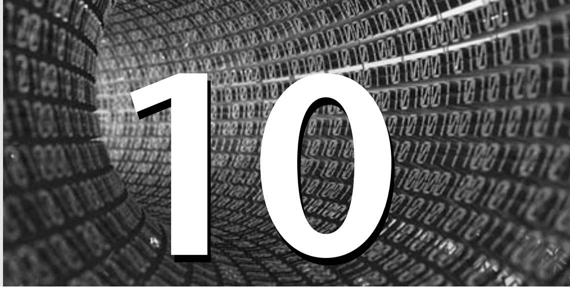

Real Inner Products and Least-Square

10.1 Introduction

To any two vectors **x** and **y** of the same dimension having real components (as distinct from complex components), we associate a scalar called the *inner product*, denoted as $\langle \mathbf{x}, \mathbf{y} \rangle$, by multiplying together the corresponding elements of **x** and **y**, and then summing the results. Students already familiar with the dot product of two- and three-dimensional vectors will undoubtedly recognize the inner product as an extension of the dot product to real vectors of all dimensions.

Example 1 Find $\langle \mathbf{x}, \mathbf{y} \rangle$ if

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} 4 \\ -5 \\ 6 \end{bmatrix}.$$

Solution $\langle \mathbf{x}, \mathbf{y} \rangle = 1(4) + 2(-5) + 3(6) = 12.$ ■

Example 2 Find $\langle \mathbf{u}, \mathbf{v} \rangle$ if $\mathbf{u} = [20 \quad -4 \quad 30 \quad 10]$ and $\mathbf{v} = [10 \quad -5 \quad -8 \quad -6]$.

Solution $\langle \mathbf{u}, \mathbf{v} \rangle = 20(10) + (-4)(-5) + 30(-8) + 10(-6) = -80.$ ■

It follows immediately from the definition that the inner product of real vectors satisfies the following properties:

(**I**1) $\langle \mathbf{x}, \mathbf{x} \rangle$ is positive if $\mathbf{x} \neq \mathbf{0}$; $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

(**I**2) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$.

(**I**3) $\langle \lambda\mathbf{x}, \mathbf{y} \rangle = \lambda\langle \mathbf{x}, \mathbf{y} \rangle$, for any real scalar λ.

(**I**4) $\langle \mathbf{x}+\mathbf{z}, \mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle + \langle \mathbf{z}, \mathbf{y}\rangle$.

(**I**5) $\langle \mathbf{0}, \mathbf{y}\rangle = 0$.

We will only prove (**I**1) here and leave the proofs of the other properties as exercises for the students (see Problems 29 through 32). Let $\mathbf{x} = [x_1\ x_2\ x_3\ \cdots\ x_n]$ be an n-dimensional row vector whose components $x_1, x_2, x_3, \ldots, x_n$ are all real. Then,

$$\langle \mathbf{x}, \mathbf{x}\rangle = (x_1)^2 + (x_2)^2 + (x_3)^2 + \cdots + (x_n)^2.$$

This sum of squares is zero if and only if $x_1 = x_2 = x_3 = \cdots = x_n = 0$, which in turn implies $\mathbf{x} = \mathbf{0}$. If any one component is not zero, that is, if $\mathbf{x}$ is not the zero vector, then the sum of squares must be positive.

The inner product of real vectors is related to the magnitude of a vector as defined in Section 1.6. In particular,

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}.$$

Example 3 Find the magnitude of $\mathbf{x} = [2\ \ -3\ \ -4]$.

Solution $\langle \mathbf{x}, \mathbf{x}\rangle = 2(2) + (-3)(-3) + (-4)(-4) = 29$, so the magnitude of $\mathbf{x}$ is

$$\|\mathbf{x}\| = \sqrt{29}. \quad \blacksquare$$

The concepts of a normalized vector and a unit vector are identical to the definitions given in Section 1.6. A nonzero vector is *normalized* if it is divided by its magnitude. A *unit vector* is a vector whose magnitude is unity. Thus, if $\mathbf{x}$ is any nonzero vector, then $(1/\|\mathbf{x}\|)\,\mathbf{x}$ is normalized. Furthermore,

$$\begin{aligned}
\left\langle \frac{1}{\|\mathbf{x}\|}\mathbf{x}, \frac{1}{\|\mathbf{x}\|}\mathbf{x}\right\rangle &= \frac{1}{\|\mathbf{x}\|}\left\langle \mathbf{x}, \frac{1}{\|\mathbf{x}\|}\mathbf{x}\right\rangle && \text{(Property } \mathbf{I}3)\\
&= \frac{1}{\|\mathbf{x}\|}\left\langle \frac{1}{\|\mathbf{x}\|}\mathbf{x}, \mathbf{x}\right\rangle && \text{(Property } \mathbf{I}2)\\
&= \left(\frac{1}{\|\mathbf{x}\|}\right)^2 \langle \mathbf{x}, \mathbf{x}\rangle && \text{(Property } \mathbf{I}3)\\
&= \left(\frac{1}{\|\mathbf{x}\|}\right)^2 \|\mathbf{x}\|^2 = 1,
\end{aligned}$$

so a normalized vector is always a unit vector.

Problems 10.1

In Problems 1 through 17, find (a) $\langle \mathbf{x}, \mathbf{y} \rangle$ and (b) $\langle \mathbf{x}, \mathbf{x} \rangle$ for the given vectors.

1. $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$.

2. $\mathbf{x} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 4 \\ -5 \end{bmatrix}$.

3. $\mathbf{x} = \begin{bmatrix} -5 \\ 7 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 3 \\ -5 \end{bmatrix}$.

4. $\mathbf{x} = [3 \quad 14]$ and $\mathbf{y} = [7 \quad 3]$.

5. $\mathbf{x} = [-2 \quad -8]$ and $\mathbf{y} = [-4 \quad -7]$.

6. $\mathbf{x} = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$.

7. $\mathbf{x} = \begin{bmatrix} -2 \\ 2 \\ -4 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} -4 \\ 3 \\ -3 \end{bmatrix}$.

8. $\mathbf{x} = \begin{bmatrix} -3 \\ -2 \\ 5 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 6 \\ -4 \\ -4 \end{bmatrix}$.

9. $\mathbf{x} = [\frac{1}{2} \quad \frac{1}{3} \quad \frac{1}{6}]$ and $\mathbf{y} = [\frac{1}{3} \quad \frac{3}{2} \quad 1]$.

10. $\mathbf{x} = [1/\sqrt{2} \quad 1/\sqrt{3} \quad 1/\sqrt{6}]$ and $\mathbf{y} = [1/\sqrt{3} \quad 3/\sqrt{2} \quad 1]$.

11. $\mathbf{x} = [\frac{1}{3} \quad \frac{1}{3} \quad \frac{1}{3}]$ and $\mathbf{y} = [\frac{1}{4} \quad \frac{1}{2} \quad \frac{1}{8}]$.

12. $\mathbf{x} = [10 \quad 20 \quad 30]$ and $\mathbf{y} = [5 \quad -7 \quad 3]$.

13. $\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}$.

14. $\mathbf{x} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ -4 \end{bmatrix}$.

15. $\mathbf{x} = \begin{bmatrix} 3 \\ 5 \\ -7 \\ -8 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 4 \\ -6 \\ -9 \\ 8 \end{bmatrix}$.

16. $\mathbf{x} = \left[\frac{1}{5} \quad \frac{1}{5} \quad \frac{1}{5} \quad \frac{1}{5} \quad \frac{1}{5}\right]$ and $\mathbf{y} = \left[1 \quad 2 \quad -3 \quad 4 \quad -5\right]$.

17. $\mathbf{x} = [1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1]$ and $\mathbf{y} = [-3 \quad 8 \quad 11 \quad -4 \quad 7]$.

18. Normalize **y** as given in Problem 1.

19. Normalize **y** as given in Problem 2.

20. Normalize **y** as given in Problem 4.

21. Normalize **y** as given in Problem 7.

22. Normalize **y** as given in Problem 8.

23. Normalize **y** as given in Problem 11.

24. Normalize **y** as given in Problem 15.

25. Normalize **y** as given in Problem 16.

26. Normalize **y** as given in Problem 17.

27. Find **x** if $\langle \mathbf{x}, \mathbf{a} \rangle \mathbf{b} = \mathbf{c}$, where

$$\mathbf{a} = \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}.$$

28. Determine whether it is possible for two nonzero vectors to have an inner product that is zero.

29. Prove Property **I**2.

30. Prove Property **I**3.

31. Prove Property **I**4.

32. Prove Property **I**5.

33. Prove that $\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{y}\|^2$.

34. Prove the *parallelogram law*:

$$\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2.$$

35. Prove that, for any scalar λ,

$$0 \le \|\lambda\mathbf{x} - \mathbf{y}\|^2 = \lambda^2\|\mathbf{x}\|^2 - 2\lambda\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{y}\|^2.$$

36. (Problem 35 continued) Take $\lambda = \langle \mathbf{x}, \mathbf{y} \rangle / \|\mathbf{x}\|^2$ and show that

$$0 \le \frac{-\langle \mathbf{x}, \mathbf{y} \rangle^2}{\|\mathbf{x}\|^2} + \|\mathbf{y}\|^2.$$

From this, deduce that

$$\langle \mathbf{x}, \mathbf{y} \rangle^2 \leq \|\mathbf{x}\|^2 \|\mathbf{y}\|^2,$$

and that

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \; \|\mathbf{y}\|.$$

This last inequality is known as the *Cauchy–Schwarz inequality*.

37. Using the results of Problem 33 and the Cauchy–Schwarz inequality, show that

$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|^2 &\leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 \\ &= (\|\mathbf{x}\| + \|\mathbf{y}\|)^2. \end{aligned}$$

From this, deduce that

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|.$$

38. Determine whether there exists a relationship between $\langle \mathbf{x}, \mathbf{y} \rangle$ and $\mathbf{x}^\mathsf{T}\mathbf{y}$, when both $\mathbf{x}$ and $\mathbf{y}$ are column vectors of identical dimension with real components.

39. Use the results of Problem 38 to prove that $\langle \mathbf{A}\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{A}^\mathsf{T}\mathbf{y} \rangle$, when $\mathbf{A}$, $\mathbf{x}$, and $\mathbf{y}$ are real matrices of dimensions $n \times n$, $n \times 1$, and $n \times 1$, respectively.

40. A generalization of the inner product for n-dimensional column vectors with real components is $\langle \mathbf{x}, \mathbf{y} \rangle_\mathbf{A} = \langle \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{y} \rangle$ for any real $n \times n$ nonsingular matrix $\mathbf{A}$. This definition reduces to the usual one when $\mathbf{A} = \mathbf{I}$.

Compute $\langle \mathbf{x}, \mathbf{y} \rangle_\mathbf{A}$ for the vectors given in Problem 1 when

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

41. Compute $\langle \mathbf{x}, \mathbf{y} \rangle_\mathbf{A}$ for the vectors given in Problem 6 when

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

42. Redo Problem 41 with

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

10.2 Orthonormal Vectors

Definition 1 Two vectors $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal* (or perpendicular) if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.

Thus, given the vectors

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix},$$

we see that $\mathbf{x}$ is orthogonal to $\mathbf{y}$ and $\mathbf{y}$ is orthogonal to $\mathbf{z}$ since $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{z} \rangle = 0$; but the vectors $\mathbf{x}$ and $\mathbf{z}$ are not orthogonal since $\langle \mathbf{x}, \mathbf{z} \rangle = 1 + 1 \neq 0$. In particular, as a direct consequence of Property (**I**5) of Section 10.1 we have that the zero vector is orthogonal to every vector.

A set of vectors is called an *orthogonal set* if each vector in the set is orthogonal to every other vector in the set. The set given above is not an orthogonal set since $\mathbf{z}$ is not orthogonal to $\mathbf{x}$ whereas the set given by $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$,

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix},$$

is an orthogonal set because each vector is orthogonal to every other vector.

Definition 2 A set of vectors is *orthonormal* if it is an orthogonal set having the property that every vector is a unit vector (a vector of magnitude 1).

The set of vectors

$$\left\{ \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is an example of an orthonormal set.

Definition 2 can be simplified if we make use of the Kronecker delta, δ_{ij}, defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j, \\ 0 & \text{if} \quad i \neq j. \end{cases} \tag{1}$$

A set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is an orthonormal set if and only if

$$\left\langle \mathbf{x}_i, \mathbf{x}_j \right\rangle = \delta_{ij} \quad \text{for all } i \text{ and } j, \quad i, j = 1, 2, \ldots, n. \tag{2}$$

The importance of orthonormal sets is that they are almost equivalent to linearly independent sets. However, since orthonormal sets have associated with them the additional structure of an inner product, they are often more convenient. We devote the remaining portion of this section to showing the equivalence of these two concepts. The utility of orthonormality will become self-evident in later sections.

Theorem 1 *An orthonormal set of vectors is linearly independent.*

Proof. Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be an orthonormal set and consider the vector equation

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_n\mathbf{x}_n = \mathbf{0} \tag{3}$$

where the c_j's $(j = 1, 2, \ldots, n)$ are constants. The set of vectors will be linearly independent if the only constants that satisfy (3) are $c_1 = c_2 = \cdots = c_n = 0$. Take the inner product of both sides of (3) with $\mathbf{x}_1$. Thus,

$$\langle c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_n\mathbf{x}_n, \mathbf{x}_1\rangle = \langle \mathbf{0}, \mathbf{x}_1\rangle .$$

Using properties (**I**3), (**I**4), and (**I**5) of Section 10.1, we have

$$c_1\langle\mathbf{x}_1, \mathbf{x}_1\rangle + c_2\langle\mathbf{x}_2, \mathbf{x}_1\rangle + \cdots + c_n\langle\mathbf{x}_n, \mathbf{x}_1\rangle = 0.$$

Finally, noting that $\langle\mathbf{x}_i, \mathbf{x}_1\rangle = \delta_{i1}$, we obtain $c_1 = 0$. Now taking the inner product of both sides of (3) with $\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n$, successively, we obtain $c_2 = 0$, $c_3 = 0, \ldots, c_n = 0$. Combining these results, we find that $c_1 = c_2 = \cdots c_n = 0$, which implies the theorem. □

Theorem 2 *For every linearly independent set of vectors* $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$*, there exists an orthonormal set of vectors* $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n\}$ *such that each* $\mathbf{q}_j (j = 1, 2, \ldots, n)$ *is a linear combination of* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_j$.

Proof. First define new vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ by

$$\mathbf{y}_1 = \mathbf{x}_1$$
$$\mathbf{y}_2 = \mathbf{x}_2 - \frac{\langle\mathbf{x}_2, \mathbf{y}_1\rangle}{\langle\mathbf{y}_1, \mathbf{y}_1\rangle}\mathbf{y}_1$$
$$\mathbf{y}_3 = \mathbf{x}_3 - \frac{\langle\mathbf{x}_3, \mathbf{y}_1\rangle}{\langle\mathbf{y}_1, \mathbf{y}_1\rangle}\mathbf{y}_1 - \frac{\langle\mathbf{x}_3, \mathbf{y}_2\rangle}{\langle\mathbf{y}_2, \mathbf{y}_2\rangle}\mathbf{y}_2$$

and, in general,

$$\mathbf{y}_j = \mathbf{x}_j - \sum_{k=1}^{j-1}\frac{\langle\mathbf{x}_j, \mathbf{y}_k\rangle}{\langle\mathbf{y}_k, \mathbf{y}_k\rangle}\mathbf{y}_k \quad (j = 2, 3, \ldots, n). \tag{4}$$

Each $\mathbf{y}_j$ is a linear combination of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_j (j = 1, 2, \ldots, n)$. Since the $\mathbf{x}$'s are linearly independent, and the coefficient of the $\mathbf{x}_j$ term in (4) is unity, it follows that $\mathbf{y}_j$ is not the zero vector (see Problem 19). Furthermore, it can be shown that the $\mathbf{y}_j$ terms form an orthogonal set (see Problem 20), hence the only property that the $\mathbf{y}_j$ terms lack in order to be the required set is that their magnitudes may not be one. We remedy this situation by defining

$$\mathbf{q}_j = \frac{\mathbf{y}_j}{\|\mathbf{y}_j\|}. \tag{5}$$

The desired set is $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n\}$.

The process used to construct the $\mathbf{q}_j$ terms is called the *Gram–Schmidt orthonormalization process.* □

Example 1 Use the Gram–Schmidt orthonormalization process to construct an orthonormal set of vectors from the linearly independent set $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$, where

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

Solution

$$\mathbf{y}_1 = \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Now $\langle \mathbf{x}_2, \mathbf{y}_1 \rangle = 0(1) + 1(1) + 1(0) = 1$, and $\langle \mathbf{y}_1, \mathbf{y}_1 \rangle = 1(1) + 1(1) + 0(0) = 2$; hence,

$$\mathbf{y}_2 = \mathbf{x}_2 - \frac{\langle \mathbf{x}_2, \mathbf{y}_1 \rangle}{\langle \mathbf{y}_1, \mathbf{y}_1 \rangle}\mathbf{y}_1 = \mathbf{x}_2 - \frac{1}{2}\mathbf{y}_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 1/2 \\ 1 \end{bmatrix}.$$

Then,

$$\begin{aligned} \langle \mathbf{x}_3, \mathbf{y}_1 \rangle &= 1(1) + 0(1) + 1(0) = 1, \\ \langle \mathbf{x}_3, \mathbf{y}_2 \rangle &= 1(-1/2) + 0\,(1/2) + 1(1) = 1/2, \\ \langle \mathbf{y}_2, \mathbf{y}_2 \rangle &= (-1/2)^2 + (1/2)^2 + (1)^2 = 3/2, \end{aligned}$$

so

$$\begin{aligned} \mathbf{y}_3 &= \mathbf{x}_3 - \frac{\langle \mathbf{x}_3, \mathbf{y}_1 \rangle}{\langle \mathbf{y}_1, \mathbf{y}_1 \rangle}\mathbf{y}_1 - \frac{\langle \mathbf{x}_3, \mathbf{y}_2 \rangle}{\langle \mathbf{y}_2, \mathbf{y}_2 \rangle}\mathbf{y}_2 = \mathbf{x}_3 - \frac{1}{2}\mathbf{y}_1 - \frac{1/2}{3/2}\mathbf{y}_2 \\ &= \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} -1/2 \\ 1/2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2/3 \\ -2/3 \\ 2/3 \end{bmatrix}. \end{aligned}$$

The vectors $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$ form an orthogonal set. To make this set orthonormal, we note that $\langle \mathbf{y}_1, \mathbf{y}_1 \rangle = 2$, $\langle \mathbf{y}_2, \mathbf{y}_2 \rangle = 3/2$, and $\langle \mathbf{y}_3, \mathbf{y}_3 \rangle = (2/3)(2/3) + (-2/3)(-2/3) + (2/3)(2/3) = 4/3$. Therefore,

$$\|\mathbf{y}_1\| = \sqrt{\langle \mathbf{y}_1, \mathbf{y}_1 \rangle} = \sqrt{2} \qquad \|\mathbf{y}_2\| = \sqrt{\langle \mathbf{y}_2, \mathbf{y}_2 \rangle} = \sqrt{3/2},$$
$$\|\mathbf{y}_3\| = \sqrt{\langle \mathbf{y}_3, \mathbf{y}_3 \rangle} = 2/\sqrt{3},$$

and

$$\mathbf{q}_1 = \frac{\mathbf{y}_1}{\|\mathbf{y}_1\|} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix},$$

$$\mathbf{q}_2 = \frac{\mathbf{y}_2}{\|\mathbf{y}_2\|} = \frac{1}{\sqrt{3/2}}\begin{bmatrix} -1/2 \\ 1/2 \\ 1 \end{bmatrix} = \begin{bmatrix} -1/\sqrt{6} \\ 1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix},$$

$$\mathbf{q}_3 = \frac{\mathbf{y}_3}{\|\mathbf{y}_3\|} = \frac{1}{2/\sqrt{3}}\begin{bmatrix} 2/3 \\ -2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}. \quad \blacksquare$$

Example 2 Use the Gram–Schmidt orthonormalization process to construct an orthonormal set of vectors from the linearly independent set $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$, where

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

Solution

$$\mathbf{y}_1 = \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix},$$

$$\langle \mathbf{y}_1, \mathbf{y}_1 \rangle = 1(1) + 1(1) + 0(0) + 1(1) = 3,$$

$$\langle \mathbf{x}_2, \mathbf{y}_1 \rangle = 1(1) + 2(1) + 1(0) + 0(1) = 3,$$

$$\mathbf{y}_2 = \mathbf{x}_2 - \frac{\langle \mathbf{x}_2, \mathbf{y}_1 \rangle}{\langle \mathbf{y}_1, \mathbf{y}_1 \rangle}\mathbf{y}_1 = \mathbf{x}_2 - \frac{3}{3}\mathbf{y}_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ -1 \end{bmatrix};$$

$$\langle \mathbf{y}_2, \mathbf{y}_2 \rangle = 0(0) + 1(1) + 1(1) + (-1)(-1) = 3,$$

$$\langle \mathbf{x}_3, \mathbf{y}_1 \rangle = 0(1) + 1(1) + 2(0) + 1(1) = 2,$$

$$\langle \mathbf{x}_3, \mathbf{y}_2 \rangle = 0(0) + 1(1) + 2(1) + 1(-1) = 2,$$

$$\mathbf{y}_3 = \mathbf{x}_3 - \frac{\langle \mathbf{x}_3, \mathbf{y}_1 \rangle}{\langle \mathbf{y}_1, \mathbf{y}_1 \rangle}\mathbf{y}_1 - \frac{\langle \mathbf{x}_3, \mathbf{y}_2 \rangle}{\langle \mathbf{y}_2, \mathbf{y}_2 \rangle}\mathbf{y}_2$$

$$= \mathbf{x}_3 - \frac{2}{3}\mathbf{y}_1 - \frac{2}{3}\mathbf{y}_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 4/3 \\ 1 \end{bmatrix};$$

$$\langle \mathbf{y}_3, \mathbf{y}_3 \rangle = \left(\frac{-2}{3}\right)^2 + \left(\frac{-1}{3}\right)^2 + \left(\frac{4}{3}\right)^2 + (1)^2 = \frac{10}{3},$$

$$\langle \mathbf{x}_4, \mathbf{y}_1 \rangle = 1(1) + 0(1) + 1(0) + 1(1) = 2,$$

$$\langle \mathbf{x}_4, \mathbf{y}_2 \rangle = 1(0) + 0(1) + 1(1) + 1(-1) = 0,$$

$$\langle \mathbf{x}_4, \mathbf{y}_3 \rangle = 1\left(\frac{-2}{3}\right) + 0\left(\frac{-1}{3}\right) + 1\left(\frac{4}{3}\right) + 1(1) = \frac{5}{3},$$

$$\mathbf{y}_4 = \mathbf{x}_4 - \frac{\langle \mathbf{x}_4, \mathbf{y}_1 \rangle}{\langle \mathbf{y}_1, \mathbf{y}_1 \rangle}\mathbf{y}_1 - \frac{\langle \mathbf{x}_4, \mathbf{y}_2 \rangle}{\langle \mathbf{y}_2, \mathbf{y}_2 \rangle}\mathbf{y}_2 - \frac{\langle \mathbf{x}_4, \mathbf{y}_3 \rangle}{\langle \mathbf{y}_3, \mathbf{y}_3 \rangle}\mathbf{y}_3$$

$$= \mathbf{x}_4 - \frac{2}{3}\mathbf{y}_1 - \frac{0}{3}\mathbf{y}_2 - \frac{5/3}{10/3}\mathbf{y}_3 = \begin{bmatrix} 2/3 \\ -1/2 \\ 1/3 \\ -1/6 \end{bmatrix}.$$

Then

$$\langle \mathbf{y}_4, \mathbf{y}_4 \rangle = (2/3)(2/3) + (-1/2)(-1/2) + (1/3)(1/3) + (-1/6)(-1/6)$$
$$= 5/6,$$

and

$$\mathbf{q}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 0 \\ 1/\sqrt{3} \end{bmatrix},$$

$$\mathbf{q}_2 = \frac{1}{\sqrt{3}}\begin{bmatrix} 0 \\ 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1/\sqrt{3} \\ 1/\sqrt{3} \\ -1/\sqrt{3} \end{bmatrix},$$

$$\mathbf{q}_3 = \frac{1}{\sqrt{10/3}}\begin{bmatrix} -2/3 \\ -1/3 \\ 4/3 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/\sqrt{30} \\ -1/\sqrt{30} \\ 4/\sqrt{30} \\ 3/\sqrt{30} \end{bmatrix},$$

$$\mathbf{q}_4 = \frac{1}{\sqrt{5/6}}\begin{bmatrix} 2/3 \\ -1/2 \\ 1/3 \\ -1/6 \end{bmatrix} = \begin{bmatrix} 4/\sqrt{30} \\ -3/\sqrt{30} \\ 2/\sqrt{30} \\ -1/\sqrt{30} \end{bmatrix}. \quad \blacksquare$$

Problems 10.2

1. Determine which of the following vectors are orthogonal:

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \ \mathbf{z} = \begin{bmatrix} -2 \\ -1 \end{bmatrix}, \ \mathbf{u} = \begin{bmatrix} -4 \\ 2 \end{bmatrix}, \ \mathbf{v} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}.$$

2. Determine which of the following vectors are orthogonal:

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{z} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \ \mathbf{u} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \ \mathbf{v} = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}.$$

3. Find x so that

$$\begin{bmatrix} 3 \\ 5 \end{bmatrix} \text{ is orthogonal to } \begin{bmatrix} x \\ 4 \end{bmatrix}.$$

4. Find x so that

$$\begin{bmatrix} -1 \\ x \\ 3 \end{bmatrix} \text{ is orthogonal to } \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

5. Find x so that $[x \ \ x \ \ 2]$ is orthogonal to $[1 \ \ 3 \ \ -1]$.

6. Find x and y so that $[x \ \ y]$ is orthogonal to $[1 \ \ 3]$.

7. Find x and y so that

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \text{ is orthogonal to both } \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \text{ and } \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

8. Find x, y, and z so that

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \text{ is orthogonal to both } \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}.$$

9. Redo Problem 8 with the additional stipulation that $[x \quad y \quad z]^{\mathsf{T}}$ be a unit vector.

In Problems 10 through 18, use the Gram–Schmidt orthonormalization process to construct an orthonormal set from the given set of linearly independent vectors.

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

11. $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 3 \\ 5 \end{bmatrix}.$

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

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

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

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

16. $\mathbf{x}_1 = \begin{bmatrix} 0 \\ 3 \\ 4 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 3 \\ 5 \\ 0 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 2 \\ 5 \\ 5 \end{bmatrix}.$

17. $\mathbf{x}_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$

18. $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}.$

19. Prove that no **y**-vector in the Gram–Schmidt orthonormalization process is zero.

20. Prove that the **y**-vectors in the Gram–Schmidt orthonormalization process form an orthogonal set. (Hint: first show that $\langle \mathbf{y}_2, \mathbf{y}_1 \rangle = 0$, hence $\mathbf{y}_2$ must be orthogonal to $\mathbf{y}_1$. Then use induction.)

21. With $\mathbf{q}_j$ defined by Eq. (5), show that Eq. (4) can be simplified to $\mathbf{y}_j = \mathbf{x}_j - \sum_{k=1}^{j-1} \langle \mathbf{x}_j, \mathbf{q}_k \rangle \mathbf{q}_k$.

22. The vectors

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

are linearly dependent. Apply the Gram–Schmidt process to it, and use the results to deduce what occurs whenever the process is applied to a linearly dependent set of vectors.

23. Prove that if **x** and **y** are orthogonal, then

$$||\mathbf{x} - \mathbf{y}||^2 = ||\mathbf{x}||^2 + ||\mathbf{y}||^2.$$

24. Prove that if **x** and **y** are orthonormal, then

$$||s\mathbf{x} + t\mathbf{y}||^2 = s^2 + t^2$$

for any two scalars s and t.

25. Let **Q** be any $n \times n$ matrix whose columns, when considered as n-dimensional vectors, form an orthonormal set. What can you say about the product $\mathbf{Q}^\mathsf{T}\mathbf{Q}$?

26. Prove that if $\langle \mathbf{y}, \mathbf{x} \rangle = 0$ for every n-dimensional vector **y**, then $\mathbf{x} = \mathbf{0}$.

27. Let **x** and **y** be any two vectors of the same dimension. Prove that $\mathbf{x} + \mathbf{y}$ is orthogonal to $\mathbf{x} - \mathbf{y}$ if and only if $||\mathbf{x}|| = ||\mathbf{y}||$.

28. Let **A** be an $n \times n$ real matrix and **p** be a real n-dimensional column vector. Show that if **p** is orthogonal to the columns of **A**, then $\langle \mathbf{Ay}, \mathbf{p} \rangle = 0$ for any n-dimensional real column vector **y**.

10.3 Projections and QR-Decompositions

As with other vector operations, the inner product has a geometrical interpretation in two or three dimensions. For simplicity, we consider two-dimensional vectors here; the extension to three dimensions is straightforward.

Let **u** and **v** be two nonzero vectors, considered as directed line segments (see Section 1.7), positioned so that their initial points coincide. The *angle between* **u** *and* **v** is the angle θ between the two line segments satisfying $0 \leq \theta \leq \pi$. See Figure 10.1.

Definition 1 If **u** and **v** are two-dimensional vectors and θ is the angle between them, then the *dot product* of these two vectors is $\mathbf{u} \cdot \mathbf{v} = ||\mathbf{u}||\,||\mathbf{v}|| \cos\theta$.

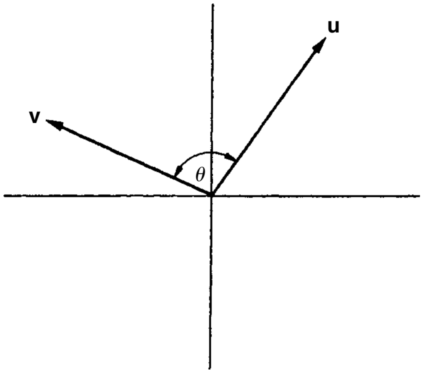

Figure 10.1

To use Definition 1, we need the cosine of the angle between two vectors, which requires us to measure the angle. We shall take another approach.

The vectors $\mathbf{u}$ and $\mathbf{v}$ along with their difference $\mathbf{u}-\mathbf{v}$ form a triangle (see Figure 10.2) having sides $||\mathbf{u}||$, $||\mathbf{v}||$, and $||\mathbf{u}-\mathbf{v}||$. It follows from the law of cosines that

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

whereupon

$$\begin{aligned} ||\mathbf{u}||\,||\mathbf{v}||\cos\theta &= \frac{1}{2}[||\mathbf{u}||^2 + ||\mathbf{v}||^2 - ||\mathbf{u}-\mathbf{v}||^2] \\ &= \frac{1}{2}\left[\langle\mathbf{u},\mathbf{u}\rangle + \langle\mathbf{v},\mathbf{v}\rangle - \langle\mathbf{u}-\mathbf{v},\mathbf{u}-\mathbf{v}\rangle\right] \\ &= \langle\mathbf{u},\mathbf{v}\rangle. \end{aligned}$$

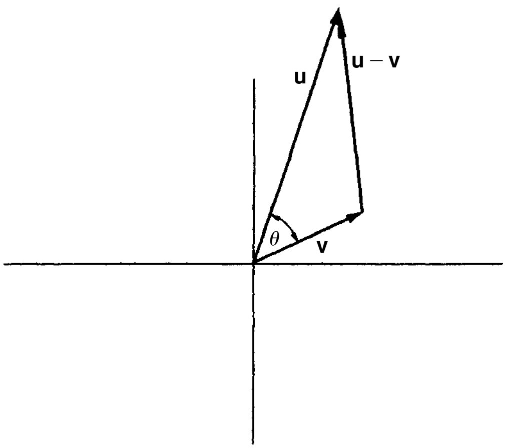

Figure 10.2

Thus, the dot product of two-dimensional vectors is the inner product of those vectors. That is,

$$\mathbf{u} \cdot \mathbf{v} = ||\mathbf{u}||\,||\mathbf{v}|| \cos\theta = \langle \mathbf{u}, \mathbf{v} \rangle. \tag{6}$$

The dot product of nonzero vectors is zero if and only if $\cos\theta = 0$, or $\theta = 90°$. Consequently, the dot product of two nonzero vectors is zero if and only if the vectors are perpendicular. This, with Eq. (6), establishes the equivalence between orthogonality and perpendicularity for two-dimensional vectors. In addition, we may rewrite Eq. (6) as

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{||\mathbf{u}||\,||\mathbf{v}||}, \tag{7}$$

and use Eq. (7) to calculate the angle between two vectors.

Example 1 Find the angle between the vectors

$$\mathbf{u} = \begin{bmatrix} 2 \\ 5 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} -3 \\ 4 \end{bmatrix}.$$

Solution $\langle \mathbf{u}, \mathbf{v} \rangle = 2(-3) + 5(4) = 14$, $\|\mathbf{u}\| = \sqrt{4+25} = \sqrt{29}$, $\|\mathbf{v}\| = \sqrt{9+16} = 5$, so $\cos\theta = 14/(5\sqrt{29}) = 0.1599$, and $\theta = 58.7°$. ■

Eq. (7) is used to define the angle between any two vectors of the same, but arbitrary dimension, even though the geometrical significance of an angle becomes meaningless for dimensions greater than three. (See Problems 9 and 10.)

A problem that occurs often in the applied sciences and that has important ramifications for us in matrices involves a given nonzero vector $\mathbf{x}$ and a nonzero reference vector $\mathbf{a}$. The problem is to decompose $\mathbf{x}$ into the sum of two vectors, $\mathbf{u} + \mathbf{v}$, where $\mathbf{u}$ is parallel to $\mathbf{a}$ and $\mathbf{v}$ is perpendicular to $\mathbf{a}$. This situation is illustrated in Figure 10.3. In physics, $\mathbf{u}$ is called the parallel component of $\mathbf{x}$ and $\mathbf{v}$ is called the

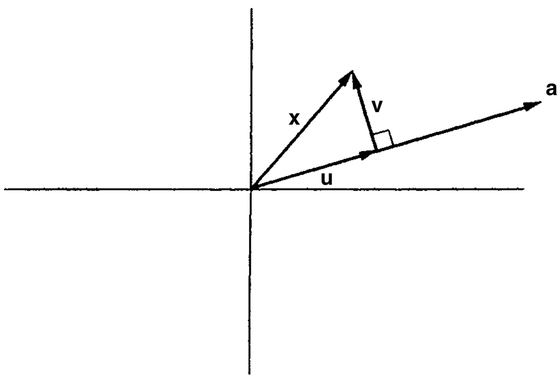

Figure 10.3

perpendicular component of $\mathbf{x}$, where parallel and perpendicular are understood to be with respect to the reference vector $\mathbf{a}$.

If $\mathbf{u}$ is to be parallel to $\mathbf{a}$, it must be a scalar multiple of $\mathbf{a}$, in particular $\mathbf{u} = \lambda\mathbf{a}$. Since we want $\mathbf{x} = \mathbf{u} + \mathbf{v}$, it follows that $\mathbf{v} = \mathbf{x} - \mathbf{u} = \mathbf{x} - \lambda\mathbf{a}$. Finally, if $\mathbf{u}$ and $\mathbf{v}$ are to be perpendicular, we require that

$$\begin{aligned} 0 &= \langle \mathbf{u}, \mathbf{v} \rangle = \langle \lambda\mathbf{a}, \mathbf{x} - \lambda\mathbf{a} \rangle \\ &= \lambda\langle \mathbf{a}, \mathbf{x} \rangle - \lambda^2 \langle \mathbf{a}, \mathbf{a} \rangle \\ &= \lambda[\langle \mathbf{a}, \mathbf{x} \rangle - \lambda\langle \mathbf{a}, \mathbf{a} \rangle]. \end{aligned}$$

Thus, either $\lambda = 0$ or $\lambda = \langle \mathbf{a}, \mathbf{x} \rangle / \langle \mathbf{a}, \mathbf{a} \rangle$. If $\lambda = 0$, then $\mathbf{u} = \lambda\mathbf{a} = \mathbf{0}$, and $\mathbf{x} = \mathbf{u} + \mathbf{v} = \mathbf{v}$, which means that $\mathbf{x}$ and $\mathbf{a}$ are perpendicular. In such a case, $\langle \mathbf{a}, \mathbf{x} \rangle = 0$. Thus, we may always infer that $\lambda = \langle \mathbf{a}, \mathbf{x} \rangle / \langle \mathbf{a}, \mathbf{a} \rangle$, with

$$\mathbf{u} = \frac{\langle \mathbf{a}, \mathbf{x} \rangle}{\langle \mathbf{a}, \mathbf{a} \rangle}\mathbf{a} \quad \text{and} \quad \mathbf{v} = \mathbf{x} - \frac{\langle \mathbf{a}, \mathbf{x} \rangle}{\langle \mathbf{a}, \mathbf{a} \rangle}\mathbf{a}.$$

In this context, $\mathbf{u}$ is the *projection of* $\mathbf{x}$ *onto* $\mathbf{a}$, and $\mathbf{v}$ is the *orthogonal complement*.

Example 2 Decompose the vector

$$\mathbf{x} = \begin{bmatrix} 2 \\ 7 \end{bmatrix}$$

into the sum of two vectors, one of which is parallel to

$$\mathbf{a} = \begin{bmatrix} -3 \\ 4 \end{bmatrix},$$

and one of which is perpendicular to $\mathbf{a}$.

Solution

$$\mathbf{u} = \frac{\langle \mathbf{a}, \mathbf{x} \rangle}{\langle \mathbf{a}, \mathbf{a} \rangle}\mathbf{a} = \frac{22}{25}\begin{bmatrix} -3 \\ 4 \end{bmatrix} = \begin{bmatrix} -2.64 \\ 3.52 \end{bmatrix},$$

$$\mathbf{v} = \mathbf{x} - \mathbf{u} = \begin{bmatrix} 2 \\ 7 \end{bmatrix} - \begin{bmatrix} -2.64 \\ 3.52 \end{bmatrix} = \begin{bmatrix} 4.64 \\ 3.48 \end{bmatrix}.$$

Then, $\mathbf{x} = \mathbf{u} + \mathbf{v}$, with $\mathbf{u}$ parallel to $\mathbf{a}$ and $\mathbf{v}$ perpendicular to $\mathbf{a}$. ■

We now extend the relationships developed in two dimensions to vectors in higher dimensions. Given a nonzero vector $\mathbf{x}$ and another nonzero reference vector $\mathbf{a}$, we define the projection of $\mathbf{x}$ onto $\mathbf{a}$ as

$$\text{proj}_{\mathbf{a}}\mathbf{x} = \frac{\langle \mathbf{a}, \mathbf{x} \rangle}{\langle \mathbf{a}, \mathbf{a} \rangle}\mathbf{a}. \tag{8}$$

As a result, we obtain the very important relationship that

$$\mathbf{x} - \frac{\langle \mathbf{a}, \mathbf{x} \rangle}{\langle \mathbf{a}, \mathbf{a} \rangle} \mathbf{a} \text{ is orthogonal to } \mathbf{a}. \tag{9}$$

That is, if we subtract from a nonzero vector $\mathbf{x}$ its projection onto another nonzero vector $\mathbf{a}$, we are left with a vector that is orthogonal to $\mathbf{a}$. (See Problem 23.)

In this context, the Gram–Schmidt process, described in Section 10.2, is almost obvious. Consider Eq. (4) from that section:

$$\mathbf{y}_j = \mathbf{x}_j - \sum_{k=1}^{j-1} \frac{\langle \mathbf{x}_j, \mathbf{y}_k \rangle}{\langle \mathbf{y}_k, \mathbf{y}_k \rangle} \mathbf{y}_k \qquad \text{(4 repeated)}$$

The quantity inside the summation sign is the projection of $\mathbf{x}_j$ onto $\mathbf{y}_k$. Thus for each k ($k = 1, 2, \dots, j-1$), we are sequentially subtracting from $\mathbf{x}_j$ its projection onto $\mathbf{y}_k$, leaving a vector that is orthogonal to $\mathbf{y}_k$.

We now propose to alter slightly the steps of the Gram–Schmidt orthonormalization process. First, we shall normalize the orthogonal vectors as soon as they are obtained, rather than waiting until the end. This will make for messier hand calculations, but for a more efficient computer algorithm. Observe that if the $\mathbf{y}_k$ vectors in Eq. (4) are unit vectors, then the denominator is unity, and need not be calculated.

Once we have fully determined a $\mathbf{y}_k$ vector, we shall immediately subtract the various projections onto this vector from all succeeding $\mathbf{x}$ vectors. In particular, once $\mathbf{y}_1$ is determined, we shall subtract the projection of $\mathbf{x}_2$ onto $\mathbf{y}_1$ from $\mathbf{x}_2$, then we shall subtract the projection of $\mathbf{x}_3$ onto $\mathbf{y}_1$ from $\mathbf{x}_3$, and continue until we have subtracted the projection of $\mathbf{x}_n$ onto $\mathbf{y}_1$ from $\mathbf{x}_n$. Only then will we return to $\mathbf{x}_2$ and normalize it to obtain $\mathbf{y}_2$. Then, we shall subtract from $\mathbf{x}_3, \mathbf{x}_4, \dots, \mathbf{x}_n$ the projections onto $\mathbf{y}_2$ from $\mathbf{x}_3, \mathbf{x}_4, \dots, \mathbf{x}_n$, respectively, before returning to $\mathbf{x}_3$ and normalizing it, thus obtaining $\mathbf{y}_3$. As a result, once we have $\mathbf{y}_1$, we alter $\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n$ so each is orthogonal to $\mathbf{y}_1$; once we have $\mathbf{y}_2$, we alter again $\mathbf{x}_3, \mathbf{x}_4, \dots, \mathbf{x}_n$ so each is also orthogonal to $\mathbf{y}_2$; and so on.

These changes are known as the *revised Gram–Schmidt algorithm*. Given a set of linearly independent vectors $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, the algorithm may be formalized as follows: Begin with $k = 1$ and, sequentially moving through $k = n$;

(i) calculate $r_{kk} = \sqrt{\langle \mathbf{x}_k, \mathbf{x}_k \rangle}$,

(ii) set $\mathbf{q}_k = (1/r_{kk})\mathbf{x}_k$,

(iii) for $j = k+1, k+2, \dots, n$, calculate $r_{kj} = \langle \mathbf{x}_j, \mathbf{q}_k \rangle$,

(iv) for $j = k+1, k+2, \dots, n$, replace $\mathbf{x}_j$ by $\mathbf{x}_j - \mathbf{r}_{kj}\mathbf{q}_k$.

The first two steps normalize, the third and fourth steps subtract projections from vectors, thereby generating orthogonality.

Example 3 Use the revised Gram–Schmidt algorithm to construct an orthonormal set of vectors from the linearly independent set $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$, where

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

Solution

FIRST ITERATION $(k = 1)$

$$r_{11} = \sqrt{\langle \mathbf{x}_1, \mathbf{x}_1 \rangle} = \sqrt{2},$$

$$\mathbf{q}_1 = \frac{1}{r_{11}}\mathbf{x}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix},$$

$$r_{12} = \langle \mathbf{x}_2, \mathbf{q}_1 \rangle = \frac{1}{\sqrt{2}},$$

$$r_{13} = \langle \mathbf{x}_3, \mathbf{q}_1 \rangle = \frac{1}{\sqrt{2}},$$

$$\mathbf{x}_2 \leftarrow \mathbf{x}_2 - r_{12}\mathbf{q}_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{\sqrt{2}}\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 1/2 \\ 1 \end{bmatrix},$$

$$\mathbf{x}_3 \leftarrow \mathbf{x}_3 - r_{13}\mathbf{q}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{\sqrt{2}}\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} = \begin{bmatrix} 1/2 \\ -1/2 \\ 1 \end{bmatrix}.$$

Note that both $\mathbf{x}_2$ and $\mathbf{x}_3$ are now orthogonal to $\mathbf{q}_1$.

SECOND ITERATION $(k = 2)$

Using vectors from the first iteration, we compute

$$r_{22} = \sqrt{\langle \mathbf{x}_2, \mathbf{x}_2 \rangle} = \sqrt{3/2},$$

$$\mathbf{q}_2 = \frac{1}{r_{22}}\mathbf{x}_2 = \frac{1}{\sqrt{3/2}}\begin{bmatrix} -1/2 \\ 1/2 \\ 1 \end{bmatrix} = \begin{bmatrix} -1/\sqrt{6} \\ 1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix},$$

$$r_{23} = \langle \mathbf{x}_3, \mathbf{q}_2 \rangle = \frac{1}{\sqrt{6}},$$

$$\mathbf{x}_3 \leftarrow \mathbf{x}_3 - r_{23}\mathbf{q}_2 = \begin{bmatrix} 1/2 \\ -1/2 \\ 1 \end{bmatrix} - \frac{1}{\sqrt{6}}\begin{bmatrix} -1/\sqrt{6} \\ 1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix} = \begin{bmatrix} 2/3 \\ -2/3 \\ 2/3 \end{bmatrix}.$$

THIRD ITERATION ($k = 3$)
Using vectors from the second iteration, we compute

$$r_{33} = \sqrt{\langle x_3, \mathbf{x}_3 \rangle} = \frac{2}{\sqrt{3}},$$

$$\mathbf{q}_3 = \frac{1}{r_{33}}\mathbf{x}_3 = \frac{1}{2/\sqrt{3}}\begin{bmatrix} 2/3 \\ -2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}.$$

The orthonormal set is $\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$. Compare with Example 1 of Section 10.2. ■

The revised Gram–Schmidt algorithm has two advantages over the Gram–Schmidt process developed in the previous section. First, it is less effected by roundoff errors, and second, the inverse process—recapturing the **x**-vectors from the **q**-vectors—becomes trivial. To understand this second advantage, let us redo Example 3 symbolically. In the first iteration, we calculated

$$\mathbf{q}_1 = \frac{1}{r_{11}}\mathbf{x}_1,$$

so, we immediately have,

$$\mathbf{x}_1 = r_{11}\mathbf{q}_1. \tag{10}$$

We then replaced $\mathbf{x}_2$ and $\mathbf{x}_3$ with vectors that were orthogonal to $\mathbf{q}_1$. If we denote these replacement vectors as $\mathbf{x}'_2$ and $\mathbf{x}'_3$, respectively, we have

$$\mathbf{x}'_2 = \mathbf{x}_2 - r_{12}\mathbf{q}_1 \text{ and } \mathbf{x}'_3 = \mathbf{x}_3 - r_{13}\mathbf{q}_1.$$

With the second iteration, we calculated

$$\mathbf{q}_2 = \frac{1}{r_{22}}\mathbf{x}'_2 = \frac{1}{r_{22}}(\mathbf{x}_2 - r_{12}\mathbf{q}_1).$$

Solving for $\mathbf{x}_2$, we get

$$\mathbf{x}_2 = r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2. \tag{11}$$

We then replaced $\mathbf{x}_3$ with a vector that was orthogonal to $\mathbf{q}_2$. If we denote this replacement vector as $\mathbf{x}''_3$, we have

$$\mathbf{x}''_3 = \mathbf{x}'_3 - r_{23}\mathbf{q}_2 = \left(\mathbf{x}_3 - r_{13}\mathbf{q}_1\right) - r_{23}\mathbf{q}_2.$$

With the third iteration, we calculated

$$\mathbf{q}_3 = \frac{1}{r_{33}}\mathbf{x}''_3 = \frac{1}{r_{33}}(\mathbf{x}_3 - r_{13}\mathbf{q}_1 - r_{23}\mathbf{q}_2).$$

Solving for $\mathbf{x}_3$, we obtain

$$\mathbf{x}_3 = r_{13}\mathbf{q}_1 + r_{23}\mathbf{q}_2 + r_{33}\mathbf{q}_3. \tag{12}$$

Eqs. (10) through (12) form a pattern that is easily extended. Begin with linearly independent vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, and use the revised Gram–Schmidt algorithm to form $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$. Then, for any $k(k = 1, 2, \ldots, n)$.

$$\mathbf{x}_k = r_{1k}\mathbf{q}_1 + r_{2k}\mathbf{q}_2 + r_{3k}\mathbf{q}_3 + \cdots + r_{kk}\mathbf{q}_k.$$

If we set $\mathbf{X} = [\mathbf{x}_1 \ \ \mathbf{x}_2 \ \ \ldots \ \ \mathbf{x}_n]$,

$$\mathbf{Q} = [\mathbf{q}_1 \ \ \mathbf{q}_2 \ \ \cdots \ \ \mathbf{q}_n] \tag{13}$$

and

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & \ldots & r_{1n} \\ 0 & r_{22} & r_{23} & \cdots & r_{2n} \\ 0 & 0 & r_{33} & \cdots & r_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & r_{nn} \end{bmatrix}; \tag{14}$$

we have the matrix representation

$$\mathbf{X} = \mathbf{QR},$$

which is known as the **QR**-decomposition of the matrix **X**. The columns of **Q** form an orthonormal set of column vectors, and **R** is upper (or right) triangular.

In general, we are given a matrix **X** and are asked to generate its **QR**-decomposition. This is accomplished by applying the revised Gram–Schmidt algorithm to the columns of **X**, providing those columns are linearly independent. Then Eqs. (13) and (14) yield the desired factorization.

Example 4 Construct a **QR**-decomposition for

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

Solution The columns of **X** are the vectors $\mathbf{x}_1, \mathbf{x}_2$, and $\mathbf{x}_3$ of Example 3. Using the results of that problem, we generate

$$\mathbf{Q} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & 1/\sqrt{6} & -1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix} \quad \text{and} \quad \mathbf{R} = \begin{bmatrix} \sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & \sqrt{3/2} & 1/\sqrt{6} \\ 0 & 0 & 2/\sqrt{3} \end{bmatrix}. \quad \blacksquare$$

Example 5 Construct a **QR**-decomposition for

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}.$$

Solution The columns of **X** are the vectors

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

We apply the revised Gram–Schmidt algorithm to these vectors. Carrying eight significant figures through all computations but rounding to four decimals for presentation purposes, we get

First Iteration ($k = 1$)

$$r_{11} = \sqrt{\langle \mathbf{x}_1, \mathbf{x}_1 \rangle} = \sqrt{3} = 1.7321,$$

$$\mathbf{q}_1 = \frac{1}{r_{11}}\mathbf{x}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5774 \\ 0.5774 \\ 0.0000 \\ 0.5774 \end{bmatrix},$$

$$r_{12} = \langle \mathbf{x}_2, \mathbf{q}_1 \rangle = 1.7321,$$

$$r_{13} = \langle \mathbf{x}_3, \mathbf{q}_1 \rangle = 1.1547,$$

$$r_{14} = \langle \mathbf{x}_4, \mathbf{q}_1 \rangle = 1.1547,$$

$$\mathbf{x}_2 \leftarrow \mathbf{x}_2 - r_{12}\mathbf{q}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix} - 1.7321 \begin{bmatrix} 0.5774 \\ 0.5774 \\ 0.0000 \\ 0.5774 \end{bmatrix} = \begin{bmatrix} 0.0000 \\ 1.0000 \\ 1.0000 \\ -1.0000 \end{bmatrix},$$

$$\mathbf{x}_3 \leftarrow \mathbf{x}_3 - r_{13}\mathbf{q}_1 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \end{bmatrix} - 1.1547 \begin{bmatrix} 0.5774 \\ 0.5774 \\ 0.0000 \\ 0.5774 \end{bmatrix} = \begin{bmatrix} -0.6667 \\ 0.3333 \\ 2.0000 \\ 0.3333 \end{bmatrix},$$

$$\mathbf{x}_4 \leftarrow \mathbf{x}_4 - r_{14}\mathbf{q}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} - 1.1547 \begin{bmatrix} 0.5774 \\ 0.5774 \\ 0.0000 \\ 0.5774 \end{bmatrix} = \begin{bmatrix} 0.3333 \\ -0.6667 \\ 1.0000 \\ 0.3333 \end{bmatrix}.$$

SECOND ITERATION ($k = 2$)
Using vectors from the first iteration, we compute

$$r_{22} = \sqrt{\langle \mathbf{x}_2, \mathbf{x}_2 \rangle} = 1.7321,$$

$$\mathbf{q}_2 = \frac{1}{r_{22}}\mathbf{x}_2 = \frac{1}{1.7321}\begin{bmatrix} 0.0000 \\ 1.0000 \\ 1.0000 \\ -1.0000 \end{bmatrix} = \begin{bmatrix} 0.0000 \\ 0.5774 \\ 0.5774 \\ -0.5774 \end{bmatrix},$$

$$r_{23} = \langle \mathbf{x}_3, \mathbf{q}_2 \rangle = 1.1547,$$

$$r_{24} = \langle \mathbf{x}_4, \mathbf{q}_2 \rangle = 0.0000,$$

$$\mathbf{x}_3 \leftarrow \mathbf{x}_3 - r_{23}\mathbf{q}_2 = \begin{bmatrix} -0.6667 \\ 0.3333 \\ 2.0000 \\ 0.3333 \end{bmatrix} - 1.1547 \begin{bmatrix} 0.0000 \\ 0.5774 \\ 0.5774 \\ -0.5774 \end{bmatrix} = \begin{bmatrix} -0.6667 \\ -0.3333 \\ 1.3333 \\ 1.0000 \end{bmatrix},$$

$$\mathbf{x}_4 \leftarrow \mathbf{x}_4 - r_{24}\mathbf{q}_2 = \begin{bmatrix} 0.3333 \\ -0.6667 \\ 1.0000 \\ 0.3333 \end{bmatrix} - 0.0000 \begin{bmatrix} 0.0000 \\ 0.5774 \\ 0.5774 \\ -0.5774 \end{bmatrix} = \begin{bmatrix} 0.3333 \\ -0.6667 \\ 1.0000 \\ 0.3333 \end{bmatrix}.$$

THIRD ITERATION ($k = 3$)
Using vectors from the second iteration, we compute

$$r_{33} = \sqrt{\langle \mathbf{x}_3, \mathbf{x}_3 \rangle} = 1.8257,$$

$$\mathbf{q}_3 = \frac{1}{r_{33}}\mathbf{x}_3 = \frac{1}{1.8257}\begin{bmatrix} -0.6667 \\ -0.3333 \\ 1.3333 \\ 1.0000 \end{bmatrix} = \begin{bmatrix} -0.3651 \\ -0.1826 \\ 0.7303 \\ 0.5477 \end{bmatrix},$$

$$r_{34} = \langle \mathbf{x}_4, \mathbf{q}_3 \rangle = 0.9129,$$

$$\mathbf{x}_4 \leftarrow \mathbf{x}_4 - r_{34}\mathbf{q}_3 = \begin{bmatrix} 0.3333 \\ -0.6667 \\ 1.0000 \\ 0.3333 \end{bmatrix} - 0.9129 \begin{bmatrix} -0.3651 \\ -0.1826 \\ 0.7303 \\ 0.5477 \end{bmatrix} = \begin{bmatrix} 0.6667 \\ -0.5000 \\ 0.3333 \\ -0.1667 \end{bmatrix}.$$

FOURTH ITERATION ($k = 4$)
Using vectors from the third iteration, we compute

$$r_{44} = \sqrt{\langle \mathbf{x}_4, \mathbf{x}_4 \rangle} = 0.9129,$$

$$\mathbf{q}_4 = \frac{1}{r_{44}}\mathbf{x}_4 = \frac{1}{0.9129}\begin{bmatrix} 0.6667 \\ -0.5000 \\ 0.3333 \\ -0.1667 \end{bmatrix} = \begin{bmatrix} 0.7303 \\ -0.5477 \\ 0.3651 \\ -0.1826 \end{bmatrix}.$$

With these entries calculated (compare with Example 2 of Section 10.2), we form

$$\mathbf{Q} = \begin{bmatrix} 0.5774 & 0.0000 & -0.3651 & 0.7303 \\ 0.5774 & 0.5774 & -0.1826 & -0.5477 \\ 0.0000 & 0.5774 & 0.7303 & 0.3651 \\ 0.5774 & -0.5774 & 0.5477 & -0.1826 \end{bmatrix}$$

and

$$\mathbf{R} = \begin{bmatrix} 1.7321 & 1.7321 & 1.1547 & 1.1547 \\ 0 & 1.7321 & 1.1547 & 0.0000 \\ 0 & 0 & 1.8257 & 0.9129 \\ 0 & 0 & 0 & 0.9129 \end{bmatrix}. \quad \blacksquare$$

Finally, we note that in contrast to **LU**-decompositions, **QR**-decompositions are applicable to nonsquare matrices as well. In particular, if we consider a matrix containing just the first two columns of the matrix **X** in Example 5, and calculate $r_{11}, r_{12}, r_{22}, \mathbf{q}_1$, and $\mathbf{q}_2$ as we did there, we have the decomposition

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0.5774 & 0.0000 \\ 0.5774 & 0.5774 \\ 0.0000 & 0.5774 \\ 0.5774 & -0.5774 \end{bmatrix} \begin{bmatrix} 1.7321 & 1.7321 \\ 0 & 1.7321 \end{bmatrix}.$$

Problems 10.3

In Problems 1 through 10, determine the (a) the angle between the given vectors, (b) the projection of $\mathbf{x}_1$ onto $\mathbf{x}_2$, and (c) its orthogonal component.

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

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

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

4. $\mathbf{x}_1 = \begin{bmatrix} 4 \\ -1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ 8 \end{bmatrix}.$

5. $\mathbf{x}_1 = \begin{bmatrix} -7 \\ -2 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 2 \\ 9 \end{bmatrix}$. **6.** $\mathbf{x}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}$.

7. $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$. **8.** $\mathbf{x}_1 = \begin{bmatrix} 0 \\ 3 \\ 4 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 2 \\ 5 \\ 5 \end{bmatrix}$.

9. $\mathbf{x}_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}$. **10.** $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 1 \\ -2 \\ 0 \\ -1 \end{bmatrix}$.

In Problems 11 through 21, determine **QR**-decompositions for the given matrices.

11. $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. **12.** $\begin{bmatrix} 1 & 3 \\ 1 & 5 \end{bmatrix}$. **13.** $\begin{bmatrix} 3 & 3 \\ -2 & 3 \end{bmatrix}$.

14. $\begin{bmatrix} 1 & 2 \\ 2 & 2 \\ 2 & 1 \end{bmatrix}$. **15.** $\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 3 & 5 \end{bmatrix}$. **16.** $\begin{bmatrix} 3 & 1 \\ -2 & 1 \\ 1 & 1 \\ -1 & 1 \end{bmatrix}$.

17. $\begin{bmatrix} 2 & 0 & 2 \\ 1 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$. **18.** $\begin{bmatrix} 1 & 2 & 2 \\ 1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix}$. **19.** $\begin{bmatrix} 0 & 3 & 2 \\ 3 & 5 & 5 \\ 4 & 0 & 5 \end{bmatrix}$.

20. $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$. **21.** $\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & -1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$.

22. Show that

$$\left\| \frac{\langle \mathbf{x}, \mathbf{a} \rangle}{\langle \mathbf{a}, \mathbf{a} \rangle} \mathbf{a} \right\| = \|\mathbf{x}\| |\cos\theta|,$$

where θ is the angle between **x** and **a**.

23. Prove directly that

$$\mathbf{x} - \frac{\langle \mathbf{a}, \mathbf{x} \rangle}{\langle \mathbf{a}, \mathbf{a} \rangle} \mathbf{a}$$

is orthogonal to **a**.

24. Discuss what is likely to occur in a **QR**-decomposition if the columns are not linearly independent, and all calculations are rounded.

10.4 The QR-Algorithm

The **QR**-algorithm is one of the more powerful numerical methods developed for computing eigenvalues of real matrices. In contrast to the power methods described in Section 6.6, which converge only to a single dominant real eigenvalue of a matrix, the **QR**-algorithm generally locates all eigenvalues, both real and complex, regardless of multiplicity.

Although a proof of the **QR**-algorithm is beyond the scope of this book, the algorithm itself is deceptively simple. As its name suggests, the algorithm is based on **QR**-decompositions. Not surprisingly then, the algorithm involves numerous arithmetic calculations, making it unattractive for hand computations but ideal for implementation on a computer.

Like many numerical methods, the **QR**-algorithm is iterative. We begin with a square real matrix $\mathbf{A}_0$. To determine its eigenvalues, we create a sequence of new matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_{k-1}, \mathbf{A}_k, \ldots$, having the property that each new matrix has the same eigenvalues as $\mathbf{A}_0$, and that these eigenvalues become increasingly obvious as the sequence progresses. To calculate $\mathbf{A}_k$ $(k = 1, 2, 3, \ldots)$ once $\mathbf{A}_{k-1}$ is known, first construct a **QR**-decomposition of $\mathbf{A}_{k-1}$:

$$\mathbf{A}_{k-1} = \mathbf{Q}_{k-1}\mathbf{R}_{k-1},$$

and then reverse the order fo the product to define

$$\mathbf{A}_k = \mathbf{R}_{k-1}\mathbf{Q}_{k-1}.$$

It can be shown that each matrix in the sequence $\{\mathbf{A}_k\}$ $(k = 1, 2, 3, \ldots)$ has identical eigenvalues. For now, we just note that the sequence generally converges to one of the following two partitioned forms:

$$\left[\begin{array}{ccccc:c} & & \mathbf{S} & & & T \\ \hdashline 0 & 0 & 0 & \cdots & 0 & a \end{array}\right] \tag{15}$$

or

$$\left[\begin{array}{ccccc:cc} & & \mathbf{U} & & & \multicolumn{2}{c}{\mathbf{V}} \\ \hdashline 0 & 0 & 0 & \cdots & 0 & b & c \\ 0 & 0 & 0 & \cdots & 0 & d & e \end{array}\right]. \tag{16}$$

If matrix (15) occurs, then the element a is an eigenvalue, and the remaining eigenvalues are found by applying the **QR**-algorithm a new to the submatrix **S**. If, on the other hand, matrix (16) occurs, then two eigenvalues are determined by solving for the roots of the characteristic equation of the 2×2 matrix in the lower right partition, namely

$$\lambda^2 - (b + e)\lambda + (be - cd) = 0.$$

The remaining eigenvalues are found by applying the **QR**-algorithm anew to the submatrix **U**.

Convergence of the algorithm is accelerated by performing a shift at each iteration. If the orders of all matrices are $n \times n$, we denote the element in the (n, n)-position of the matrix $\mathbf{A}_{k-1}$ as w_{k-1}, and construct a **QR**-decomposition for the shifted matrix $\mathbf{A}_{k-1} - w_{k-1}\mathbf{I}$. That is,

$$\mathbf{A}_{k-1} - w_{k-1}\mathbf{I} = \mathbf{Q}_{k-1}\mathbf{R}_{k-1}. \tag{17}$$

We define

$$\mathbf{A}_k = \mathbf{R}_{k-1}\mathbf{Q}_{k-1} + w_{k-1}\mathbf{I}. \tag{18}$$

Example 1 Find the eigenvalues of

$$\mathbf{A}_0 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 18 & -1 & -7 \end{bmatrix}.$$

Solution Using the **QR**-algorithm with shifting, carrying all calculations to eight significant figures but rounding to four decimals for presentation, we compute

$$\mathbf{A}_0 - (-7)\mathbf{I} = \begin{bmatrix} 7 & 1 & 0 \\ 0 & 7 & 1 \\ 18 & -1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0.3624 & 0.1695 & -0.9165 \\ 0.0000 & 0.9833 & 0.1818 \\ 0.9320 & -0.0659 & 0.3564 \end{bmatrix} \begin{bmatrix} 19.3132 & -0.5696 & 0.0000 \\ 0.0000 & 7.1187 & 0.9833 \\ 0.0000 & 0.0000 & 0.1818 \end{bmatrix}$$

$$= \mathbf{Q}_0\mathbf{R}_0,$$

$$\mathbf{A}_1 = \mathbf{R}_0\mathbf{Q}_0 + (-7)\mathbf{I}$$

$$= \begin{bmatrix} 19.3132 & -0.5696 & 0.0000 \\ 0.0000 & 7.1187 & 0.9833 \\ 0.0000 & 0.0000 & 0.1818 \end{bmatrix} \begin{bmatrix} 0.3624 & 0.1695 & -0.9165 \\ 0.0000 & 0.9833 & 0.1818 \\ 0.9320 & -0.0659 & 0.3564 \end{bmatrix}$$

$$+ \begin{bmatrix} -7 & 0 & 0 \\ 0 & -7 & 0 \\ 0 & 0 & -7 \end{bmatrix}$$

$$= \begin{bmatrix} 0.0000 & 2.7130 & -17.8035 \\ 0.9165 & -0.0648 & 1.6449 \\ 0.1695 & -0.0120 & -6.9352 \end{bmatrix},$$

$$\mathbf{A}_1 - (-6.9352)\mathbf{I} = \begin{bmatrix} 6.9352 & 2.7130 & -17.8035 \\ 0.9165 & 6.8704 & 1.6449 \\ 0.1695 & -0.0120 & 0.0000 \end{bmatrix}$$

$$= \begin{bmatrix} 0.9911 & -0.1306 & -0.0260 \\ 0.1310 & 0.9913 & 0.0120 \\ 0.0242 & -0.0153 & 0.9996 \end{bmatrix} \begin{bmatrix} 6.9975 & 3.5884 & -17.4294 \\ 0.0000 & 6.4565 & 3.9562 \\ 0.0000 & 0.0000 & 0.4829 \end{bmatrix}$$

$$= \mathbf{Q}_1\mathbf{R}_1,$$

$$\mathbf{A}_2 = \mathbf{R}_1\mathbf{Q}_1 + (-6.9352)\mathbf{I} = \begin{bmatrix} 0.0478 & 2.9101 & -17.5612 \\ 0.9414 & -0.5954 & 4.0322 \\ 0.0117 & -0.0074 & -6.4525 \end{bmatrix}.$$

Continuing in this manner, we generate sequentially

$$\mathbf{A}_3 = \begin{bmatrix} 0.5511 & 2.7835 & -16.8072 \\ 0.7826 & -1.1455 & 6.5200 \\ 0.0001 & -0.0001 & -6.4056 \end{bmatrix}$$

and

$$\mathbf{A}_4 = \begin{bmatrix} 0.9259 & 2.5510 & -15.9729 \\ 0.5497 & -1.5207 & 8.3583 \\ 0.0000 & -0.0000 & -6.4051 \end{bmatrix}.$$

$\mathbf{A}_4$ has form (15) with

$$\mathbf{S} = \begin{bmatrix} 0.9259 & 2.5510 \\ 0.5497 & -1.5207 \end{bmatrix} \quad \text{and} \quad a = -6.4051.$$

One eigenvalue is -6.4051, which is identical to the value obtained in Example 2 of Section 6.6. In addition, the characteristic equation of $\mathbf{R}$ is $\lambda^2 + 0.5948\lambda - 2.8103 = 0$, which admits both -2 and 1.4052 as roots. These are the other two eigenvalues of $\mathbf{A}_0$. ■

Example 2 Find the eigenvalues of

$$\mathbf{A}_0 = \begin{bmatrix} 0 & 0 & 0 & -25 \\ 1 & 0 & 0 & 30 \\ 0 & 1 & 0 & -18 \\ 0 & 0 & 1 & 6 \end{bmatrix}.$$

Solution Using the **QR**-algorithm with shifting, carrying all calculations to eight significant figures but rounding to four decimals for presentation, we compute

$$\mathbf{A}_0 - (6)\mathbf{I} = \begin{bmatrix} -6 & 0 & 0 & -25 \\ 1 & -6 & 0 & 30 \\ 0 & 1 & -6 & -18 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -0.9864 & -0.1621 & -0.0270 & -0.0046 \\ 0.1644 & -0.9726 & -0.1620 & -0.0274 \\ 0.0000 & 0.1666 & -0.9722 & -0.1643 \\ 0.0000 & 0.0000 & 0.1667 & -0.9860 \end{bmatrix}$$

$$\times \begin{bmatrix} 6.0828 & -0.9864 & 0.0000 & -29.5918 \\ 0.0000 & 6.0023 & -0.9996 & -28.1246 \\ 0.0000 & 0.0000 & 6.0001 & 13.3142 \\ 0.0000 & 0.0000 & 0.0000 & 2.2505 \end{bmatrix}$$

$$= \mathbf{Q}_0\mathbf{R}_0,$$

$$\mathbf{A}_1 = \mathbf{R}_0\mathbf{Q}_0 + (6)\mathbf{I} = \begin{bmatrix} -0.1622 & -0.0266 & 4.9275 & -29.1787 \\ 0.9868 & -0.0044 & -4.6881 & 27.7311 \\ 0.0000 & 0.9996 & 2.3856 & -14.1140 \\ 0.0000 & 0.0000 & 0.3751 & 3.7810 \end{bmatrix},$$

$$\mathbf{A}_1 - (3.7810)\mathbf{I} = \begin{bmatrix} -3.9432 & -0.0266 & 4.9275 & -29.1787 \\ 0.9868 & -3.7854 & -4.6881 & 27.7311 \\ 0.0000 & 0.9996 & -1.3954 & -14.1140 \\ 0.0000 & 0.0000 & 0.3751 & 0.0000 \end{bmatrix}$$

$$= \begin{bmatrix} -0.9701 & -0.2343 & -0.0628 & -0.0106 \\ 0.2428 & -0.9361 & -0.2509 & -0.0423 \\ 0.0000 & 0.2622 & -0.9516 & -0.1604 \\ 0.0000 & 0.0000 & 0.1662 & -0.9861 \end{bmatrix}$$

$$\times \begin{bmatrix} 4.0647 & -0.8931 & -5.9182 & 35.0379 \\ 0.0000 & 3.8120 & 2.8684 & -22.8257 \\ 0.0000 & 0.0000 & 2.2569 & 8.3060 \\ 0.0000 & 0.0000 & 0.0000 & 1.3998 \end{bmatrix}$$

$$= \mathbf{Q}_1\mathbf{R}_1,$$

$$\mathbf{A}_2 = \mathbf{R}_1\mathbf{Q}_1 + (3.7810)\mathbf{I} = \begin{bmatrix} -0.3790 & -1.6681 & 11.4235 & -33.6068 \\ 0.9254 & 0.9646 & -7.4792 & 21.8871 \\ 0.0000 & 0.5918 & 3.0137 & -8.5524 \\ 0.0000 & 0.0000 & 0.2326 & 2.4006 \end{bmatrix}.$$

Continuing in this manner, we generate, after 25 iterations,

$$\mathbf{A}_{25} = \begin{bmatrix} 4.8641 & -4.4404 & 18.1956 & -28.7675 \\ 4.2635 & -2.8641 & 13.3357 & -21.3371 \\ 0.0000 & 0.0000 & 2.7641 & -4.1438 \\ 0.0000 & 0.0000 & 0.3822 & 1.2359 \end{bmatrix},$$

which has form (16) with

$$\mathbf{U} = \begin{bmatrix} 4.8641 & -4.4404 \\ 4.2635 & -2.8641 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} b & c \\ d & e \end{bmatrix} = \begin{bmatrix} 2.7641 & -4.1438 \\ 0.3822 & 1.2359 \end{bmatrix}.$$

The characteristic equation of $\mathbf{U}$ is $\lambda^2 - 2\lambda + 5 = 0$, which has as its roots $1 \pm 2i$; the characteristic equation of the other 2×2 matrix is $\lambda^2 - 4\lambda + 4.9999 = 0$, which has as its roots $2 \pm i$. These roots are the four eigenvalues of $\mathbf{A}_0$. ■

Problems 10.4

1. Use one iteration of the **QR**-algorithm to calculate $\mathbf{A}_1$ when

$$\mathbf{A}_0 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 18 & -1 & 7 \end{bmatrix}.$$

Note that this matrix differs from the one in Example 1 by a single sign.

2. Use one iteration of the **QR**-algorithm to calculate $\mathbf{A}_1$ when

$$\mathbf{A}_0 = \begin{bmatrix} 2 & -17 & 7 \\ -17 & -4 & 1 \\ 7 & 1 & -14 \end{bmatrix}.$$

3. Use one iteration of the **QR**-algorithm to calculate $\mathbf{A}_1$ when

$$\mathbf{A}_0 = \begin{bmatrix} 0 & 0 & 0 & -13 \\ 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & -14 \\ 0 & 0 & 1 & 4 \end{bmatrix}.$$

In Problems 4 through 14, use the **QR**-algorithm to calculate the eigenvalues of the given matrices:

4. The matrix defined in Problem 1.

5. The matrix defined in Problem 2.

6. $\begin{bmatrix} 3 & 0 & 0 \\ 2 & 6 & 4 \\ 2 & 3 & 5 \end{bmatrix}$. **7.** $\begin{bmatrix} 7 & 2 & 0 \\ 2 & 1 & 6 \\ 0 & 6 & 7 \end{bmatrix}$. **8.** $\begin{bmatrix} 3 & 2 & 3 \\ 2 & 6 & 6 \\ 3 & 6 & 11 \end{bmatrix}$.

9. $\begin{bmatrix} 2 & 0 & -1 \\ 2 & 3 & 2 \\ -1 & 0 & 2 \end{bmatrix}$. **10.** $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 5 & -9 & 6 \end{bmatrix}$. **11.** $\begin{bmatrix} 3 & 0 & 5 \\ 1 & 1 & 1 \\ -2 & 0 & -3 \end{bmatrix}$.

12. The matrix in Problem 3.

13. $\begin{bmatrix} 0 & 3 & 2 & -1 \\ 1 & 0 & 2 & -3 \\ 3 & 1 & 0 & -1 \\ 2 & -2 & 1 & 1 \end{bmatrix}$. **14.** $\begin{bmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{bmatrix}$.

10.5 Least-Squares

Analyzing data for forecasting and predicting future events is common to business, engineering, and the sciences, both physical and social. If such data are plotted, as in Figure 10.4, they constitute a *scatter diagram*, which may provide insight into the underlying relationship between system variables. For example, the data in Figure 10.4 appears to follow a straight line relationship reasonably well. The problem then is to determine the equation of the straight line that best fits the data.

A straight line in the variables x and y having the equation

$$y = mx + c, \tag{19}$$

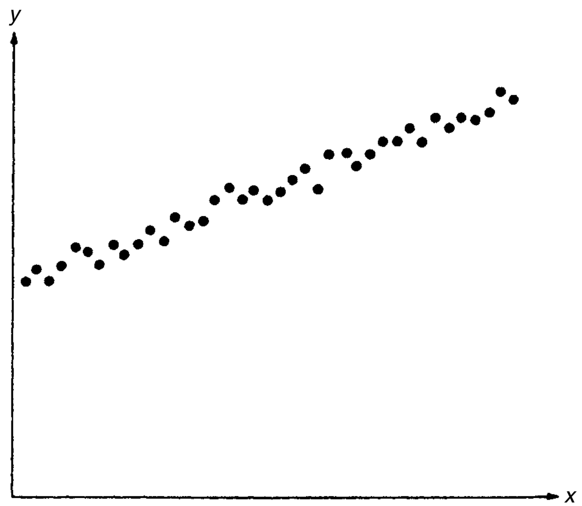

Figure 10.4

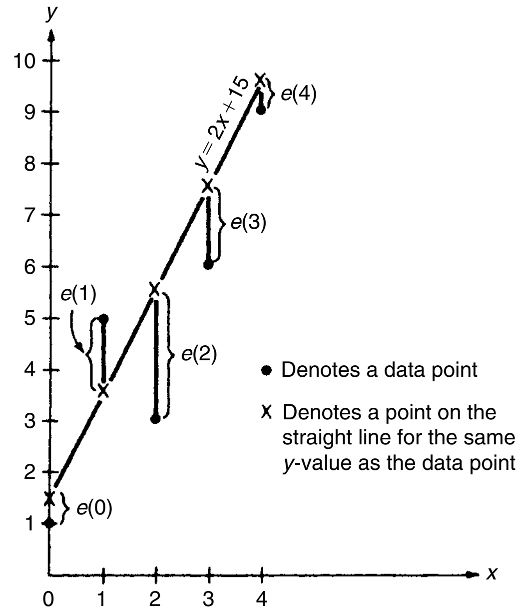

Figure 10.5

where m and c are constants, will have one y-value on the line for each value of x. This y-value may or may not agree with the data at the same value of x. Thus, for values of x at which data are available, we generally have two values of y, one value from the data and a second value from the straight line approximation to the data. This situation is illustrated in Figure 10.5. The error at each x, designated as $e(x)$, is the difference between the y-value of the data and the y-value obtained from the straight-line approximation.

Example 1 Calculate the errors made in approximating the data given in Figure 10.5 by the line $y = 2x + 1.5$.

Solution The line and the given data points are plotted in Figure 10.5. There are errors at $x = 0, x = 1, x = 2, x = 3,$ and $x = 4$. Evaluating the equation $y = 2x + 1.5$ at these values of x, we compute Table 10.1.

It now follows that

$$e(0) = 1 - 1.5 = -0.5,$$

$$e(1) = 5 - 3.5 = 1.5,$$

$$e(2) = 3 - 5.5 = -2.5,$$

$$e(3) = 6 - 7.5 = -1.5,$$

Table 10.1

Given data		Evaluated from $y = 2x + 1.5$
x	y	y
0	1	1.5
1	5	3.5
2	3	5.5
3	6	7.5
4	9	9.5

and

$$e(4) = 9 - 9.5 = -0.5.$$

Note that these errors could have been read directly from the graph. ■

We can extend this concept of error to the more general situation involving N data points. Let $(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_N, y_N)$ be a set of N data points for a particular situation. Any straight-line approximation to this data generates errors $e(x_1),\ e(x_2),\ e(x_3), \ldots, e(x_N)$ which individually can be positive, negative, or zero. The latter case occurs when the approximation agrees with the data at a particular point. We define the overall error as follows.

Definition 1 The *least-squares error* E is the sum of the squares of the individual errors. That is,

$$E = [e(x_1)]^2 + [e(x_2)]^2 + [e(x_3)]^2 + \cdots + [e(x_N)]^2.$$

The only way the total error E can be zero is for each of the individual errors to be zero. Since each term of E is squared, an equal number of positive and negative individual errors cannot sum to zero.

Example 2 Compute the least-squares error for the approximation used in Example 1.

Solution

$$\begin{aligned} E &= [e(0)]^2 + [e(1)]^2 + [e(2)]^2 + [e(3)]^2 + [e(4)]^2 \\ &= (-0.5)^2 + (1.5)^2 + (-2.5)^2 + (-1.5)^2 + (-0.5)^2 \\ &= 0.25 + 2.25 + 6.25 + 2.25 + 0.25 \\ &= 11.25. \end{aligned}$$ ■

Definition 2 The *least-squares straight line* is the line that minimizes the least-squares error.

We seek values of m and c in (19) that minimize the least-squares error. For such a line,

$$e(x_i) = y_i - (mx_i + c),$$

so we want the values for m and c that minimize

$$E = \sum_{i=1}^{N} (y_i - mx_i - c)^2.$$

This occurs when

$$\frac{\partial E}{\partial m} = \sum_{i=1}^{N} 2(y_i - mx_i - c)(-x_i) = 0$$

and

$$\frac{\partial E}{\partial c} = \sum_{i=1}^{N} 2(y_i - mx_i - c)(-1) = 0,$$

or, upon simplifying, when

$$\left(\sum_{i=1}^{N} x_i^2\right) m + \left(\sum_{i=1}^{N} x_i\right) c = \sum_{i=1}^{N} x_i y_i, \tag{20}$$

$$\left(\sum_{i=1}^{N} x_i\right) m + Nc = \sum_{i=1}^{N} y_i.$$

System (20) makes up the *normal equations* for a least-squares fit in two variables.

Example 3 Find the least-squares straight line for the following $x - y$ data:

x	0	1	2	3	4
y	1	5	3	6	9

Solution Table 10.2 contains the required summations.

For this data, the normal equations become

$$30m + 10c = 65,$$

$$10m + 5c = 24,$$

which has as its solution $m = 1.7$ and $c = 1.4$. The least-squares straight line is $y = 1.7x + 1.4$.

Table 10.2

	x_i	y_i	$(x_i)^2$	$x_i y_i$
	0	1	0	0
	1	5	1	5
	2	3	4	6
	3	6	9	18
	4	9	16	36
Sum	$\sum_{i=1}^{5} x_i = 10$	$\sum_{i=1}^{5} y_i = 24$	$\sum_{i=1}^{5} (x_i)^2 = 30$	$\sum_{i=1}^{5} x_i y_i = 65$

■

The normal equations have a simple matrix representation. Ideally, we would like to choose m and c for (19) so that

$$y_i = mx_i + c$$

for all data pairs $(x_i, y_i),\ i = 1, 2, \ldots, N$. That is, we want the constants m and c to solve the system

$$\begin{aligned} mx_1 + c &= y_1, \\ mx_2 + c &= y_2, \\ mx_3 + c &= y_3, \\ &\vdots \\ mx_N + c &= y_N, \end{aligned}$$

or, equivalently, the matrix equation

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{bmatrix}.$$

This system has the standard form $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is defined as a matrix having two columns, the first being the data vector $\begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_N \end{bmatrix}^{\mathsf{T}}$, and the second containing all ones, $\mathbf{x} = \begin{bmatrix} m & c \end{bmatrix}^{\mathsf{T}}$, and $\mathbf{b}$ is the data vector $\begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_N \end{bmatrix}^{\mathsf{T}}$. In this context, $\mathbf{Ax} = \mathbf{b}$ has a solution for $\mathbf{x}$ if and only if the data falls on a straight line. If not, then the matrix system is inconsistent, and we seek the least-squares solution. That is, we seek the vector $\mathbf{x}$ that minimizes the least-squares error as stipulated in Definition 2, having the matrix form

$$E = \|\mathbf{Ax} - \mathbf{b}\|^2. \tag{21}$$

The solution is the vector $\mathbf{x}$ satisfying the normal equations, which take the matrix form

$$\mathbf{A}^{\mathsf{T}}\mathbf{A}\mathbf{x} = \mathbf{A}^{\mathsf{T}}\mathbf{b}. \tag{22}$$

System (22) is identical to system (20) when $\mathbf{A}$ and $\mathbf{b}$ are defined as above.

We now generalize to all linear systems of the form $\mathbf{Ax} = \mathbf{b}$. We are primarily interested in cases where the system is inconsistent (rendering the methods developed in Chapter 2 useless), ands this generally occurs when $\mathbf{A}$ has more rows than columns. We shall place no restrictions on the number of columns in $\mathbf{A}$, but we will assume that *the columns are linearly independent*. We seek the vector $\mathbf{x}$ that minimizes the least-squares error defined by Eq. (21).

Theorem 1 *If* $\mathbf{x}$ *has the property that* $\mathbf{Ax} - \mathbf{b}$ *is orthogonal to the columns of* $\mathbf{A}$, *then* $\mathbf{x}$ *minimizes* $\|\mathbf{Ax} - \mathbf{b}\|^2$.

Proof. For any vector $\mathbf{x}_0$ of appropriate dimension,

$$\begin{aligned}
\|\mathbf{Ax}_0 - \mathbf{b}\|^2 &= \|(\mathbf{Ax}_0 - \mathbf{Ax}) + (\mathbf{Ax} - \mathbf{b})\|^2 \\
&= \langle(\mathbf{Ax}_0 - \mathbf{Ax}) + (\mathbf{Ax} - \mathbf{b})\,,\,(\mathbf{Ax}_0 - \mathbf{Ax}) + (\mathbf{Ax} - \mathbf{b})\rangle \\
&= \langle(\mathbf{Ax}_0 - \mathbf{Ax})\,,\,(\mathbf{Ax}_0 - \mathbf{Ax})\rangle + \langle(\mathbf{Ax} - \mathbf{b})\,,\,(\mathbf{Ax} - \mathbf{b})\rangle \\
&= +2\,\langle(\mathbf{Ax}_0 - \mathbf{Ax})\,,\,(\mathbf{Ax} - \mathbf{b})\rangle \\
&= \|(\mathbf{Ax}_0 - \mathbf{Ax})\|^2 + \|(\mathbf{Ax} - \mathbf{b})\|^2 \\
&= +2\,\langle\mathbf{Ax}_0, (\mathbf{Ax} - \mathbf{b})\rangle - 2\,\langle\mathbf{Ax}, (\mathbf{Ax} - \mathbf{b})\rangle\,.
\end{aligned}$$

It follows directly from Problem 28 of Section 10.2 that the last two inner products are both zero (take $\mathbf{p} = \mathbf{Ax} - \mathbf{b}$). Therefore,

$$\begin{aligned}
\|\mathbf{Ax}_0 - \mathbf{b}\|^2 &= \|(\mathbf{Ax}_0 - \mathbf{Ax})\|^2 + \|(\mathbf{Ax} - \mathbf{b})\|^2 \\
&\geq \|(\mathbf{Ax} - \mathbf{b})\|^2,
\end{aligned}$$

and $\mathbf{x}$ minimizes Eq. (21). □

As a consequence of Theorem 1, we seek a vector $\mathbf{x}$ having the property that $\mathbf{Ax} - \mathbf{b}$ is orthogonal to the columns of $\mathbf{A}$. Denoting the columns of $\mathbf{A}$ as $\mathbf{A_1}, \mathbf{A_2}, \ldots, \mathbf{A}_n$, respectively, we require

$$\langle\mathbf{A}_i,\ \mathbf{Ax} - \mathbf{b}\rangle = 0 \quad (i = 1, 2, \ldots, n).$$

If $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^{\mathsf{T}}$ denotes an arbitrary vector of appropriate dimension, then

$$\mathbf{Ay} = \mathbf{A}_1 y_1 + \mathbf{A}_2 y_2 + \cdots + \mathbf{A}_n y_n,$$

and

$$\begin{aligned}\langle \mathbf{Ay},\ (\mathbf{Ax}-\mathbf{b})\rangle &= \left\langle \sum_{i=1}^{n} \mathbf{A}_i y_i,\ (\mathbf{Ax}-\mathbf{b})\right\rangle \\ &= \sum_{i=1}^{n} \langle \mathbf{A}_i y_i,\ (\mathbf{Ax}-\mathbf{b})\rangle \\ &= \sum_{i=1}^{n} y_i \langle \mathbf{A_i},\ (\mathbf{Ax}-\mathbf{b})\rangle \\ &= 0.\end{aligned} \tag{23}$$

It also follows from Problem 39 of Section 6.1 that

$$\langle \mathbf{Ay}, (\mathbf{Ax}-\mathbf{b})\rangle = \langle \mathbf{y},\ \mathbf{A}^{\mathsf{T}}(\mathbf{Ax}-\mathbf{b})\rangle = \langle \mathbf{y},\ (\mathbf{A}^{\mathsf{T}}\mathbf{Ax}-\mathbf{A}^{\mathsf{T}}\mathbf{b})\rangle. \tag{24}$$

Eqs. (23) and (24) imply that $\langle \mathbf{y}, (\mathbf{A}^{\mathsf{T}}\mathbf{Ax}-\mathbf{A}^{\mathsf{T}}\mathbf{b})\rangle = 0$ for any $\mathbf{y}$. We may deduce from Problem 26 of Section 10.2 that $\mathbf{A}^{\mathsf{T}}\mathbf{Ax}-\mathbf{A}^{\mathsf{T}}\mathbf{b} = 0$, or $\mathbf{A}^{\mathsf{T}}\mathbf{Ax} = \mathbf{A}^{\mathsf{T}}\mathbf{b}$, *which has the same form as Eq*. (22)! Therefore, a vector $\mathbf{x}$ is the least-squares solution to $\mathbf{Ax} = \mathbf{b}$ if and only if it is the solution to $\mathbf{A}^{\mathsf{T}}\mathbf{Ax} = \mathbf{A}^{\mathsf{T}}\mathbf{b}$. This set of normal equations is guaranteed to have a unique solution whenever the columns of $\mathbf{A}$ are linearly independent, *and it may be solved using any of the methods described in the previous chapters*!

Example 4 Find the least-squares solution to

$$\begin{aligned} x + 2y + z &= 1,\\ 3x - y \phantom{{}+2z} &= 2,\\ 2x + y - z &= 2,\\ x + 2y + 2z &= 1.\end{aligned}$$

Solution This system takes the matrix form $\mathbf{Ax} = \mathbf{b}$, with

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 1\\ 3 & -1 & 0\\ 2 & 1 & -1\\ 1 & 2 & 2\end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x\\ y\\ z\end{bmatrix}, \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 1\\ 2\\ 2\\ 1\end{bmatrix}.$$

Then,

$$\mathbf{A}^{\mathsf{T}}\mathbf{A} = \begin{bmatrix} 15 & 3 & 1\\ 3 & 10 & 5\\ 1 & 5 & 6\end{bmatrix} \quad \text{and} \quad \mathbf{A}^{\mathsf{T}}\mathbf{b} = \begin{bmatrix} 12\\ 4\\ 1\end{bmatrix},$$

and the normal equations become

$$\begin{bmatrix} 15 & 3 & 1 \\ 3 & 10 & 5 \\ 1 & 5 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 12 \\ 4 \\ 1 \end{bmatrix}.$$

Using Gaussian elimination, we obtain as the unique solution to this set of equations $x = 0.7597$, $y = 0.2607$, and $z = -0.1772$, which is also the least-squares solution to the original system. ■

Example 5 Find the least-squares solution to

$$\begin{aligned} 0x + 3y &= 80, \\ 2x + 5y &= 100, \\ 5x - 2y &= 60, \\ -x + 8y &= 130, \\ 10x - y &= 150. \end{aligned}$$

Solution This system takes the matrix form $\mathbf{Ax} = \mathbf{b}$, with

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 5 \\ 5 & -2 \\ -1 & 8 \\ 10 & -1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 80 \\ 100 \\ 60 \\ 130 \\ 150 \end{bmatrix}.$$

Then,

$$\mathbf{A}^{\mathsf{T}}\mathbf{A} = \begin{bmatrix} 131 & -15 \\ -15 & 103 \end{bmatrix} \quad \text{and} \quad \mathbf{A}^{\mathsf{T}}\mathbf{b} = \begin{bmatrix} 1950 \\ 1510 \end{bmatrix},$$

and the normal equations become

$$\begin{bmatrix} 131 & -15 \\ -15 & 103 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1950 \\ 1510 \end{bmatrix}.$$

The unique solution to this set of equations is $x = 16.8450$, and $y = 17.1134$, rounded to four decimals, which is also the least-squares solution to the original system. ■

Problems 10.5

In Problems 1 through 8, find the least-squares solution to the given systems of equations:

1. $\begin{aligned} 2x + 3y &= 8, \\ 3x - y &= 5, \\ x + y &= 6. \end{aligned}$

2. $\begin{aligned} 2x + y &= 8, \\ x + y &= 4, \\ -x + y &= 0, \\ 3x + y &= 13. \end{aligned}$

3. $\begin{aligned} x + 3y &= 65, \\ 2x - y &= 0, \\ 3x + y &= 50, \\ 2x + 2y &= 55. \end{aligned}$

4. $\begin{aligned} 2x + y &= 6, \\ x + y &= 8, \\ -2x + y &= 11, \\ -x + y &= 8, \\ 3x + y &= 4. \end{aligned}$

5. $\begin{aligned} 2x + 3y - 4z &= 1, \\ x - 2y + 3z &= 3, \\ x + 4y + 2z &= 6, \\ 2x + y - 3z &= 1. \end{aligned}$

6. $\begin{aligned} 2x + 3y + 2z &= 25, \\ 2x - y + 3z &= 30, \\ 3x + 4y - 2z &= 20, \\ 3x + 5y + 4z &= 55. \end{aligned}$

7. $\begin{aligned} x + y - z &= 90, \\ 2x + y + z &= 200, \\ x + 2y + 2z &= 320, \\ 3x - 2y - 4z &= 10, \\ 3x + 2y - 3z &= 220. \end{aligned}$

8. $\begin{aligned} x + 2y + 2z &= 1, \\ 2x + 3y + 2z &= 2, \\ 2x + 4y + 4z &= -2, \\ 3x + 5y + 4z &= 1, \\ x + 3y + 2z &= -1. \end{aligned}$

9. Which of the systems, if any, given in Problems 1 through 8 represent a least-squares, straight line fit to data?

10. The monthly sales figures (in thousands of dollars) for a newly opened shoe store are:

month	1	2	3	4	5
sales	9	16	14	15	21

(a) Plot a scatter diagram for this data.

(b) Find the least-squares straight line that best fits this data.

(c) Use this line to predict sales revenue for month 6.

11. The number of new cars sold at a new car dealership over the first 8 weeks of the new season are:

week	1	2	3	4	5	6	7	8
sales	51	50	45	46	43	39	35	34

(a) Plot a scatter diagram for this data.

(b) Find the least-squares straight line that best fits this data.

(c) Use this line to predict sales for weeks 9 and 10.

12. Annual rainfall data (in inches) for a given town over the last seven years are:

year	1	2	3	4	5	6	7
rainfall	10.5	10.8	10.9	11.7	11.4	11.8	12.2

(a) Find the least-squares straight line that best fits this data.

(b) Use this line to predict next year's rainfall.

13. Solve system (20) algebraically and explain why the solution would be susceptible to round-off error.

14. **(Coding)** To minimize the round-off error associated with solving the normal equations for a least-squares straight line fit, the (x_i, y_i)-data are coded before using them in calculations. Each x_i-value is replaced by the difference between x_i and the average of all x_i-data. That is, if

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i, \text{ then set } x_i' = x_i - \overline{X},$$

and fit a straight line to the (x_i', y_i)-data instead.

Explain why this coding scheme avoids the round-off errors associated with uncoded data.

15. (a) Code the data given in Problem 10 using the procedure described in Problem 14.

(b) Find the least-squares straight line fit for this coded data.

16. (a) Code the data given in Problem 11 using the procedure described in Problem 14.

(b) Find the least-squares straight line fit for this coded data.

17. Census figures for the population (in millions of people) for a particular region of the country are as follows:

year	1950	1960	1970	1980	1990
population	25.3	23.5	20.6	18.7	17.8

(a) Code this data using the procedure described in Problem 14, and then find the least-squares straight line that best fits it.

(b) Use this line to predict the population in 2000.

18. Show that if $\mathbf{A} = \mathbf{QR}$ is a $\mathbf{QR}$-decomposition of $\mathbf{A}$, then the normal equations given by Eq. (22) can be written as $\mathbf{R}^{\mathsf{T}}\mathbf{R}\mathbf{x} = \mathbf{R}^{\mathsf{T}}\mathbf{Q}^{\mathsf{T}}\mathbf{b}$, which reduces to $\mathbf{R}\mathbf{x} = \mathbf{Q}^{\mathsf{T}}\mathbf{b}$. This is a numerically stable set of equations to solve, not subject to the same round-off errors associated with solving the normal equations directly.

19. Use the procedure described in Problem 18 to solve Problem 1.

20. Use the procedure described in Problem 18 to solve Problem 2.

21. Use the procedure described in Problem 18 to solve Problem 5.

22. Use the procedure described in Problem 18 to solve Problem 6.

23. Determine the error vector associated with the least-squares solution of Problem 1, and then calculate the inner product of this vector with each of the columns of the coefficient matrix associated with the given set of equations.

24. Determine the error vector associated with the least-squares solution of Problem 5, and then calculate the inner product of this vector with each of the columns of the coefficient matrix associated with the given set of equations.