

Low

**PROOF** If  $\mathbf{0} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$  for some scalars  $c_1, \dots, c_p$ , then

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

because  $\mathbf{u}_1$  is orthogonal to  $\mathbf{u}_2, \dots, \mathbf{u}_p$ . Since  $\mathbf{u}_1$  is nonzero,  $\mathbf{u}_1 \cdot \mathbf{u}_1$  is not zero and so  $c_1 = 0$ . Similarly,  $c_2, \dots, c_p$  must be zero. Thus  $S$  is linearly independent. ■

**DEFINITION**

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

The next theorem suggests why an orthogonal basis is much nicer than other bases. The weights in a linear combination can be computed easily.

**THEOREM 5**

Let  $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$  be an orthogonal basis for a subspace  $W$  of  $\mathbb{R}^n$ . For each  $\mathbf{y}$  in  $W$ , the weights in the linear combination

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

are given by

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

**PROOF** As in the preceding proof, the orthogonality of  $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$  shows that

$$\mathbf{y} \cdot \mathbf{u}_1 = (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p) \cdot \mathbf{u}_1 = c_1(\mathbf{u}_1 \cdot \mathbf{u}_1)$$

Since  $\mathbf{u}_1 \cdot \mathbf{u}_1$  is not zero, the equation above can be solved for  $c_1$ . To find  $c_j$  for  $j = 2, \dots, p$ , compute  $\mathbf{y} \cdot \mathbf{u}_j$  and solve for  $c_j$ . ■

**EXAMPLE 2** The set  $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$  in Example 1 is an orthogonal basis for  $\mathbb{R}^3$ .

Express the vector  $\mathbf{y} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}$  as a linear combination of the vectors in  $S$ .

**SOLUTION** Compute

$$\begin{aligned} \mathbf{y} \cdot \mathbf{u}_1 &= 11, & \mathbf{y} \cdot \mathbf{u}_2 &= -12, & \mathbf{y} \cdot \mathbf{u}_3 &= -33 \\ \mathbf{u}_1 \cdot \mathbf{u}_1 &= 11, & \mathbf{u}_2 \cdot \mathbf{u}_2 &= 6, & \mathbf{u}_3 \cdot \mathbf{u}_3 &= 33/2 \end{aligned}$$

By Theorem 5,

$$\begin{aligned} \mathbf{y} &= \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 + \frac{\mathbf{y} \cdot \mathbf{u}_3}{\mathbf{u}_3 \cdot \mathbf{u}_3} \mathbf{u}_3 \\ &= \frac{11}{11} \mathbf{u}_1 + \frac{-12}{6} \mathbf{u}_2 + \frac{-33}{33/2} \mathbf{u}_3 \\ &= \mathbf{u}_1 - 2\mathbf{u}_2 - 2\mathbf{u}_3 \end{aligned}$$

Notice how easy it is to compute the weights needed to build  $\mathbf{y}$  from an orthogonal basis. If the basis were not orthogonal, it would be necessary to solve a system of linear equations in order to find the weights, as in Chapter 1.

We turn next to a construction that will become a key step in many calculations involving orthogonality, and it will lead to a geometric interpretation of Theorem 5.

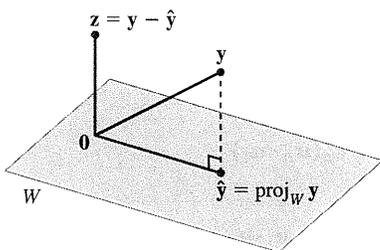


FIGURE 2

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

## An Orthogonal Projection

Given a nonzero vector  $\mathbf{u}$  in  $\mathbb{R}^n$ , consider the problem of decomposing a vector  $\mathbf{y}$  in  $\mathbb{R}^n$  into the sum of two vectors, one a multiple of  $\mathbf{u}$  and the other orthogonal to  $\mathbf{u}$ . We wish to write

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

where  $\hat{\mathbf{y}} = \alpha\mathbf{u}$  for some scalar  $\alpha$  and  $\mathbf{z}$  is some vector orthogonal to  $\mathbf{u}$ . See Figure 2. Given any scalar  $\alpha$ , let  $\mathbf{z} = \mathbf{y} - \alpha\mathbf{u}$ , so that (1) is satisfied. Then  $\mathbf{y} - \hat{\mathbf{y}}$  is orthogonal to  $\mathbf{u}$  if and only if

$$0 = (\mathbf{y} - \alpha\mathbf{u}) \cdot \mathbf{u} = \mathbf{y} \cdot \mathbf{u} - (\alpha\mathbf{u}) \cdot \mathbf{u} = \mathbf{y} \cdot \mathbf{u} - \alpha(\mathbf{u} \cdot \mathbf{u})$$

That is, (1) is satisfied with  $\mathbf{z}$  orthogonal to  $\mathbf{u}$  if and only if  $\alpha = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}$  and  $\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}$ . The vector  $\hat{\mathbf{y}}$  is called the **orthogonal projection of  $\mathbf{y}$  onto  $\mathbf{u}$** , and the vector  $\mathbf{z}$  is called the **component of  $\mathbf{y}$  orthogonal to  $\mathbf{u}$** .

If  $c$  is any nonzero scalar and if  $\mathbf{u}$  is replaced by  $c\mathbf{u}$  in the definition of  $\hat{\mathbf{y}}$ , then the orthogonal projection of  $\mathbf{y}$  onto  $c\mathbf{u}$  is exactly the same as the orthogonal projection of  $\mathbf{y}$  onto  $\mathbf{u}$  (Exercise 39). Hence this projection is determined by the *subspace*  $L$  spanned by  $\mathbf{u}$  (the line through  $\mathbf{u}$  and  $\mathbf{0}$ ). Sometimes  $\hat{\mathbf{y}}$  is denoted by  $\text{proj}_L \mathbf{y}$  and is called the **orthogonal projection of  $\mathbf{y}$  onto  $L$** . That is,

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

**EXAMPLE 3** Let  $\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$  and  $\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ . Find the orthogonal projection of  $\mathbf{y}$  onto  $\mathbf{u}$ . Then write  $\mathbf{y}$  as the sum of two orthogonal vectors, one in  $\text{Span}\{\mathbf{u}\}$  and one orthogonal to  $\mathbf{u}$ .

**SOLUTION** Compute

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

The orthogonal projection of  $\mathbf{y}$  onto  $\mathbf{u}$  is

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u} = \frac{40}{20} \mathbf{u} = 2 \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

and the component of  $\mathbf{y}$  orthogonal to  $\mathbf{u}$  is

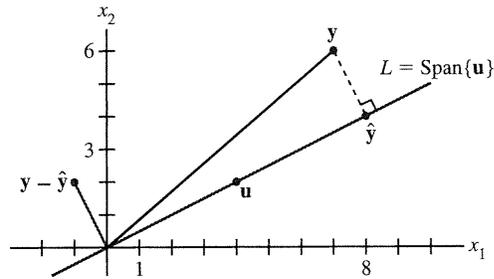
$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

The sum of these two vectors is  $\mathbf{y}$ . That is,

$$\begin{array}{ccc} \begin{bmatrix} 7 \\ 6 \end{bmatrix} &= \begin{bmatrix} 8 \\ 4 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix} \\ \uparrow & \quad \uparrow & \quad \uparrow \\ \mathbf{y} & \quad \hat{\mathbf{y}} & \quad (\mathbf{y} - \hat{\mathbf{y}}) \end{array}$$

This decomposition of  $\mathbf{y}$  is illustrated in Figure 3. *Note:* If the calculations above are correct, then  $\{\hat{\mathbf{y}}, \mathbf{y} - \hat{\mathbf{y}}\}$  will be an orthogonal set. As a check, compute

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



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

Since the line segment in Figure 3 between  $\mathbf{y}$  and  $\hat{\mathbf{y}}$  is perpendicular to  $L$ , by construction of  $\hat{\mathbf{y}}$ , the point identified with  $\hat{\mathbf{y}}$  is the closest point of  $L$  to  $\mathbf{y}$ . (This can be proved from geometry. We will assume this for  $\mathbb{R}^2$  now and prove it for  $\mathbb{R}^n$  in Section 6.3.)

**EXAMPLE 4** Find the distance in Figure 3 from  $\mathbf{y}$  to  $L$ .

**SOLUTION** The distance from  $\mathbf{y}$  to  $L$  is the length of the perpendicular line segment from  $\mathbf{y}$  to the orthogonal projection  $\hat{\mathbf{y}}$ . This length equals the length of  $\mathbf{y} - \hat{\mathbf{y}}$ . Thus the distance is

$$\|\mathbf{y} - \hat{\mathbf{y}}\| = \sqrt{(-1)^2 + 2^2} = \sqrt{5} \quad \blacksquare$$

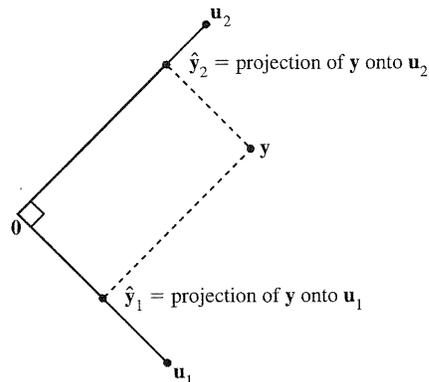
## A Geometric Interpretation of Theorem 5

The formula for the orthogonal projection  $\hat{\mathbf{y}}$  in (2) has the same appearance as each of the terms in Theorem 5. Thus Theorem 5 decomposes a vector  $\mathbf{y}$  into a sum of orthogonal projections onto one-dimensional subspaces.

It is easy to visualize the case in which  $W = \mathbb{R}^2 = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ , with  $\mathbf{u}_1$  and  $\mathbf{u}_2$  orthogonal. Any  $\mathbf{y}$  in  $\mathbb{R}^2$  can be written in the form

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

The first term in (3) is the projection of  $\mathbf{y}$  onto the subspace spanned by  $\mathbf{u}_1$  (the line through  $\mathbf{u}_1$  and the origin), and the second term is the projection of  $\mathbf{y}$  onto the subspace spanned by  $\mathbf{u}_2$ . Thus (3) expresses  $\mathbf{y}$  as the sum of its projections onto the (orthogonal) axes determined by  $\mathbf{u}_1$  and  $\mathbf{u}_2$ . See Figure 4.



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

**STUDY GUIDE** offers additional resources for mastering the concepts around an orthogonal basis.

4. Since  $U$  is an  $n \times n$  matrix with orthonormal columns, by Theorem 6,  $U^T U = I$ . Taking the determinant of the left side of this equation, and applying Theorems 5 and 6 from Section 3.2 results in  $\det U^T U = (\det U^T)(\det U) = (\det U)(\det U) = (\det U)^2$ . Recall  $\det I = 1$ . Putting the two sides of the equation back together results in  $(\det U)^2 = 1$  and hence  $\det U = \pm 1$ .

## 6.3 Orthogonal Projections

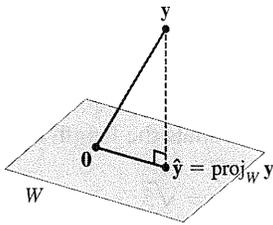


FIGURE 1

The orthogonal projection of a point in  $\mathbb{R}^2$  onto a line through the origin has an important analogue in  $\mathbb{R}^n$ . Given a vector  $\mathbf{y}$  and a subspace  $W$  in  $\mathbb{R}^n$ , there is a vector  $\hat{\mathbf{y}}$  in  $W$  such that (1)  $\hat{\mathbf{y}}$  is the unique vector in  $W$  for which  $\mathbf{y} - \hat{\mathbf{y}}$  is orthogonal to  $W$ , and (2)  $\hat{\mathbf{y}}$  is the unique vector in  $W$  closest to  $\mathbf{y}$ . See Figure 1. These two properties of  $\hat{\mathbf{y}}$  provide the key to finding least-squares solutions of linear systems.

To prepare for the first theorem, observe that whenever a vector  $\mathbf{y}$  is written as a linear combination of vectors  $\mathbf{u}_1, \dots, \mathbf{u}_n$  in  $\mathbb{R}^n$ , the terms in the sum for  $\mathbf{y}$  can be grouped into two parts so that  $\mathbf{y}$  can be written as

$$\mathbf{y} = \mathbf{z}_1 + \mathbf{z}_2$$

where  $\mathbf{z}_1$  is a linear combination of some of the  $\mathbf{u}_i$  and  $\mathbf{z}_2$  is a linear combination of the rest of the  $\mathbf{u}_i$ . This idea is particularly useful when  $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$  is an orthogonal basis. Recall from Section 6.1 that  $W^\perp$  denotes the set of all vectors orthogonal to a subspace  $W$ .

**EXAMPLE 1** Let  $\{\mathbf{u}_1, \dots, \mathbf{u}_5\}$  be an orthogonal basis for  $\mathbb{R}^5$  and let

$$\mathbf{y} = c_1 \mathbf{u}_1 + \dots + c_5 \mathbf{u}_5$$

Consider the subspace  $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ , and write  $\mathbf{y}$  as the sum of a vector  $\mathbf{z}_1$  in  $W$  and a vector  $\mathbf{z}_2$  in  $W^\perp$ .

**SOLUTION** Write

$$\mathbf{y} = \underbrace{c_1 \mathbf{u}_1 + c_2 \mathbf{u}_2}_{\mathbf{z}_1} + \underbrace{c_3 \mathbf{u}_3 + c_4 \mathbf{u}_4 + c_5 \mathbf{u}_5}_{\mathbf{z}_2}$$

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

and  $\mathbf{z}_2 = c_3 \mathbf{u}_3 + c_4 \mathbf{u}_4 + c_5 \mathbf{u}_5$  is in  $\text{Span}\{\mathbf{u}_3, \mathbf{u}_4, \mathbf{u}_5\}$ .

To show that  $\mathbf{z}_2$  is in  $W^\perp$ , it suffices to show that  $\mathbf{z}_2$  is orthogonal to the vectors in the basis  $\{\mathbf{u}_1, \mathbf{u}_2\}$  for  $W$ . (See Section 6.1.) Using properties of the inner product, compute

$$\begin{aligned} \mathbf{z}_2 \cdot \mathbf{u}_1 &= (c_3 \mathbf{u}_3 + c_4 \mathbf{u}_4 + c_5 \mathbf{u}_5) \cdot \mathbf{u}_1 \\ &= c_3 \mathbf{u}_3 \cdot \mathbf{u}_1 + c_4 \mathbf{u}_4 \cdot \mathbf{u}_1 + c_5 \mathbf{u}_5 \cdot \mathbf{u}_1 \\ &= 0 \end{aligned}$$

because  $\mathbf{u}_1$  is orthogonal to  $\mathbf{u}_3, \mathbf{u}_4$ , and  $\mathbf{u}_5$ . A similar calculation shows that  $\mathbf{z}_2 \cdot \mathbf{u}_2 = 0$ . Thus  $\mathbf{z}_2$  is in  $W^\perp$ . ■

The next theorem shows that the decomposition  $\mathbf{y} = \mathbf{z}_1 + \mathbf{z}_2$  in Example 1 can be computed without having an orthogonal basis for  $\mathbb{R}^n$ . It is enough to have an orthogonal basis only for  $W$ .

## THEOREM 8

**The Orthogonal Decomposition Theorem**

Let  $W$  be a subspace of  $\mathbb{R}^n$ . Then each  $\mathbf{y}$  in  $\mathbb{R}^n$  can be written uniquely in the form

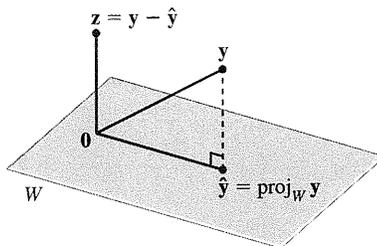
$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \quad (1)$$

where  $\hat{\mathbf{y}}$  is in  $W$  and  $\mathbf{z}$  is in  $W^\perp$ . In fact, if  $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$  is any orthogonal basis of  $W$ , then

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

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

The vector  $\hat{\mathbf{y}}$  in (2) is called the **orthogonal projection of  $\mathbf{y}$  onto  $W$**  and often is written as  $\text{proj}_W \mathbf{y}$ . See Figure 2. When  $W$  is a one-dimensional subspace, the formula for  $\hat{\mathbf{y}}$  matches the formula given in Section 6.2.



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

**PROOF** Let  $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$  be any orthogonal basis for  $W$ , and define  $\hat{\mathbf{y}}$  by (2).<sup>1</sup> Then  $\hat{\mathbf{y}}$  is in  $W$  because  $\hat{\mathbf{y}}$  is a linear combination of the basis  $\mathbf{u}_1, \dots, \mathbf{u}_p$ . Let  $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$ . Since  $\mathbf{u}_1$  is orthogonal to  $\mathbf{u}_2, \dots, \mathbf{u}_p$ , it follows from (2) that

$$\begin{aligned} \mathbf{z} \cdot \mathbf{u}_1 &= (\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{u}_1 = \mathbf{y} \cdot \mathbf{u}_1 - \left( \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \right) \mathbf{u}_1 \cdot \mathbf{u}_1 - 0 - \cdots - 0 \\ &= \mathbf{y} \cdot \mathbf{u}_1 - \mathbf{y} \cdot \mathbf{u}_1 = 0 \end{aligned}$$

Thus  $\mathbf{z}$  is orthogonal to  $\mathbf{u}_1$ . Similarly,  $\mathbf{z}$  is orthogonal to each  $\mathbf{u}_j$  in the basis for  $W$ . Hence  $\mathbf{z}$  is orthogonal to every vector in  $W$ . That is,  $\mathbf{z}$  is in  $W^\perp$ .

To show that the decomposition in (1) is unique, suppose  $\mathbf{y}$  can also be written as  $\mathbf{y} = \hat{\mathbf{y}}_1 + \mathbf{z}_1$ , with  $\hat{\mathbf{y}}_1$  in  $W$  and  $\mathbf{z}_1$  in  $W^\perp$ . Then  $\hat{\mathbf{y}} + \mathbf{z} = \hat{\mathbf{y}}_1 + \mathbf{z}_1$  (since both sides equal  $\mathbf{y}$ ), and so

$$\hat{\mathbf{y}} - \hat{\mathbf{y}}_1 = \mathbf{z}_1 - \mathbf{z}$$

This equality shows that the vector  $\mathbf{v} = \hat{\mathbf{y}} - \hat{\mathbf{y}}_1$  is in  $W$  and in  $W^\perp$  (because  $\mathbf{z}_1$  and  $\mathbf{z}$  are both in  $W^\perp$ , and  $W^\perp$  is a subspace). Hence  $\mathbf{v} \cdot \mathbf{v} = 0$ , which shows that  $\mathbf{v} = \mathbf{0}$ . This proves that  $\hat{\mathbf{y}} = \hat{\mathbf{y}}_1$  and also  $\mathbf{z}_1 = \mathbf{z}$ . ■

The uniqueness of the decomposition (1) shows that the orthogonal projection  $\hat{\mathbf{y}}$  depends only on  $W$  and not on the particular basis used in (2).

<sup>1</sup> We may assume that  $W$  is not the zero subspace, for otherwise  $W^\perp = \mathbb{R}^n$  and (1) is simply  $\mathbf{y} = \mathbf{0} + \mathbf{y}$ . The next section will show that any nonzero subspace of  $\mathbb{R}^n$  has an orthogonal basis.

**EXAMPLE 2** Let  $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$ ,  $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$ , and  $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ . Observe that  $\{\mathbf{u}_1, \mathbf{u}_2\}$  is an orthogonal basis for  $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ . Write  $\mathbf{y}$  as the sum of a vector in  $W$  and a vector orthogonal to  $W$ .

**SOLUTION** The orthogonal projection of  $\mathbf{y}$  onto  $W$  is

$$\begin{aligned} \hat{\mathbf{y}} &= \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 \\ &= \frac{9}{30} \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \frac{9}{30} \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{15}{30} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} \end{aligned}$$

Also

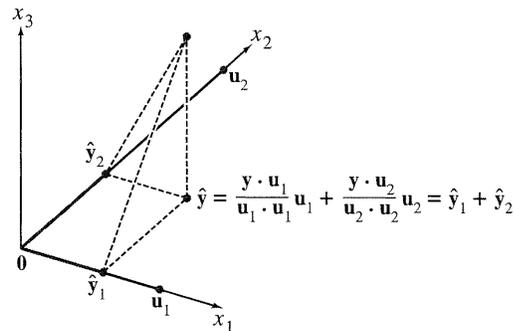
$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} = \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$

Theorem 8 ensures that  $\mathbf{y} - \hat{\mathbf{y}}$  is in  $W^\perp$ . To check the calculations, however, it is a good idea to verify that  $\mathbf{y} - \hat{\mathbf{y}}$  is orthogonal to both  $\mathbf{u}_1$  and  $\mathbf{u}_2$  and hence to all of  $W$ . The desired decomposition of  $\mathbf{y}$  is

$$\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} + \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$

## A Geometric Interpretation of the Orthogonal Projection

When  $W$  is a one-dimensional subspace, the formula (2) for  $\text{proj}_W \mathbf{y}$  contains just one term. Thus, when  $\dim W > 1$ , each term in (2) is itself an orthogonal projection of  $\mathbf{y}$  onto a one-dimensional subspace spanned by one of the  $\mathbf{u}$ 's in the basis for  $W$ . Figure 3 illustrates this when  $W$  is a subspace of  $\mathbb{R}^3$  spanned by  $\mathbf{u}_1$  and  $\mathbf{u}_2$ . Here  $\hat{\mathbf{y}}_1$  and  $\hat{\mathbf{y}}_2$  denote the projections of  $\mathbf{y}$  onto the lines spanned by  $\mathbf{u}_1$  and  $\mathbf{u}_2$ , respectively. The orthogonal projection  $\hat{\mathbf{y}}$  of  $\mathbf{y}$  onto  $W$  is the sum of the projections of  $\mathbf{y}$  onto one-dimensional subspaces that are orthogonal to each other. The vector  $\hat{\mathbf{y}}$  in Figure 3 corresponds to the vector  $\hat{\mathbf{y}}$  in Figure 4 of Section 6.2, because now it is  $\hat{\mathbf{y}}$  that is in  $W$ .



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

## Properties of Orthogonal Projections

If  $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$  is an orthogonal basis for  $W$  and if  $\mathbf{y}$  happens to be in  $W$ , then the formula for  $\text{proj}_W \mathbf{y}$  is exactly the same as the representation of  $\mathbf{y}$  given in Theorem 5 in Section 6.2. In this case,  $\text{proj}_W \mathbf{y} = \mathbf{y}$ .

If  $\mathbf{y}$  is in  $W = \text{Span}\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ , then  $\text{proj}_W \mathbf{y} = \mathbf{y}$ .

This fact also follows from the next theorem.

### THEOREM 9

#### The Best Approximation Theorem

Let  $W$  be a subspace of  $\mathbb{R}^n$ , let  $\mathbf{y}$  be any vector in  $\mathbb{R}^n$ , and let  $\hat{\mathbf{y}}$  be the orthogonal projection of  $\mathbf{y}$  onto  $W$ . Then  $\hat{\mathbf{y}}$  is the closest point in  $W$  to  $\mathbf{y}$ , in the sense that

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

for all  $\mathbf{v}$  in  $W$  distinct from  $\hat{\mathbf{y}}$ .

The vector  $\hat{\mathbf{y}}$  in Theorem 9 is called **the best approximation to  $\mathbf{y}$  by elements of  $W$** . Later sections in the text will examine problems where a given  $\mathbf{y}$  must be replaced, or *approximated*, by a vector  $\mathbf{v}$  in some fixed subspace  $W$ . The distance from  $\mathbf{y}$  to  $\mathbf{v}$ , given by  $\|\mathbf{y} - \mathbf{v}\|$ , can be regarded as the “error” of using  $\mathbf{v}$  in place of  $\mathbf{y}$ . Theorem 9 says that this error is minimized when  $\mathbf{v} = \hat{\mathbf{y}}$ .

Inequality (3) leads to a new proof that  $\hat{\mathbf{y}}$  does not depend on the particular orthogonal basis used to compute it. If a different orthogonal basis for  $W$  was used to construct an orthogonal projection of  $\mathbf{y}$ , then this projection would also be the closest point in  $W$  to  $\mathbf{y}$ , namely  $\hat{\mathbf{y}}$ .

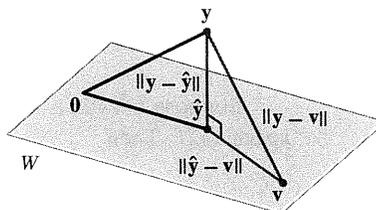
**PROOF** Take  $\mathbf{v}$  in  $W$  distinct from  $\hat{\mathbf{y}}$ . See Figure 4. Then  $\hat{\mathbf{y}} - \mathbf{v}$  is in  $W$ . By the Orthogonal Decomposition Theorem,  $\mathbf{y} - \hat{\mathbf{y}}$  is orthogonal to  $W$ . In particular,  $\mathbf{y} - \hat{\mathbf{y}}$  is orthogonal to  $\hat{\mathbf{y}} - \mathbf{v}$  (which is in  $W$ ). Since

$$\mathbf{y} - \mathbf{v} = (\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \mathbf{v})$$

the Pythagorean Theorem gives

$$\|\mathbf{y} - \mathbf{v}\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}} - \mathbf{v}\|^2$$

(See the right triangle outlined in teal in Figure 4. The length of each side is labeled.) Now  $\|\hat{\mathbf{y}} - \mathbf{v}\|^2 > 0$  because  $\hat{\mathbf{y}} - \mathbf{v} \neq \mathbf{0}$ , and so inequality (3) follows immediately. ■



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

**EXAMPLE 3** If  $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$ ,  $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$ ,  $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ , and  $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ , as in Example 2, then the closest point in  $W$  to  $\mathbf{y}$  is

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} \quad \blacksquare$$

**EXAMPLE 4** The distance from a point  $\mathbf{y}$  in  $\mathbb{R}^n$  to a subspace  $W$  is defined as the distance from  $\mathbf{y}$  to the nearest point in  $W$ . Find the distance from  $\mathbf{y}$  to  $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ , where

$$\mathbf{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \quad \mathbf{u}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

**SOLUTION** By the Best Approximation Theorem, the distance from  $\mathbf{y}$  to  $W$  is  $\|\mathbf{y} - \hat{\mathbf{y}}\|$ , where  $\hat{\mathbf{y}} = \text{proj}_W \mathbf{y}$ . Since  $\{\mathbf{u}_1, \mathbf{u}_2\}$  is an orthogonal basis for  $W$ ,

$$\hat{\mathbf{y}} = \frac{15}{30} \mathbf{u}_1 + \frac{-21}{6} \mathbf{u}_2 = \frac{1}{2} \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} - \frac{7}{2} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}$$

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix} - \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 6 \end{bmatrix}$$

$$\|\mathbf{y} - \hat{\mathbf{y}}\|^2 = 3^2 + 6^2 = 45$$

The distance from  $\mathbf{y}$  to  $W$  is  $\sqrt{45} = 3\sqrt{5}$ . \blacksquare

The final theorem in this section shows how formula (2) for  $\text{proj}_W \mathbf{y}$  is simplified when the basis for  $W$  is an orthonormal set.

### THEOREM 10

If  $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$  is an orthonormal basis for a subspace  $W$  of  $\mathbb{R}^n$ , then

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

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

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

**PROOF** Formula (4) follows immediately from (2) in Theorem 8. Also, (4) shows that  $\text{proj}_W \mathbf{y}$  is a linear combination of the columns of  $U$  using the weights  $\mathbf{y} \cdot \mathbf{u}_1, \mathbf{y} \cdot \mathbf{u}_2, \dots, \mathbf{y} \cdot \mathbf{u}_p$ . The weights can be written as  $\mathbf{u}_1^T \mathbf{y}, \mathbf{u}_2^T \mathbf{y}, \dots, \mathbf{u}_p^T \mathbf{y}$ , showing that they are the entries in  $U^T \mathbf{y}$  and justifying (5). \blacksquare

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

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

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

If  $U$  is an  $n \times n$  (square) matrix with orthonormal columns, then  $U$  is an *orthogonal* matrix, the column space  $W$  is all of  $\mathbb{R}^n$ , and  $UU^T \mathbf{y} = I \mathbf{y} = \mathbf{y}$  for all  $\mathbf{y}$  in  $\mathbb{R}^n$ .

## Solution to Practice Problems

1. Let  $\mathbf{v}_1 = \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$  and  $\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 = \mathbf{x}_2 - 0\mathbf{v}_1 = \mathbf{x}_2$ . So  $\{\mathbf{x}_1, \mathbf{x}_2\}$  is already orthogonal. All that is needed is to normalize the vectors. Let

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|} \mathbf{v}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

Instead of normalizing  $\mathbf{v}_2$  directly, normalize  $\mathbf{v}'_2 = 3\mathbf{v}_2$  instead:

$$\mathbf{u}_2 = \frac{1}{\|\mathbf{v}'_2\|} \mathbf{v}'_2 = \frac{1}{\sqrt{1^2 + 1^2 + (-2)^2}} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{bmatrix}$$

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

2. Since the columns of  $A$  are linearly dependent, there is a nontrivial vector  $\mathbf{x}$  such that  $A\mathbf{x} = \mathbf{0}$ . But then  $QR\mathbf{x} = \mathbf{0}$ . Applying Theorem 7 from Section 6.2 results in  $\|R\mathbf{x}\| = \|QR\mathbf{x}\| = \|\mathbf{0}\| = 0$ . But  $\|R\mathbf{x}\| = 0$  implies  $R\mathbf{x} = \mathbf{0}$ , by Theorem 1 from Section 6.1. Thus there is a nontrivial vector  $\mathbf{x}$  such that  $R\mathbf{x} = \mathbf{0}$  and hence, by the Invertible Matrix Theorem,  $R$  cannot be invertible.

## 6.5 Least-Squares Problems

Inconsistent systems arise often in applications. When a solution is demanded and none exists, the best one can do is to find an  $\mathbf{x}$  that makes  $A\mathbf{x}$  as close as possible to  $\mathbf{b}$ .

Think of  $A\mathbf{x}$  as an *approximation* to  $\mathbf{b}$ . The smaller the distance between  $\mathbf{b}$  and  $A\mathbf{x}$ , given by  $\|\mathbf{b} - A\mathbf{x}\|$ , the better the approximation. The **general least-squares problem** is to find an  $\mathbf{x}$  that makes  $\|\mathbf{b} - A\mathbf{x}\|$  as small as possible. The adjective “least-squares” arises from the fact that  $\|\mathbf{b} - A\mathbf{x}\|$  is the square root of a sum of squares.

### DEFINITION

If  $A$  is  $m \times n$  and  $\mathbf{b}$  is in  $\mathbb{R}^m$ , a **least-squares solution** of  $A\mathbf{x} = \mathbf{b}$  is an  $\hat{\mathbf{x}}$  in  $\mathbb{R}^n$  such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

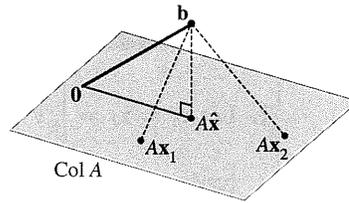
for all  $\mathbf{x}$  in  $\mathbb{R}^n$ .

The most important aspect of the least-squares problem is that no matter what  $\mathbf{x}$  we select, the vector  $A\mathbf{x}$  will necessarily be in the column space,  $\text{Col } A$ . So we seek an  $\mathbf{x}$  that makes  $A\mathbf{x}$  the closest point in  $\text{Col } A$  to  $\mathbf{b}$ . See Figure 1. (Of course, if  $\mathbf{b}$  happens to be in  $\text{Col } A$ , then  $\mathbf{b}$  is  $A\mathbf{x}$  for some  $\mathbf{x}$ , and such an  $\mathbf{x}$  is a “least-squares solution.”)

### Solution of the General Least-Squares Problem

Given  $A$  and  $\mathbf{b}$  as above, apply the Best Approximation Theorem in Section 6.3 to the subspace  $\text{Col } A$ . Let

$$\hat{\mathbf{b}} = \text{proj}_{\text{Col } A} \mathbf{b}$$

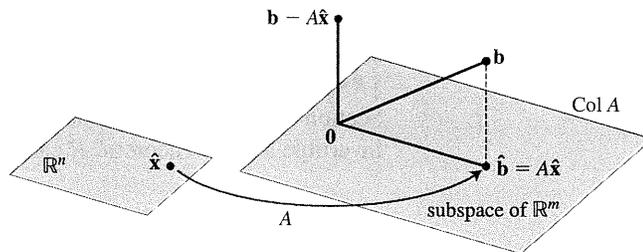


**FIGURE 1** The vector  $\mathbf{b}$  is closer to  $A\hat{\mathbf{x}}$  than to  $A\mathbf{x}$  for other  $\mathbf{x}$ .

Because  $\hat{\mathbf{b}}$  is in the column space of  $A$ , the equation  $A\mathbf{x} = \hat{\mathbf{b}}$  is consistent, and there is an  $\hat{\mathbf{x}}$  in  $\mathbb{R}^n$  such that

$$A\hat{\mathbf{x}} = \hat{\mathbf{b}} \quad (1)$$

Since  $\hat{\mathbf{b}}$  is the closest point in  $\text{Col } A$  to  $\mathbf{b}$ , a vector  $\hat{\mathbf{x}}$  is a least-squares solution of  $A\mathbf{x} = \mathbf{b}$  if and only if  $\hat{\mathbf{x}}$  satisfies (1). Such an  $\hat{\mathbf{x}}$  in  $\mathbb{R}^n$  is a list of weights that will build  $\hat{\mathbf{b}}$  out of the columns of  $A$ . See Figure 2. [There are many solutions of (1) if the equation has free variables.]



**FIGURE 2** The least-squares solution  $\hat{\mathbf{x}}$  is in  $\mathbb{R}^n$ .

Suppose  $\hat{\mathbf{x}}$  satisfies  $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ . By the Orthogonal Decomposition Theorem in Section 6.3, the projection  $\hat{\mathbf{b}}$  has the property that  $\mathbf{b} - \hat{\mathbf{b}}$  is orthogonal to  $\text{Col } A$ , so  $\mathbf{b} - A\hat{\mathbf{x}}$  is orthogonal to each column of  $A$ . If  $\mathbf{a}_j$  is any column of  $A$ , then  $\mathbf{a}_j \cdot (\mathbf{b} - A\hat{\mathbf{x}}) = 0$ , and  $\mathbf{a}_j^T (\mathbf{b} - A\hat{\mathbf{x}}) = 0$ . Since each  $\mathbf{a}_j^T$  is a row of  $A^T$ ,

$$A^T (\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0} \quad (2)$$

(This equation also follows from Theorem 3 in Section 6.1.) Thus

$$\begin{aligned} A^T \mathbf{b} - A^T A \hat{\mathbf{x}} &= \mathbf{0} \\ A^T A \hat{\mathbf{x}} &= A^T \mathbf{b} \end{aligned}$$

These calculations show that each least-squares solution of  $A\mathbf{x} = \mathbf{b}$  satisfies the equation

$$A^T A \mathbf{x} = A^T \mathbf{b} \quad (3)$$

The matrix equation (3) represents a system of equations called the **normal equations** for  $A\mathbf{x} = \mathbf{b}$ . A solution of (3) is often denoted by  $\hat{\mathbf{x}}$ .

### THEOREM 13

The set of least-squares solutions of  $A\mathbf{x} = \mathbf{b}$  coincides with the nonempty set of solutions of the normal equations  $A^T A \mathbf{x} = A^T \mathbf{b}$ .

**PROOF** As shown, the set of least-squares solutions is nonempty and each least-squares solution  $\hat{\mathbf{x}}$  satisfies the normal equations. Conversely, suppose  $\hat{\mathbf{x}}$  satisfies  $A^T A \hat{\mathbf{x}} = A^T \mathbf{b}$ . Then  $\hat{\mathbf{x}}$  satisfies (2), which shows that  $\mathbf{b} - A\hat{\mathbf{x}}$  is orthogonal to the rows of  $A^T$  and hence is orthogonal to the columns of  $A$ . Since the columns of  $A$  span  $\text{Col } A$ , the vector  $\mathbf{b} - A\hat{\mathbf{x}}$  is orthogonal to all of  $\text{Col } A$ . Hence the equation

$$\mathbf{b} = A\hat{\mathbf{x}} + (\mathbf{b} - A\hat{\mathbf{x}})$$

is a decomposition of  $\mathbf{b}$  into the sum of a vector in  $\text{Col } A$  and a vector orthogonal to  $\text{Col } A$ . By the uniqueness of the orthogonal decomposition,  $A\hat{\mathbf{x}}$  must be the orthogonal projection of  $\mathbf{b}$  onto  $\text{Col } A$ . That is,  $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ , and  $\hat{\mathbf{x}}$  is a least-squares solution. ■

**EXAMPLE 1** Find a least-squares solution of the inconsistent system  $A\mathbf{x} = \mathbf{b}$  for

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

**SOLUTION** To use normal equations (3), compute:

$$A^T A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$A^T \mathbf{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

Then the equation  $A^T A \mathbf{x} = A^T \mathbf{b}$  becomes

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

Row operations can be used to solve this system, but since  $A^T A$  is invertible and  $2 \times 2$ , it is probably faster to compute

$$(A^T A)^{-1} = \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$$

and then to solve  $A^T A \mathbf{x} = A^T \mathbf{b}$  as

$$\begin{aligned} \hat{\mathbf{x}} &= (A^T A)^{-1} A^T \mathbf{b} \\ &= \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix} \begin{bmatrix} 19 \\ 11 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 84 \\ 168 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \end{aligned} \quad \blacksquare$$

In many calculations,  $A^T A$  is invertible, but this is not always the case. The next example involves a matrix of the sort that appears in what are called *analysis of variance* problems in statistics.

**EXAMPLE 2** Find a least-squares solution of  $A\mathbf{x} = \mathbf{b}$  for

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

SOLUTION Compute

$$A^T A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}$$

$$A^T \mathbf{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}$$

The augmented matrix for  $A^T A \mathbf{x} = A^T \mathbf{b}$  is

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

The general solution is  $x_1 = 3 - x_4$ ,  $x_2 = -5 + x_4$ ,  $x_3 = -2 + x_4$ , and  $x_4$  is free. So the general least-squares solution of  $A\mathbf{x} = \mathbf{b}$  has the form

$$\hat{\mathbf{x}} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \blacksquare$$

The next theorem gives useful criteria for determining when there is only one least-squares solution of  $A\mathbf{x} = \mathbf{b}$ . (Of course, the orthogonal projection  $\hat{\mathbf{b}}$  is always unique.)

#### THEOREM 14

Let  $A$  be an  $m \times n$  matrix. The following statements are logically equivalent:

- The equation  $A\mathbf{x} = \mathbf{b}$  has a unique least-squares solution for each  $\mathbf{b}$  in  $\mathbb{R}^m$ .
- The columns of  $A$  are linearly independent.
- The matrix  $A^T A$  is invertible.

When these statements are true, the least-squares solution  $\hat{\mathbf{x}}$  is given by

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b} \quad (4)$$

The main elements of a proof of Theorem 14 are outlined in Exercises 27–29, which also review concepts from Chapter 4. Formula (4) for  $\hat{\mathbf{x}}$  is useful mainly for theoretical purposes and for hand calculations when  $A^T A$  is a  $2 \times 2$  invertible matrix.

When a least-squares solution  $\hat{\mathbf{x}}$  is used to produce  $A\hat{\mathbf{x}}$  as an approximation to  $\mathbf{b}$ , the distance from  $\mathbf{b}$  to  $A\hat{\mathbf{x}}$  is called the **least-squares error** of this approximation.

**EXAMPLE 3** Given  $A$  and  $\mathbf{b}$  as in Example 1, determine the least-squares error in the least-squares solution of  $A\mathbf{x} = \mathbf{b}$ .

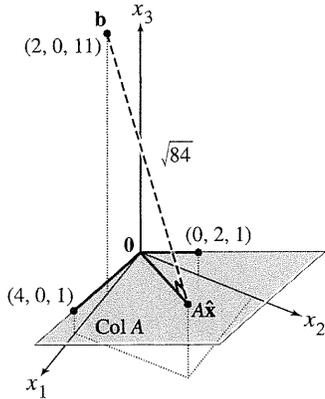


FIGURE 3

SOLUTION From Example 1,

$$\mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} \quad \text{and} \quad A\hat{\mathbf{x}} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$$

Hence

$$\mathbf{b} - A\hat{\mathbf{x}} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ -4 \\ 8 \end{bmatrix}$$

and

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| = \sqrt{(-2)^2 + (-4)^2 + 8^2} = \sqrt{84}$$

The least-squares error is  $\sqrt{84}$ . For any  $\mathbf{x}$  in  $\mathbb{R}^2$ , the distance between  $\mathbf{b}$  and the vector  $A\mathbf{x}$  is at least  $\sqrt{84}$ . See Figure 3. Note that the least-squares solution  $\hat{\mathbf{x}}$  itself does not appear in the figure. ■

### Alternative Calculations of Least-Squares Solutions

The next example shows how to find a least-squares solution of  $A\mathbf{x} = \mathbf{b}$  when the columns of  $A$  are orthogonal. Such matrices often appear in linear regression problems, discussed in the next section.

**EXAMPLE 4** Find a least-squares solution of  $A\mathbf{x} = \mathbf{b}$  for

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

SOLUTION Because the columns  $\mathbf{a}_1$  and  $\mathbf{a}_2$  of  $A$  are orthogonal, the orthogonal projection of  $\mathbf{b}$  onto  $\text{Col } A$  is given by

$$\begin{aligned} \hat{\mathbf{b}} &= \frac{\mathbf{b} \cdot \mathbf{a}_1}{\mathbf{a}_1 \cdot \mathbf{a}_1} \mathbf{a}_1 + \frac{\mathbf{b} \cdot \mathbf{a}_2}{\mathbf{a}_2 \cdot \mathbf{a}_2} \mathbf{a}_2 = \frac{8}{4} \mathbf{a}_1 + \frac{45}{90} \mathbf{a}_2 \\ &= \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} + \begin{bmatrix} -3 \\ -1 \\ 1/2 \\ 7/2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 5/2 \\ 11/2 \end{bmatrix} \end{aligned} \tag{5}$$

Now that  $\hat{\mathbf{b}}$  is known, we can solve  $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ . But this is trivial, since we already know what weights to place on the columns of  $A$  to produce  $\hat{\mathbf{b}}$ . It is clear from (5) that

$$\hat{\mathbf{x}} = \begin{bmatrix} 8/4 \\ 45/90 \end{bmatrix} = \begin{bmatrix} 2 \\ 1/2 \end{bmatrix} \quad \blacksquare$$

In some cases, the normal equations for a least-squares problem can be *ill-conditioned*; that is, small errors in the calculations of the entries of  $A^T A$  can sometimes cause relatively large errors in the solution  $\hat{\mathbf{x}}$ . If the columns of  $A$  are linearly independent, the least-squares solution can often be computed more reliably through a QR factorization of  $A$  (described in Section 6.4).<sup>1</sup>

<sup>1</sup> The QR method is compared with the standard normal equation method in G. Golub and C. Van Loan, *Matrix Computations*, 3rd ed. (Baltimore: Johns Hopkins Press, 1996), pp. 230–231.



## 7.1 Diagonalization of Symmetric Matrices

A **symmetric** matrix is a matrix  $A$  such that  $A^T = A$ . Such a matrix is necessarily square. Its main diagonal entries are arbitrary, but its other entries occur in pairs—on opposite sides of the main diagonal.

**EXAMPLE 1** Of the following matrices, only the first three are symmetric:

$$\begin{aligned} \text{Symmetric: } & \begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix}, \quad \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & 8 \\ 0 & 8 & -7 \end{bmatrix}, \quad \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} \\ \text{Nonsymmetric: } & \begin{bmatrix} 1 & -3 \\ 3 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & -4 & 0 \\ -6 & 1 & -4 \\ 0 & -6 & 1 \end{bmatrix}, \quad \begin{bmatrix} 5 & 4 & 3 & 2 \\ 4 & 3 & 2 & 1 \\ 3 & 2 & 1 & 0 \end{bmatrix} \quad \blacksquare \end{aligned}$$

To begin the study of symmetric matrices, it is helpful to review the diagonalization process of Section 5.3.

**EXAMPLE 2** If possible, diagonalize the matrix  $A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}$ .

**SOLUTION** The characteristic equation of  $A$  is

$$0 = -\lambda^3 + 17\lambda^2 - 90\lambda + 144 = -(\lambda - 8)(\lambda - 6)(\lambda - 3)$$

Standard calculations produce a basis for each eigenspace:

$$\lambda = 8: \mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}; \quad \lambda = 6: \mathbf{v}_2 = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}; \quad \lambda = 3: \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

These three vectors form a basis for  $\mathbb{R}^3$ . In fact, it is easy to check that  $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$  is an *orthogonal* basis for  $\mathbb{R}^3$ . Experience from Chapter 6 suggests that an *orthonormal* basis might be useful for calculations, so here are the normalized (unit) eigenvectors.

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

Let

$$P = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}, \quad D = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Then  $A = PDP^{-1}$ , as usual. But this time, since  $P$  is square and has orthonormal columns,  $P$  is an *orthogonal* matrix, and  $P^{-1}$  is simply  $P^T$ . (See Section 6.2.)  $\blacksquare$

Theorem 1 explains why the eigenvectors in Example 2 are orthogonal—they correspond to distinct eigenvalues.

## THEOREM 1

If  $A$  is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.

**PROOF** Let  $\mathbf{v}_1$  and  $\mathbf{v}_2$  be eigenvectors that correspond to distinct eigenvalues, say,  $\lambda_1$  and  $\lambda_2$ . To show that  $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ , compute

$$\begin{aligned}\lambda_1 \mathbf{v}_1 \cdot \mathbf{v}_2 &= (\lambda_1 \mathbf{v}_1)^T \mathbf{v}_2 = (A\mathbf{v}_1)^T \mathbf{v}_2 && \text{Since } \mathbf{v}_1 \text{ is an eigenvector} \\ &= (\mathbf{v}_1^T A^T) \mathbf{v}_2 = \mathbf{v}_1^T (A\mathbf{v}_2) && \text{Since } A^T = A \\ &= \mathbf{v}_1^T (\lambda_2 \mathbf{v}_2) && \text{Since } \mathbf{v}_2 \text{ is an eigenvector} \\ &= \lambda_2 \mathbf{v}_1^T \mathbf{v}_2 = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2\end{aligned}$$

Hence  $(\lambda_1 - \lambda_2)\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ . But  $\lambda_1 - \lambda_2 \neq 0$ , so  $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ . ■

The special type of diagonalization in Example 2 is crucial for the theory of symmetric matrices. An  $n \times n$  matrix  $A$  is said to be **orthogonally diagonalizable** if there are an orthogonal matrix  $P$  (with  $P^{-1} = P^T$ ) and a diagonal matrix  $D$  such that

$$A = PDP^T = PDP^{-1} \quad (1)$$

Such a diagonalization requires  $n$  linearly independent and orthonormal eigenvectors. When is this possible? If  $A$  is orthogonally diagonalizable as in (1), then

$$A^T = (PDP^T)^T = P^{TT} D^T P^T = PDP^T = A$$

Thus  $A$  is symmetric! Theorem 2 below shows that, conversely, every symmetric matrix is orthogonally diagonalizable. The proof is much harder and is omitted; the main idea for a proof will be given after Theorem 3.

## THEOREM 2

An  $n \times n$  matrix  $A$  is orthogonally diagonalizable if and only if  $A$  is a symmetric matrix.

This theorem is rather amazing, because the work in Chapter 5 would suggest that it is usually impossible to tell when a matrix is diagonalizable. But this is not the case for symmetric matrices.

The next example treats a matrix whose eigenvalues are not all distinct.

**EXAMPLE 3** Orthogonally diagonalize the matrix  $A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$ , whose

characteristic equation is

$$0 = -\lambda^3 + 12\lambda^2 - 21\lambda - 98 = -(\lambda - 7)^2(\lambda + 2)$$

**SOLUTION** The usual calculations produce bases for the eigenspaces:

$$\lambda = 7: \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix}; \quad \lambda = -2: \mathbf{v}_3 = \begin{bmatrix} -1 \\ -1/2 \\ 1 \end{bmatrix}$$

Although  $\mathbf{v}_1$  and  $\mathbf{v}_2$  are linearly independent, they are not orthogonal. Recall from Section 6.2 that the projection of  $\mathbf{v}_2$  onto  $\mathbf{v}_1$  is  $\frac{\mathbf{v}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1$ , and the component of  $\mathbf{v}_2$  orthogonal to  $\mathbf{v}_1$  is

$$\mathbf{z}_2 = \mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 = \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix} - \frac{-1/2}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1/4 \\ 1 \\ 1/4 \end{bmatrix}$$

Then  $\{\mathbf{v}_1, \mathbf{z}_2\}$  is an orthogonal set in the eigenspace for  $\lambda = 7$ . (Note that  $\mathbf{z}_2$  is a linear combination of the eigenvectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$ , so  $\mathbf{z}_2$  is in the eigenspace. This construction of  $\mathbf{z}_2$  is just the Gram–Schmidt process of Section 6.4.) Since the eigenspace is two-dimensional (with basis  $\mathbf{v}_1, \mathbf{v}_2$ ), the orthogonal set  $\{\mathbf{v}_1, \mathbf{z}_2\}$  is an *orthogonal basis* for the eigenspace, by the Basis Theorem. (See Section 2.9 or 4.5.)

Normalize  $\mathbf{v}_1$  and  $\mathbf{z}_2$  to obtain the following orthonormal basis for the eigenspace for  $\lambda = 7$ :

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1/\sqrt{18} \\ 4/\sqrt{18} \\ 1/\sqrt{18} \end{bmatrix}$$

An orthonormal basis for the eigenspace for  $\lambda = -2$  is

$$\mathbf{u}_3 = \frac{1}{\|2\mathbf{v}_3\|} 2\mathbf{v}_3 = \frac{1}{3} \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}$$

By Theorem 1,  $\mathbf{u}_3$  is orthogonal to the other eigenvectors  $\mathbf{u}_1$  and  $\mathbf{u}_2$ . Hence  $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$  is an orthonormal set. Let

$$P = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3] = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{18} & -2/3 \\ 0 & 4/\sqrt{18} & -1/3 \\ 1/\sqrt{2} & 1/\sqrt{18} & 2/3 \end{bmatrix}, \quad D = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

Then  $P$  orthogonally diagonalizes  $A$ , and  $A = PDP^{-1}$ . ■

In Example 3, the eigenvalue 7 has multiplicity two and the eigenspace is two-dimensional. This fact is not accidental, as the next theorem shows.

## The Spectral Theorem

The set of eigenvalues of a matrix  $A$  is sometimes called the *spectrum* of  $A$ , and the following description of the eigenvalues is called a *spectral theorem*.

### THEOREM 3

#### The Spectral Theorem for Symmetric Matrices

An  $n \times n$  symmetric matrix  $A$  has the following properties:

- $A$  has  $n$  real eigenvalues, counting multiplicities.
- The dimension of the eigenspace for each eigenvalue  $\lambda$  equals the multiplicity of  $\lambda$  as a root of the characteristic equation.
- The eigenspaces are mutually orthogonal, in the sense that eigenvectors corresponding to different eigenvalues are orthogonal.
- $A$  is orthogonally diagonalizable.

Part (a) follows from Exercise 28 in Section 5.5. Part (b) follows easily from part (d). (See Exercise 37.) Part (c) is Theorem 1. Because of (a), a proof of (d) can be given using Exercise 38 and the Schur factorization discussed in Supplementary Exercise 34 in Chapter 6. The details are omitted.

## Spectral Decomposition

Suppose  $A = PDP^{-1}$ , where the columns of  $P$  are orthonormal eigenvectors  $\mathbf{u}_1, \dots, \mathbf{u}_n$  of  $A$  and the corresponding eigenvalues  $\lambda_1, \dots, \lambda_n$  are in the diagonal matrix  $D$ . Then, since  $P^{-1} = P^T$ ,

$$\begin{aligned} A = PDP^T &= [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_n] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \\ &= [\lambda_1 \mathbf{u}_1 \quad \cdots \quad \lambda_n \mathbf{u}_n] \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \end{aligned}$$

Using the column–row expansion of a product (Theorem 10 in Section 2.4), we can write

$$A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T \quad (2)$$

This representation of  $A$  is called a **spectral decomposition** of  $A$  because it breaks up  $A$  into pieces determined by the spectrum (eigenvalues) of  $A$ . Each term in (2) is an  $n \times n$  matrix of rank 1. For example, every column of  $\lambda_1 \mathbf{u}_1 \mathbf{u}_1^T$  is a multiple of  $\mathbf{u}_1$ . Furthermore, each matrix  $\mathbf{u}_j \mathbf{u}_j^T$  is a **projection matrix** in the sense that for each  $\mathbf{x}$  in  $\mathbb{R}^n$ , the vector  $(\mathbf{u}_j \mathbf{u}_j^T) \mathbf{x}$  is the orthogonal projection of  $\mathbf{x}$  onto the subspace spanned by  $\mathbf{u}_j$ . (See Exercise 41.)

**EXAMPLE 4** Construct a spectral decomposition of the matrix  $A$  that has the orthogonal diagonalization

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

**SOLUTION** Denote the columns of  $P$  by  $\mathbf{u}_1$  and  $\mathbf{u}_2$ . Then

$$A = 8\mathbf{u}_1 \mathbf{u}_1^T + 3\mathbf{u}_2 \mathbf{u}_2^T$$

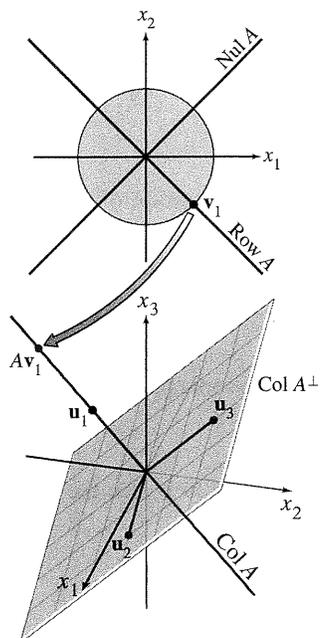
To verify this decomposition of  $A$ , compute

$$\mathbf{u}_1 \mathbf{u}_1^T = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 4/5 & 2/5 \\ 2/5 & 1/5 \end{bmatrix}$$

$$\mathbf{u}_2 \mathbf{u}_2^T = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 1/5 & -2/5 \\ -2/5 & 4/5 \end{bmatrix}$$

and

$$8\mathbf{u}_1 \mathbf{u}_1^T + 3\mathbf{u}_2 \mathbf{u}_2^T = \begin{bmatrix} 32/5 & 16/5 \\ 16/5 & 8/5 \end{bmatrix} + \begin{bmatrix} 3/5 & -6/5 \\ -6/5 & 12/5 \end{bmatrix} = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix} = A \quad \blacksquare$$



The fundamental subspaces in Example 4.

From (5) and (6), the orthogonal complement of  $\text{Nul } A^T$  is  $\text{Col } A$ . Interchanging  $A$  and  $A^T$ , note that  $(\text{Nul } A)^\perp = \text{Col } A^T = \text{Row } A$ . Hence, from (7),

$$\{\mathbf{v}_1, \dots, \mathbf{v}_r\} \tag{8}$$

is an orthonormal basis for  $\text{Row } A$ .

Figure 4 summarizes (5)–(8), but shows the orthogonal basis  $\{\sigma_1 \mathbf{u}_1, \dots, \sigma_r \mathbf{u}_r\}$  for  $\text{Col } A$  instead of the normalized basis, to remind you that  $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$  for  $1 \leq i \leq r$ . Explicit orthonormal bases for the four fundamental subspaces determined by  $A$  are useful in some calculations, particularly in constrained optimization problems. ■

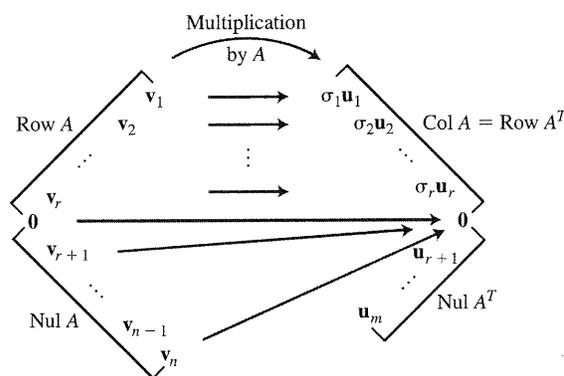


FIGURE 4 The four fundamental subspaces and the action of  $A$ .

The four fundamental subspaces and the concept of singular values provide the final statements of the Invertible Matrix Theorem. (Recall that statements about  $A^T$  have been omitted from the theorem to avoid nearly doubling the number of statements.) The other statements were given in Sections 2.3, 2.9, 3.2, 4.5, and 5.2.

**THEOREM**

**The Invertible Matrix Theorem (concluded)**

Let  $A$  be an  $n \times n$  matrix. Then the following statements are each equivalent to the statement that  $A$  is an invertible matrix:

- s.  $(\text{Col } A)^\perp = \{\mathbf{0}\}$ .
- t.  $(\text{Nul } A)^\perp = \mathbb{R}^n$ .
- u.  $\text{Row } A = \mathbb{R}^n$ .
- v.  $A$  has  $n$  nonzero singular values.

**EXAMPLE 7** (Reduced SVD and the Pseudoinverse of  $A$ ) When  $\Sigma$  contains rows or columns of zeros, a more compact decomposition of  $A$  is possible. Using the notation established above, let  $r = \text{rank } A$ , and partition  $U$  and  $V$  into submatrices whose first blocks contain  $r$  columns:

$$U = [U_r \quad U_{m-r}], \quad \text{where } U_r = [\mathbf{u}_1 \quad \dots \quad \mathbf{u}_r]$$

$$V = [V_r \quad V_{n-r}], \quad \text{where } V_r = [\mathbf{v}_1 \quad \dots \quad \mathbf{v}_r]$$

Then  $U_r$  is  $m \times r$  and  $V_r$  is  $n \times r$ . (To simplify notation, we consider  $U_{m-r}$  or  $V_{n-r}$  even though one of them may have no columns.) Then partitioned matrix multiplication shows that

$$A = [U_r \quad U_{m-r}] \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^T \\ V_{n-r}^T \end{bmatrix} = U_r D V_r^T \quad (9)$$

This factorization of  $A$  is called a **reduced singular value decomposition** of  $A$ . Since the diagonal entries in  $D$  are nonzero,  $D$  is invertible. The following matrix is called the **pseudoinverse** (also, the **Moore–Penrose inverse**) of  $A$ :

$$A^+ = V_r D^{-1} U_r^T \quad (10)$$

Supplementary Exercises 28–30 at the end of the chapter explore some of the properties of the reduced singular value decomposition and the pseudoinverse. ■

**EXAMPLE 8** (Least-Squares Solution) Given the equation  $A\mathbf{x} = \mathbf{b}$ , use the pseudoinverse of  $A$  in (10) to define

$$\hat{\mathbf{x}} = A^+ \mathbf{b} = V_r D^{-1} U_r^T \mathbf{b}$$

Then, from the SVD in (9),

$$\begin{aligned} A\hat{\mathbf{x}} &= (U_r D V_r^T)(V_r D^{-1} U_r^T \mathbf{b}) \\ &= U_r D D^{-1} U_r^T \mathbf{b} \quad \text{Because } V_r^T V_r = I_r \\ &= U_r U_r^T \mathbf{b} \end{aligned}$$

It follows from (5) that  $U_r U_r^T \mathbf{b}$  is the orthogonal projection  $\hat{\mathbf{b}}$  of  $\mathbf{b}$  onto  $\text{Col } A$ . (See Theorem 10 in Section 6.3.) Thus  $\hat{\mathbf{x}}$  is a least-squares solution of  $A\mathbf{x} = \mathbf{b}$ . In fact, this  $\hat{\mathbf{x}}$  has the smallest length among all least-squares solutions of  $A\mathbf{x} = \mathbf{b}$ . See Supplementary Exercise 30. ■

#### Numerical Notes

Examples 1–4 and the exercises illustrate the concept of singular values and suggest how to perform calculations by hand. In practice, the computation of  $A^T A$  should be avoided, since any errors in the entries of  $A$  are squared in the entries of  $A^T A$ . There exist fast iterative methods that produce the singular values and singular vectors of  $A$  accurately to many decimal places.

#### Practice Problems

1. Given a singular value decomposition,  $A = U \Sigma V^T$ , find an SVD of  $A^T$ . How are the singular values of  $A$  and  $A^T$  related?
2. For any  $n \times n$  matrix  $A$ , use the SVD to show that there is an  $n \times n$  orthogonal matrix  $Q$  such that  $A^T A = Q^T (A^T A) Q$ .

*Remark:* Practice Problem 2 establishes that for any  $n \times n$  matrix  $A$ , the matrices  $A A^T$  and  $A^T A$  are *orthogonally similar*.