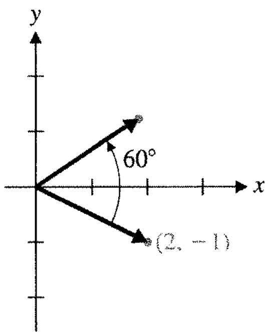

Figure 3.10
A 60° rotation

rotation, while a negative angle is clockwise.) Since $\cos 60° = 1/2$ and $\sin 60° = \sqrt{3}/2$, we compute

$$R_{60}\begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} \cos 60° & -\sin 60° \\ \sin 60° & \cos 60° \end{bmatrix}\begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \end{bmatrix}$$
$$= \begin{bmatrix} (2+\sqrt{3})/2 \\ (2\sqrt{3}-1)/2 \end{bmatrix}$$

Thus, the image of the point $(2, -1)$ under this rotation is the point $((2+\sqrt{3})/2, (2\sqrt{3}-1)/2) \approx (1.87, 1.23)$, as shown in Figure 3.10.

Example 3.59

(a) Show that the transformation $P : \mathbb{R}^2 \to \mathbb{R}^2$ that projects a point onto the x-axis is a linear transformation and find its standard matrix.
(b) More generally, if ℓ is a line through the origin in $\mathbb{R}^2$, show that the transformation $P_\ell : \mathbb{R}^2 \to \mathbb{R}^2$ that projects a point onto ℓ is a linear transformation and find its standard matrix.

Solution (a) As Figure 3.11 shows, P sends the point (x, y) to the point $(x, 0)$. Thus,

$$P\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ 0 \end{bmatrix} = x\begin{bmatrix} 1 \\ 0 \end{bmatrix} + y\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

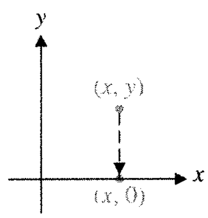

Figure 3.11
A projection

It follows that P is a matrix transformation (and hence a linear transformation) with standard matrix $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.

(b) Let the line ℓ have direction vector $\mathbf{d}$ and let $\mathbf{v}$ be an arbitrary vector. Then P_ℓ is given by $\text{proj}_{\mathbf{d}}(\mathbf{v})$, the projection of $\mathbf{v}$ onto $\mathbf{d}$, which you'll recall from Section 1.2 has the formula

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

Thus, to show that P_ℓ is linear, we proceed as follows:

$$P_\ell(\mathbf{u}+\mathbf{v}) = \left(\frac{\mathbf{d}\cdot(\mathbf{u}+\mathbf{v})}{\mathbf{d}\cdot\mathbf{d}}\right)\mathbf{d}$$
$$= \left(\frac{\mathbf{d}\cdot\mathbf{u}+\mathbf{d}\cdot\mathbf{v}}{\mathbf{d}\cdot\mathbf{d}}\right)\mathbf{d}$$
$$= \left(\frac{\mathbf{d}\cdot\mathbf{u}}{\mathbf{d}\cdot\mathbf{d}} + \frac{\mathbf{d}\cdot\mathbf{v}}{\mathbf{d}\cdot\mathbf{d}}\right)\mathbf{d}$$
$$= \left(\frac{\mathbf{d}\cdot\mathbf{u}}{\mathbf{d}\cdot\mathbf{d}}\right)\mathbf{d} + \left(\frac{\mathbf{d}\cdot\mathbf{v}}{\mathbf{d}\cdot\mathbf{d}}\right)\mathbf{d} = P_\ell(\mathbf{u}) + P_\ell(\mathbf{v})$$

Similarly, $P_\ell(c\mathbf{v}) = cP_\ell(\mathbf{v})$ for any scalar c (Exercise 52). Hence, P_ℓ is a linear transformation.

To find the standard matrix of P_ℓ, we apply Theorem 3.31. If we let $\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$, then

$$P_\ell(\mathbf{e}_1) = \left(\frac{\mathbf{d}\cdot\mathbf{e}_1}{\mathbf{d}\cdot\mathbf{d}}\right)\mathbf{d} = \frac{d_1}{d_1^2 + d_2^2}\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \frac{1}{d_1^2 + d_2^2}\begin{bmatrix} d_1^2 \\ d_1 d_2 \end{bmatrix}$$

and
$$P_\ell(\mathbf{e}_2) = \left(\frac{\mathbf{d}\cdot\mathbf{e}_2}{\mathbf{d}\cdot\mathbf{d}}\right)\mathbf{d} = \frac{d_2}{d_1^2 + d_2^2}\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \frac{1}{d_1^2 + d_2^2}\begin{bmatrix} d_1 d_2 \\ d_2^2 \end{bmatrix}$$

Thus, the standard matrix of the projection is

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

As a check, note that in part (a) we could take $\mathbf{d} = \mathbf{e}_1$ as a direction vector for the x-axis. Therefore, $d_1 = 1$ and $d_2 = 0$, and we obtain $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, as before.

New Linear Transformations from Old

If $T : \mathbb{R}^m \to \mathbb{R}^n$ and $S : \mathbb{R}^n \to \mathbb{R}^p$ are linear transformations, then we may follow T by S to form the ***composition*** of the two transformations, denoted $S \circ T$. Notice that, in order for $S \circ T$ to make sense, the codomain of T and the domain of S must match (in this case, they are both $\mathbb{R}^n$) and the resulting composite transformation $S \circ T$ goes from the domain of T to the codomain of S (in this case, $S \circ T : \mathbb{R}^m \to \mathbb{R}^p$). Figure 3.12 shows schematically how this composition works. The formal definition of composition of transformations is taken directly from this figure and is the same as the corresponding definition of composition of ordinary functions:

$$(S \circ T)(\mathbf{v}) = S(T(\mathbf{v}))$$

Of course, we would like $S \circ T$ to be a linear transformation too, and happily we find that it is. We can demonstrate this by showing that $S \circ T$ satisfies the definition of a linear transformation (which we will do in Chapter 6), but, since for the time being we are assuming that linear transformations and matrix transformations are the same thing, it is enough to show that $S \circ T$ is a matrix transformation. We will use the notation $[T]$ for the standard matrix of a linear transformation T.

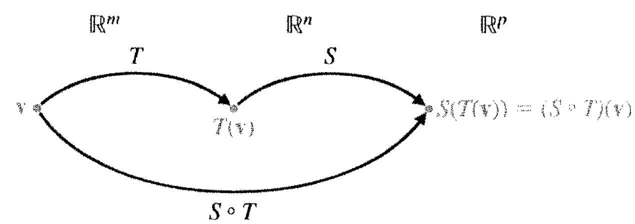

Figure 3.12
The composition of transformations

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.

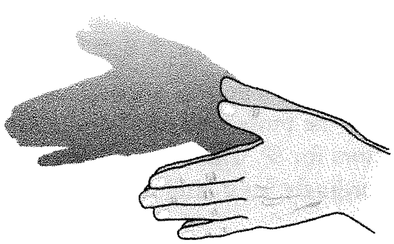

Figure 5.1
Shadows on a wall are projections

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

$$P = \frac{1}{d_1^2 + d_2^2}\begin{bmatrix} d_1^2 & d_1d_2 \\ d_1d_2 & d_2^2 \end{bmatrix} = \begin{bmatrix} d_1^2/(d_1^2 + d_2^2) & d_1d_2/(d_1^2 + d_2^2) \\ d_1d_2/(d_1^2 + d_2^2) & d_2^2/(d_1^2 + d_2^2) \end{bmatrix}$$

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

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.

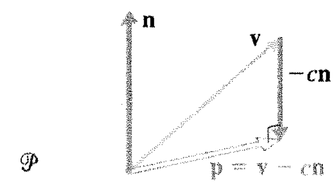

Figure 5.2
Projection onto a plane

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:

$$\text{(a) } x + y + z = 0 \quad \text{(b) } x - 2z = 0 \quad \text{(c) } 2x - 3y + z = 0$$

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

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

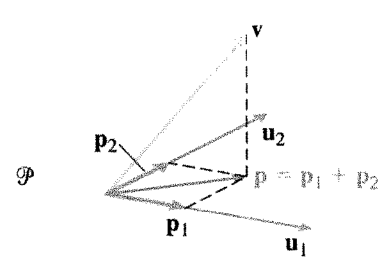

Figure 5.3

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

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

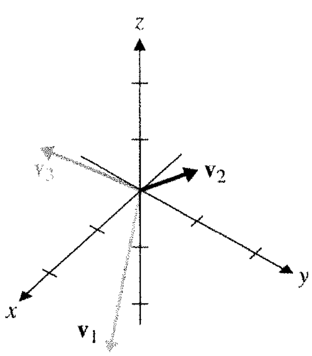

Figure 5.4
An orthogonal set of vectors

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.

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

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

or, equivalently,

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

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

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

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 $S = \{\mathbf{q}_1, \dots, \mathbf{q}_k\}$ 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

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

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

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

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.

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.

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

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

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

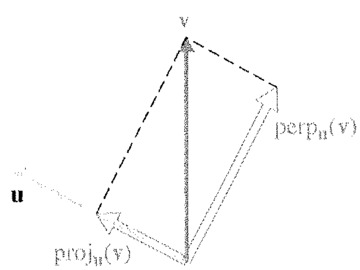

Figure 5.7
$\mathbf{v} = \text{proj}_{\mathbf{u}}(\mathbf{v}) + \text{perp}_{\mathbf{u}}(\mathbf{v})$

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.

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* $\mathbf{v}$ *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* $\mathbf{v}$ *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

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

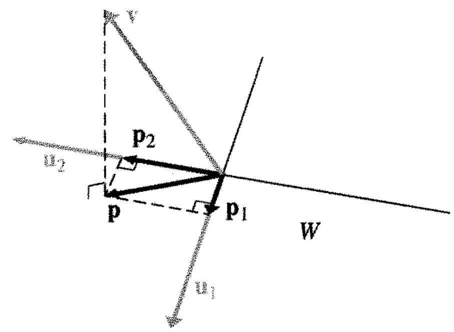

Figure 5.8
$\mathbf{p} = \mathbf{p}_1 + \mathbf{p}_2$

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

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

Therefore,

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

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.)

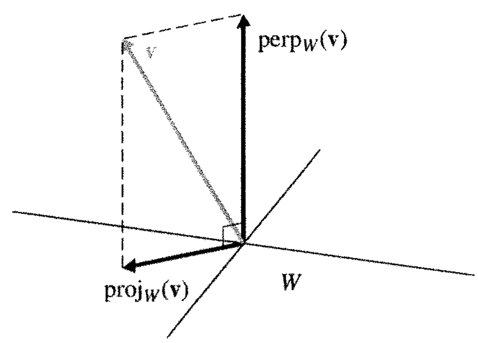

Figure 5.9
$\mathbf{v} = \text{proj}_W(\mathbf{v}) + \text{perp}_W(\mathbf{v})$

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

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

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

$$A = \lambda_1\mathbf{q}_1\mathbf{q}_1^T + \lambda_2\mathbf{q}_2\mathbf{q}_2^T + \lambda_3\mathbf{q}_3\mathbf{q}_3^T$$

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

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:

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