


Figure 5.8
$\mathbf{p} = \mathbf{p}_1 + \mathbf{p}_2$

Since the vectors $\mathbf{u}_i$ are orthogonal, the orthogonal projection of $\mathbf{v}$ onto W is the sum of its projections onto one-dimensional subspaces that are mutually orthogonal. Figure 5.8 illustrates this situation with $W = \text{span}(\mathbf{u}_1, \mathbf{u}_2)$, $\mathbf{p} = \text{proj}_W(\mathbf{v})$, $\mathbf{p}_1 = \text{proj}_{\mathbf{u}_1}(\mathbf{v})$, and $\mathbf{p}_2 = \text{proj}_{\mathbf{u}_2}(\mathbf{v})$.

As a special case of the definition of $\text{proj}_W(\mathbf{v})$, we now also have a nice geometric interpretation of Theorem 5.2. In terms of our present notation and terminology, that theorem states that if $\mathbf{w}$ is in the subspace W of $\mathbb{R}^n$, which has orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$, then

$$\begin{aligned}\mathbf{w} &= \left(\frac{\mathbf{w}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\right)\mathbf{v}_1 + \cdots + \left(\frac{\mathbf{w}\cdot\mathbf{v}_k}{\mathbf{v}_k\cdot\mathbf{v}_k}\right)\mathbf{v}_k \\ &= \text{proj}_{\mathbf{v}_1}(\mathbf{w}) + \cdots + \text{proj}_{\mathbf{v}_k}(\mathbf{w})\end{aligned}$$

Thus, $\mathbf{w}$ is decomposed into a sum of orthogonal projections onto mutually orthogonal one-dimensional subspaces of W.

The definition above seems to depend on the choice of orthogonal basis; that is, a different basis $\{\mathbf{u}_1', \ldots, \mathbf{u}_k'\}$ for W would appear to give a "different" $\text{proj}_W(\mathbf{v})$ and $\text{perp}_W(\mathbf{v})$. Fortunately, this is not the case, as we will soon prove. For now, let's be content with an example.

Example 5.11

Let W be the plane in $\mathbb{R}^3$ with equation $x - y + 2z = 0$, and let $\mathbf{v} = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$. Find the orthogonal projection of $\mathbf{v}$ onto W and the component of $\mathbf{v}$ orthogonal to W.

Solution In Example 5.3, we found an orthogonal basis for W. Taking

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

we have

$$\begin{aligned}\mathbf{u}_1\cdot\mathbf{v} &= 2 \qquad \mathbf{u}_2\cdot\mathbf{v} = -2 \\ \mathbf{u}_1\cdot\mathbf{u}_1 &= 2 \qquad \mathbf{u}_2\cdot\mathbf{u}_2 = 3\end{aligned}$$

Therefore,

$$\begin{aligned}\text{proj}_W(\mathbf{v}) &= \left(\frac{\mathbf{u}_1\cdot\mathbf{v}}{\mathbf{u}_1\cdot\mathbf{u}_1}\right)\mathbf{u}_1 + \left(\frac{\mathbf{u}_2\cdot\mathbf{v}}{\mathbf{u}_2\cdot\mathbf{u}_2}\right)\mathbf{u}_2 \\ &= \tfrac{2}{2}\begin{bmatrix}1\\1\\0\end{bmatrix} - \tfrac{2}{3}\begin{bmatrix}-1\\1\\1\end{bmatrix} = \begin{bmatrix}\frac{5}{3}\\ \frac{1}{3}\\ -\frac{2}{3}\end{bmatrix}\end{aligned}$$

and $$\text{perp}_W(\mathbf{v}) = \mathbf{v} - \text{proj}_W(\mathbf{v}) = \begin{bmatrix}3\\-1\\2\end{bmatrix} - \begin{bmatrix}\frac{5}{3}\\ \frac{1}{3}\\ -\frac{2}{3}\end{bmatrix} = \begin{bmatrix}\frac{4}{3}\\ -\frac{4}{3}\\ \frac{8}{3}\end{bmatrix}$$

It is easy to see that $\text{proj}_W(\mathbf{v})$ is in W, since it satisfies the equation of the plane. It is equally easy to see that $\text{perp}_W(\mathbf{v})$ is orthogonal to W, since it is a scalar multiple of the normal vector $\begin{bmatrix}1\\-1\\2\end{bmatrix}$ to W. (See Figure 5.9.)

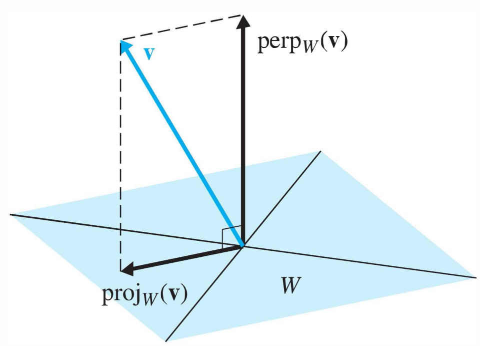


Figure 5.9
$\mathbf{v} = \text{proj}_W(\mathbf{v}) + \text{perp}_W(\mathbf{v})$

The next theorem shows that we can always find a decomposition of a vector with respect to a subspace and its orthogonal complement.

Theorem 5.11 **The Orthogonal Decomposition Theorem**

Let W be a subspace of $\mathbb{R}^n$ and let $\mathbf{v}$ be a vector in $\mathbb{R}^n$. Then there are unique vectors $\mathbf{w}$ in W and $\mathbf{w}^\perp$ in $W^\perp$ such that

$$\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$$

Proof We need to show two things: that such a decomposition *exists* and that it is *unique*.

To show existence, we choose an orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ for W. Let $\mathbf{w} = \text{proj}_W(\mathbf{v})$ and let $\mathbf{w}^\perp = \text{perp}_W(\mathbf{v})$. Then

$$\mathbf{w} + \mathbf{w}^\perp = \text{proj}_W(\mathbf{v}) + \text{perp}_W(\mathbf{v}) = \text{proj}_W(\mathbf{v}) + (\mathbf{v} - \text{proj}_W(\mathbf{v})) = \mathbf{v}$$

Clearly, $\mathbf{w} = \text{proj}_W(\mathbf{v})$ is in W, since it is a linear combination of the basis vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$. To show that $\mathbf{w}^\perp$ is in $W^\perp$, it is enough to show that $\mathbf{w}^\perp$ is orthogonal to each of the basis vectors $\mathbf{u}_i$, by Theorem 5.9(d). We compute

$$\begin{aligned}
\mathbf{u}_i \cdot \mathbf{w}^\perp &= u_i \cdot \text{perp}_W(\mathbf{v}) \\
&= \mathbf{u}_i \cdot (\mathbf{v} - \text{proj}_W(\mathbf{v})) \\
&= \mathbf{u}_i \cdot \left(\mathbf{v} - \left(\frac{\mathbf{u}_1 \cdot \mathbf{v}}{\mathbf{u}_1 \cdot \mathbf{u}_1}\right)\mathbf{u}_1 - \cdots - \left(\frac{\mathbf{u}_k \cdot \mathbf{v}}{\mathbf{u}_k \cdot \mathbf{u}_k}\right)\mathbf{u}_k\right) \\
&= \mathbf{u}_i \cdot \mathbf{v} - \left(\frac{\mathbf{u}_1 \cdot \mathbf{v}}{\mathbf{u}_1 \cdot \mathbf{u}_1}\right)(\mathbf{u}_i \cdot \mathbf{u}_1) - \cdots - \left(\frac{\mathbf{u}_i \cdot \mathbf{v}}{\mathbf{u}_i \cdot \mathbf{u}_i}\right)(\mathbf{u}_i \cdot \mathbf{u}_i) - \cdots \\
&\quad - \left(\frac{\mathbf{u}_k \cdot \mathbf{v}}{\mathbf{u}_k \cdot \mathbf{u}_k}\right)(\mathbf{u}_i \cdot \mathbf{u}_k) \\
&= \mathbf{u}_i \cdot \mathbf{v} - 0 - \cdots - \left(\frac{\mathbf{u}_i \cdot \mathbf{v}}{\mathbf{u}_i \cdot \mathbf{u}_i}\right)(\mathbf{u}_i \cdot \mathbf{u}_i) - \cdots - 0 \\
&= \mathbf{u}_i \cdot \mathbf{v} - \mathbf{u}_i \cdot \mathbf{v} = 0
\end{aligned}$$

since $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ for $j \neq i$. This proves that $\mathbf{w}^\perp$ is in $W^\perp$ and completes the existence part of the proof.

To show the uniqueness of this decomposition, let's suppose we have another decomposition $\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_1^\perp$, where $\mathbf{w}_1$ is in W and $\mathbf{w}_1^\perp$ is in $W^\perp$. Then $\mathbf{w} + \mathbf{w}^\perp = \mathbf{w}_1 + \mathbf{w}_1^\perp$, so

$$\mathbf{w} - \mathbf{w}_1 = \mathbf{w}_1^\perp - \mathbf{w}^\perp$$

But since $\mathbf{w} - \mathbf{w}_1$ is in W and $\mathbf{w}_1^\perp - \mathbf{w}^\perp$ is in $W^\perp$ (because these are subspaces), we know that this common vector is in $W \cap W^\perp = \{\mathbf{0}\}$ [using Theorem 5.9(c)]. Thus,

$$\mathbf{w} - \mathbf{w}_1 = \mathbf{w}_1^\perp - \mathbf{w}^\perp = 0$$

so $\mathbf{w}_1 = \mathbf{w}$ and $\mathbf{w}_1^\perp = \mathbf{w}_1$.

Example 5.11 illustrated the Orthogonal Decomposition Theorem. When W is the subspace of $\mathbb{R}^3$ given by the plane with equation $x - y + 2z = 0$, the orthogonal decomposition of $\mathbf{v} = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$ with respect to W is $\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$, where

$$\mathbf{w} = \text{proj}_W(\mathbf{v}) = \begin{bmatrix} \frac{5}{3} \\ \frac{1}{3} \\ -\frac{2}{3} \end{bmatrix} \quad \text{and} \quad \mathbf{w}^\perp = \text{perp}_W(\mathbf{v}) = \begin{bmatrix} \frac{4}{3} \\ -\frac{4}{3} \\ \frac{8}{3} \end{bmatrix}$$

The uniqueness of the orthogonal decomposition guarantees that the definitions of $\text{proj}_W(\mathbf{v})$ and $\text{perp}_W(\mathbf{v})$ do not depend on the choice of orthogonal basis. The Orthogonal Decomposition Theorem also allows us to prove property (b) of Theorem 5.9. We state that property here as a corollary to the Orthogonal Decomposition Theorem.

Corollary 5.12 If W is a subspace of $\mathbb{R}^n$, then

$$(W^\perp)^\perp = W$$

Proof If $\mathbf{w}$ is in W and $\mathbf{x}$ is in $W^{\perp}$, then $\mathbf{w} \cdot \mathbf{x} = 0$. But this now implies that $\mathbf{w}$ is in $(W^{\perp})^{\perp}$. Hence, $W \subseteq (W^{\perp})^{\perp}$. Now let $\mathbf{v}$ be in $(W^{\perp})^{\perp}$. By Theorem 5.11, we can write $\mathbf{v} = \mathbf{w} + \mathbf{w}^{\perp}$ for (unique) vectors $\mathbf{w}$ in W and $\mathbf{w}^{\perp}$ in $W^{\perp}$. But now

$$0 = \mathbf{v} \cdot \mathbf{w}^{\perp} = (\mathbf{w} + \mathbf{w}^{\perp}) \cdot \mathbf{w}^{\perp} = \mathbf{w} \cdot \mathbf{w}^{\perp} + \mathbf{w}^{\perp} \cdot \mathbf{w}^{\perp} = 0 + \mathbf{w}^{\perp} \cdot \mathbf{w}^{\perp} = \mathbf{w}^{\perp} \cdot \mathbf{w}^{\perp}$$

so $\mathbf{w}^{\perp} = \mathbf{0}$. Therefore, $\mathbf{v} = \mathbf{w} + \mathbf{w}^{\perp} = \mathbf{w}$, and thus $\mathbf{v}$ is in W. This shows that $(W^{\perp})^{\perp} \subseteq W$ and, since the reverse inclusion is also true, we conclude that $(W^{\perp})^{\perp} = W$, as required.

There is also a nice relationship between the dimensions of W and $W^{\perp}$, expressed in Theorem 5.13.

Theorem 5.13 If W is a subspace of $\mathbb{R}^n$, then

$$\dim W + \dim W^{\perp} = n$$

Proof Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be an orthogonal basis for W and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_l\}$ be an orthogonal basis for $W^{\perp}$. Then $\dim W = k$ and $\dim W^{\perp} = l$. Let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{v}_1, \ldots, \mathbf{v}_l\}$. We claim that $\mathcal{B}$ is an orthogonal basis for $\mathbb{R}^n$.

We first note that, since each $\mathbf{u}_i$ is in W and each $\mathbf{v}_j$ is in $W^{\perp}$,

$$\mathbf{u}_i \cdot \mathbf{v}_j = 0 \quad \text{for } i = 1, \ldots, k \text{ and } j = 1, \ldots, l$$

Thus, $\mathcal{B}$ is an orthogonal set and, hence, is linearly independent, by Theorem 5.1. Next, if $\mathbf{v}$ is a vector in $\mathbb{R}^n$, the Orthogonal Decomposition Theorem tells us that $\mathbf{v} = \mathbf{w} + \mathbf{w}^{\perp}$ for some $\mathbf{w}$ in W and $\mathbf{w}^{\perp}$ in $W^{\perp}$. Since $\mathbf{w}$ can be written as a linear combination of the vectors $\mathbf{u}_i$ and $\mathbf{w}^{\perp}$ can be written as a linear combination of the vectors $\mathbf{v}_j$, $\mathbf{v}$ can be written as a linear combination of the vectors in $\mathcal{B}$. Therefore, $\mathcal{B}$ spans $\mathbb{R}^n$ also and so is a basis for $\mathbb{R}^n$. It follows that $k + l = \dim \mathbb{R}^n$, or

$$\dim W + \dim W^{\perp} = n$$

As a lovely bonus, when we apply this result to the fundamental subspaces of a matrix, we get a quick proof of the Rank Theorem (Theorem 3.26), restated here as Corollary 5.14.

Corollary 5.14 **The Rank Theorem**

If A is an $m \times n$ matrix, then

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = n$$

Proof In Theorem 5.13, take $W = \operatorname{row}(A)$. Then $W^{\perp} = \operatorname{null}(A)$, by Theorem 5.10, so $\dim W = \operatorname{rank}(A)$ and $\dim W^{\perp} = \operatorname{nullity}(A)$. The result follows.

Note that we get a counterpart identity by taking $W = \operatorname{col}(A)$ [and therefore $W^{\perp} = \operatorname{null}(A^T)$]:

$$\operatorname{rank}(A) + \operatorname{nullity}(A^T) = m$$

Sections 5.1 and 5.2 have illustrated some of the advantages of working with orthogonal bases. However, we have not established that every subspace *has* an orthogonal basis, nor have we given a method for constructing such a basis (except in particular examples, such as Example 5.3). These issues are the subject of the next section.

Exercises 5.2

In Exercises 1–6, find the orthogonal complement $W^\perp$ of W and give a basis for $W^\perp$.

1. $W = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : 2x - y = 0 \right\}$

2. $W = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : 3x + 4y = 0 \right\}$

3. $W = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : x + y - z = 0 \right\}$

4. $W = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : 2x - y + 3z = 0 \right\}$

5. $W = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : x = t, y = -t, z = 3t \right\}$

6. $W = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : x = \frac{1}{2}t, y = -\frac{1}{2}t, z = 2t \right\}$

In Exercises 7 and 8, find bases for the row space and null space of A. Verify that every vector in row(A) is orthogonal to every vector in null(A).

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

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

In Exercises 9 and 10, find bases for the column space of A and the null space of A^T for the given exercise. Verify that every vector in col(A) is orthogonal to every vector in null(A^T).

9. Exercise 7 **10.** Exercise 8

In Exercises 11–14, let W be the subspace spanned by the given vectors. Find a basis for $W^\perp$.

11. $\mathbf{w}_1 = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}$

12. $\mathbf{w}_1 = \begin{bmatrix} 1 \\ -1 \\ 3 \\ -2 \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \end{bmatrix}$

13. $\mathbf{w}_1 = \begin{bmatrix} 2 \\ -1 \\ 6 \\ 3 \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} -1 \\ 2 \\ -3 \\ -2 \end{bmatrix}, \mathbf{w}_3 = \begin{bmatrix} 2 \\ 5 \\ 6 \\ 1 \end{bmatrix}$

14. $\mathbf{w}_1 = \begin{bmatrix} 4 \\ 6 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \\ -3 \end{bmatrix}, \mathbf{w}_3 = \begin{bmatrix} 2 \\ 2 \\ 2 \\ -1 \\ 2 \end{bmatrix}$

In Exercises 15–18, find the orthogonal projection of **v** *onto the subspace W spanned by the vectors* $\mathbf{u}_i$. *(You may assume that the vectors* $\mathbf{u}_i$ *are orthogonal.)*

15. $\mathbf{v} = \begin{bmatrix} 7 \\ -4 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

16. $\mathbf{v} = \begin{bmatrix} 3 \\ 1 \\ -2 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$

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

18. $\mathbf{v} = \begin{bmatrix} 3 \\ -2 \\ 4 \\ -3 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$

In Exercises 19–22, find the orthogonal decomposition of **v** *with respect to W.*

19. $\mathbf{v} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}, W = \text{span}\left(\begin{bmatrix} 1 \\ 3 \end{bmatrix}\right)$

20. $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix}, W = \text{span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right)$

21. $\mathbf{v} = \begin{bmatrix} 4 \\ -2 \\ 3 \end{bmatrix}, W = \text{span}\left(\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}\right)$

22. $\mathbf{v} = \begin{bmatrix} 2 \\ -1 \\ 5 \\ 6 \end{bmatrix}, W = \text{span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}\right)$

23. Prove Theorem 5.9(c).

24. Prove Theorem 5.9(d).

25. Let W be a subspace of $\mathbb{R}^n$ and $\mathbf{v}$ a vector in $\mathbb{R}^n$. Suppose that $\mathbf{w}$ and $\mathbf{w}'$ are orthogonal vectors with $\mathbf{w}$ in W and that $\mathbf{v} = \mathbf{w} + \mathbf{w}'$. Is it necessarily true that $\mathbf{w}'$ is in $W^\perp$? Either prove that it is true or find a counterexample.

26. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an orthogonal basis for $\mathbb{R}^n$ and let $W = \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$. Is it necessarily true that $W^\perp = \text{span}(\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n)$? Either prove that it is true or find a counterexample.

In Exercises 27–29, let W be a subspace of $\mathbb{R}^n$*, and let* **x** *be a vector in* $\mathbb{R}^n$.

27. Prove that $\mathbf{x}$ is in W if and only if $\text{proj}_W(\mathbf{x}) = \mathbf{x}$.

28. Prove that $\mathbf{x}$ is orthogonal to W if and only if $\text{proj}_W(\mathbf{x}) = \mathbf{0}$.

29. Prove that $\text{proj}_W(\text{proj}_W(\mathbf{x})) = \text{proj}_W(\mathbf{x})$.

30. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be an orthonormal set in $\mathbb{R}^n$, and let $\mathbf{x}$ be a vector in $\mathbb{R}^n$.

(a) Prove that

$$\|\mathbf{x}\|^2 \geq |\mathbf{x}\cdot\mathbf{v}_1|^2 + |\mathbf{x}\cdot\mathbf{v}_2|^2 + \cdots + |\mathbf{x}\cdot\mathbf{v}_k|^2$$

(This inequality is called ***Bessel's Inequality.***)

(b) Prove that Bessel's Inequality is an equality if and only if $\mathbf{x}$ is in span(S).

5.3 The Gram-Schmidt Process and the *QR* Factorization

In this section, we present a simple method for constructing an orthogonal (or orthonormal) basis for any subspace of $\mathbb{R}^n$. This method will then lead us to one of the most useful of all matrix factorizations.

The Gram-Schmidt Process

We would like to be able to find an orthogonal basis for a subspace W of $\mathbb{R}^n$. The idea is to begin with an arbitrary basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ for W and to "orthogonalize" it one vector at a time. We will illustrate the basic construction with the subspace W from Example 5.3.

Example 5.12

Let $W = \text{span}(\mathbf{x}_1, \mathbf{x}_2)$, where

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

Construct an orthogonal basis for W.

Solution Starting with $\mathbf{x}_1$, we get a second vector that is orthogonal to it by taking the component of $\mathbf{x}_2$ orthogonal to $\mathbf{x}_1$ (Figure 5.10).

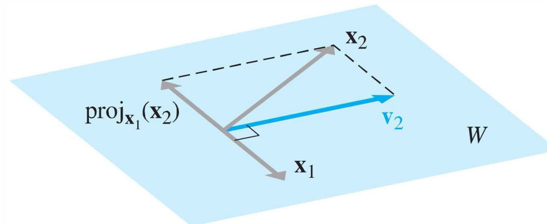


Figure 5.10
Constructing $\mathbf{v}_2$ orthogonal to $\mathbf{x}_1$

Algebraically, we set $\mathbf{v}_1 = \mathbf{x}_1$, so

$$\mathbf{v}_2 = \text{perp}_{\mathbf{x}_1}(\mathbf{x}_2) = \mathbf{x}_2 - \text{proj}_{\mathbf{x}_1}(\mathbf{x}_2)$$
$$= \mathbf{x}_2 - \left(\frac{\mathbf{x}_1 \cdot \mathbf{x}_2}{\mathbf{x}_1 \cdot \mathbf{x}_1}\right)\mathbf{x}_1$$
$$= \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} - \left(\frac{-2}{2}\right)\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

Then $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal set of vectors in W. Hence, $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a linearly independent set and therefore a basis for W, since dim $W = 2$.

Remark Observe that this method depends on the *order* of the original basis vectors. In Example 5.12, if we had taken $\mathbf{x}_1 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, we would have obtained a different orthogonal basis for W. (Verify this.)

The generalization of this method to more than two vectors begins as in Example 5.12. Then the process is to iteratively construct the components of subsequent vectors orthogonal to all of the vectors that have already been constructed. The method is known as the ***Gram-Schmidt Process.***

Theorem 5.15 **The Gram-Schmidt Process**

Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ be a basis for a subspace W of $\mathbb{R}^n$ and define the following:

$$\mathbf{v}_1 = \mathbf{x}_1, \qquad W_1 = \text{span}(\mathbf{x}_1)$$
$$\mathbf{v}_2 = \mathbf{x}_2 - \left(\frac{\mathbf{v}_1 \cdot \mathbf{x}_2}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1, \qquad W_2 = \text{span}(\mathbf{x}_1, \mathbf{x}_2)$$
$$\mathbf{v}_3 = \mathbf{x}_3 - \left(\frac{\mathbf{v}_1 \cdot \mathbf{x}_3}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{x}_3}{\mathbf{v}_2 \cdot \mathbf{v}_2}\right)\mathbf{v}_2, \qquad W_3 = \text{span}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$$
$$\vdots$$
$$\mathbf{v}_k = \mathbf{x}_k - \left(\frac{\mathbf{v}_1 \cdot \mathbf{x}_k}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{x}_k}{\mathbf{v}_2 \cdot \mathbf{v}_2}\right)\mathbf{v}_2 - \cdots$$
$$- \left(\frac{\mathbf{v}_{k-1} \cdot \mathbf{x}_k}{\mathbf{v}_{k-1} \cdot \mathbf{v}_{k-1}}\right)\mathbf{v}_{k-1}, \qquad W_k = \text{span}(\mathbf{x}_1, \ldots, \mathbf{x}_k)$$

Then for each $i = 1, \ldots, k$, $\{\mathbf{v}_1, \ldots, \mathbf{v}_i\}$ is an orthogonal basis for W_i. In particular, $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthogonal basis for W.

Jörgen Pedersen Gram (1850–1916) was a Danish actuary (insurance statistician) who was interested in the science of measurement. He first published the process that bears his name in an 1883 paper on least squares. Erhard Schmidt (1876–1959) was a German mathematician who studied under the great David Hilbert and is considered one of the founders of the branch of mathematics known as functional analysis. His contribution to the Gram-Schmidt Process came in a 1907 paper on integral equations, in which he wrote out the details of the method more explicitly than Gram had done.

Stated succinctly, Theorem 5.15 says that every subspace of $\mathbb{R}^n$ has an orthogonal basis, and it gives an algorithm for constructing such a basis.

Proof We will prove by induction that, for each $i = 1, \ldots, k$, $\{\mathbf{v}_1, \ldots, \mathbf{v}_i\}$ is an orthogonal basis for W_i.

Since $\mathbf{v}_1 = \mathbf{x}_1$, clearly $\{\mathbf{v}_1\}$ is an (orthogonal) basis for $W_1 = \text{span}(\mathbf{x}_1)$. Now assume that, for some $i < k$, $\{\mathbf{v}_1, \ldots, \mathbf{v}_i\}$ is an orthogonal basis for W_i. Then

$$\mathbf{v}_{i+1} = \mathbf{x}_{i+1} - \left(\frac{\mathbf{v}_1 \cdot \mathbf{x}_{i+1}}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{x}_{i+1}}{\mathbf{v}_2 \cdot \mathbf{v}_2}\right)\mathbf{v}_2 - \cdots - \left(\frac{\mathbf{v}_i \cdot \mathbf{x}_{i+1}}{\mathbf{v}_i \cdot \mathbf{v}_i}\right)\mathbf{v}_i$$

By the induction hypothesis, $\{\mathbf{v}_1, \ldots, \mathbf{v}_i\}$ is an orthogonal basis for $\text{span}(\mathbf{x}_1, \ldots, \mathbf{x}_i) = W_i$. Hence,

$$\mathbf{v}_{i+1} = \mathbf{x}_{i+1} - \text{proj}_{W_i}(\mathbf{x}_{i+1}) = \text{perp}_{W_i}(\mathbf{x}_{i+1})$$

So, by the Orthogonal Decomposition Theorem, $\mathbf{v}_{i+1}$ is orthogonal to W_i. By definition, $\mathbf{v}_1, \ldots, \mathbf{v}_i$ are linear combinations of $\mathbf{x}_1, \ldots, \mathbf{x}_i$ and, hence, are in W_i. Therefore, $\{\mathbf{v}_1, \ldots, \mathbf{v}_{i+1}\}$ is an orthogonal set of vectors in W_{i+1}.

Moreover, $\mathbf{v}_{i+1} \neq \mathbf{0}$, since otherwise $\mathbf{x}_{i+1} = \text{proj}_{W_i}(\mathbf{x}_{i+1})$, which in turn implies that $\mathbf{x}_{i+1}$ is in W_i. But this is impossible, since $W_i = \text{span}(\mathbf{x}_1, \ldots, \mathbf{x}_i)$ and $\{\mathbf{x}_1, \ldots, \mathbf{x}_{i+1}\}$ is linearly independent. (Why?) We conclude that $\{\mathbf{v}_1, \ldots, \mathbf{v}_{i+1}\}$ is a set of $i + 1$ linearly independent vectors in W_{i+1}. Consequently, $\{\mathbf{v}_1, \ldots, \mathbf{v}_{i+1}\}$ is a basis for W_{i+1}, since $\dim W_{i+1} = i + 1$. This completes the proof.

If we require an orthonormal basis for W, we simply need to normalize the orthogonal vectors produced by the Gram-Schmidt Process. That is, for each i, we replace $\mathbf{v}_i$ by the unit vector $\mathbf{q}_i = (1/\|\mathbf{v}_i\|)\mathbf{v}_i$.

Example 5.13

Apply the Gram-Schmidt Process to construct an orthonormal basis for the subspace $W = \text{span}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ of $\mathbb{R}^4$, where

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 2 \\ 2 \\ 1 \\ 2 \end{bmatrix}$$

Solution First we note that $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is a linearly independent set, so it forms a basis for W. We begin by setting $\mathbf{v}_1 = \mathbf{x}_1$. Next, we compute the component of $\mathbf{x}_2$ orthogonal to $W_1 = \text{span}(\mathbf{v}_1)$:

$$\mathbf{v}_2 = \text{perp}_{W_1}(\mathbf{x}_2) = \mathbf{x}_2 - \left(\frac{\mathbf{v}_1 \cdot \mathbf{x}_2}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1$$

$$= \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix} - \left(\tfrac{2}{4}\right)\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{3}{2} \\ \frac{3}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

For hand calculations, it is a good idea to "scale" $\mathbf{v}_2$ at this point to eliminate fractions. When we are finished, we can rescale the orthogonal set we are constructing to obtain an orthonormal set; thus, we can replace each $\mathbf{v}_i$ by any convenient scalar multiple without affecting the final result. Accordingly, we replace $\mathbf{v}_2$ by

$$\mathbf{v}_2' = 2\mathbf{v}_2 = \begin{bmatrix} 3 \\ 3 \\ 1 \\ 1 \end{bmatrix}$$

We now find the component of $\mathbf{x}_3$ orthogonal to

$$W_2 = \text{span}(\mathbf{x}_1, \mathbf{x}_2) = \text{span}(\mathbf{v}_1, \mathbf{v}_2) = \text{span}(\mathbf{v}_1, \mathbf{v}_2')$$

using the orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2'\}$:

$$\begin{aligned} \mathbf{v}_3 = \text{perp}_{W_2}(\mathbf{x}_3) &= \mathbf{x}_3 - \left(\frac{\mathbf{v}_1 \cdot \mathbf{x}_3}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2' \cdot \mathbf{x}_3}{\mathbf{v}_2' \cdot \mathbf{v}_2'}\right)\mathbf{v}_2' \\ &= \begin{bmatrix} 2 \\ 2 \\ 1 \\ 2 \end{bmatrix} - \left(\tfrac{1}{4}\right)\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} - \left(\tfrac{15}{20}\right)\begin{bmatrix} 3 \\ 3 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -\frac{1}{2} \\ 0 \\ \frac{1}{2} \\ 1 \end{bmatrix} \end{aligned}$$

Again, we rescale and use $\mathbf{v}_3' = 2\mathbf{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}$.

We now have an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2', \mathbf{v}_3'\}$ for W. (Check to make sure that these vectors are orthogonal.) To obtain an orthonormal basis, we normalize each vector:

$$\mathbf{q}_1 = \left(\frac{1}{\|\mathbf{v}_1\|}\right)\mathbf{v}_1 = \left(\frac{1}{2}\right)\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \end{bmatrix}$$

$$\mathbf{q}_2 = \left(\frac{1}{\|\mathbf{v}_2'\|}\right)\mathbf{v}_2' = \left(\frac{1}{2\sqrt{5}}\right)\begin{bmatrix} 3 \\ 3 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3/2\sqrt{5} \\ 3/2\sqrt{5} \\ 1/2\sqrt{5} \\ 1/2\sqrt{5} \end{bmatrix} = \begin{bmatrix} 3\sqrt{5}/10 \\ 3\sqrt{5}/10 \\ \sqrt{5}/10 \\ \sqrt{5}/10 \end{bmatrix}$$

$$\mathbf{q}_3 = \left(\frac{1}{\|\mathbf{v}_3'\|}\right)\mathbf{v}_3' = \left(\frac{1}{\sqrt{6}}\right)\begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1/\sqrt{6} \\ 0 \\ 1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix} = \begin{bmatrix} -\sqrt{6}/6 \\ 0 \\ \sqrt{6}/6 \\ \sqrt{6}/3 \end{bmatrix}$$

Then $\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$ is an orthonormal basis for W.

One of the important uses of the Gram-Schmidt Process is to construct an orthogonal basis that contains a specified vector. The next example illustrates this application.

Example 5.14 Find an orthogonal basis for $\mathbb{R}^3$ that contains the vector

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Solution We first find *any* basis for $\mathbb{R}^3$ containing $\mathbf{v}_1$. If we take

$$\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

then $\{\mathbf{v}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is clearly a basis for $\mathbb{R}^3$. (Why?) We now apply the Gram-Schmidt Process to this basis to obtain

$$\mathbf{v}_2 = \mathbf{x}_2 - \left(\frac{\mathbf{v}_1 \cdot \mathbf{x}_2}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - \left(\tfrac{2}{14}\right)\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{7} \\ \frac{5}{7} \\ -\frac{3}{7} \end{bmatrix}, \quad \mathbf{v}_2' = \begin{bmatrix} -1 \\ 5 \\ -3 \end{bmatrix}$$

and finally

$$\mathbf{v}_3 = \mathbf{x}_3 - \left(\frac{\mathbf{v}_1 \cdot \mathbf{x}_3}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2' \cdot \mathbf{x}_3}{\mathbf{v}_2' \cdot \mathbf{v}_2'}\right)\mathbf{v}_2' = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \left(\tfrac{3}{14}\right)\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \left(\tfrac{-3}{35}\right)\begin{bmatrix} -1 \\ 5 \\ -3 \end{bmatrix} = \begin{bmatrix} \frac{-3}{10} \\ 0 \\ \frac{1}{10} \end{bmatrix},$$

$$\mathbf{v}_3' = \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$$

Then $\{\mathbf{v}_1, \mathbf{v}_2', \mathbf{v}_3'\}$ is an orthogonal basis for $\mathbb{R}^3$ that contains $\mathbf{v}_1$.

Similarly, given a unit vector, we can find an orthonormal basis that contains it by using the preceding method and then normalizing the resulting orthogonal vectors.

Remark When the Gram-Schmidt Process is implemented on a computer, there is almost always some roundoff error, leading to a loss of orthogonality in the vectors $\mathbf{q}_i$. To avoid this loss of orthogonality, some modifications are usually made. The vectors $\mathbf{v}_i$ are normalized as soon as they are computed, rather than at the end, to give the vectors $\mathbf{q}_i$, and as each $\mathbf{q}_i$ is computed, the remaining vectors $\mathbf{x}_j$ are modified to be orthogonal to $\mathbf{q}_i$. This procedure is known as the ***Modified Gram-Schmidt Process.*** In practice, however, a version of the *QR* factorization is used to compute orthonormal bases.

The *QR* Factorization

If A is an $m \times n$ matrix with linearly independent columns (requiring that $m \geq n$), then applying the Gram-Schmidt Process to these columns yields a very useful factorization of A into the product of a matrix Q with orthonormal columns and an upper triangular matrix R. This is the ***QR factorization,*** and it has applications to the numerical approximation of eigenvalues, which we explore at the end of this section, and to the problem of least squares approximation, which we discuss in Chapter 7.

To see how the QR factorization arises, let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ be the (linearly independent) columns of A and let $\mathbf{q}_1, \ldots, \mathbf{q}_n$ be the orthonormal vectors obtained by applying the Gram-Schmidt Process to A with normalizations. From Theorem 5.15, we know that, for each $i = 1, \ldots, n$,

$$W_i = \text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_i) = \text{span}(\mathbf{q}_1, \ldots, \mathbf{q}_i)$$

Therefore, there are scalars $r_{1i}, r_{2i}, \ldots, r_{ii}$ such that

$$\mathbf{a}_i = r_{1i}\mathbf{q}_1 + r_{2i}\mathbf{q}_2 + \cdots + r_{ii}\mathbf{q}_i \quad \text{for } i = 1, \ldots, n$$

That is,

$$\begin{aligned} \mathbf{a}_1 &= r_{11}\mathbf{q}_1 \\ \mathbf{a}_2 &= r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2 \\ &\vdots \\ \mathbf{a}_n &= r_{1n}\mathbf{q}_1 + r_{2n}\mathbf{q}_2 + \cdots + r_{nn}\mathbf{q}_n \end{aligned}$$

which can be written in matrix form as

$$A = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n] = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \cdots \quad \mathbf{q}_n]\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{nn} \end{bmatrix} = QR$$

Clearly, the matrix Q has orthonormal columns. It is also the case that the diagonal entries of R are all nonzero. To see this, observe that if $r_{ii} = 0$, then $\mathbf{a}_i$ is a linear combination of $\mathbf{q}_1, \ldots, \mathbf{q}_{i-1}$ and, hence, is in W_{i-1}. But then $\mathbf{a}_i$ would be a linear combination of $\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}$, which is impossible, since $\mathbf{a}_1, \ldots, \mathbf{a}_i$ are linearly independent. We conclude that $r_{ii} \neq 0$ for $i = 1, \ldots, n$. Since R is upper triangular, it follows that it must be invertible. (See Exercise 23.)

We have proved the following theorem.

Theorem 5.16 **The *QR* Factorization**

Let A be an $m \times n$ matrix with linearly independent columns. Then A can be factored as $A = QR$, where Q is an $m \times n$ matrix with orthonormal columns and R is an invertible upper triangular matrix.

Remarks

- We can also arrange for the diagonal entries of R to be *positive*. If any $r_{ii} < 0$, simply replace $\mathbf{q}_i$ by $-\mathbf{q}_i$ and r_{ii} by $-r_{ii}$.
- The requirement that A have linearly independent columns is a necessary one. To prove this, suppose that A is an $m \times n$ matrix that has a QR factorization, as in Theorem 5.16. Then, since R is invertible, we have $Q = AR^{-1}$. Hence, rank(Q) = rank(A), by Exercise 61 in Section 3.5. But rank$(Q) = n$, since its columns are orthonormal and, therefore, linearly independent. So rank$(A) = n$ too, and consequently the columns of A are linearly independent, by the Fundamental Theorem.

- The QR factorization can be extended to arbitrary matrices in a slightly modified form. If A is $m \times n$, it is possible to find a sequence of orthogonal matrices $Q_1, \ldots, Q_{m-1}$ such that $Q_{m-1} \cdots Q_2 Q_1 A$ is an upper triangular $m \times n$ matrix R. Then $A = QR$, where $Q = (Q_{m-1} \cdots Q_2 Q_1)^{-1}$ is an orthogonal matrix. We will examine this approach in Exploration: The Modified QR Factorization.

Example 5.15

Find a QR factorization of

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

Solution The columns of A are just the vectors from Example 5.13. The orthonormal basis for col(A) produced by the Gram-Schmidt Process was

$$\mathbf{q}_1 = \begin{bmatrix} 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \end{bmatrix}, \quad \mathbf{q}_2 = \begin{bmatrix} 3\sqrt{5}/10 \\ 3\sqrt{5}/10 \\ \sqrt{5}/10 \\ \sqrt{5}/10 \end{bmatrix}, \quad \mathbf{q}_3 = \begin{bmatrix} -\sqrt{6}/6 \\ 0 \\ \sqrt{6}/6 \\ \sqrt{6}/3 \end{bmatrix}$$

so

$$Q = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \mathbf{q}_3] = \begin{bmatrix} 1/2 & 3\sqrt{5}/10 & -\sqrt{6}/6 \\ -1/2 & 3\sqrt{5}/10 & 0 \\ -1/2 & \sqrt{5}/10 & \sqrt{6}/6 \\ 1/2 & \sqrt{5}/10 & \sqrt{6}/3 \end{bmatrix}$$

From Theorem 5.16, $A = QR$ for some upper triangular matrix R. To find R, we use the fact that Q has orthonormal columns and, hence, $Q^TQ = I$. Therefore,

$$Q^TA = Q^TQR = IR = R$$

We compute

$$R = Q^TA = \begin{bmatrix} 1/2 & -1/2 & -1/2 & 1/2 \\ 3\sqrt{5}/10 & 3\sqrt{5}/10 & \sqrt{5}/10 & \sqrt{5}/10 \\ -\sqrt{6}/6 & 0 & \sqrt{6}/6 & \sqrt{6}/3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 \\ -1 & 1 & 2 \\ -1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 1 & 1/2 \\ 0 & \sqrt{5} & 3\sqrt{5}/2 \\ 0 & 0 & \sqrt{6}/2 \end{bmatrix}$$

Exercises 5.3

In Exercises 1–4, the given vectors form a basis for $\mathbb{R}^2$ or $\mathbb{R}^3$. Apply the Gram-Schmidt Process to obtain an orthogonal basis. Then normalize this basis to obtain an orthonormal basis.

1. $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

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

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

4. $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

In Exercises 5 and 6, the given vectors form a basis for a subspace W of $\mathbb{R}^3$ or $\mathbb{R}^4$. Apply the Gram-Schmidt Process to obtain an orthogonal basis for W.

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

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

In Exercises 7 and 8, find the orthogonal decomposition of $\mathbf{v}$ *with respect to the subspace W.*

7. $\mathbf{v} = \begin{bmatrix} 4 \\ -4 \\ 3 \end{bmatrix}$, W as in Exercise 5

8. $\mathbf{v} = \begin{bmatrix} 1 \\ 4 \\ 0 \\ 2 \end{bmatrix}$, W as in Exercise 6

Use the Gram-Schmidt Process to find an orthogonal basis for the column spaces of the matrices in Exercises 9 and 10.

9. $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

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

11. Find an orthogonal basis for $\mathbb{R}^3$ that contains the vector $\begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix}$.

12. Find an orthogonal basis for $\mathbb{R}^4$ that contains the vectors

$$\begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 1 \\ 0 \\ 1 \\ 3 \end{bmatrix}$$

In Exercises 13 and 14, fill in the missing entries of Q to make Q an orthogonal matrix.

13. $Q = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{3} & * \\ 0 & 1/\sqrt{3} & * \\ -1/\sqrt{2} & 1/\sqrt{3} & * \end{bmatrix}$

14. $Q = \begin{bmatrix} 1/2 & 2/\sqrt{14} & * & * \\ 1/2 & 1/\sqrt{14} & * & * \\ 1/2 & 0 & * & * \\ 1/2 & -3/\sqrt{14} & * & * \end{bmatrix}$

In Exercises 15 and 16, find a QR factorization of the matrix in the given exercise.

15. Exercise 9 **16.** Exercise 10

In Exercises 17 and 18, the columns of Q were obtained by applying the Gram-Schmidt Process to the columns of A. Find the upper triangular matrix R such that $A = QR$.

17. $A = \begin{bmatrix} 2 & 8 & 2 \\ 1 & 7 & -1 \\ -2 & -2 & 1 \end{bmatrix}, Q = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix}$

18. $A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \\ -1 & -1 \\ 0 & 1 \end{bmatrix}, Q = \begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{3} \\ 2/\sqrt{6} & 0 \\ -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \end{bmatrix}$

19. If A is an orthogonal matrix, find a QR factorization of A.

20. Prove that A is invertible if and only if $A = QR$, where Q is orthogonal and R is upper triangular with nonzero entries on its diagonal.

In Exercises 21 and 22, use the method suggested by Exercise 20 to compute A^{-1} for the matrix A in the given exercise.

21. Exercise 9 **22.** Exercise 15

23. Let A be an $m \times n$ matrix with linearly independent columns. Give an alternative proof that the upper triangular matrix R in a QR factorization of A must be invertible, using property (c) of the Fundamental Theorem.

24. Let A be an $m \times n$ matrix with linearly independent columns and let $A = QR$ be a QR factorization of A. Show that A and Q have the same column space.

Explorations

The Modified QR Factorization

When the matrix A does not have linearly independent columns, the Gram-Schmidt Process as we have stated it does not work and so cannot be used to develop a generalized QR factorization of A. There is a modification of the Gram-Schmidt Process that can be used, but instead we will explore a method that converts A into upper triangular form one column at a time, using a sequence of orthogonal matrices. The method is analogous to that of LU factorization, in which the matrix L is formed using a sequence of elementary matrices.

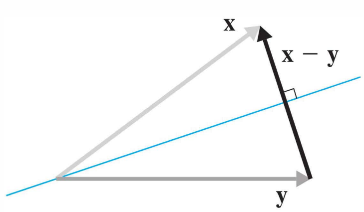


Figure 5.11

The first thing we need is the "orthogonal analogue" of an elementary matrix; that is, we need to know how to construct an orthogonal matrix Q that will transform a given column of A—call it $\mathbf{x}$—into the corresponding column of R—call it $\mathbf{y}$. By Theorem 5.6, it will be necessary that $\|\mathbf{x}\| = \|Q\mathbf{x}\| = \|\mathbf{y}\|$. Figure 5.11 suggests a way to proceed: We can reflect $\mathbf{x}$ in a line perpendicular to $\mathbf{x} - \mathbf{y}$. If

$$\mathbf{u} = \left(\frac{1}{\|\mathbf{x} - \mathbf{y}\|}\right)(\mathbf{x} - \mathbf{y}) = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

is the unit vector in the direction of $\mathbf{x} - \mathbf{y}$, then $\mathbf{u}^{\perp} = \begin{bmatrix} -d_2 \\ d_1 \end{bmatrix}$ is orthogonal to $\mathbf{u}$, and we can use Exercise 26 in Section 3.6 to find the standard matrix Q of the reflection in the line through the origin in the direction of $\mathbf{u}^{\perp}$.

1. Show that $Q = \begin{bmatrix} 1 - 2d_1^2 & -2d_1d_2 \\ -2d_1d_2 & 1 - 2d_2^2 \end{bmatrix} = I - 2\mathbf{u}\mathbf{u}^T$.

2. Compute Q for

$$\text{(a)}\ \mathbf{u} = \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix} \quad \text{(b)}\ \mathbf{x} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}$$

We can generalize the definition of Q as follows. If $\mathbf{u}$ is any unit vector in $\mathbb{R}^n$, we define an $n \times n$ matrix Q as

$$Q = I - 2\mathbf{u}\mathbf{u}^T$$

Such a matrix is called a ***Householder matrix*** (or an ***elementary reflector***).

3. Prove that every Householder matrix Q satisfies the following properties:

   (a) Q is symmetric. (b) Q is orthogonal. (c) $Q^2 = I$

4. Prove that if Q is a Householder matrix corresponding to the unit vector $\mathbf{u}$, then

$$Q\mathbf{v} = \begin{cases} -\mathbf{v} & \text{if } \mathbf{v} \text{ is in span}(\mathbf{u}) \\ \mathbf{v} & \text{if } \mathbf{v}\cdot\mathbf{u} = \mathbf{0} \end{cases}$$

5. Compute Q for $\mathbf{u} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$ and verify Problems 3 and 4.

6. Let $\mathbf{x} \neq \mathbf{y}$ with $\|\mathbf{x}\| = \|\mathbf{y}\|$ and set $\mathbf{u} = (1/\|\mathbf{x} - \mathbf{y}\|)(\mathbf{x} - \mathbf{y})$. Prove that the corresponding Householder matrix Q satisfies $Q\mathbf{x} = \mathbf{y}$. [*Hint:* Apply Exercise 57 in Section 1.2 to the result in Problem 4.]

7. Find Q and verify Problem 6 for

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$$

Alston Householder (1904–1993) was one of the pioneers in the field of numerical linear algebra. He was the first to present a systematic treatment of algorithms for solving problems involving linear systems. In addition to introducing the widely used Householder transformations that bear his name, he was one of the first to advocate the systematic use of norms in linear algebra. His 1964 book *The Theory of Matrices in Numerical Analysis* is considered a classic.

We are now ready to perform the triangularization of an $m \times n$ matrix A, column by column.

8. Let $\mathbf{x}$ be the first column of A and let

$$\mathbf{y} = \begin{bmatrix} \|\mathbf{x}\| \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Show that if Q_1 is the Householder matrix given by Problem 6, then Q_1A is a matrix with the block form

$$Q_1A = \begin{bmatrix} * & * \\ \mathbf{0} & A_1 \end{bmatrix}$$

where A_1 is $(m-1)\times(n-1)$.

If we repeat Problem 8 on the matrix A_1, we use a Householder matrix P_2 such that

$$P_2A_1 = \begin{bmatrix} * & * \\ \mathbf{0} & A_2 \end{bmatrix}$$

where A_2 is $(m-2)\times(n-2)$.

9. Set $Q_2 = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & P_2 \end{bmatrix}$. Show that Q_2 is an orthogonal matrix and that

$$Q_2Q_1A = \begin{bmatrix} * & * & * \\ 0 & * & * \\ \mathbf{0} & \mathbf{0} & A_2 \end{bmatrix}$$

10. Show that we can continue in this fashion to find a sequence of orthogonal matrices $Q_1, \ldots, Q_{m-1}$ such that $Q_{m-1} \cdots Q_2 Q_1 A = R$ is an upper triangular $m \times n$ matrix (i.e., $r_{ij} = 0$ if $i > j$).

11. Deduce that $A = QR$ with $Q = Q_1 Q_2 \cdots Q_{m-1}$ orthogonal.

12. Use the method of this exploration to find a QR factorization of

$$\text{(a) } A = \begin{bmatrix} 3 & 9 & 1 \\ -4 & 3 & 2 \end{bmatrix} \quad \text{(b) } A = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 2 & -4 & 1 & 1 \\ 2 & -5 & -1 & -2 \end{bmatrix}$$

Approximating Eigenvalues with the QR Algorithm

See G. H. Golub and C. F. Van Loan, *Matrix Computations* (Baltimore: Johns Hopkins University Press, 1983).

One of the best (and most widely used) methods for numerically approximating the eigenvalues of a matrix makes use of the QR factorization. The purpose of this exploration is to introduce this method, the ***QR* algorithm,** and to show it at work in a few examples. For a more complete treatment of this topic, consult any good text on numerical linear algebra. (You will find it helpful to use a CAS to perform the calculations in the problems below.)

Given a square matrix A, the first step is to factor it as $A = QR$ (using whichever method is appropriate). Then we define $A_1 = RQ$.

1. First prove that A_1 is similar to A. Then prove that A_1 has the same eigenvalues as A.

2. If $A = \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix}$, find A_1 and verify that it has the same eigenvalues as A.

Continuing the algorithm, we factor A_1 as $A_1 = Q_1R_1$ and set $A_2 = R_1Q_1$. Then we factor $A_2 = Q_2R_2$ and set $A_3 = R_2Q_2$, and so on. That is, for $k \geq 1$, we compute $A_k = Q_kR_k$ and then set $A_{k+1} = R_kQ_k$.

3. Prove that A_k is similar to A for all $k \geq 1$.

4. Continuing Problem 2, compute A_2, A_3, A_4, and A_5, using two-decimal-place accuracy. What do you notice?

It can be shown that if the eigenvalues of A are all real and have distinct absolute values, then the matrices A_k approach an upper triangular matrix U.

5. What will be true of the diagonal entries of this matrix U?

6. Approximate the eigenvalues of the following matrices by applying the QR algorithm. Use two-decimal-place accuracy and perform at least five iterations.

$$\text{(a) } \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} \qquad \text{(b) } \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$$

$$\text{(c) } \begin{bmatrix} 1 & 0 & -1 \\ 1 & 2 & 1 \\ -4 & 0 & 1 \end{bmatrix} \quad \text{(d) } \begin{bmatrix} 1 & 1 & -1 \\ 0 & 2 & 0 \\ -2 & 4 & 2 \end{bmatrix}$$

7. Apply the QR algorithm to the matrix $A = \begin{bmatrix} 2 & 3 \\ -1 & -2 \end{bmatrix}$. What happens? Why?

8. Shift the eigenvalues of the matrix in Problem 7 by replacing A with $B = A + 0.9I$. Apply the QR algorithm to B and then shift back by subtracting 0.9 from the (approximate) eigenvalues of B. Verify that this method approximates the eigenvalues of A.

9. Let $Q_0 = Q$ and $R_0 = R$. First show that

$$Q_0Q_1 \cdots Q_{k-1}A_k = AQ_0Q_1 \cdots Q_{k-1}$$

for all $k \geq 1$. Then show that

$$(Q_0Q_1 \cdots Q_k)(R_k \cdots R_1R_0) = A(Q_0Q_1 \cdots Q_{k-1})(R_{k-1} \cdots R_1R_0)$$

[*Hint:* Repeatedly use the same approach used for the first equation, working from the "inside out."] Finally, deduce that $(Q_0Q_1 \cdots Q_k)(R_k \cdots R_1R_0)$ is the QR factorization of A^{k+1}.

5.4 Orthogonal Diagonalization of Symmetric Matrices

We saw in Chapter 4 that a square matrix with real entries will not necessarily have real eigenvalues. Indeed, the matrix $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ has complex eigenvalues i and $-i$. We also discovered that not all square matrices are diagonalizable. The situation changes dramatically if we restrict our attention to real *symmetric* matrices. As we will show in this section, all of the eigenvalues of a real symmetric matrix are real, and such a matrix is always diagonalizable.

Recall that a symmetric matrix is one that equals its own transpose. Let's begin by studying the diagonalization process for a symmetric 2×2 matrix.

Example 5.16

If possible, diagonalize the matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}$.

Solution The characteristic polynomial of A is $\lambda^2 + \lambda - 6 = (\lambda + 3)(\lambda - 2)$, from which we see that A has eigenvalues $\lambda_1 = -3$ and $\lambda_2 = 2$. Solving for the corresponding eigenvectors, we find

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

respectively. So A is diagonalizable, and if we set $P = [\mathbf{v}_1 \;\; \mathbf{v}_2]$, then we know that $P^{-1}AP = \begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix} = D$.

However, we can do better. Observe that $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal. So, if we normalize them to get the unit eigenvectors

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{5} \\ -2/\sqrt{5} \end{bmatrix} \quad \text{and} \quad \mathbf{u}_2 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$$

and then take

$$Q = [\mathbf{u}_1 \;\; \mathbf{u}_2] = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}$$

we have $Q^{-1}AQ = D$ also. But now Q is an *orthogonal* matrix, since $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthonormal set of vectors. Therefore, $Q^{-1} = Q^T$, and we have $Q^TAQ = D$. (Note that checking is easy, since computing Q^{-1} only involves taking a transpose!)

The situation in Example 5.16 is the one that interests us. It is important enough to warrant a new definition.

Definition A square matrix A is ***orthogonally diagonalizable*** if there exists an orthogonal matrix Q and a diagonal matrix D such that $Q^TAQ = D$.

We are interested in finding conditions under which a matrix is orthogonally diagonalizable. Theorem 5.17 shows us where to look.

Theorem 5.17 If A is orthogonally diagonalizable, then A is symmetric.

Proof If A is orthogonally diagonalizable, then there exists an orthogonal matrix Q and a diagonal matrix D such that $Q^TAQ = D$. Since $Q^{-1} = Q^T$, we have $Q^TQ = I = QQ^T$, so

$$QDQ^T = QQ^TAQQ^T = IAI = A$$

But then

$$A^T = (QDQ^T)^T = (Q^T)^TD^TQ^T = QDQ^T = A$$

since every diagonal matrix is symmetric. Hence, A is symmetric.

Remark Theorem 5.17 shows that the orthogonally diagonalizable matrices are all to be found *among* the symmetric matrices. It does *not* say that every symmetric matrix must be orthogonally diagonalizable. However, it is a remarkable fact that this indeed is true! Finding a proof for this amazing result will occupy us for much of the rest of this section.

a + bi

We next prove that we don't need to worry about *complex* eigenvalues when working with symmetric matrices with *real* entries.

Theorem 5.18 If A is a real symmetric matrix, then the eigenvalues of A are real.

Recall that the *complex conjugate* of a complex number $z = a + bi$ is the number $\bar{z} = a - bi$ (see Appendix C). To show that z is real, we need to show that $b = 0$. One way to do this is to show that $z = \bar{z}$, for then $bi = -bi$ (or $2bi = 0$), from which it follows that $b = 0$.

We can also extend the notion of complex conjugate to vectors and matrices by, for example, defining $\bar{A}$ to be the matrix whose entries are the complex conjugates of the entries of A; that is, if $A = [a_{ij}]$, then $\bar{A} = [\bar{a}_{ij}]$. The rules for complex conjugation extend easily to matrices; in particular, we have $\overline{AB} = \bar{A}\bar{B}$ for compatible matrices A and B.

Proof Suppose that λ is an eigenvalue of A with corresponding eigenvector $\mathbf{v}$. Then $A\mathbf{v} = \lambda\mathbf{v}$, and, taking complex conjugates, we have $\overline{A\mathbf{v}} = \overline{\lambda\mathbf{v}}$. But then

$$A\bar{\mathbf{v}} = \bar{A}\bar{\mathbf{v}} = \overline{A\mathbf{v}} = \overline{\lambda\mathbf{v}} = \bar{\lambda}\bar{\mathbf{v}}$$

since A is real. Taking transposes and using the fact that A is symmetric, we have

$$\bar{\mathbf{v}}^TA = \bar{\mathbf{v}}^TA^T = (A\bar{\mathbf{v}})^T = (\bar{\lambda}\bar{\mathbf{v}})^T = \bar{\lambda}\bar{\mathbf{v}}^T$$

Therefore,

$$\lambda(\bar{\mathbf{v}}^T\mathbf{v}) = \bar{\mathbf{v}}^T(\lambda\mathbf{v}) = \bar{\mathbf{v}}^T(A\mathbf{v}) = (\bar{\mathbf{v}}^TA)\mathbf{v} = (\bar{\lambda}\bar{\mathbf{v}}^T)\mathbf{v} = \bar{\lambda}(\bar{\mathbf{v}}^T\mathbf{v})$$

or $(\lambda - \bar{\lambda})(\bar{\mathbf{v}}^T\mathbf{v}) = 0$.

Now if $\mathbf{v} = \begin{bmatrix} a_1 + b_1i \\ \vdots \\ a_n + b_ni \end{bmatrix}$, then $\bar{\mathbf{v}} = \begin{bmatrix} a_1 - b_1i \\ \vdots \\ a_n - b_ni \end{bmatrix}$, so

$$\bar{\mathbf{v}}^T\mathbf{v} = (a_1^2 + b_1^2) + \cdots + (a_n^2 + b_n^2) \neq 0$$

since $\mathbf{v} \neq \mathbf{0}$ (because it is an eigenvector). We conclude that $\lambda - \bar{\lambda} = 0$, or $\lambda = \bar{\lambda}$. Hence, λ is real.

Theorem 4.20 showed that, for any square matrix, eigenvectors corresponding to distinct eigenvalues are linearly independent. For symmetric matrices, something stronger is true: Such eigenvectors are *orthogonal*.

Theorem 5.19 If A is a symmetric matrix, then any two eigenvectors corresponding to distinct eigenvalues of A are orthogonal.

Proof Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be eigenvectors corresponding to the distinct eigenvalues $\lambda_1 \neq \lambda_2$ so that $A\mathbf{v}_1 = \lambda_1\mathbf{v}_1$ and $A\mathbf{v}_2 = \lambda_2\mathbf{v}_2$. Using $A^T = A$ and the fact that $\mathbf{x}\cdot\mathbf{y} = \mathbf{x}^T\mathbf{y}$ for any two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, we have

$$\begin{aligned}\lambda_1(\mathbf{v}_1\cdot\mathbf{v}_2) &= (\lambda_1\mathbf{v}_1)\cdot\mathbf{v}_2 = A\mathbf{v}_1\cdot\mathbf{v}_2 = (A\mathbf{v}_1)^T\mathbf{v}_2\\ &= (\mathbf{v}_1^TA^T)\mathbf{v}_2 = (\mathbf{v}_1^TA)\mathbf{v}_2 = \mathbf{v}_1^T(A\mathbf{v}_2)\\ &= \mathbf{v}_1^T(\lambda_2\mathbf{v}_2) = \lambda_2(\mathbf{v}_1^T\mathbf{v}_2) = \lambda_2(\mathbf{v}_1\cdot\mathbf{v}_2)\end{aligned}$$

Hence, $(\lambda_1 - \lambda_2)(\mathbf{v}_1\cdot\mathbf{v}_2) = 0$. But $\lambda_1 - \lambda_2 \neq 0$, so $\mathbf{v}_1\cdot\mathbf{v}_2 = 0$, as we wished to show.

Example 5.17 Verify the result of Theorem 5.19 for

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

Solution The characteristic polynomial of A is $-\lambda^3 + 6\lambda^2 - 9\lambda + 4 = -(\lambda - 4)\cdot(\lambda - 1)^2$, from which it follows that the eigenvalues of A are $\lambda_1 = 4$ and $\lambda_2 = 1$. The corresponding eigenspaces are

$$E_4 = \text{span}\left(\begin{bmatrix}1\\1\\1\end{bmatrix}\right) \quad\text{and}\quad E_1 = \text{span}\left(\begin{bmatrix}-1\\0\\1\end{bmatrix}, \begin{bmatrix}-1\\1\\0\end{bmatrix}\right)$$

(Check this.) We easily verify that

$$\begin{bmatrix}1\\1\\1\end{bmatrix}\cdot\begin{bmatrix}-1\\0\\1\end{bmatrix} = 0 \quad\text{and}\quad \begin{bmatrix}1\\1\\1\end{bmatrix}\cdot\begin{bmatrix}-1\\1\\0\end{bmatrix} = 0$$

from which it follows that every vector in E_4 is orthogonal to every vector in E_1. (Why?)

Remark Note that $\begin{bmatrix}-1\\0\\1\end{bmatrix}\cdot\begin{bmatrix}-1\\1\\0\end{bmatrix} = 1$. Thus, eigenvectors corresponding to the *same* eigenvalue need not be orthogonal.

We can now prove the main result of this section. It is called the Spectral Theorem, since the set of eigenvalues of a matrix is sometimes called the ***spectrum*** of the matrix. (Technically, we should call Theorem 5.20 the *Real* Spectral Theorem, since there is a corresponding result for matrices with complex entries.)

Theorem 5.20 **The Spectral Theorem**

Let A be an $n \times n$ real matrix. Then A is symmetric if and only if it is orthogonally diagonalizable.

Spectrum is a Latin word meaning "image." When atoms vibrate, they emit light. And when light passes through a prism, it spreads out into a spectrum—a band of rainbow colors. Vibration frequencies correspond to the eigenvalues of a certain operator and are visible as bright lines in the spectrum of light that is emitted from a prism. Thus, we can literally see the eigenvalues of the atom in its spectrum, and for this reason, it is appropriate that the word *spectrum* has come to be applied to the set of all eigenvalues of a matrix (or operator).

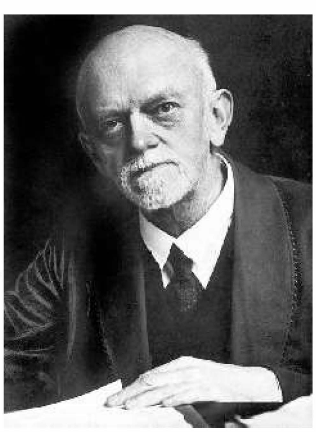
Proof We have already proved the "if" part as Theorem 5.17. To prove the "only if" implication, we proceed by induction on n. For $n = 1$, there is nothing to do, since a 1×1 matrix is already in diagonal form. Now assume that every $k \times k$ real symmetric matrix with real eigenvalues is orthogonally diagonalizable. Let $n = k + 1$ and let A be an $n \times n$ real symmetric matrix with real eigenvalues.

Let λ_1 be one of the eigenvalues of A and let $\mathbf{v}_1$ be a corresponding eigenvector. Then $\mathbf{v}_1$ is a *real* vector (why?) and we can assume that $\mathbf{v}_1$ is a unit vector, since otherwise we can normalize it and we will still have an eigenvector corresponding to λ_1. Using the Gram-Schmidt Process, we can extend $\mathbf{v}_1$ to an orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of $\mathbb{R}^n$. Now we form the matrix

$$Q_1 = [\mathbf{v}_1 \quad \mathbf{v}_2 \cdots \mathbf{v}_n]$$

Then Q_1 is orthogonal, and

$$\begin{aligned}
Q_1^T A Q_1 &= \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} A[\mathbf{v}_1 \quad \mathbf{v}_2 \cdots \mathbf{v}_n] = \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} [A\mathbf{v}_1 \quad A\mathbf{v}_2 \cdots A\mathbf{v}_n] \\
&= \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} [\lambda_1\mathbf{v}_1 \quad A\mathbf{v}_2 \cdots A\mathbf{v}_n] \\
&= \left[\begin{array}{c|c} \lambda_1 & * \\ \hline \mathbf{0} & A_1 \end{array}\right] = B
\end{aligned}$$


Baldwin H. Ward & Kathryn C. Ward/CORBIS

In a lecture he delivered at the University of Göttingen in 1905, the German mathematician David Hilbert (1862–1943) considered linear operators acting on certain infinite-dimensional vector spaces. Out of this lecture arose the notion of a quadratic form in infinitely many variables, and it was in this context that Hilbert first used the term *spectrum* to mean a complete set of eigenvalues. The spaces in question are now called *Hilbert spaces*.

Hilbert made major contributions to many areas of mathematics, among them integral equations, number theory, geometry, and the foundations of mathematics. In 1900, at the Second International Congress of Mathematicians in Paris, Hilbert gave an address entitled "The Problems of Mathematics." In it, he challenged mathematicians to solve 23 problems of fundamental importance during the coming century. Many of the problems have been solved—some were proved true, others false—and some may never be solved. Nevertheless, Hilbert's speech energized the mathematical community and is often regarded as the most influential speech ever given about mathematics.

since $\mathbf{v}_1^T(\lambda_1\mathbf{v}_1) = \lambda_1(\mathbf{v}_1^T\mathbf{v}_1) = \lambda_1(\mathbf{v}_1 \cdot \mathbf{v}_1) = \lambda_1$ and $\mathbf{v}_i^T(\lambda_1\mathbf{v}_1) = \lambda_1(\mathbf{v}_i^T\mathbf{v}_1) = \lambda_1(\mathbf{v}_i \cdot \mathbf{v}_1) = 0$ for $i \neq 1$, because $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthonormal set.

But

$$B^T = (Q_1^TAQ_1)^T = Q_1^TA^T(Q_1^T)^T = Q_1^TAQ_1 = B$$

so B is symmetric. Therefore, B has the block form

$$B = \left[\begin{array}{c|c} \lambda_1 & \mathbf{0} \\ \hline \mathbf{0} & A_1 \end{array}\right]$$

and A_1 is symmetric. Furthermore, B is similar to A (why?), so the characteristic polynomial of B is equal to the characteristic polynomial of A, by Theorem 4.22. By Exercise 39 in Section 4.3, the characteristic polynomial of A_1 divides the characteristic polynomial of A. It follows that the eigenvalues of A_1 are also eigenvalues of A and, hence, are real. We also see that A_1 has real entries. (Why?) Thus, A_1 is a $k \times k$ real symmetric matrix with real eigenvalues, so the induction hypothesis applies to it. Hence, there is an orthogonal matrix P_2 such that $P_2^TA_1P_2$ is a diagonal matrix—say, D_1. Now let

$$Q_2 = \left[\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & P_2 \end{array}\right]$$

Then Q_2 is an orthogonal $(k+1)\times(k+1)$ matrix, and therefore so is $Q = Q_1Q_2$. Consequently,

$$\begin{aligned} Q^TAQ &= (Q_1Q_2)^TA(Q_1Q_2) = (Q_2^TQ_1^T)A(Q_1Q_2) = Q_2^T(Q_1^TAQ_1)Q_2 = Q_2^TBQ_2 \\ &= \left[\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & P_2^T \end{array}\right]\left[\begin{array}{c|c} \lambda_1 & \mathbf{0} \\ \hline \mathbf{0} & A_1 \end{array}\right]\left[\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & P_2 \end{array}\right] \\ &= \left[\begin{array}{c|c} \lambda_1 & \mathbf{0} \\ \hline \mathbf{0} & P_2^TA_1P_2 \end{array}\right] \\ &= \left[\begin{array}{c|c} \lambda_1 & \mathbf{0} \\ \hline \mathbf{0} & D_1 \end{array}\right] \end{aligned}$$

which is a diagonal matrix. This completes the induction step, and we conclude that, for all $n \geq 1$, an $n \times n$ real symmetric matrix with real eigenvalues is orthogonally diagonalizable.

Example 5.18

Orthogonally diagonalize the matrix

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

Solution This is the matrix from Example 5.17. We have already found that the eigenspaces of A are

$$E_4 = \text{span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right) \quad \text{and} \quad E_1 = \text{span}\left(\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}\right)$$

We need three orthonormal eigenvectors. First, we apply the Gram-Schmidt Process to

$$\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

to obtain

$$\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} -\frac{1}{2} \\ 1 \\ -\frac{1}{2} \end{bmatrix}$$

The new vector, which has been constructed to be orthogonal to $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, is still in E_1 (why?) and so is orthogonal to $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Thus, we have three mutually orthogonal vectors, and all we need to do is normalize them and construct a matrix Q with these vectors as its columns. We find that

$$Q = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 0 & 2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \end{bmatrix}$$

and it is straightforward to verify that

$$Q^TAQ = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The Spectral Theorem allows us to write a real symmetric matrix A in the form $A = QDQ^T$, where Q is orthogonal and D is diagonal. The diagonal entries of D are just the eigenvalues of A, and if the columns of Q are the orthonormal vectors $\mathbf{q}_1, \ldots, \mathbf{q}_n$, then, using the column-row representation of the product, we have

$$\begin{aligned} A = QDQ^T &= [\mathbf{q}_1 \ \cdots \ \mathbf{q}_n] \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} \\ &= [\lambda_1\mathbf{q}_1 \ \cdots \ \lambda_n\mathbf{q}_n] \begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} \\ &= \lambda_1\mathbf{q}_1\mathbf{q}_1^T + \lambda_2\mathbf{q}_2\mathbf{q}_2^T + \cdots + \lambda_n\mathbf{q}_n\mathbf{q}_n^T \end{aligned}$$

This is called the ***spectral decomposition*** of A. Each of the terms $\lambda_i\mathbf{q}_i\mathbf{q}_i^T$ is a rank 1 matrix, by Exercise 62 in Section 3.5, and $\mathbf{q}_i\mathbf{q}_i^T$ is actually the matrix of the projection onto the subspace spanned by $\mathbf{q}_i$. (See Exercise 25.) For this reason, the spectral decomposition

$$A = \lambda_1\mathbf{q}_1\mathbf{q}_1^T + \lambda_2\mathbf{q}_2\mathbf{q}_2^T + \cdots + \lambda_n\mathbf{q}_n\mathbf{q}_n^T$$

is sometimes referred to as the ***projection form of the Spectral Theorem.***

Example 5.19

Find the spectral decomposition of the matrix A from Example 5.18.

Solution From Example 5.18, we have:

$$\lambda_1 = 4, \qquad \lambda_2 = 1, \qquad \lambda_3 = 1$$

$$\mathbf{q}_1 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}, \quad \mathbf{q}_2 = \begin{bmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{q}_3 = \begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ -1/\sqrt{6} \end{bmatrix}$$

Therefore,

$$\mathbf{q}_1\mathbf{q}_1^T = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix} [1/\sqrt{3} \quad 1/\sqrt{3} \quad 1/\sqrt{3}] = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

$$\mathbf{q}_2\mathbf{q}_2^T = \begin{bmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix} [-1/\sqrt{2} \quad 0 \quad 1/\sqrt{2}] = \begin{bmatrix} 1/2 & 0 & -1/2 \\ 0 & 0 & 0 \\ -1/2 & 0 & 1/2 \end{bmatrix}$$

$$\mathbf{q}_3\mathbf{q}_3^T = \begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ -1/\sqrt{6} \end{bmatrix} [-1/\sqrt{6} \quad 2/\sqrt{6} \quad -1/\sqrt{6}] = \begin{bmatrix} 1/6 & -1/3 & 1/6 \\ -1/3 & 2/3 & -1/3 \\ 1/6 & -1/3 & 1/6 \end{bmatrix}$$

so

$$\begin{aligned} A &= \lambda_1\mathbf{q}_1\mathbf{q}_1^T + \lambda_2\mathbf{q}_2\mathbf{q}_2^T + \lambda_3\mathbf{q}_3\mathbf{q}_3^T \\ &= 4\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} + \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} + \begin{bmatrix} \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \end{bmatrix} \end{aligned}$$

which can be easily verified.

In this example, $\lambda_2 = \lambda_3$, so we could combine the last two terms $\lambda_2\mathbf{q}_2\mathbf{q}_2^T + \lambda_3\mathbf{q}_3\mathbf{q}_3^T$ to get

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

The rank 2 matrix $\mathbf{q}_2\mathbf{q}_2^T + \mathbf{q}_3\mathbf{q}_3^T$ is the matrix of a projection onto the two-dimensional subspace (i.e., the plane) spanned by $\mathbf{q}_2$ and $\mathbf{q}_3$. (See Exercise 26.)

Observe that the spectral decomposition expresses a symmetric matrix A explicitly in terms of its eigenvalues and eigenvectors. This gives us a way of constructing a matrix with given eigenvalues and (orthonormal) eigenvectors.

Example 5.20

Find a 2×2 matrix with eigenvalues $\lambda_1 = 3$ and $\lambda_2 = -2$ and corresponding eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$$

Solution We begin by normalizing the vectors to obtain an orthonormal basis $\{\mathbf{q}_1, \mathbf{q}_2\}$, with

$$\mathbf{q}_1 = \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix} \quad \text{and} \quad \mathbf{q}_2 = \begin{bmatrix} -\frac{4}{5} \\ \frac{3}{5} \end{bmatrix}$$

Now, we compute the matrix A whose spectral decomposition is

$$\begin{aligned} A &= \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^T \\ &= 3\begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix}\begin{bmatrix} \frac{3}{5} & \frac{4}{5} \end{bmatrix} - 2\begin{bmatrix} -\frac{4}{5} \\ \frac{3}{5} \end{bmatrix}\begin{bmatrix} -\frac{4}{5} & \frac{3}{5} \end{bmatrix} \\ &= 3\begin{bmatrix} \frac{9}{25} & \frac{12}{25} \\ \frac{12}{25} & \frac{16}{25} \end{bmatrix} - 2\begin{bmatrix} \frac{16}{25} & -\frac{12}{25} \\ -\frac{12}{25} & \frac{9}{25} \end{bmatrix} \\ &= \begin{bmatrix} -\frac{1}{5} & \frac{12}{5} \\ \frac{12}{5} & \frac{6}{5} \end{bmatrix} \end{aligned}$$

It is easy to check that A has the desired properties. (Do this.)

Exercises 5.4

Orthogonally diagonalize the matrices in Exercises 1–10 by finding an orthogonal matrix Q and a diagonal matrix D such that $Q^TAQ = D$.

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

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

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

4. $A = \begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix}$

5. $A = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 1 & 3 \\ 0 & 3 & 1 \end{bmatrix}$

6. $A = \begin{bmatrix} 2 & 3 & 0 \\ 3 & 2 & 4 \\ 0 & 4 & 2 \end{bmatrix}$

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

8. $A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}$

9. $A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$

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

11. If $b \neq 0$, orthogonally diagonalize $A = \begin{bmatrix} a & b \\ b & a \end{bmatrix}$.

12. If $b \neq 0$, orthogonally diagonalize $A = \begin{bmatrix} a & 0 & b \\ 0 & a & 0 \\ b & 0 & a \end{bmatrix}$.

13. Let A and B be orthogonally diagonalizable $n \times n$ matrices and let c be a scalar. Use the Spectral Theorem to prove that the following matrices are orthogonally diagonalizable:

(a) $A + B$ **(b)** cA **(c)** A^2

14. If A is an invertible matrix that is orthogonally diagonalizable, show that A^{-1} is orthogonally diagonalizable.

15. If A and B are orthogonally diagonalizable and $AB = BA$, show that AB is orthogonally diagonalizable.

16. If A is a symmetric matrix, show that every eigenvalue of A is nonnegative if and only if $A = B^2$ for some symmetric matrix B.

In Exercises 17–20, find a spectral decomposition of the matrix in the given exercise.

17. Exercise 1

18. Exercise 2

19. Exercise 5

20. Exercise 8

In Exercises 21 and 22, find a symmetric 2×2 matrix with eigenvalues λ_1 and λ_2 and corresponding orthogonal eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$.

21. $\lambda_1 = -1, \lambda_2 = 2, \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

22. $\lambda_1 = 3, \lambda_2 = -3, \mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$

In Exercises 23 and 24, find a symmetric 3×3 matrix with eigenvalues λ_1, λ_2, and λ_3 and corresponding orthogonal eigenvectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$.

23. $\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 3, \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix},$
$\mathbf{v}_3 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$

24. $\lambda_1 = 1, \lambda_2 = -4, \lambda_3 = -4, \mathbf{v}_1 = \begin{bmatrix} 4 \\ 5 \\ -1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix},$
$\mathbf{v}_3 = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$

25. Let $\mathbf{q}$ be a unit vector in $\mathbb{R}^n$ and let W be the subspace spanned by $\mathbf{q}$. Show that the orthogonal projection of a vector $\mathbf{v}$ onto W (as defined in Sections 1.2 and 5.2) is given by

$$\text{proj}_W(\mathbf{v}) = (\mathbf{q}\mathbf{q}^T)\mathbf{v}$$

and that the matrix of this projection is thus $\mathbf{q}\mathbf{q}^T$. [*Hint:* Remember that, for $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T\mathbf{y}$.]

26. Let $\{\mathbf{q}_1, \ldots, \mathbf{q}_k\}$ be an orthonormal set of vectors in $\mathbb{R}^n$ and let W be the subspace spanned by this set.

(a) Show that the matrix of the orthogonal projection onto W is given by

$$P = \mathbf{q}_1\mathbf{q}_1^T + \cdots + \mathbf{q}_k\mathbf{q}_k^T$$

(b) Show that the projection matrix P in part (a) is symmetric and satisfies $P^2 = P$.

(c) Let $Q = [\mathbf{q}_1 \; \cdots \; \mathbf{q}_k]$ be the $n \times k$ matrix whose columns are the orthonormal basis vectors of W. Show that $P = QQ^T$ and deduce that $\text{rank}(P) = k$.

27. Let A be an $n \times n$ real matrix, all of whose eigenvalues are real. Prove that there exist an orthogonal matrix Q and an upper triangular matrix T such that $Q^TAQ = T$. This very useful result is known as ***Schur's Triangularization Theorem.*** [*Hint:* Adapt the proof of the Spectral Theorem.]

28. Let A be a nilpotent matrix (see Exercise 56 in Section 4.2). Prove that there is an orthogonal matrix Q such that $Q^T AQ$ is upper triangular with zeros on its diagonal. [*Hint:* Use Exercise 27.]

5.5 Applications

Quadratic Forms

An expression of the form

$$ax^2 + by^2 + cxy$$

is called a ***quadratic form*** in x and y. Similarly,

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz$$

is a quadratic form in x, y, and z. In words, a quadratic form is a sum of terms, each of which has total degree *two* in the variables. Therefore, $5x^2 - 3y^2 + 2xy$ is a quadratic form, but $x^2 + y^2 + x$ is not.

We can represent quadratic forms using matrices as follows:

$$ax^2 + by^2 + cxy = [x \quad y]\begin{bmatrix} a & c/2 \\ c/2 & b \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

and

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz = [x \quad y \quad z]\begin{bmatrix} a & d/2 & e/2 \\ d/2 & b & f/2 \\ e/2 & f/2 & c \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

(Verify these.) Each has the form $\mathbf{x}^T A\mathbf{x}$, where the matrix A is symmetric. This observation leads us to the following general definition.

Definition A ***quadratic form*** in n variables is a function $f : \mathbb{R}^n \to \mathbb{R}$ of the form

$$f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$$

where A is a symmetric $n \times n$ matrix and $\mathbf{x}$ is in $\mathbb{R}^n$. We refer to A as the ***matrix associated with f***.

Example 5.21

What is the quadratic form with associated matrix $A = \begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}$?

Solution If $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, then

$$f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = [x_1 \quad x_2]\begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2x_1^2 + 5x_2^2 - 6x_1x_2$$

Observe that the *off-diagonal* entries $a_{12} = a_{21} = -3$ of A are *combined* to give the coefficient -6 of x_1x_2. This is true generally. We can expand a quadratic form in n variables $\mathbf{x}^T A\mathbf{x}$ as follows:

$$\mathbf{x}^T A\mathbf{x} = a_{11}x_1^2 + a_{22}x_2^2 + \cdots + a_{nn}x_n^2 + \sum_{i<j} 2a_{ij}x_ix_j$$

Thus, if $i \neq j$, the coefficient of x_ix_j is $2a_{ij}$.

Example 5.22

Find the matrix associated with the quadratic form

$$f(x_1, x_2, x_3) = 2x_1^2 - x_2^2 + 5x_3^2 + 6x_1x_2 - 3x_1x_3$$

Solution The coefficients of the squared terms x_i^2 go on the diagonal as a_{ii}, and the coefficients of the cross-product terms x_ix_j are split between a_{ij} and a_{ji}. This gives

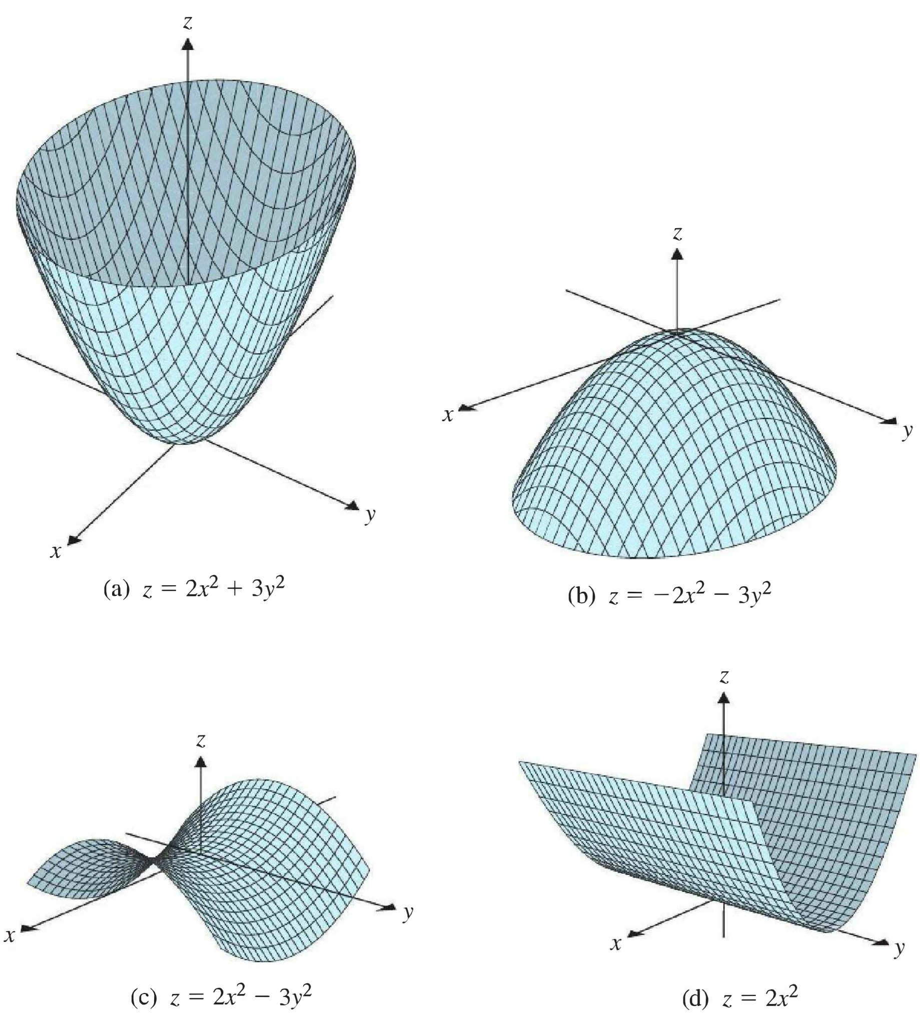
$$A = \begin{bmatrix} 2 & 3 & -\frac{3}{2} \\ 3 & -1 & 0 \\ -\frac{3}{2} & 0 & 5 \end{bmatrix}$$

so $$f(x_1, x_2, x_3) = [x_1 \quad x_2 \quad x_3] \begin{bmatrix} 2 & 3 & -\frac{3}{2} \\ 3 & -1 & 0 \\ -\frac{3}{2} & 0 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

as you can easily check.

In the case of a quadratic form $f(x, y)$ in two variables, the graph of $z = f(x, y)$ is a surface in $\mathbb{R}^3$. Some examples are shown in Figure 5.12.

Observe that the effect of holding x or y constant is to take a cross section of the graph parallel to the yz or xz planes, respectively. For the graphs in Figure 5.12, all of these cross sections are easy to identify. For example, in Figure 5.12(a), the cross sections we get by holding x or y constant are all parabolas opening upward, so $f(x, y) \geq 0$ for all values of x and y. In Figure 5.12(c), holding x constant gives parabolas opening downward and holding y constant gives parabolas opening upward, producing a *saddle point*.



(a) $z = 2x^2 + 3y^2$

(b) $z = -2x^2 - 3y^2$

(c) $z = 2x^2 - 3y^2$

(d) $z = 2x^2$

Figure 5.12

Graphs of quadratic forms $f(x, y)$

What makes this type of analysis quite easy is the fact that these quadratic forms have no cross-product terms. The matrix associated with such a quadratic form is a diagonal matrix. For example,

$$2x^2 - 3y^2 = [x \quad y]\begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

In general, the matrix of a quadratic form is a symmetric matrix, and we saw in Section 5.4 that such matrices can always be diagonalized. We will now use this fact to show that, for *every* quadratic form, we can eliminate the cross-product terms by means of a suitable change of variable.

Let $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ be a quadratic form in n variables, with A a symmetric $n \times n$ matrix. By the Spectral Theorem, there is an orthogonal matrix Q that diagonalizes A; that is, $Q^TAQ = D$, where D is a diagonal matrix displaying the eigenvalues of A. We now set

$$\mathbf{x} = Q\mathbf{y} \quad \text{or, equivalently,} \quad \mathbf{y} = Q^{-1}\mathbf{x} = Q^T\mathbf{x}$$

Substitution into the quadratic form yields

$$\begin{aligned} \mathbf{x}^T A\mathbf{x} &= (Q\mathbf{y})^T A(Q\mathbf{y}) \\ &= \mathbf{y}^T Q^T A Q\mathbf{y} \\ &= \mathbf{y}^T D\mathbf{y} \end{aligned}$$

which is a quadratic form without cross-product terms, since D is diagonal. Furthermore, if the eigenvalues of A are $\lambda_1, \ldots, \lambda_n$, then Q can be chosen so that

$$D = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

If $\mathbf{y} = [y_1 \quad \cdots \quad y_n]^T$, then, with respect to these new variables, the quadratic form becomes

$$\mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$$

This process is called ***diagonalizing a quadratic form.*** We have just proved the following theorem, known as the ***Principal Axes Theorem.*** (The reason for this name will become clear in the next subsection.)

Theorem 5.21 **The Principal Axes Theorem**

Every quadratic form can be diagonalized. Specifically, if A is the $n \times n$ symmetric matrix associated with the quadratic form $\mathbf{x}^T A\mathbf{x}$ and if Q is an orthogonal matrix such that $Q^TAQ = D$ is a diagonal matrix, then the change of variable $\mathbf{x} = Q\mathbf{y}$ transforms the quadratic form $\mathbf{x}^T A\mathbf{x}$ into the quadratic form $\mathbf{y}^T D\mathbf{y}$, which has no cross-product terms. If the eigenvalues of A are $\lambda_1, \ldots, \lambda_n$ and $\mathbf{y} = [y_1 \quad \cdots \quad y_n]^T$, then

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$$

Example 5.23 Find a change of variable that transforms the quadratic form

$$f(x_1, x_2) = 5x_1^2 + 4x_1x_2 + 2x_2^2$$

into one with no cross-product terms.

Solution The matrix of f is

$$A = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$$

with eigenvalues $\lambda_1 = 6$ and $\lambda_2 = 1$. Corresponding unit eigenvectors are

$$\mathbf{q}_1 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix} \quad \text{and} \quad \mathbf{q}_2 = \begin{bmatrix} 1/\sqrt{5} \\ -2/\sqrt{5} \end{bmatrix}$$

(Check this.) If we set

$$Q = \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ 1/\sqrt{5} & -2/\sqrt{5} \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 6 & 0 \\ 0 & 1 \end{bmatrix}$$

then $Q^TAQ = D$. The change of variable $\mathbf{x} = Q\mathbf{y}$, where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

converts f into

$$f(\mathbf{y}) = f(y_1, y_2) = [y_1 \quad y_2]\begin{bmatrix} 6 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 6y_1^2 + y_2^2$$

The original quadratic form $\mathbf{x}^TA\mathbf{x}$ and the new one $\mathbf{y}^TD\mathbf{y}$ (referred to in the Principal Axes Theorem) are *equal* in the following sense. In Example 5.23, suppose we want to evaluate $f(\mathbf{x}) = \mathbf{x}^TA\mathbf{x}$ at $\mathbf{x} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$. We have

$$f(-1, 3) = 5(-1)^2 + 4(-1)(3) + 2(3)^2 = 11$$

In terms of the new variables,

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{y} = Q^T\mathbf{x} = \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ 1/\sqrt{5} & -2/\sqrt{5} \end{bmatrix}\begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} \\ -7/\sqrt{5} \end{bmatrix}$$

so

$$f(y_1, y_2) = 6y_1^2 + y_2^2 = 6(1/\sqrt{5})^2 + (-7/\sqrt{5})^2 = 55/5 = 11$$

exactly as before.

The Principal Axes Theorem has some interesting and important consequences. We will consider two of these. The first relates to the possible *values* that a quadratic form can take on.

Definition A quadratic form $f(\mathbf{x}) = \mathbf{x}^TA\mathbf{x}$ is classified as one of the following:

1. ***positive definite*** if $f(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$
2. ***positive semidefinite*** if $f(\mathbf{x}) \geq 0$ for all $\mathbf{x}$
3. ***negative definite*** if $f(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$
4. ***negative semidefinite*** if $f(\mathbf{x}) \leq 0$ for all $\mathbf{x}$
5. ***indefinite*** if $f(\mathbf{x})$ takes on both positive and negative values

A symmetric matrix A is called ***positive definite, positive semidefinite, negative definite, negative semidefinite,*** or ***indefinite*** if the associated quadratic form $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ has the corresponding property.

The quadratic forms in parts (a), (b), (c), and (d) of Figure 5.12 are positive definite, negative definite, indefinite, and positive semidefinite, respectively. The Principal Axes Theorem makes it easy to tell if a quadratic form has one of these properties.

Theorem 5.22

Let A be an $n \times n$ symmetric matrix. The quadratic form $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is

a. positive definite if and only if all of the eigenvalues of A are positive.
b. positive semidefinite if and only if all of the eigenvalues of A are nonnegative.
c. negative definite if and only if all of the eigenvalues of A are negative.
d. negative semidefinite if and only if all of the eigenvalues of A are nonpositive.
e. indefinite if and only if A has both positive and negative eigenvalues.

You are asked to prove Theorem 5.22 in Exercise 27.

Example 5.24

Classify $f(x, y, z) = 3x^2 + 3y^2 + 3z^2 - 2xy - 2xz - 2yz$ as positive definite, negative definite, indefinite, or none of these.

Solution The matrix associated with f is

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

which has eigenvalues 1, 4, and 4. (Verify this.) Since all of these eigenvalues are positive, f is a positive definite quadratic form.

If a quadratic form $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is positive definite, then, since $f(\mathbf{0}) = 0$, the *minimum* value of $f(\mathbf{x})$ is 0 and it occurs at the origin. Similarly, a negative definite quadratic form has a maximum at the origin. Thus, Theorem 5.22 allows us to solve certain types of maxima/minima problems easily, without resorting to calculus. A type of problem that falls into this category is the ***constrained optimization problem.***

It is often important to know the maximum or minimum values of a quadratic form subject to certain constraints. (Such problems arise not only in mathematics but also in statistics, physics, engineering, and economics.) We will be interested in finding the extreme values of $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ subject to the constraint that $\|\mathbf{x}\| = 1$. In the case of a quadratic form in two variables, we can visualize what the problem means. The graph of $z = f(x, y)$ is a surface in $\mathbb{R}^3$, and the constraint $\|\mathbf{x}\| = 1$ restricts the point (x, y) to the unit circle in the xy-plane. Thus, we are considering those points that lie simultaneously on the surface and on the unit cylinder perpendicular to the xy plane. These points form a curve lying on the surface, and we want the highest and lowest points on this curve. Figure 5.13 shows this situation for the quadratic form and corresponding surface in Figure 5.12(c).

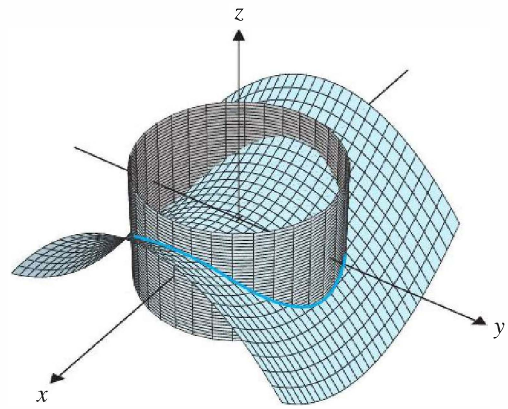


Figure 5.13
The intersection of $z = 2x^2 - 3y^2$ with the cylinder $x^2 + y^2 = 1$

In this case, the maximum and minimum values of $f(x, y) = 2x^2 - 3y^2$ (the highest and lowest points on the curve of intersection) are 2 and -3, respectively, which are just the eigenvalues of the associated matrix. Theorem 5.23 shows that this is always the case.

Theorem 5.23 Let $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ be a quadratic form with associated $n \times n$ symmetric matrix A. Let the eigenvalues of A be $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then the following are true, subject to the constraint $\|\mathbf{x}\| = 1$:

a. $\lambda_1 \geq f(\mathbf{x}) \geq \lambda_n$
b. The maximum value of $f(\mathbf{x})$ is λ_1, and it occurs when $\mathbf{x}$ is a unit eigenvector corresponding to λ_1.
c. The minimum value of $f(\mathbf{x})$ is λ_n, and it occurs when $\mathbf{x}$ is a unit eigenvector corresponding to λ_n.

Proof As usual, we begin by orthogonally diagonalizing A. Accordingly, let Q be an orthogonal matrix such that $Q^T AQ$ is the diagonal matrix

$$D = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

Then, by the Principal Axes Theorem, the change of variable $\mathbf{x} = Q\mathbf{y}$ gives $\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y}$. Now note that $\mathbf{y} = Q^T\mathbf{x}$ implies that

$$\mathbf{y}^T\mathbf{y} = (Q^T\mathbf{x})^T(Q^T\mathbf{x}) = \mathbf{x}^T(Q^T)^T Q^T\mathbf{x} = \mathbf{x}^T QQ^T\mathbf{x} = \mathbf{x}^T\mathbf{x}$$

since $Q^T = Q^{-1}$. Hence, using $\mathbf{x} \cdot \mathbf{x} = \mathbf{x}^T\mathbf{x}$, we see that $\|\mathbf{y}\| = \sqrt{\mathbf{y}^T\mathbf{y}} = \sqrt{\mathbf{x}^T\mathbf{x}} = \|\mathbf{x}\| = 1$. Thus, if $\mathbf{x}$ is a unit vector, so is the corresponding $\mathbf{y}$, and the values of $\mathbf{x}^T A\mathbf{x}$ and $\mathbf{y}^T D\mathbf{y}$ are the same.

(a) To prove property (a), we observe that if $\mathbf{y} = [y_1 \ \cdots \ y_n]^T$, then

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{x}^T A \mathbf{x} = \mathbf{y}^T D \mathbf{y} \\ &= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \\ &\le \lambda_1 y_1^2 + \lambda_1 y_2^2 + \cdots + \lambda_1 y_n^2 \\ &= \lambda_1 (y_1^2 + y_2^2 + \cdots + y_n^2) \\ &= \lambda_1 \|\mathbf{y}\|^2 \\ &= \lambda_1 \end{aligned}$$

Thus, $f(\mathbf{x}) \le \lambda_1$ for all $\mathbf{x}$ such that $\|\mathbf{x}\| = 1$. The proof that $f(\mathbf{x}) \ge \lambda_n$ is similar. (See Exercise 37.)

(b) If $\mathbf{q}_1$ is a unit eigenvector corresponding to λ_1, then $A\mathbf{q}_1 = \lambda_1 \mathbf{q}_1$ and

$$f(\mathbf{q}_1) = \mathbf{q}_1^T A \mathbf{q}_1 = \mathbf{q}_1^T \lambda_1 \mathbf{q}_1 = \lambda_1 (\mathbf{q}_1^T \mathbf{q}_1) = \lambda_1$$

This shows that the quadratic form actually takes on the value λ_1, and so, by property (a), it is the maximum value of $f(\mathbf{x})$ and it occurs when $\mathbf{x} = \mathbf{q}_1$.

(c) You are asked to prove this property in Exercise 38.

Example 5.25

Find the maximum and minimum values of the quadratic form $f(x_1, x_2) = 5x_1^2 + 4x_1x_2 + 2x_2^2$ subject to the constraint $x_1^2 + x_2^2 = 1$, and determine values of x_1 and x_2 for which each of these occurs.

Solution In Example 5.23, we found that f has the associated eigenvalues $\lambda_1 = 6$ and $\lambda_2 = 1$, with corresponding unit eigenvectors

$$\mathbf{q}_1 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix} \quad \text{and} \quad \mathbf{q}_2 = \begin{bmatrix} 1/\sqrt{5} \\ -2/\sqrt{5} \end{bmatrix}$$

Therefore, the maximum value of f is 6 when $x_1 = 2/\sqrt{5}$ and $x_2 = 1/\sqrt{5}$. The minimum value of f is 1 when $x_1 = 1/\sqrt{5}$ and $x_2 = -2/\sqrt{5}$. (Observe that these extreme values occur twice—in opposite directions—since $-\mathbf{q}_1$ and $-\mathbf{q}_2$ are also unit eigenvectors for λ_1 and λ_2, respectively.)

Graphing Quadratic Equations

The general form of a quadratic equation in two variables x and y is

$$ax^2 + by^2 + cxy + dx + ey + f = 0$$

where at least one of a, b, and c is nonzero. The graphs of such quadratic equations are called ***conic sections*** (or ***conics***), since they can be obtained by taking cross sections of a (double) cone (i.e., slicing it with a plane). The most important of the conic sections are the ellipses (with circles as a special case), hyperbolas, and parabolas. These are called the ***nondegenerate*** conics. Figure 5.14 shows how they arise.

It is also possible for a cross section of a cone to result in a single point, a straight line, or a pair of lines. These are called ***degenerate*** conics. (See Exercises 59–64.)

The graph of a nondegenerate conic is said to be in ***standard position*** relative to the coordinate axes if its equation can be expressed in one of the forms in Figure 5.15.

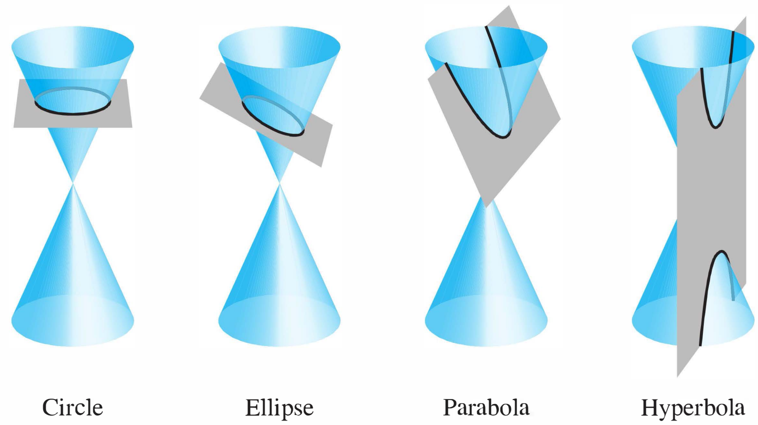


Figure 5.14
The nondegenerate conics

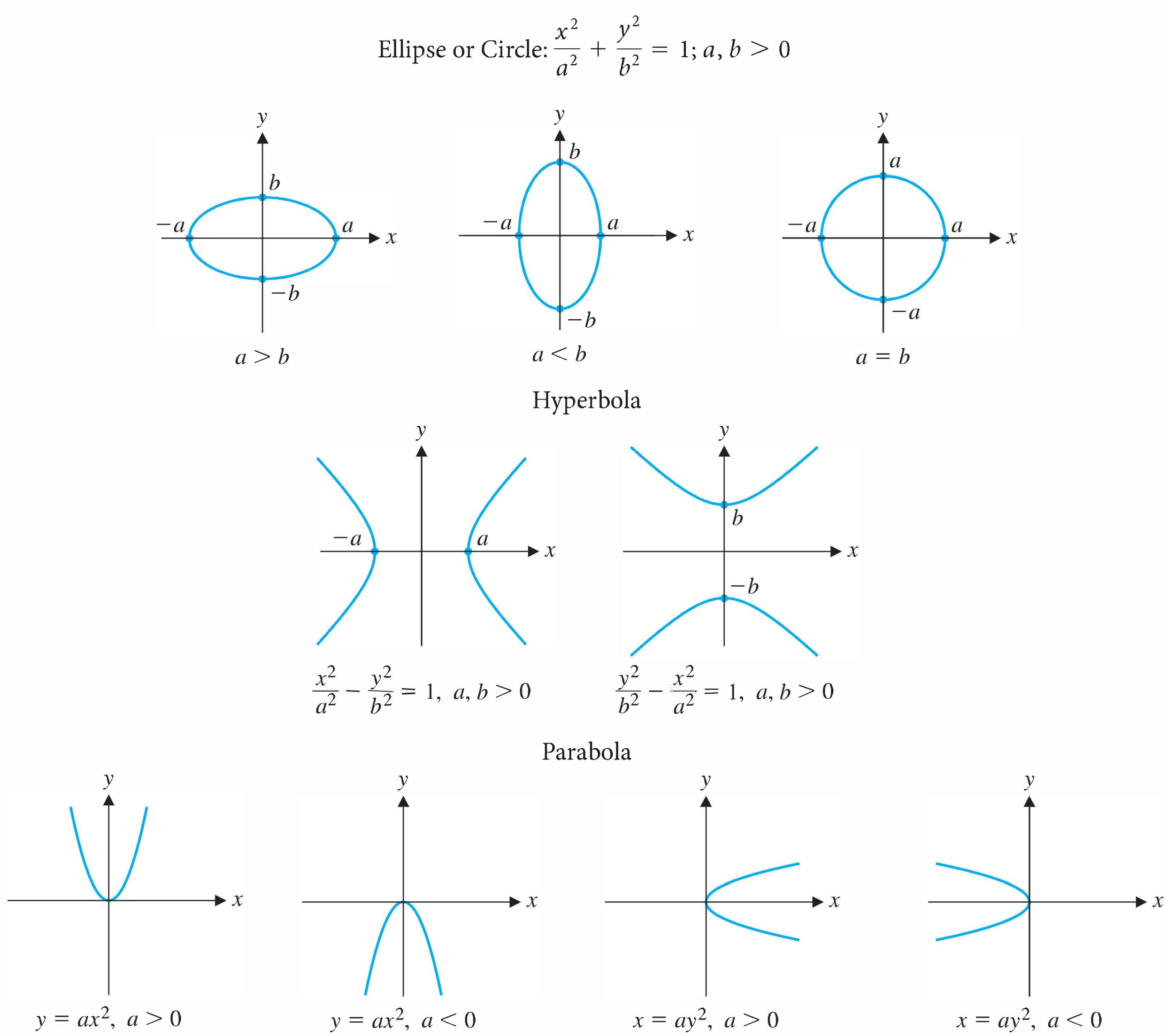


Figure 5.15
Nondegenerate conics in standard position

Example 5.26

If possible, write each of the following quadratic equations in the form of a conic in standard position and identify the resulting graph.

(a) $4x^2 + 9y^2 = 36$ (b) $4x^2 - 9y^2 + 1 = 0$ (c) $4x^2 - 9y = 0$

Solution (a) The equation $4x^2 + 9y^2 = 36$ can be written in the form

$$\frac{x^2}{9} + \frac{y^2}{4} = 1$$

so its graph is an ellipse intersecting the x-axis at $(\pm 3, 0)$ and the y-axis at $(0, \pm 2)$.

(b) The equation $4x^2 - 9y^2 + 1 = 0$ can be written in the form

$$\frac{y^2}{\frac{1}{9}} - \frac{x^2}{\frac{1}{4}} = 1$$

so its graph is a hyperbola, opening up and down, intersecting the y-axis at $(0, \pm\frac{1}{3})$.

(c) The equation $4x^2 - 9y = 0$ can be written in the form

$$y = \frac{4}{9}x^2$$

so its graph is a parabola opening upward.

If a quadratic equation contains too many terms to be written in one of the forms in Figure 5.15, then its graph is not in standard position. When there are additional terms but no xy term, the graph of the conic has been *translated* out of standard position.

Example 5.27

Identify and graph the conic whose equation is

$$x^2 + 2y^2 - 6x + 8y + 9 = 0$$

Solution We begin by grouping the x and y terms separately to get

$$(x^2 - 6x) + (2y^2 + 8y) = -9$$

or

$$(x^2 - 6x) + 2(y^2 + 4y) = -9$$

Next, we complete the squares on the two expressions in parentheses to obtain

$$(x^2 - 6x + 9) + 2(y^2 + 4y + 4) = -9 + 9 + 8$$

or

$$(x - 3)^2 + 2(y + 2)^2 = 8$$

We now make the substitutions $x' = x - 3$ and $y' = y + 2$, turning the above equation into

$$(x')^2 + 2(y')^2 = 8 \quad \text{or} \quad \frac{(x')^2}{8} + \frac{(y')^2}{4} = 1$$

This is the equation of an ellipse in standard position in the $x'y'$ coordinate system, intersecting the x'-axis at $(\pm 2\sqrt{2}, 0)$ and the y'-axis at $(0, \pm 2)$. The origin in the $x'y'$ coordinate system is at $x = 3$, $y = -2$, so the ellipse has been translated out of standard position 3 units to the right and 2 units down. Its graph is shown in Figure 5.16.

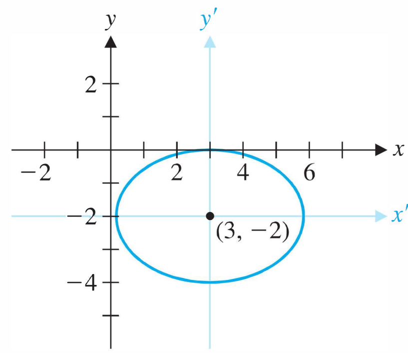


Figure 5.16
A translated ellipse

If a quadratic equation contains a cross-product term, then it represents a conic that has been *rotated*.

Example 5.28

Identify and graph the conic whose equation is

$$5x^2 + 4xy + 2y^2 = 6$$

Solution The left-hand side of the equation is a quadratic form, so we can write it in matrix form as $\mathbf{x}^T A\mathbf{x} = 6$, where

$$A = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$$

In Example 5.23, we found that the eigenvalues of A are 6 and 1, and a matrix Q that orthogonally diagonalizes A is

$$Q = \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ 1/\sqrt{5} & -2/\sqrt{5} \end{bmatrix}$$

Observe that $\det Q = -1$. In this example, we will interchange the columns of this matrix to make the determinant equal to $+1$. Then Q will be the matrix of a *rotation*, by Exercise 28 in Section 5.1. It is always possible to rearrange the columns of an orthogonal matrix Q to make its determinant equal to $+1$. (Why?) We set

$$Q = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}$$

instead, so that

$$Q^T A Q = \begin{bmatrix} 1 & 0 \\ 0 & 6 \end{bmatrix} = D$$

The change of variable $\mathbf{x} = Q\mathbf{x}'$ converts the given equation into the form $(\mathbf{x}')^T D\mathbf{x}' = 6$ by means of a rotation. If $\mathbf{x}' = \begin{bmatrix} x' \\ y' \end{bmatrix}$, then this equation is just

$$(x')^2 + 6(y')^2 = 6 \quad \text{or} \quad \frac{(x')^2}{6} + (y')^2 = 1$$

which represents an ellipse in the $x'y'$ coordinate system.

To graph this ellipse, we need to know which vectors play the roles of $\mathbf{e}_1' = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2' = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ in the new coordinate system. (These two vectors locate the positions of the x' and y' axes.) But, from $\mathbf{x} = Q\mathbf{x}'$, we have

$$Q\mathbf{e}_1' = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} \\ -2/\sqrt{5} \end{bmatrix}$$

and

$$Q\mathbf{e}_2' = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$$

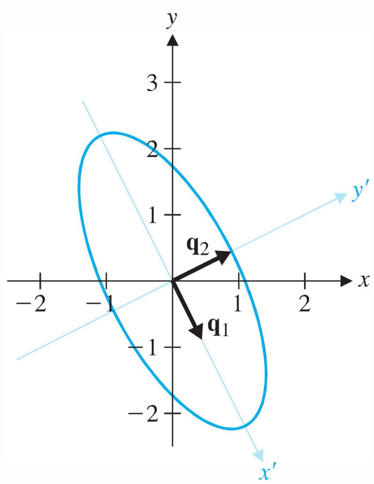


Figure 5.17
A rotated ellipse

These are just the columns $\mathbf{q}_1$ and $\mathbf{q}_2$ of Q, which are the eigenvectors of A! The fact that these are orthonormal vectors agrees perfectly with the fact that the change of variable is just a rotation. The graph is shown in Figure 5.17.

You can now see why the Principal Axes Theorem is so named. If a real symmetric matrix A arises as the coefficient matrix of a quadratic equation, the eigenvectors of A give the directions of the principal axes of the corresponding graph.

It is possible for the graph of a conic to be both rotated and translated out of standard position, as illustrated in Example 5.29.

Example 5.29

Identify and graph the conic whose equation is

$$5x^2 + 4xy + 2y^2 - \frac{28}{\sqrt{5}}x - \frac{4}{\sqrt{5}}y + 4 = 0$$

Solution The strategy is to eliminate the cross-product term first. In matrix form, the equation is $\mathbf{x}^T A\mathbf{x} + B\mathbf{x} + 4 = 0$, where

$$A = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -\frac{28}{\sqrt{5}} & -\frac{4}{\sqrt{5}} \end{bmatrix}$$

The cross-product term comes from the quadratic form $\mathbf{x}^T A\mathbf{x}$, which we diagonalize as in Example 5.28 by setting $\mathbf{x} = Q\mathbf{x}'$, where

$$Q = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}$$

Then, as in Example 5.28,

$$\mathbf{x}^T A\mathbf{x} = (\mathbf{x}')^T D\mathbf{x}' = (x')^2 + 6(y')^2$$

But now we also have

$$B\mathbf{x} = BQ\mathbf{x}' = \begin{bmatrix} -\frac{28}{\sqrt{5}} & -\frac{4}{\sqrt{5}} \end{bmatrix}\begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}\begin{bmatrix} x' \\ y' \end{bmatrix} = -4x' - 12y'$$

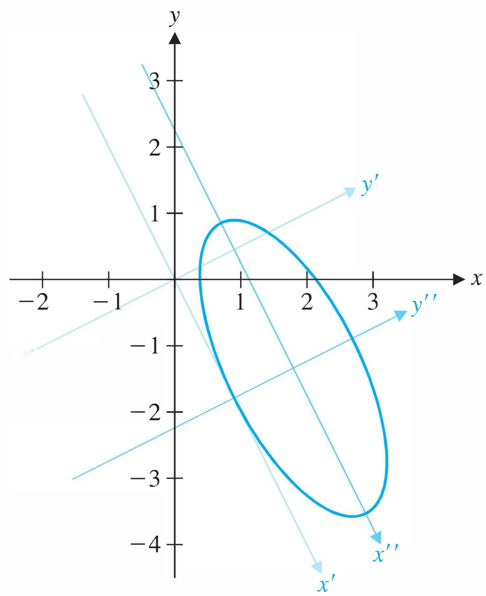


Figure 5.18

Thus, in terms of x' and y', the given equation becomes

$$(x')^2 + 6(y')^2 - 4x' - 12y' + 4 = 0$$

To bring the conic represented by this equation into standard position, we need to *translate* the $x'y'$ axes. We do so by completing the squares, as in Example 5.27. We have

$$((x')^2 - 4x' + 4) + 6((y')^2 - 2y' + 1) = -4 + 4 + 6 = 6$$

or

$$(x' - 2)^2 + 6(y' - 1)^2 = 6$$

This gives us the translation equations

$$x'' = x' - 2 \quad \text{and} \quad y'' = y' - 1$$

In the $x''y''$ coordinate system, the equation is simply

$$(x'')^2 + 6(y'')^2 = 6$$

which is the equation of an ellipse (as in Example 5.28). We can sketch this ellipse by first rotating and then translating. The resulting graph is shown in Figure 5.18.

The general form of a quadratic equation in three variables x, y, and z is

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j = 0$$

where at least one of $a, b, \ldots, f$ is nonzero. The graph of such a quadratic equation is called a ***quadric surface*** (or ***quadric***). Once again, to recognize a quadric we need

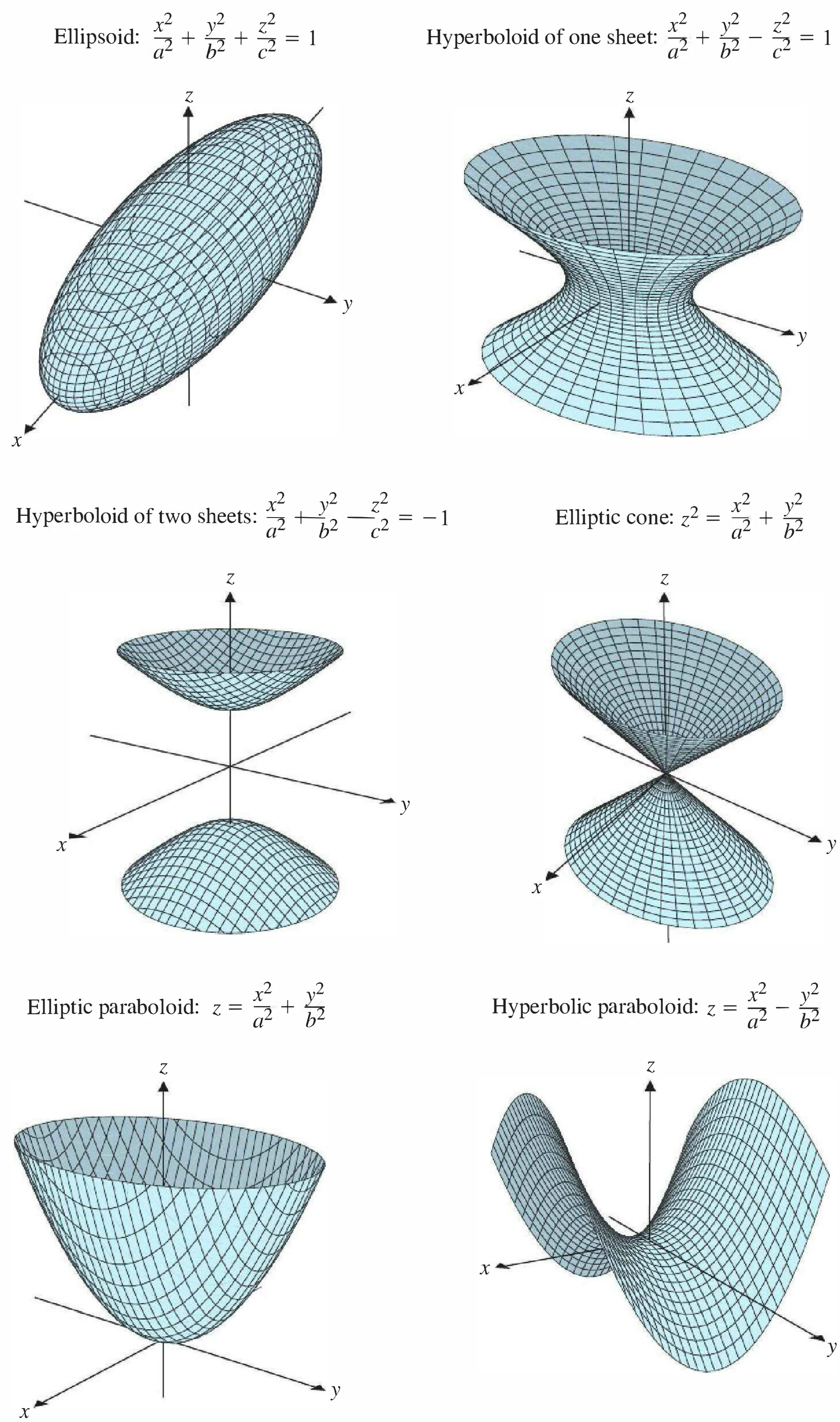


Figure 5.19
Quadric surfaces

to put it into standard position. Some quadrics in standard position are shown in Figure 5.19; others are obtained by permuting the variables.

Example 5.30

Identify the quadric surface whose equation is

$$5x^2 + 11y^2 + 2z^2 + 16xy + 20xz - 4yz = 36$$

Solution The equation can be written in matrix form as $\mathbf{x}^TA\mathbf{x} = 36$, where

$$A = \begin{bmatrix} 5 & 8 & 10 \\ 8 & 11 & -2 \\ 10 & -2 & 2 \end{bmatrix}$$

We find the eigenvalues of A to be 18, 9, and -9, with corresponding orthogonal eigenvectors

$$\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 2 \\ -1 \\ -2 \end{bmatrix}$$

respectively. We normalize them to obtain

$$\mathbf{q}_1 = \begin{bmatrix} \frac{2}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}, \quad \mathbf{q}_2 = \begin{bmatrix} \frac{1}{3} \\ -\frac{2}{3} \\ \frac{2}{3} \end{bmatrix}, \quad \text{and} \quad \mathbf{q}_3 = \begin{bmatrix} \frac{2}{3} \\ -\frac{1}{3} \\ -\frac{2}{3} \end{bmatrix}$$

and form the orthogonal matrix

$$Q = [\mathbf{q}_1 \ \ \mathbf{q}_2 \ \ \mathbf{q}_3] = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \end{bmatrix}$$

Note that in order for Q to be the matrix of a rotation, we require $\det Q = 1$, which is true in this case. (Otherwise, $\det Q = -1$, and swapping two columns changes the sign of the determinant.) Therefore,

$$Q^TAQ = D = \begin{bmatrix} 18 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & -9 \end{bmatrix}$$

and, with the change of variable $\mathbf{x} = Q\mathbf{x}'$, we get $\mathbf{x}^TA\mathbf{x} = (\mathbf{x}')D\mathbf{x}' = 36$, so

$$18(x')^2 + 9(y')^2 - 9(z')^2 = 36 \quad \text{or} \quad \frac{(x')^2}{2} + \frac{(y')^2}{4} - \frac{(z')^2}{4} = 1$$

From Figure 5.19, we recognize this equation as the equation of a hyperboloid of one sheet. The x', y', and z' axes are in the directions of the eigenvectors $\mathbf{q}_1$, $\mathbf{q}_2$, and $\mathbf{q}_3$, respectively. The graph is shown in Figure 5.20.

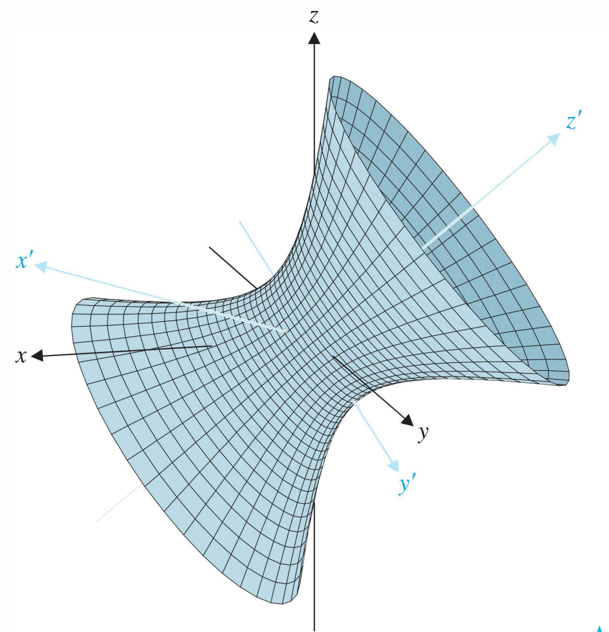


Figure 5.20
A hyperboloid of one sheet in nonstandard position

We can also identify and graph quadrics that have been translated out of standard position using the "complete-the-squares method" of Examples 5.27 and 5.29. You will be asked to do so in the exercises.

Exercises 5.5

Quadratic Forms

In Exercises 1–6, evaluate the quadratic form $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ for the given A and $\mathbf{x}$.

1. $A = \begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$

2. $A = \begin{bmatrix} 5 & 1 \\ 1 & -1 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

3. $A = \begin{bmatrix} 3 & -2 \\ -2 & 4 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$

4. $A = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 2 & 1 \\ -3 & 1 & 3 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$

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

6. $A = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

In Exercises 7–12, find the symmetric matrix A associated with the given quadratic form.

7. $x_1^2 + 2x_2^2 + 6x_1x_2$

8. x_1x_2

9. $3x^2 - 3xy - y^2$

10. $x_1^2 - x_3^2 + 8x_1x_2 - 6x_2x_3$

11. $5x_1^2 - x_2^2 + 2x_3^2 + 2x_1x_2 - 4x_1x_3 + 4x_2x_3$

12. $2x^2 - 3y^2 + z^2 - 4xz$

Diagonalize the quadratic forms in Exercises 13–18 by finding an orthogonal matrix Q such that the change of variable $\mathbf{x} = Q\mathbf{y}$ transforms the given form into one with no cross-product terms. Give Q and the new quadratic form.

13. $2x_1^2 + 5x_2^2 - 4x_1x_2$

14. $x^2 + 8xy + y^2$

15. $7x_1^2 + x_2^2 + x_3^2 + 8x_1x_2 + 8x_1x_3 - 16x_2x_3$

16. $x_1^2 + x_2^2 + 3x_3^2 - 4x_1x_2$

17. $x^2 + z^2 - 2xy + 2yz$

18. $2xy + 2xz + 2yz$

Classify each of the quadratic forms in Exercises 19–26 as positive definite, positive semidefinite, negative definite, negative semidefinite, or indefinite.

19. $x_1^2 + 2x_2^2$ **20.** $x_1^2 + x_2^2 - 2x_1x_2$

21. $-2x^2 - 2y^2 + 2xy$ **22.** $x^2 + y^2 + 4xy$

23. $2x_1^2 + 2x_2^2 + 2x_3^2 + 2x_1x_2 + 2x_1x_3 + 2x_2x_3$

24. $x_1^2 + x_2^2 + x_3^2 + 2x_1x_3$ **25.** $x_1^2 + x_2^2 - x_3^2 + 4x_1x_2$

26. $-x^2 - y^2 - z^2 - 2xy - 2xz - 2yz$

27. Prove Theorem 5.22.

28. Let $A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$ be a symmetric 2×2 matrix. Prove that A is positive definite if and only if $a > 0$ and $\det A > 0$. [*Hint:* $ax^2 + 2bxy + dy^2 = a\left(x + \frac{b}{a}y\right)^2 + \left(d - \frac{b^2}{a}\right)y^2$.]

29. Let B be an invertible matrix. Show that $A = B^TB$ is positive definite.

30. Let A be a positive definite symmetric matrix. Show that there exists an invertible matrix B such that $A = B^TB$. [*Hint:* Use the Spectral Theorem to write $A = QDQ^T$. Then show that D can be factored as C^TC for some invertible matrix C.]

31. Let A and B be positive definite symmetric $n \times n$ matrices and let c be a positive scalar. Show that the following matrices are positive definite.

(a) cA **(b)** A^2 **(c)** $A + B$

(d) A^{-1} (First show that A is necessarily invertible.)

32. Let A be a positive definite symmetric matrix. Show that there is a positive definite symmetric matrix B such that $A = B^2$. (Such a matrix B is called a ***square root*** of A.)

In Exercises 33–36, find the maximum and minimum values of the quadratic form $f(\mathbf{x})$ in the given exercise, subject to the constraint $\|\mathbf{x}\| = 1$, and determine the values of $\mathbf{x}$ for which these occur.

33. Exercise 20 **34.** Exercise 22

35. Exercise 23 **36.** Exercise 24

37. Finish the proof of Theorem 5.23(a).

38. Prove Theorem 5.23(c).

Graphing Quadratic Equations

In Exercises 39–44, identify the graph of the given equation.

39. $x^2 + 5y^2 = 25$ **40.** $x^2 - y^2 - 4 = 0$

41. $x^2 - y - 1 = 0$ **42.** $2x^2 + y^2 - 8 = 0$

43. $3x^2 = y^2 - 1$ **44.** $x = -2y^2$

In Exercises 45–50, use a translation of axes to put the conic in standard position. Identify the graph, give its equation in the translated coordinate system, and sketch the curve.

45. $x^2 + y^2 - 4x - 4y + 4 = 0$

46. $4x^2 + 2y^2 - 8x + 12y + 6 = 0$

47. $9x^2 - 4y^2 - 4y = 37$ **48.** $x^2 + 10x - 3y = -13$

49. $2y^2 + 4x + 8y = 0$

50. $2y^2 - 3x^2 - 18x - 20y + 11 = 0$

In Exercises 51–54, use a rotation of axes to put the conic in standard position. Identify the graph, give its equation in the rotated coordinate system, and sketch the curve.

51. $x^2 + xy + y^2 = 6$ **52.** $4x^2 + 10xy + 4y^2 = 9$

53. $4x^2 + 6xy - 4y^2 = 5$ **54.** $3x^2 - 2xy + 3y^2 = 8$

In Exercises 55–58, identify the conic with the given equation and give its equation in standard form.

55. $3x^2 - 4xy + 3y^2 - 28\sqrt{2}x + 22\sqrt{2}y + 84 = 0$

56. $6x^2 - 4xy + 9y^2 - 20x - 10y - 5 = 0$

57. $2xy + 2\sqrt{2}x - 1 = 0$

58. $x^2 - 2xy + y^2 + 4\sqrt{2}x - 4 = 0$

Sometimes the graph of a quadratic equation is a straight line, a pair of straight lines, or a single point. We refer to such a graph as a ***degenerate conic.*** *It is also possible that the equation is not satisfied for any values of the variables, in which case there is no graph at all and we refer to the conic as an* ***imaginary conic.*** *In Exercises 59–64, identify the conic with the given equation as either degenerate or imaginary and, where possible, sketch the graph.*

59. $x^2 - y^2 = 0$ **60.** $x^2 + 2y^2 + 2 = 0$

61. $3x^2 + y^2 = 0$ **62.** $x^2 + 2xy + y^2 = 0$

63. $x^2 - 2xy + y^2 + 2\sqrt{2}x - 2\sqrt{2}y = 0$

64. $2x^2 + 2xy + 2y^2 + 2\sqrt{2}x - 2\sqrt{2}y + 6 = 0$

65. Let A be a symmetric 2×2 matrix and let k be a scalar. Prove that the graph of the quadratic equation $\mathbf{x}^TA\mathbf{x} = k$ is

(a) a hyperbola if $k \neq 0$ and $\det A < 0$

(b) an ellipse, circle, or imaginary conic if $k \neq 0$ and $\det A > 0$

(c) a pair of straight lines or an imaginary conic if $k \neq 0$ and $\det A = 0$

(d) a pair of straight lines or a single point if $k = 0$ and $\det A \neq 0$

(e) a straight line if $k = 0$ and $\det A = 0$

[*Hint:* Use the Principal Axes Theorem.]

In Exercises 66–73, identify the quadric with the given equation and give its equation in standard form.

66. $4x^2 + 4y^2 + 4z^2 + 4xy + 4xz + 4yz = 8$

67. $x^2 + y^2 + z^2 - 4yz = 1$

68. $-x^2 - y^2 - z^2 + 4xy + 4xz + 4yz = 12$

69. $2xy + z = 0$

70. $16x^2 + 100y^2 + 9z^2 - 24xz - 60x - 80z = 0$

71. $x^2 + y^2 - 2z^2 + 4xy - 2xz + 2yz - x + y + z = 0$

72. $10x^2 + 25y^2 + 10z^2 - 40xz + 20\sqrt{2}x + 50y + 20\sqrt{2}z = 15$

73. $11x^2 + 11y^2 + 14z^2 + 2xy + 8xz - 8yz - 12x + 12y + 12z = 6$

74. Let A be a real 2×2 matrix with complex eigenvalues $\lambda = a \pm bi$ such that $b \neq 0$ and $|\lambda| = 1$. Prove that every trajectory of the dynamical system $\mathbf{x}_{k+1} = A\mathbf{x}_k$ lies on an ellipse. [*Hint:* Theorem 4.43 shows that if $\mathbf{v}$ is an eigenvector corresponding to $\lambda = a - bi$, then the matrix $P = [\text{Re}\,\mathbf{v} \quad \text{Im}\,\mathbf{v}]$ is invertible and $A = P\begin{bmatrix} a & -b \\ b & a \end{bmatrix}P^{-1}$. Set $B = (PP^T)^{-1}$. Show that the quadratic $\mathbf{x}^TB\mathbf{x} = k$ defines an ellipse for all $k > 0$, and prove that if $\mathbf{x}$ lies on this ellipse, so does $A\mathbf{x}$.]

Chapter Review

Key Definitions and Concepts

fundamental subspaces of a matrix, 380
Gram-Schmidt Process, 389
orthogonal basis, 370
orthogonal complement of a subspace, 378
orthogonal matrix, 374
orthogonal projection, 382
orthogonal set of vectors, 369
Orthogonal Decomposition Theorem, 384
orthogonally diagonalizable matrix, 400
orthonormal basis, 372
orthonormal set of vectors, 372
properties of orthogonal matrices, 374–376
QR factorization, 393
Rank Theorem, 386
spectral decomposition, 405
Spectral Theorem, 403

Review Questions

1. Mark each of the following statements true or false:

(a) Every orthonormal set of vectors is linearly independent.

(b) Every nonzero subspace of $\mathbb{R}^n$ has an orthogonal basis.

(c) If A is a square matrix with orthonormal rows, then A is an orthogonal matrix.

(d) Every orthogonal matrix is invertible.

(e) If A is a matrix with $\det A = 1$, then A is an orthogonal matrix.

(f) If A is an $m \times n$ matrix such that $(\text{row}(A))^\perp = \mathbb{R}^n$, then A must be the zero matrix.

(g) If W is a subspace of $\mathbb{R}^n$ and $\mathbf{v}$ is a vector in $\mathbb{R}^n$ such that $\text{proj}_W(\mathbf{v}) = \mathbf{0}$, then $\mathbf{v}$ must be the zero vector.

(h) If A is a symmetric, orthogonal matrix, then $A^2 = I$.

(i) Every orthogonally diagonalizable matrix is invertible.

(j) Given any n real numbers $\lambda_1, \ldots, \lambda_n$, there exists a symmetric $n \times n$ matrix with $\lambda_1, \ldots, \lambda_n$ as its eigenvalues.

2. Find all values of a and b such that

$$\left\{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \\ -2 \end{bmatrix}, \begin{bmatrix} a \\ b \\ 3 \end{bmatrix}\right\}$$ is an orthogonal set of vectors.

3. Find the coordinate vector $[\mathbf{v}]_\mathcal{B}$ of $\mathbf{v} = \begin{bmatrix} 7 \\ -3 \\ 2 \end{bmatrix}$ with respect to the orthogonal basis

$$\mathcal{B} = \left\{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}\right\} \text{ of } \mathbb{R}^3$$

4. The coordinate vector of a vector $\mathbf{v}$ with respect to an orthonormal basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$ of $\mathbb{R}^2$ is $[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} -3 \\ 1/2 \end{bmatrix}$. If $\mathbf{v}_1 = \begin{bmatrix} 3/5 \\ 4/5 \end{bmatrix}$, find all possible vectors $\mathbf{v}$.

5. Show that $\begin{bmatrix} 6/7 & 2/7 & 3/7 \\ -1/\sqrt{5} & 0 & 2/\sqrt{5} \\ 4/7\sqrt{5} & -15/7\sqrt{5} & 2/7\sqrt{5} \end{bmatrix}$ is an orthogonal matrix.

6. If $\begin{bmatrix} 1/2 & a \\ b & c \end{bmatrix}$ is an orthogonal matrix, find all possible values of a, b, and c.

7. If Q is an orthogonal $n \times n$ matrix and $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthonormal set in $\mathbb{R}^n$, prove that $\{Q\mathbf{v}_1, \ldots, Q\mathbf{v}_k\}$ is an orthonormal set.

8. If Q is an $n \times n$ matrix such that the angles $\angle(Q\mathbf{x}, Q\mathbf{y})$ and $\angle(\mathbf{x}, \mathbf{y})$ are equal for all vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, prove that Q is an orthogonal matrix.

In Questions 9–12, find a basis for $W^{\perp}$.

9. W is the line in $\mathbb{R}^2$ with general equation $2x - 5y = 0$

10. W is the line in $\mathbb{R}^3$ with parametric equations
$x = t$
$y = 2t$
$z = -t$

11. $W = \text{span}\left\{ \begin{bmatrix} 1 \\ -1 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -3 \end{bmatrix} \right\}$

12. $W = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \\ 2 \end{bmatrix} \right\}$

13. Find bases for each of the four fundamental subspaces of

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

14. Find the orthogonal decomposition of

$$\mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 2 \end{bmatrix}$$

with respect to

$$W = \text{span}\left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ -2 \\ 1 \end{bmatrix} \right\}$$

15. **(a)** Apply the Gram-Schmidt Process to

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

to find an orthogonal basis for $W = \text{span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$.

(b) Use the result of part (a) to find a QR factorization of $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$.

16. Find an orthogonal basis for $\mathbb{R}^4$ that contains the vectors $\begin{bmatrix} 1 \\ 0 \\ 2 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ 1 \\ -1 \end{bmatrix}$.

17. Find an orthogonal basis for the subspace

$$W = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} : x_1 + x_2 + x_3 + x_4 = 0 \right\} \text{ of } \mathbb{R}^4$$

18. Let $A = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix}$.

(a) Orthogonally diagonalize A.
(b) Give the spectral decomposition of A.

19. Find a symmetric matrix with eigenvalues $\lambda_1 = \lambda_2 = 1$, $\lambda_3 = -2$ and eigenspaces

$$E_1 = \text{span}\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right), E_{-2} = \text{span}\left(\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right)$$

20. If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthonormal basis for $\mathbb{R}^n$ and

$$A = c_1\mathbf{v}_1\mathbf{v}_1^T + c_2\mathbf{v}_2\mathbf{v}_2^T + \cdots + c_n\mathbf{v}_n\mathbf{v}_n^T$$

prove that A is a symmetric matrix with eigenvalues $c_1, c_2, \ldots, c_n$ and corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.