$$A=\begin{bmatrix}-1 & 7 & -1\\ 0 & 1 & 0\\ 0 & 15 & -2\end{bmatrix}$$

Answer:

$$\begin{bmatrix}-1 & 10237 & -2047\\ 0 & 1 & 0\\ 0 & 10245 & -2048\end{bmatrix}$$

24. In each part, compute the stated power of

$$A=\begin{bmatrix}1 & -2 & 8\\ 0 & -1 & 0\\ 0 & 0 & -1\end{bmatrix}$$

(a) A^{1000} (b) A^{-1000} (c) A^{2301} (d) A^{-2301}

25. Find A^n if n is a positive integer and

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

Answer:

$$A^n=PD^nP^{-1}=\begin{bmatrix}1 & 1 & 1\\ 2 & 0 & -1\\ 1 & -1 & 1\end{bmatrix}\begin{bmatrix}1^n & 0 & 0\\ 0 & 3^n & 0\\ 0 & 0 & 4^n\end{bmatrix}\begin{bmatrix}\frac{1}{6} & \frac{1}{3} & \frac{1}{6}\\ \frac{1}{2} & 0 & -\frac{1}{2}\\ \frac{1}{3} & -\frac{1}{3} & \frac{1}{3}\end{bmatrix}$$

26. Let

$$A=\begin{bmatrix}a & b\\ c & d\end{bmatrix}$$

Show that

(a) A is diagonalizable if $(a-d)^2+4bc>0$.

(b) A is not diagonalizable if $(a-d)^2+4bc<0$.

[*Hint:* See Exercise 19 of Section 5.1.]

27. In the case where the matrix A in Exercise 26 is diagonalizable, find a matrix P that diagonalizes A. [*Hint:* See Exercise 20 of Section 5.1.]

Answer:

On possibility is $P=\begin{bmatrix}-b & -b\\ a-\lambda_1 & a-\lambda_2\end{bmatrix}$ where λ_1 and λ_2 are as in Exercise 20 of Section 5.1.

28. Prove that similar matrices have the same rank.

29. Prove that similar matrices have the same nullity.

30. Prove that similar matrices have the same trace.

31. Prove that if A is diagonalizable, then so is A^k for every positive integer k.

32. Prove that if A is a diagonalizable matrix, then the rank of A is the number of nonzero eigenvalues of A.

33. Suppose that the characteristic polynomial of some matrix A is found to be $p(\lambda) = (\lambda - 1)(\lambda - 3)^2(\lambda - 4)^3$. In each part, answer the question and explain your reasoning.

(a) What can you say about the dimensions of the eigenspaces of A?

(b) What can you say about the dimensions of the eigenspaces if you know that A is diagonalizable?

(c) If $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly independent set of eigenvectors of A all of which correspond to the same eigenvalue of A, what can you say about the eigenvalue?

Answer:

(a) $\lambda = 1$: dimension $= 1$; $\lambda = 3$: dimension ≤ 2; $\lambda = 4$: dimension ≤ 3

(b) Dimensions will be exactly 1, 2, and 3.

(c) $\lambda = 4$

34. This problem will lead you through a proof of the fact that the algebraic multiplicity of an eigenvalue of an $n \times n$ matrix A is greater than or equal to the geometric multiplicity. For this purpose, assume that λ_0 is an eigenvalue with geometric multiplicity k.

(a) Prove that there is a basis $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ for R^n in which the first k vectors of B form a basis for the eigenspace corresponding to λ_0.

(b) Let P be the matrix having the vectors in B as columns. Prove that the product AP can be expressed as

$$AP = P\begin{bmatrix} \lambda_0 I_k & X \\ 0 & Y \end{bmatrix}$$

[*Hint:* Compare the first k column vectors on both sides.]

(c) Use the result in part (b) to prove that A is similar to

$$C = \begin{bmatrix} \lambda_0 I_k & X \\ 0 & Y \end{bmatrix}$$

and hence that A and C have the same characteristic polynomial.

(d) By considering $\det(\lambda I - C)$, prove that the characteristic polynomial of C (and hence A) contains the factor $(\lambda - \lambda_0)$ at least k times, thereby proving that the algebraic multiplicity of λ_0 is greater than or equal to the geometric multiplicity k.

True-False Exercises

In parts (a)–(h) determine whether the statement is true or false, and justify your answer.

(a) Every square matrix is similar to itself.

Answer:

True

(b) If A, B, and C are matrices for which A is similar to B and B is similar to C, then A is similar to C.

Answer:

True

(c) If A and B are similar invertible matrices, then A^{-1} and B^{-1} are similar.

Answer:

True

(d) If A is diagonalizable, then there is a unique matrix P such that $P^{-1}AP$ is diagonal.

Answer:

False

(e) If A is diagonalizable and invertible, then A^{-1} is diagonalizable.

Answer:

True

(f) If A is diagonalizable, then A^T is diagonalizable.

Answer:

True

(g) If there is a basis for R^n consisting of eigenvectors of an $n \times n$ matrix A, then A is diagonalizable.

Answer:

True

(h) If every eigenvalue of a matrix A has algebraic multiplicity 1, then A is diagonalizable.

Answer:

True

Copyright © 2010 John Wiley & Sons, Inc. All rights reserved.

5.3 Complex Vector Spaces

Because the characteristic equation of any square matrix can have complex solutions, the notions of complex eigenvalues and eigenvectors arise naturally, even within the context of matrices with real entries. In this section we will discuss this idea and apply our results to study symmetric matrices in more detail. A review of the essentials of complex numbers appears in the back of this text.

Review of Complex Numbers

Recall that if $z = a + bi$ is a complex number, then:

- $\mathrm{Re}(z) = a$ and $\mathrm{Im}(z) = b$ are called the ***real part*** of z and the ***imaginary part*** of z, respectively,
- $|z| = \sqrt{a^2 + b^2}$ is called the ***modulus*** (or ***absolute value***) of z,
- $\bar{z} = a - bi$ is called the ***complex conjugate*** of z,
- $z\bar{z} = a^2 + b^2 = |z|^2$,
- the angle ϕ in Figure 5.3.1 is called an ***argument*** of z,
- $\mathrm{Re}(z) = |z| \cos \phi$
- $\mathrm{Im}(z) = |z| \sin \phi$
- $z = |z|(\cos \phi + i \sin \phi)$ is called the ***polar form*** of z.

Figure 5.3.1

Complex Eigenvalues

In Formula 3 of Section 5.1 we observed that the characteristic equation of a general $n \times n$ matrix A has the form

$$\lambda^n + c_1\lambda^{n-1} + \ldots + c_n = 0 \tag{1}$$

in which the highest power of λ has a coefficient of 1. Up to now we have limited our discussion to matrices in which the solutions of 1 are real numbers. However, it is possible for the characteristic equation of a matrix A with real entries to have imaginary solutions; for example, the characteristic equation of the matrix

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

is

$$\begin{vmatrix} \lambda + 2 & 1 \\ -5 & \lambda - 2 \end{vmatrix} = \lambda^2 + 1 = 0$$

which has the imaginary solutions $\lambda = i$ and $\lambda = -i$. To deal with this case we will need to explore the notion of a complex vector space and some related ideas.

Vectors in C^n

A vector space in which scalars are allowed to be complex numbers is called a ***complex vector space***. In this section we will be concerned only with the following complex generalization of the real vector space R^n.

DEFINITION 1

If n is a positive integer, then a ***complex n-tuple*** is a sequence of n complex numbers $(v_1, v_2, \ldots, v_n)$. The set of all complex n-tuples is called ***complex n-space*** and is denoted by C^n. Scalars are complex numbers, and the operations of addition, subtraction, and scalar multiplication are performed componentwise.

The terminology used for n-tuples of real numbers applies to complex n-tuples without change. Thus, if $v_1, v_2, \ldots, v_n$ are complex numbers, then we call $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ a ***vector*** in C^n and $v_1, v_2, \ldots, v_n$ its ***components***. Some examples of vectors in C^3 are

$$\mathbf{u} = (1+i,\ -4i,\ 3+2i), \quad \mathbf{v} = (0, i, 5), \quad \mathbf{w} = \left(6-\sqrt{2}i,\ 9+\frac{1}{2}i,\ \pi i\right)$$

Every vector

$$\mathbf{v} = (v_1, v_2, \ldots, v_n) = (a_1 + b_1 i, a_2 + b_2 i, \ldots, a_n + b_n i)$$

in C^n can be split into ***real*** and ***imaginary parts*** as

$$\mathbf{v} = (a_1, a_2, \ldots, a_n) + i(b_1, b_2, \ldots, b_n)$$

which we also denote as

$$\mathbf{v} = \mathrm{Re}(\mathbf{v}) + i\ \mathrm{Im}(\mathbf{v})$$

where

$$\mathrm{Re}(\mathbf{v}) = (a_1, a_2, \ldots, a_n) \qquad \text{and} \qquad \mathrm{Im}(\mathbf{v}) = (b_1, b_2, \ldots, b_n)$$

The vector

$$\overline{\mathbf{v}} = (\overline{v}_1, \overline{v}_2, \ldots, \overline{v}_n) = (a_1 - b_1 i, a_2 - b_2 i, \ldots, a_n - b_n i)$$

is called the ***complex conjugate*** of $\mathbf{v}$ and can be expressed in terms of $\mathrm{Re}(\mathbf{v})$ and $\mathrm{Im}(\mathbf{v})$ as

$$\overline{\mathbf{v}} = (a_1, a_2, \ldots, a_n) - i(b_1, b_2, \ldots, b_n) = \mathrm{Re}(\mathbf{v}) - i\ \mathrm{Im}(\mathbf{v}) \tag{2}$$

It follows that the vectors in R^n can be viewed as those vectors in C^n whose imaginary part is zero; or stated another way, a vector $\mathbf{v}$ in C^n is in R^n if and only if $\overline{\mathbf{v}} = \mathbf{v}$.

In this section we will also need to consider matrices with complex entries, so henceforth we will call a matrix A a ***real matrix*** if its entries are required to be real numbers and a ***complex matrix*** if its entries are allowed to be complex numbers. The standard operations on real matrices carry over to complex matrices without change, and all of the familiar properties of matrices continue to hold.

If A is a complex matrix, then $\mathrm{Re}(A)$ and $\mathrm{Im}(A)$ are the matrices formed from the real and imaginary parts of the entries of A, and $\overline{A}$ is the matrix formed by taking the complex conjugate of each entry in A.

EXAMPLE 1 Real and Imaginary Parts of Vectors and Matrices

Let

$$\mathbf{v} = (3+i,\ -2i,\ 5) \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} 1+i & -i \\ 4 & 6-2i \end{bmatrix}$$

Then

$$\overline{\mathbf{v}} = (3-i,\ 2i,\ 5), \quad \operatorname{Re}(\mathbf{v}) = (3,\ 0,\ 5), \quad \operatorname{Im}(\mathbf{v}) = (1,\ -2,\ 0)$$

$$\overline{A} = \begin{bmatrix} 1-i & i \\ 4 & 6+2i \end{bmatrix}, \quad \operatorname{Re}(A) = \begin{bmatrix} 1 & 0 \\ 4 & 6 \end{bmatrix}, \quad \operatorname{Im}(A) = \begin{bmatrix} 1 & -1 \\ 0 & -2 \end{bmatrix}$$

$$\det(A) = \begin{vmatrix} 1+i & -i \\ 4 & 6-2i \end{vmatrix} = (1+i)(6-2i) - (-i)(4) = 8+8i$$

Algebraic Properties of the Complex Conjugate

The next two theorems list some properties of complex vectors and matrices that we will need in this section. Some of the proofs are given as exercises.

THEOREM 5.3.1

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in C^n, and if k is a scalar, then:

(a) $\overline{\overline{\mathbf{u}}} = \mathbf{u}$

(b) $\overline{k\mathbf{u}} = \overline{k}\overline{\mathbf{u}}$

(c) $\overline{\mathbf{u}+\mathbf{v}} = \overline{\mathbf{u}} + \overline{\mathbf{v}}$

(d) $\overline{\mathbf{u}-\mathbf{v}} = \overline{\mathbf{u}} - \overline{\mathbf{v}}$

THEOREM 5.3.2

If A is an $m \times k$ complex matrix and B is a $k \times n$ complex matrix, then:

(a) $\overline{\overline{A}} = A$

(b) $\left(\overline{A^T}\right) = \left(\overline{A}\right)^T$

(c) $\overline{AB} = \overline{A}\ \overline{B}$

The Complex Euclidean Inner Product

The following definition extends the notions of dot product and norm to C^n.

DEFINITION 2

If $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ are vectors in C^n, then the ***complex Euclidean inner product of*** of $\mathbf{u}$ and $\mathbf{v}$ (also called the ***complex dot product***) is denoted by $\mathbf{u} \cdot \mathbf{v}$ and is defined as

$$\mathbf{u} \cdot \mathbf{v} = u_1\overline{v}_1 + u_2\overline{v}_2 + \ldots + u_n\overline{v}_n \tag{3}$$

We also define the ***Euclidean norm*** on C^n to be

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{|v_1|^2 + |v_2|^2 + \ldots + |v_n|^2} \tag{4}$$

As in the real case, we call **v** a ***unit vector*** in C^n if $\|\mathbf{v}\| = 1$, and we say two vectors **u** and **v** are ***orthogonal*** if $\mathbf{u} \cdot \mathbf{v} = 0$.

The complex conjugates in 3 ensure that $\|\mathbf{v}\|$ is a real number, for without them the quantity $\mathbf{v} \cdot \mathbf{v}$ in 4 might be imaginary.

EXAMPLE 2 Complex Euclidean Inner Product and Norm

Find $\mathbf{u} \cdot \mathbf{v}$, $\mathbf{v} \cdot \mathbf{u}$, $\|\mathbf{u}\|$, and $\|\mathbf{v}\|$ for the vectors

$$\mathbf{u} = (1+i, i, 3-i) \quad \text{and} \quad \mathbf{v} = (1+i, 2, 4i)$$

Solution

$$\mathbf{u} \cdot \mathbf{v} = (1+i)\overline{(1+i)} + i\overline{(2)} + (3-i)\overline{(4i)} = (1+i)(1-i) + 2i + (3-i)(-4i) = -2-10i$$

$$\mathbf{v} \cdot \mathbf{u} = (1+i)\overline{(1+i)} + 2\overline{(i)} + (4i)\overline{(3-i)} = (1+i)(1-i) - 2i + 4i(3+i) = -2+10i$$

$$\|\mathbf{u}\| = \sqrt{|1+i|^2 + |i|^2 + |3-i|^2} = \sqrt{2+1+10} = \sqrt{13}$$

$$\|\mathbf{v}\| = \sqrt{|1+i|^2 + |2|^2 + |4i|^2} = \sqrt{2+4+16} = \sqrt{22}$$

Recall from Table 1 of Section 3.2 that if **u** and **v** are *column vectors* in R^n, then their dot product can be expressed as

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T\mathbf{v} = \mathbf{v}^T\mathbf{u}$$

The analogous formulas in C^n are (verify)

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T\overline{\mathbf{v}} = \overline{\mathbf{v}}^T\mathbf{u} \tag{5}$$

Example 2 reveals a major difference between the dot product on R^n and the complex dot product on C^n. For the dot product on R^n we always have $\mathbf{v} \cdot \mathbf{u} = \mathbf{u} \cdot \mathbf{v}$ (the *symmetry property*), but for the complex dot product the corresponding relationship is given by $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}}$, which is called its ***antisymmetry*** property. The following theorem is an analog of Theorem 3.2.2.

THEOREM 5.3.3

If **u**, **v**, and **w** are vectors in C^n, and if k is a scalar, then the complex Euclidean inner product has the following properties:

(a) $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}}$ **[Antisymmetry property]**

(b) $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$ **[Distributive property]**

(c) $k(\mathbf{u} \cdot \mathbf{v}) = (k\mathbf{u}) \cdot \mathbf{v}$ **[Homogeneity property]**

(d) $\mathbf{u} \cdot k\mathbf{v} = \overline{k}(\mathbf{u} \cdot \mathbf{v})$ **[Antihomogeneity property]**

(e) $\mathbf{v} \cdot \mathbf{v} \geq 0$ *and* $\mathbf{v} \cdot \mathbf{v} = 0$ *if and only if* $\mathbf{v} = \mathbf{0}$. **[Positivity property]**

Parts (*c*) and (*d*) of this theorem state that a scalar multiplying a complex Euclidean inner product can be regrouped with the first vector, but to regroup it with the second vector you must first take its complex conjugate. We will prove part (*d*), and leave the others as exercises.

Proof (d)

$$k(\mathbf{u} \cdot \mathbf{v}) = k(\overline{\mathbf{v} \cdot \mathbf{u}}) = \overline{\overline{k}}(\overline{\mathbf{v} \cdot \mathbf{u}}) = \overline{\overline{k}(\mathbf{v} \cdot \mathbf{u})} = \overline{(\overline{k}\mathbf{v}) \cdot \mathbf{u}} = \mathbf{u} \cdot (\overline{k}v)$$

To complete the proof. substitute $\overline{k}$ for k and use the fact that $\overline{\overline{k}} = k$.

Vector Concepts in C^n

Except for the use of complex scalars, the notions of linear combination, linear independence, subspace, spanning, basis, and dimension carry over without change to C^n.

Is R^n a subspace of C^n? Explain.

Eigenvalues and eigenvectors are defined for complex matrices exactly as for real matrices. If A is an $n \times n$ matrix with complex entries, then the complex roots of the characteristic equation $\det(\lambda I - A) = 0$ are called ***complex eigenvalues*** of A. As in the real case, λ is a complex eigenvalue of A if and only if there exists a nonzero vector $\mathbf{x}$ in C^n such that $A\mathbf{x} = \lambda\mathbf{x}$. Each such $\mathbf{x}$ is called a ***complex eigenvector*** of A corresponding to λ. The complex eigenvectors of A corresponding to λ are the nonzero solutions of the linear system $(\lambda I - A)\mathbf{x} = 0$, and the set of all such solutions is a subspace of C^n, called the ***eigenspace*** of A corresponding to λ.

The following theorem states that if a *real matrix* has complex eigenvalues, then those eigenvalues and their corresponding eigenvectors occur in conjugate pairs.

THEOREM 5.3.4

If λ is an eigenvalue of a real $n \times n$ matrix A, and if $\mathbf{x}$ is a corresponding eigenvector, then $\overline{\lambda}$ is also an eigenvalue of A, and $\overline{\mathbf{x}}$ is a corresponding eigenvector.

Proof Since λ is an eigenvalue of A and $\mathbf{x}$ is a corresponding eigenvector, we have

$$\overline{A\mathbf{x}} = \overline{\lambda\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}} \tag{6}$$

However, $\overline{A} = A$, since A has real entries, so it follows from part (*c*) of Theorem 5.3.2 that

$$\overline{A\mathbf{x}} = \overline{A}\overline{\mathbf{x}} = A\overline{\mathbf{x}} \tag{7}$$

Equations 6 and 7 together imply that

$$A_x = \overline{A\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}$$

in which $\bar{\mathbf{x}} \neq \mathbf{0}$ (why?); this tells us that $\bar{\lambda}$ is an eigenvalue of A and $\bar{\mathbf{x}}$ is a corresponding eigenvector.

EXAMPLE 3 Complex Eigenvalues and Eigenvectors

Find the eigenvalues and bases for the eigenspaces of

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

Solution The characteristic polynomial of A is

$$\begin{vmatrix} \lambda+2 & 1 \\ -5 & \lambda-2 \end{vmatrix} = \lambda^2 + 1 = (\lambda - i)(\lambda + i)$$

so the eigenvalues of A are $\lambda = i$ and $\lambda = -i$. Note that these eigenvalues are complex conjugates, as guaranteed by Theorem 5.3.4.

To find the eigenvectors we must solve the system

$$\begin{bmatrix} \lambda+2 & 1 \\ -5 & \lambda-2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

with $\lambda = i$ and then with $\lambda = -i$. With $\lambda = i$, this system becomes

$$\begin{bmatrix} i+2 & 1 \\ -5 & i-2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{8}$$

We could solve this system by reducing the augmented matrix

$$\begin{bmatrix} i+2 & 1 & 0 \\ -5 & i-2 & 0 \end{bmatrix} \tag{9}$$

to reduced row echelon form by Gauss-Jordan elimination, though the complex arithmetic is somewhat tedious. A simpler procedure here is first to observe that the reduced row echelon form of 9 must have a row of zeros because 8 has nontrivial solutions. This being the case, each row of 9 must be a scalar multiple of the other, and hence the first row can be made into a row of zeros by adding a suitable multiple of the second row to it. Accordingly, we can simply set the entries in the first row to zero, then interchange the rows, and then multiply the new first row by $-\frac{1}{5}$ to obtain the reduced row echelon form

$$\begin{bmatrix} 1 & \frac{2}{5} - \frac{1}{5}i & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus, a general solution of the system is

$$x_1 = \left(-\frac{2}{5} + \frac{1}{5}i\right)t, \quad x_2 = t$$

This tells us that the eigenspace corresponding to $\lambda = i$ is one-dimensional and consists of all complex scalar multiples of the basis vector

$$\mathbf{x} = \begin{bmatrix} -\frac{2}{5} + \frac{1}{5}\mathbf{i} \\ 1 \end{bmatrix} \tag{10}$$

As a check, let us confirm that $A\mathbf{x} = i\mathbf{x}$. We obtain

$$A\mathbf{x} = \begin{bmatrix} -2 & -1 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} -\frac{2}{5} + \frac{1}{5}i \\ 1 \end{bmatrix} = \begin{bmatrix} -2\left(-\frac{2}{5} + \frac{1}{5}i\right) - 1 \\ 5\left(-\frac{2}{5} + \frac{1}{5}i\right) + 2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{5} - \frac{2}{5}i \\ i \end{bmatrix} = i\mathbf{x}$$

We could find a basis for the eigenspace corresponding to $\lambda = -i$ in a similar way, but the work is unnecessary,

since Theorem 5.3.4 implies that

$$\bar{\mathbf{x}} = \begin{bmatrix} -\frac{2}{5} - \frac{1}{5}i \\ 1 \end{bmatrix} \tag{11}$$

must be a basis for this eigenspace. The following computations confirm that $\bar{\mathbf{x}}$ is an eigenvector of A corresponding to $\lambda = -i$:

$$\begin{aligned} A\bar{\mathbf{x}} &= \begin{bmatrix} -2 & -1 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} -\frac{2}{5} - \frac{1}{5}i \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -2\left(-\frac{2}{5} - \frac{1}{5}i\right) - 1 \\ 5\left(-\frac{2}{5} - \frac{1}{5}i\right) + 2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{5} + \frac{2}{5}i \\ -i \end{bmatrix} = -i\bar{\mathbf{x}} \end{aligned}$$

Since a number of our subsequent examples will involve 2×2 matrices with real entries, it will be useful to discuss some general results about the eigenvalues of such matrices. Observe first that the characteristic polynomial of the matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is

$$\det(\lambda I - A) = \begin{vmatrix} \lambda - a & -b \\ -c & \lambda - d \end{vmatrix} = (\lambda - a)(\lambda - d) - bc = \lambda^2 - (a + d)\lambda + (ad - bc)$$

We can express this in terms of the trace and determinant of A as

$$\det(\lambda I - A) = \lambda^2 - \operatorname{tr}(A)\lambda + \det(A) \tag{12}$$

from which it follows that the characteristic equation of A is

$$\lambda^2 - \operatorname{tr}(A)\lambda + \det(A) = 0 \tag{13}$$

Now recall from algebra that if $ax^2 + bx + c = 0$ is a quadratic equation with real coefficients, then the ***discriminant*** $b^2 - 4ac$ determines the nature of the roots:

$b^2 - 4ac > 0$ [**Two distinct real roots**]

$b^2 - 4ac = 0$ [**One repeated real root**]

$b^2 - 4ac < 0$ [**Two conjugate imaginary roots**]

Applying this to 13 with $a = 1$, $b = -\operatorname{tr}(A)$, and $c = \det(A)$ yields the following theorem.

Olga Taussky-Todd (1906–1995)

Historical Note Olga Taussky-Todd was one of the pioneering women in matrix analysis and the first woman appointed to the faculty at the California Institute of Technology. She worked at the National Physical Laboratory in London during World War II, where she was assigned to study flutter in supersonic aircraft. While there, she realized that some results about the eigenvalues of a certain 6×6 complex matrix could be used to answer key questions about the flutter problem that would otherwise have required laborious calculation. After World War II Olga Taussky-Todd continued her work on matrix-related subjects and helped to draw many known but disparate results about matrices into the coherent subject that we now call matrix theory.
[*Image: Courtesy of the Archives, California Institute of Technology*]

THEOREM 5.3.5

If A is a 2×2 matrix with real entries, then the characteristic equation of A is $\lambda^2 - \operatorname{tr}(A)\lambda + \det(A) = 0$ and

(a) A has two distinct real eigenvalues if $\operatorname{tr}(A)^2 - 4\det(A) > 0$;

(b) A has one repeated real eigenvalue if $\operatorname{tr}(A)^2 - 4\det(A) = 0$;

(c) A has two complex conjugate eigenvalues if $\operatorname{tr}(A)^2 - 4\det(A) < 0$.

EXAMPLE 4 Eigenvalues of a 2 × 2 Matrix

In each part, use Formula 13 for the characteristic equation to find the eigenvalues of

(a) $A = \begin{bmatrix} 2 & 2 \\ -1 & 5 \end{bmatrix}$

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

(c) $A = \begin{bmatrix} 2 & 3 \\ -3 & 2 \end{bmatrix}$

Solution

(a) We have $\operatorname{tr}(A) = 7$ and $\det(A) = 12$, so the characteristic equation of A is

$$\lambda^2 - 7\lambda + 12 = 0$$

Factoring yields $(\lambda - 4)(\lambda - 3) = 0$, so the eigenvalues of A are $\lambda = 4$ and $\lambda = 3$.

(b) We have $\operatorname{tr}(A) = 2$ and $\det(A) = 1$, so the characteristic equation of A is

$$\lambda^2 - 2\lambda + 1 = 0$$

Factoring this equation yields $(\lambda - 1)^2 = 0$, so $\lambda = 1$ is the only eigenvalue of A; it has algebraic multiplicity 2.

(c) We have $\operatorname{tr}(A) = 4$ and $\det(A) = 13$, so the characteristic equation of A is

$$\lambda^2 - 4\lambda + 13 = 0$$

Solving this equation by the quadratic formula yields

$$\lambda = \frac{4 \pm \sqrt{(-4)^2 - 4(13)}}{2} = \frac{4 \pm \sqrt{-36}}{2} = 2 \pm 3i$$

Thus, the eigenvalues of A are $\lambda = 2 + 3i$ and $\lambda = 2 - 3i$.

Symmetric Matrices Have Real Eigenvalues

Our next result, which is concerned with the eigenvalues of real symmetric matrices, is important in a wide variety of applications. The key to its proof is to think of a real symmetric matrix as a complex matrix whose entries have an imaginary part of zero.

THEOREM 5.3.6

If A is a real symmetric matrix, then A has real eigenvalues.

Proof Suppose that λ is an eigenvalue of A and $\mathbf{x}$ is a corresponding eigenvector, where we allow for the possibility that λ is complex and $\mathbf{x}$ is in C^n. Thus,

$$A\mathbf{x} = \lambda\mathbf{x}$$

where $\mathbf{x} \neq \mathbf{0}$. If we multiply both sides of this equation by $\overline{\mathbf{x}}^T$ and use the fact that

$$\overline{\mathbf{x}}^T A\mathbf{x} = \overline{\mathbf{x}}^T(\lambda\mathbf{x}) = \lambda\left(\overline{\mathbf{x}}^T\mathbf{x}\right) = \lambda(\mathbf{x} \cdot \mathbf{x}) = \lambda\|\mathbf{x}\|^2$$

then we obtain

$$\lambda = \frac{\overline{\mathbf{x}}^T A\mathbf{x}}{\|\mathbf{x}\|^2}$$

Since the denominator in this expression is real, we can prove that λ is real by showing that

$$\overline{\overline{\mathbf{x}}^T A\mathbf{x}} = \overline{\mathbf{x}}^T A\mathbf{x} \tag{14}$$

But, A is symmetric and has real entries, so it follows from the second equality in 14 and properties of the conjugate that

$$\overline{\overline{\mathbf{x}}^T A\mathbf{x}} = \overline{\overline{\mathbf{x}}}^T \overline{A\mathbf{x}} = \mathbf{x}^T \overline{A\mathbf{x}} = \left(\overline{A\mathbf{x}}\right)^T \mathbf{x} = \left(\overline{A}\overline{\mathbf{x}}\right)^T \mathbf{x} = (A\overline{\mathbf{x}})^T \mathbf{x} = \overline{\mathbf{x}}^T A^T \mathbf{x} = \overline{\mathbf{x}}^T A\mathbf{x}$$

A Geometric Interpretation of Complex Eigenvalues

The following theorem is the key to understanding the geometric significance of complex eigenvalues of real 2×2 matrices.

THEOREM 5.3.7

The eigenvalues of the real matrix

$$C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \tag{15}$$

are $\lambda = a \pm bi$. If a and b are not both zero, then this matrix can be factored as

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \tag{16}$$

where φ is the angle from the positive x-axis to the ray that joins the origin to the point (a, b) (Figure 5.3.2).

Figure 5.3.2

Geometrically, this theorem states that multiplication by a matrix of form 15 can be viewed as a rotation through the angle φ followed by a scaling with factor $|\lambda|$ (Figure 5.3.3).

Figure 5.3.3

Proof The characteristic equation of C is $(\lambda - a)^2 + b^2 = 0$ (verify), from which it follows that the eigenvalues of C are $\lambda = a \pm bi$. Assuming that a and b are not both zero, let φ be the angle from the positive x-axis to the ray that joins the origin to the point (a, b). The angle φ is an argument of the eigenvalue $\lambda = a + bi$, so we see from Figure 5.3.2 that

$$a = |\lambda| \cos\phi \quad \text{and} \quad b = |\lambda| \sin\phi$$

It follows from this that the matrix in 15 can be written as

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{bmatrix} \begin{bmatrix} \frac{a}{|\lambda|} & -\frac{b}{|\lambda|} \\ \frac{b}{|\lambda|} & \frac{a}{|\lambda|} \end{bmatrix} = \begin{bmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$

The following theorem, whose proof is considered in the exercises, shows that every real 2×2 matrix with complex eigenvalues is similar to a matrix of form 15.

THEOREM 5.3.8

Let A be a real 2×2 matrix with complex eigenvalues $\lambda = a \pm bi$ (where $b \neq 0$). If $\mathbf{x}$ is an eigenvector of A corresponding to $\lambda = a - bi$, then the matrix $P = \begin{bmatrix} \mathrm{Re}(\mathbf{x}) & \mathrm{Im}(\mathbf{x}) \end{bmatrix}$ is invertible and

$$A = P\begin{bmatrix} a & -b \\ b & a \end{bmatrix}P^{-1} \tag{17}$$

EXAMPLE 5 A Matrix Factorization Using Complex Eigenvalues

Factor the matrix in Example 3 into form 17 using the eigenvalue $\lambda = -i$ and the corresponding eigenvector that was given in 11.

Solution For consistency with the notation in Theorem 5.3.8, let us denote the eigenvector in 11 that corresponds to $\lambda = -i$ by $\mathbf{x}$ (rather than $\overline{\mathbf{x}}$ as before). For this λ and $\mathbf{x}$ we have

$$a = 0, \quad b = 1, \quad \mathrm{Re}(\mathbf{x}) = \begin{bmatrix} -\frac{2}{5} \\ 1 \end{bmatrix}, \quad \mathrm{Im}(\mathbf{x}) = \begin{bmatrix} -\frac{1}{5} \\ 0 \end{bmatrix}$$

Thus,

$$P = \begin{bmatrix} \mathrm{Re}(\mathbf{x}) & \mathrm{Im}(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} -\frac{2}{5} & -\frac{1}{5} \\ 1 & 0 \end{bmatrix}$$

so A can be factored in form 17 as

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

You may want to confirm this by multiplying out the right side.

A Geometric Interpretation of Theorem 5.3.8

To clarify what Theorem 5.3.8 says geometrically, let us denote the matrices on the right side of 16 by S and R_ϕ, respectively, and then use 16 to rewrite 17 as

$$A = PSR_\phi P^{-1} = P\begin{bmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{bmatrix}\begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}P^{-1} \tag{18}$$

If we now view P as the transition matrix from the basis $B = \{\mathrm{Re}(\mathbf{x}), \mathrm{Im}(\mathbf{x})\}$ to the standard basis, then 18 tells us that computing a product $A\mathbf{x}_0$ can be broken down into a three-step process:

Step 1 Map $\mathbf{x}_0$ from standard coordinates into B-coordinates by forming the product $P^{-1}\mathbf{x}_0$.

Step 2 Rotate and scale the vector $P^{-1}\mathbf{x}_0$ by forming the product $SR_\phi P^{-1}\mathbf{x}_0$.

Step 3 Map the rotated and scaled vector back to standard coordinates to obtain $A\mathbf{x}_0 = PSR_\phi P^{-1}\mathbf{x}_0$.

Power Sequences

There are many problems in which one is interested in how successive applications of a matrix transformation affect a specific vector. For example, if A is the standard matrix for an operator on R^n and $\mathbf{x}_0$ is some fixed vector in R^n, then one might be interested in the behavior of the power sequence

$$\mathbf{x}_0, \quad A\mathbf{x}_0, \quad A^2\mathbf{x}_0, \ldots, \quad A^k\mathbf{x}_0, \ldots$$

For example, if

$$A = \begin{bmatrix} \frac{1}{2} & \frac{3}{4} \\ -\frac{3}{5} & \frac{11}{10} \end{bmatrix} \quad \text{and} \quad \mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

then with the help of a computer or calculator one can show that the first four terms in the power sequence are

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad A\mathbf{x}_0 = \begin{bmatrix} 1.25 \\ 0.5 \end{bmatrix}, \quad A^2\mathbf{x}_0 = \begin{bmatrix} 1.0 \\ -0.2 \end{bmatrix}, \quad A^3\mathbf{x}_0 = \begin{bmatrix} 0.35 \\ -0.82 \end{bmatrix}$$

With the help of MATLAB or a computer algebra system one can show that if the first 100 terms are plotted as ordered pairs (x, y), then the points move along the elliptical path shown in Figure 5.3.4*a*.

Figure 5.3.4

To understand why the points move along an elliptical path, we will need to examine the eigenvalues and eigenvectors of A. We leave it for you to show that the eigenvalues of A are $\lambda = \frac{4}{5} \pm \frac{3}{5}i$ and that the corresponding eigenvectors are

$$\lambda_1 = \frac{4}{5} - \frac{3}{5}i: \quad \mathbf{v}_1 = \left(\frac{1}{2} + i, 1\right) \quad \text{and} \quad \lambda_2 = \frac{4}{5} + \frac{3}{5}i: \quad \mathbf{v}_2 = \left(\frac{1}{2} - i, 1\right)$$

If we take $\lambda = \lambda_1 = \frac{4}{5} - \frac{3}{5}i$ and $\mathbf{x} = \mathbf{v}_1 = \left(\frac{1}{2} + i, 1\right)$ in 17 and use the fact that $|\lambda| = 1$, then we obtain the factorization

$$\begin{bmatrix} \frac{1}{2} & \frac{3}{4} \\ -\frac{3}{5} & \frac{11}{10} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & -\frac{1}{2} \end{bmatrix} \tag{19}$$

$$A \quad = \quad P \quad R_\phi \quad P^{-1}$$

where R_ϕ is a rotation about the origin through the angle φ whose tangent is

$$\tan\phi = \frac{\sin\phi}{\cos\phi} = \frac{3/5}{4/5} = \frac{3}{4} \qquad \left(\phi = \tan^{-1}\frac{3}{4} \approx 36.9°\right)$$

The matrix P in 19 is the transition matrix from the basis

$$B = \{\text{Re}(\mathbf{x}), \text{Im}(\mathbf{x})\} = \left\{\left(\frac{1}{2}, 1\right), (1, 0)\right\}$$

to the standard basis, and P^{-1} is the transition matrix from the standard basis to the basis B (Figure 5.3.5). Next, observe that if n is a positive integer, then 19 implies that

$$A^n\mathbf{x}_0 = \left(PR_\phi P^{-1}\right)^n \mathbf{x}_0 = PR_\phi^n P^{-1}\mathbf{x}_0$$

so the product $A^n\mathbf{x}_0$ can be computed by first mapping $\mathbf{x}_0$ into the point $P^{-1}\mathbf{x}_0$ in B-coordinates, then multiplying by R_ϕ^n to rotate this point about the origin through the angle $n\phi$, and then multiplying $R_\phi^n P^{-1}\mathbf{x}_0$ by P to map the resulting point back to

standard coordinates. We can now see what is happening geometrically: In B-coordinates each successive multiplication by A causes the point $P^{-1}\mathbf{x}_0$ to advance through an angle φ, thereby tracing a circular orbit about the origin. However, the basis B is *skewed* (not orthogonal), so when the points on the circular orbit are transformed back to standard coordinates, the effect is to distort the circular orbit into the elliptical orbit traced by $A^n\mathbf{x}_0$ (Figure 5.3.4*b*). Here are the computations for the first step (successive steps are illustrated in Figure 5.3.4*c*):

$$\begin{bmatrix} \frac{1}{2} & \frac{3}{4} \\ -\frac{3}{5} & \frac{11}{10} \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & -\frac{1}{2} \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{2} & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix}\begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix} \quad [\mathbf{x_0}\text{ is mapped to } B-\text{coordinates .}]$$

$$= \begin{bmatrix} \frac{1}{2} & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix} \quad \left[\text{The point}\left(1, \frac{1}{2}\right)\text{is rotated through the angle } \phi\text{ .}\right]$$

$$= \begin{bmatrix} \frac{5}{4} \\ \frac{1}{2} \end{bmatrix} \quad \left[\text{The point}\left(\frac{1}{2}, 1\right)\text{is mapped to standard coordinates .}\right]$$

Figure 5.3.5

Concept Review

- Real part of z
- Imaginary part of z
- Modulus of z
- Complex conjugate of z
- Argument of z
- Polar form of z
- Complex vector space
- Complex n-tuple
- Complex n-space
- Real matrix
- Complex matrix
- Complex Euclidean inner product
- Euclidean norm on C^n

- Antisymmetry property
- Complex eigenvalue
- Complex eigenvector
- Eigenspace in C^n
- Discriminant

Skills

- Find the real part, imaginary part, and complex conjugate of a complex matrix or vector.
- Find the determinant of a complex matrix.
- Find complex inner products and norms of complex vectors.
- Find the eigenvalues and bases for the eigenspaces of complex matrices.
- Factor a 2×2 real matrix with complex eigenvalues into a product of a scaling matrix and a rotation matrix.

Exercise Set 5.3

In Exercises 1–2, find $\overline{\mathbf{u}}$, $\mathrm{Re}(\mathbf{u})$, $\mathrm{Im}(\mathbf{u})$, and $\|\mathbf{u}\|$.

1. $\mathbf{u} = (2 - i, 4i, 1 + i)$

Answer:

$\overline{\mathbf{u}} = (2 + i, \ -4i, 1 - i), \ \mathrm{Re}\ (\mathbf{u}) = (2, 0, 1), \ \mathrm{Im}\ (\mathbf{u}) = (-1, 4, 1), \ \|\mathbf{u}\| = \sqrt{23}$

2. $\mathbf{u} = (6, 1 + 4i, 6 - 2i)$

In Exercises 3–4, show that $\mathbf{u}$, $\mathbf{v}$, and k satisfy Theorem 5.3.1.

3. $\mathbf{u} = (3 - 4i, 2 + i, \ -6i), \mathbf{v} = (1 + i, 2 - i, 4), k = i$

4. $\mathbf{u} = (6, 1 + 4i, 6 - 2i), \mathbf{v} = (4, 3 + 2i, i - 3), k = -i$

5. Solve the equation $i\mathbf{x} - 3\mathbf{v} = \overline{\mathbf{u}}$ for $\mathbf{x}$, where $\mathbf{u}$ and $\mathbf{v}$ are the vectors in Exercise 3.

Answer:

$\mathbf{x} = (7 - 6i, \ -4 - 8i, 6 - 12i)$

6. Solve the equation $(1 + i)\mathbf{x} + 2\mathbf{u} = \overline{\mathbf{v}}$ for $\mathbf{x}$, where $\mathbf{u}$ and $\mathbf{v}$ are the vectors in Exercise 4.

In Exercises 7–8, find $\overline{A}$, $\mathrm{Re}(A)$, $\mathrm{Im}(A)$, $\det(A)$, and $\mathrm{tr}(A)$.

7. $A = \begin{bmatrix} -5i & 4 \\ 2 - i & 1 + 5i \end{bmatrix}$

Answer:

$\overline{A} = \begin{bmatrix} 5i & 4 \\ 2 + i & 1 - 5i \end{bmatrix}, \ \mathrm{Re}\ (A) = \begin{bmatrix} 0 & 4 \\ 2 & 1 \end{bmatrix}, \ \mathrm{Im}\ (A) = \begin{bmatrix} -5 & 0 \\ -1 & 5 \end{bmatrix}, \ \det(A) = 17 - i, \ \mathrm{tr}(A) = 1$

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

9. Let A be the matrix given in Exercise 7, and let B be the matrix

$$B = \begin{bmatrix} 1 - i \\ 2i \end{bmatrix}$$

Confirm that these matrices have the properties stated in Theorem 5.3.2.

10. Let A be the matrix given in Exercise 8, and let B be the matrix

$$B = \begin{bmatrix} 5i \\ 1-4i \end{bmatrix}$$

Confirm that these matrices have the properties stated in Theorem 5.3.2.

In Exercises 11–12, compute $\mathbf{u} \cdot \mathbf{v}$, $\mathbf{u} \cdot \mathbf{w}$, and $\mathbf{v} \cdot \mathbf{w}$, and show that the vectors satisfy Formula 5 and parts (a), (b), and (c) of Theorem 5.3.3.

11. $\mathbf{u} = (i, 2i, 3),\ \mathbf{v} = (4, -2i, 1+i),\ \mathbf{w} = (2-i, 2i, 5+3i),\ k = 2i$

Answer:

$\mathbf{u} \cdot \mathbf{v} = -1+i,\ \mathbf{u} \cdot \mathbf{w} = 18-7i,\ \mathbf{v} \cdot \mathbf{w} = 12+6i$

12. $\mathbf{u} = (1+i, 4, 3i), \quad \mathbf{v} = (3, -4i, 2+3i), \quad \mathbf{w} = (1-i, 4i, 4-5i), \quad k = 1+i$

13. Compute $\overline{(\mathbf{u} \cdot \overline{\mathbf{v}}) - \overline{\mathbf{w} \cdot \mathbf{u}}}$ for the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in Exercise 11.

Answer:

$-11-14i$

14. Compute $\overline{\left(i\overline{\mathbf{u}} \cdot \mathbf{w}\right) + (\|\mathbf{u}\|\mathbf{v}) \cdot \mathbf{u}}$ for the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in Exercise 12.

In Exercises 15–18, find the eigenvalues and bases for the eigenspaces of A.

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

Answer:

$\lambda_1 = 2-i,\ \mathbf{x}_1 = \begin{bmatrix} 2-i \\ 1 \end{bmatrix};\ \lambda_2 = 2+i,\ \mathbf{x}_1 = \begin{bmatrix} 2+i \\ 1 \end{bmatrix}$

16. $A = \begin{bmatrix} -1 & -5 \\ 4 & 7 \end{bmatrix}$

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

Answer:

$\lambda_1 = 4-i,\ \mathbf{x}_1 = \begin{bmatrix} 1-i \\ 1 \end{bmatrix};\ \lambda_2 = 4+i,\ \mathbf{x}_1 = \begin{bmatrix} 1+i \\ 1 \end{bmatrix}$

18. $A = \begin{bmatrix} 8 & 6 \\ -3 & 2 \end{bmatrix}$

In Exercises 19–22, each matrix C has form 15. Theorem 5.3.7 implies that C is the product of a scaling matrix with factor $|\lambda|$ and a rotation matrix with angle φ. Find $|\lambda|$ and φ for which $-\pi < \phi \leq \pi$.

19. $C = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$

Answer:

$|\lambda| = \sqrt{2},\ \phi = \frac{\pi}{4}$

20. $C = \begin{bmatrix} 0 & 5 \\ -5 & 0 \end{bmatrix}$

21. $C = \begin{bmatrix} 1 & \sqrt{3} \\ -\sqrt{3} & 1 \end{bmatrix}$

Answer:

$|\lambda| = 2,\ \phi = -\frac{\pi}{3}$

22. $C = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ -\sqrt{2} & \sqrt{2} \end{bmatrix}$

In Exercises 23–26, find an invertible matrix P and a matrix C of form 15 such that $A = PCP^{-1}$.

23. $A = \begin{bmatrix} -1 & -5 \\ 4 & 7 \end{bmatrix}$

Answer:

$P = \begin{bmatrix} -2 & -1 \\ 2 & 0 \end{bmatrix},\ C = \begin{bmatrix} 3 & -2 \\ 2 & 3 \end{bmatrix}$

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

25. $A = \begin{bmatrix} 8 & 6 \\ -3 & 2 \end{bmatrix}$

Answer:

$P = \begin{bmatrix} 1 & -1 \\ -1 & 0 \end{bmatrix},\ C = \begin{bmatrix} 5 & -3 \\ 3 & 5 \end{bmatrix}$

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

27. Find all complex scalars k, if any, for which **u** and **v** are orthogonal in C^3.

(a) $\mathbf{u} = (2i, i, 3i), \quad \mathbf{v} = (i, 6i, k)$

(b) $\mathbf{u} = (k, k, 1+i), \quad \mathbf{v} = (1, -1, 1-i)$

Answer:

(a) $k = -\frac{8}{3}i$

(b) None

28. Show that if A is a real $n \times n$ matrix and **x** is a column vector in C^n, then $\mathrm{Re}(A\mathbf{x}) = A(\mathrm{Re}(\mathbf{x}))$ and $\mathrm{Im}(A\mathbf{x}) = A(\mathrm{Im}(\mathbf{x}))$.

29. The matrices

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

called ***Pauli spin matrices***, are used in quantum mechanics to study particle spin. The ***Dirac matrices***, which are also used in quantum mechanics, are expressed in terms of the Pauli spin matrices and the 2×2 identity matrix I_2 as

$$\beta = \begin{bmatrix} I_2 & 0 \\ 0 & -I_2 \end{bmatrix}, \quad \alpha_x = \begin{bmatrix} 0 & \sigma_1 \\ \sigma_1 & 0 \end{bmatrix},$$

$$\alpha_y = \begin{bmatrix} 0 & \sigma_2 \\ \sigma_2 & 0 \end{bmatrix}, \quad \alpha_z = \begin{bmatrix} 0 & \sigma_3 \\ \sigma_3 & 0 \end{bmatrix}$$

(a) Show that $\beta^2 = \alpha_x^2 = \alpha_y^2 = \alpha_z^2$.

(b) Matrices A and B for which $AB = -BA$ are said to be ***anticommutative***. Show that the Dirac matrices are anticommutative.

30. If k is a real scalar and $\mathbf{v}$ is a vector in R^n, then Theorem 3.2.1 states that $\|k\mathbf{v}\| = |k|\|\mathbf{v}\|$. Is this relationship also true if k is a complex scalar and $\mathbf{v}$ is a vector in C^n? Justify your answer.

31. Prove part (c) of Theorem 5.3.1.

32. Prove Theorem 5.3.2.

33. Prove that if $\mathbf{u}$ and $\mathbf{v}$ are vectors in C^n, then

$$\begin{aligned}\mathbf{u}\cdot\mathbf{v} = \quad & \frac{1}{4}\|\mathbf{u}+\mathbf{v}\|^2 - \frac{1}{4}\|\mathbf{u}-\mathbf{v}\|^2 \\ & + \frac{i}{4}\|\mathbf{u}+i\mathbf{v}\|^2 - \frac{i}{4}\|\mathbf{u}-i\mathbf{v}\|^2\end{aligned}$$

34. It follows from Theorem 5.3.7 that the eigenvalues of the rotation matrix

$$R_\phi = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$

are $\lambda = \cos\phi \pm i\sin\phi$. Prove that if $\mathbf{x}$ is an eigenvector corresponding to either eigenvalue, then $\mathrm{Re}(\mathbf{x})$ and $\mathrm{Im}(\mathbf{x})$ are orthogonal and have the same length. [*Note:* This implies that $P = [\mathrm{Re}(\mathbf{x})\mathrm{Im}(\mathbf{x})]$ is a real scalar multiple of an orthogonal matrix.]

35. The two parts of this exercise lead you through a proof of Theorem 5.3.8.

(a) For notational simplicity, let

$$M = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

and let $\mathbf{u} = \mathrm{Re}(\mathbf{x})$ and $\mathbf{v} = \mathrm{Im}(\mathbf{x})$, so $P = [\mathbf{u}|\mathbf{v}]$. Show that the relationship $A\mathbf{x} = \lambda\mathbf{x}$ implies that

$$A\mathbf{x} = (a\mathbf{u} + b\mathbf{v}) + i(-b\mathbf{u} + a\mathbf{v})$$

and then equate real and imaginary parts in this equation to show that

$$AP = [A\mathbf{u}|A\mathbf{v}] = [a\mathbf{u} + b\mathbf{v}\,|-b\mathbf{u} + a\mathbf{v}] = PM$$

(b) Show that P is invertible, thereby completing the proof, since the result in part (a) implies that $A = PMP^{-1}$. [*Hint:* If P is not invertible, then one of its column vectors is a real scalar multiple of the other, say $\mathbf{v} = c\mathbf{u}$. Substitute this into the equations $A\mathbf{u} = a\mathbf{u} + b\mathbf{v}$ and $A\mathbf{v} = -b\mathbf{u} + a\mathbf{v}$ obtained in part (a), and show that $(1 + c^2)b\mathbf{u} = \mathbf{0}$. Finally, show that this leads to a contradiction, thereby proving that P is invertible.]

36. In this problem you will prove the complex analog of the Cauchy-Schwarz inequality.

(a) Prove: If k is a complex number, and $\mathbf{u}$ and $\mathbf{v}$ are vectors in C^n, then

$$(\mathbf{u} - k\mathbf{v})\cdot(\mathbf{u} - k\mathbf{v}) = \mathbf{u}\cdot\mathbf{u} - \bar{k}(\mathbf{u}\cdot\mathbf{v}) - k\overline{(\mathbf{u}\cdot\mathbf{v})} + k\bar{k}(\mathbf{v}\cdot\mathbf{v})$$

(b) Use the result in part (a) to prove that

$$0 \le \mathbf{u}\cdot\mathbf{u} - \bar{k}(\mathbf{u}\cdot\mathbf{v}) - k\overline{(\mathbf{u}\cdot\mathbf{v})} + k\bar{k}(\mathbf{v}\cdot\mathbf{v})$$

(c) Take $k = (\mathbf{u}\cdot\mathbf{v}) / (\mathbf{v}\cdot\mathbf{v})$ in part (b) to prove that

$$|\mathbf{u}\cdot\mathbf{v}| \le \|\mathbf{u}\|\ \|\mathbf{v}\|$$

True-False Exercises

In parts (a)–(f) determine whether the statement is true or false, and justify your answer.

(a) There is a real 5×5 matrix with no real eigenvalues.

Answer:

False

(b) The eigenvalues of a 2×2 complex matrix are the solutions of the equation $\lambda^2 - \mathrm{tr}(A)\lambda + \det(A) = 0$.

Answer:

True

(c) Matrices that have the same complex eigenvalues with the same algebraic multiplicities have the same trace.

Answer:

False

(d) If λ is a complex eigenvalue of a real matrix A with a corresponding complex eigenvector $\mathbf{v}$, then $\overline{\lambda}$ is a complex eigenvalue of A and $\overline{\mathbf{v}}$ is a complex eigenvector of A corresponding to $\overline{\lambda}$.

Answer:

True

(e) Every eigenvalue of a complex symmetric matrix is real.

Answer:

False

(f) If a 2×2 real matrix A has complex eigenvalues and $\mathbf{x}_0$ is a vector in R^2, then the vectors $\mathbf{x}_0, A\mathbf{x}_0, A^2\mathbf{x}_0, \ldots, A^n\mathbf{x}_0, \ldots$ lie on an ellipse.

Answer:

False

Copyright © 2010 John Wiley & Sons, Inc. All rights reserved.

5.4 Differential Equations

Many laws of physics, chemistry, biology, engineering, and economics are described in terms of "differential equations"—that is, equations involving functions and their derivatives. In this section we will illustrate one way in which linear algebra, eigenvalues and eigenvectors can be applied to solving systems of differential equations. Calculus is a prerequisite for this section.

Terminology

Recall from calculus that a ***differential equation*** is an equation involving unknown functions and their derivatives. The ***order*** of a differential equation is the order of the highest derivative it contains. The simplest differential equations are the first-order equations of the form

$$y' = ay \tag{1}$$

where $y = f(x)$ is an unknown differentiable function to be determined, $y' = dy / dx$ is its derivative, and a is a constant. As with most differential equations, this equation has infinitely many solutions; they are the functions of the form

$$y = ce^{ax} \tag{2}$$

where c is an arbitrary constant. That every function of this form is a solution of 1 follows from the computation

$$y' = cae^{ax} = ay$$

and that these are the only solution is shown in the exercises. Accordingly, we call 2 the ***general solution*** of 1. As an example, the general solution of the differential equation $y' = 5y$ is

$$y = ce^{5x} \tag{3}$$

Often, a physical problem that leads to a differential equation imposes some conditions that enable us to isolate one particular solution from the general solution. For example, if we require that solution 3 of the equation $y' = 5y$ satisfy the added condition

$$y(0) = 6 \tag{4}$$

(that is, $y = 6$ when $x = 0$), then on substituting these values in 3, we obtain $6 = ce^0 = c,$ from which we conclude that

$$y = 6e^{5x}$$

is the only solution $y' = 5y$ that satisfies 4.

A condition such as 4, which specifies the value of the general solution at a point is called an ***initial condition***, and the problem of solving a differential equation subject to an initial condition is called an ***initial-value problem***.

First-Order Linear Systems

In this section we will be concerned with solving systems of differential equations of the form

$$\begin{array}{rcl} y'_1 & = & a_{11}y_1 + a_{12}y_2 + \ldots + a_{1n}y_n \\ y'_2 & = & a_{21}y_1 + a_{22}y_2 + \ldots + a_{2n}y_n \\ \vdots & & \vdots \\ y'_n & = & a_{n1}y_1 + a_{n2}y_2 + \ldots + a_{nn}y_n \end{array} \tag{5}$$

where $y_1 = f_1(x), y_2 = f_2(x), \ldots, y_n = f_n(x)$ are functions to be determined, and the a_{ij}'s are constants. In matrix notation, 5 can be written as

$$\begin{bmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

or, more briefly as

A system of differential equations of form 5 is called a ***first-order linear system***.

$$\mathbf{y}' = A\mathbf{y} \tag{6}$$

where the notation $\mathbf{y}'$ denotes the vector obtained by differentiating each component of y.

EXAMPLE 1 Solution of a Linear System with Initial Conditions

(a) Write the following system in matrix form:

$$\begin{array}{rcl} y'_1 & = & 3y_1 \\ y'_2 & = & -2y_2 \\ y'_3 & = & 5y_3 \end{array} \tag{7}$$

(b) Solve the system.

(c) Find a solution of the system that satisfies the initial conditions $y_1(0) = 1$, $y_2(0) = 4$, and $y_3(0) = -2$.

Solution

(a)

$$\begin{bmatrix} y'_1 \\ y'_2 \\ y'_3 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \tag{8}$$

or

$$\mathbf{y}' = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 5 \end{bmatrix} \mathbf{y} \tag{9}$$

(b) Because each equation in 7 involves only one unknown function, we can solve the equations individually. It follows from 2 that these solutions are

$$\begin{aligned} y_1 &= c_1 e^{3x} \\ y_2 &= c_2 e^{-2x} \\ y_3 &= c_3 e^{5x} \end{aligned}$$

or, in matrix notation,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} c_1 e^{3x} \\ c_2 e^{-2x} \\ c_3 e^{5x} \end{bmatrix} \tag{10}$$

(c) From the given initial conditions, we obtain

$$\begin{aligned} 1 &= y_1(0) = c_1 e^0 = c_1 \\ 4 &= y_2(0) = c_2 e^0 = c_2 \\ -2 &= y_3(0) = c_3 e^0 = c_3 \end{aligned}$$

so the solution satisfying these conditions is

$$y_1 = e^{3x},\ y_2 = 4e^{-2x},\ y_3 = -2e^{5x}$$

or, in matrix notation,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} e^{3x} \\ 4e^{-2x} \\ -2e^{5x} \end{bmatrix}$$

Solution by Diagonalization

What made the system in Example 1 easy to solve was the fact that each equation involved only one of the unknown functions, so its matrix formulation, $\mathbf{y}' = A\mathbf{y}$, had a *diagonal* coefficient matrix A [Formula 9]. A more complicated situation occurs when some or all of the equations in the system involve more than one of the unknown functions, for in this case the coefficient matrix is not diagonal. Let us now consider how we might solve such a system.

The basic idea for solving a system $\mathbf{y}' = A\mathbf{y}$ whose coefficient matrix A is not diagonal is to introduce a new unknown vector u that is related to the unknown vector y by an equation of the form $\mathbf{y} = P\mathbf{u}$ in which P is an invertible matrix that diagonalizes A. Of course, such a matrix may or may not exist, but if it does then we can rewrite the equation $\mathbf{y}' = A\mathbf{y}$ as

$$P\mathbf{u}' = A(P\mathbf{u})$$

or alternatively as

$$\mathbf{u}' = \left(P^{-1}AP\right)\mathbf{u}$$

Since P is assumed to diagonalize A, this equation has the form

$$\mathbf{u}' = D\mathbf{u}$$

where D is diagonal. We can now solve this equation for $\mathbf{u}$ using the method of Example 1, and then obtain y by matrix multiplication using the relationship $\mathbf{y} = P\mathbf{u}$.

In summary, we have the following procedure for solving a system $\mathbf{y}' = A\mathbf{y}$ in the case were A is diagonalizable.

A Procedure for Solving y′ = *A*y if *A* is Diagonalizable

Step 1. Find a matrix P that diagonalizes A.

Step 2. Make the substitutions $\mathbf{y} = P\mathbf{u}$ and $\mathbf{y}' = P\mathbf{u}'$ to obtain a new "diagonal system" $\mathbf{u}' = D\mathbf{u}$, where $D = P^{-1}AP$.

Step 3. Solve $\mathbf{u}' = D\mathbf{u}$.

Step 4. Determine y from the equation $\mathbf{y} = P\mathbf{u}$.

EXAMPLE 2 Solution Using Diagonalization

(a) Solve the system

$$\begin{aligned} y_1' &= y_1 + y_2 \\ y_2' &= 4y_1 - 2y_2 \end{aligned}$$

(b) Find the solution that satisfies the initial conditions $y_1(0) = 1, y_2(0) = 6$.

Solution

(a) The coefficient matrix for the system is

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

As discussed in Section 5.2, A will be diagonalized by any matrix P whose columns are linearly independent eigenvectors of A. Since

$$\det(\lambda I - A) = \begin{vmatrix} \lambda - 1 & -1 \\ -4 & \lambda + 2 \end{vmatrix} = \lambda^2 + \lambda - 6 = (\lambda + 3)(\lambda - 2)$$

the eigenvalues of A are $\lambda = 2$ and $\lambda = -3$. By definition,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

is an eigenvector of A corresponding to λ if and only if $\mathbf{x}$ is a nontrivial solution of

$$\begin{bmatrix} \lambda - 1 & -1 \\ -4 & \lambda + 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

If $\lambda = 2$, this system becomes

$$\begin{bmatrix} 1 & -1 \\ -4 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Solving this system yields $x_1 = t, x_2 = t,$ so

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} t \\ t \end{bmatrix} = t\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Thus,

$$\mathbf{p}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

is a basis for the eigenspace corresponding to $\lambda = 2$. Similarly, you can show that

$$\mathbf{p}_2 = \begin{bmatrix} -\frac{1}{4} \\ 1 \end{bmatrix}$$

is a basis for the eigenspace corresponding to $\lambda = -3$. Thus,

$$P = \begin{bmatrix} 1 & -\frac{1}{4} \\ 1 & 1 \end{bmatrix}$$

diagonalizes A, and

$$D = P^{-1}AP = \begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}$$

Thus, as noted in Step 2 of the procedure stated above, the substitution

$$\mathbf{y} = P\mathbf{u} \text{ and } \mathbf{y}' = P\mathbf{u}'$$

yields the "diagonal system"

$$\mathbf{u}' = D\mathbf{u} = \begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}\mathbf{u} \quad \text{or} \quad \begin{aligned} u_1' &= 2u_1 \\ u_2' &= -3u_2 \end{aligned}$$

From 2 the solution of this system is

$$\begin{aligned} u_1 &= c_1e^{2x} \\ u_2 &= c_2e^{-3x} \end{aligned} \quad \text{or} \quad \mathbf{u} = \begin{bmatrix} c_1e^{2x} \\ c_2e^{-3x} \end{bmatrix}$$

so the equation $\mathbf{y} = P\mathbf{u}$ yields, as the solution for y,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{4} \\ 1 & 1 \end{bmatrix}\begin{bmatrix} c_1e^{2x} \\ c_2e^{-3x} \end{bmatrix} = \begin{bmatrix} c_1e^{2x} - \frac{1}{4}c_2e^{-3x} \\ c_1e^{2x} + c_2e^{-3x} \end{bmatrix}$$

or

$$\begin{aligned} y_1 &= c_1e^{2x} - \frac{1}{4}c_2e^{-3x} \\ y_2 &= c_1e^{2x} + c_2e^{-3x} \end{aligned} \tag{11}$$

(b) If we substitute the given initial conditions in 11, we obtain

$$\begin{aligned} c_1 - \frac{1}{4}c_2 &= 1 \\ c_1 + c_2 &= 6 \end{aligned}$$

Solving this system, we obtain $c_1 = 2, c_2 = 4,$ so it follows from 11 that the solution satisfying the initial conditions is

$$\begin{aligned} y_1 &= 2e^{2x} - e^{-3x} \\ y_2 &= 2e^{2x} + 4e^{-3x} \end{aligned}$$

Remark Keep in mind that the method of Example 2 works because the coefficient matrix of the system can be diagonalized. In cases where this is not so, other methods are required. These are typically discussed in books devoted to differential equations.

Concept Review

- Differential equation
- Order of a differential equation
- General solution
- Particular solution
- Initial condition
- Initial-value problem
- First-order linear system

Skills

- Find the matrix form of a system of linear differential equations.
- Find the general solution of a system of linear differential equations by diagonalization.
- Find the particular solution of a system of linear differential equations satisfying an initial condition.

Exercise Set 5.4

1. (a) Solve the system

$$\begin{aligned} y_1' &= y_1 + 4y_2 \\ y_2' &= 2y_1 + 3y_2 \end{aligned}$$

(b) Find the solution that satisfies the initial conditions $y_1(0) = 0, y_2(0) = 0$.

Answer:

(a) $y_1 = c_1e^{5x} - 2c_2e^{-x}$

$y_2 = c_1e^{5x} + c_2e^{-x}$

(b) $y_1 = 0$

$y_2 = 0$

2. (a) Solve the system

$$\begin{aligned} y_1' &= y_1 + 3y_2 \\ y_2' &= 4y_1 + 5y_2 \end{aligned}$$

(b) Find the solution that satisfies the conditions $y_1(0) = 2, y_2(0) = 1$.

3. (a) Solve the system

$$\begin{aligned} y_1' &= 4y_1 + y_3 \\ y_2' &= -2y_1 + y_2 \\ y_3' &= -2y_1 + y_3 \end{aligned}$$

(b) Find the solution that satisfies the initial conditions $y_1(0) = -1, y_2(0) = 1, y_3(0) = 0$.

Answer:

(a) $y_1 = -c_2e^{2x} + c_3e^{3x}$

$y_2 = c_1e^x + 2c_2e^{2x} - c_3e^{3x}$

$y_3 = 2c_2e^{2x} - c_3e^{3x}$

(b) $y_1 = e^{2x} - 2e^{3x}$

$y_2 = e^x - 2e^{2x} + 2e^{3x}$

$y_3 = -2e^{2x} + 2e^{3x}$

4. Solve the system

$$\begin{aligned} y_1' &= 4y_1 + 2y_2 + 2y_3 \\ y_2' &= 2y_1 + 4y_2 + 2y_3 \\ y_3' &= 2y_1 + 2y_2 + 4y_3 \end{aligned}$$

5. Show that every solution of $y' = ay$ has the form $y = ce^{ax}$.

[*Hint:* Let $y = f(x)$ be a solution of the equation, and show that $f(x)e^{-ax}$ is constant.]

6. Show that if A is diagonalizable and

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

is a solution of the system $\mathbf{y}' = A\mathbf{y}$, then each y_i is a linear combination of $e^{\lambda_1 x}, e^{\lambda_2 x}, \ldots, e^{\lambda_n x}$, where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A.

7. Sometimes it is possible to solve a single higher-order linear differential equation with constant coefficients by expressing it as a system and applying the methods of this section. For the differential equation $y'' - y' - 6y = 0$, show that the substitutions $y_1 = y$ and $y_2 = y'$ lead to the system

$$\begin{aligned} y_1' &= y_2 \\ y_2' &= 6y_1 + y_2 \end{aligned}$$

Solve this system, and use the result to solve the original differential equation.

Answer:

$y = c_1 e^{3x} + c_2 e^{-2x}$

8. Use the procedure in Exercise 7 to solve $y'' + y' - 12y = 0$.

9. Explain how you might use the procedure in Exercise 7 to solve $y''' - 6y'' + 11y' - 6y = 0$. Use your procedure to solve the equation.

Answer:

$y = c_1 e^{x} + c_2 e^{2x} + c_3 e^{3x}$

10. (a) By rewriting 11 in matrix form, show that the solution of the system in Example 2 can be expressed as

$$\mathbf{y} = c_1 e^{2x} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 e^{-3x} \begin{bmatrix} -\frac{1}{4} \\ 1 \end{bmatrix}$$

This is called the ***general solution*** of the system.

(b) Note that in part (a), the vector in the first term is an eigenvector corresponding to the eigenvalue $\lambda_1 = 2$, and the vector in the second term is an eigenvector corresponding to the eigenvalue $\lambda_2 = -3$. This is a special case of the following general result:

Theorem. If the coefficient matrix A of the system $\mathbf{y}' = A\mathbf{y}$ is diagonalizable, then the general solution of the system can be expressed as

$$\mathbf{y} = c_1 e^{\lambda_1 x}\mathbf{x}_1 + c_2 e^{\lambda_2 x}\mathbf{x}_2 + \ldots + c_n e^{\lambda_n x}\mathbf{x}_n$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A, and $\mathbf{x}_i$ is an eigenvector of A corresponding to λ_i.

Prove this result by tracing through the four-step procedure preceding Example 2 with

$$D = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix} \text{ and } P = [\mathbf{x}_1 | \mathbf{x}_2 | \ldots \mathbf{x}_n]$$

11. Consider the system of differential equations $\mathbf{y}' = A\mathbf{y}$, where A is a 2×2 matrix. For what values of $a_{11}, a_{12}, a_{21}, a_{22}$ do the component solutions $y_1(t), y_2(t)$ tend to zero as $t \to \infty$? In particular, what must be true about the determinant and the trace of A for this to happen?

12. Solve the nondiagonalizable system

$$\begin{aligned} y_1' &= y_1 + y_2 \\ y_2' &= y_2 \end{aligned}$$

True-False Exercises

In parts (a)–(e) determine whether the statement is true or false, and justify your answer.

(a) Every system of differential equations $\mathbf{y}' = A\mathbf{y}$ has a solution.

Answer:

False

(b) If $\mathbf{x}' = A\mathbf{x}$ and $\mathbf{y}' = A\mathbf{y}$, then $\mathbf{x} = \mathbf{y}$.

Answer:

False

(c) If $\mathbf{x}' = A\mathbf{x}$ and $\mathbf{y}' = A\mathbf{y}$, then $(c\mathbf{x} + d\mathbf{y})' = A(c\mathbf{x} + d\mathbf{y})$ for all scalars c and d.

Answer:

True

(d) If A is a square matrix with distinct real eigenvalues, then it is possible to solve $\mathbf{x}' = A\mathbf{x}$ by diagonalization.

Answer:

True

(e) If A and P are similar matrices, then $\mathbf{y}' = A\mathbf{y}$ and $\mathbf{u}' = P\mathbf{u}$ have the same solutions.

Answer:

False

Copyright © 2010 John Wiley & Sons, Inc. All rights reserved.

Chapter 5 Supplementary Exercises

1. (a) Show that if $0 < \theta < \pi$, then

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

has no eigenvalues and consequently no eigenvectors.

(b) Give a geometric explanation of the result in part (a).

Answer:

(b) The transformation rotates vectors through the angle θ; therefore, if $0 < \theta < \pi$, then no nonzero vector is transformed into a vector in the same or opposite direction.

2. Find the eigenvalues of

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ k^3 & -3k^2 & 3k \end{bmatrix}$$

3. (a) Show that if D is a diagonal matrix with nonnegative entries on the main diagonal, then there is a matrix S such that $S^2 = D$.

(b) Show that if A is a diagonalizable matrix with nonnegative eigenvalues, then there is a matrix S such that $S^2 = A$.

(c) Find a matrix S such that $S^2 = A$, given that

$$A = \begin{bmatrix} 1 & 3 & 1 \\ 0 & 4 & 5 \\ 0 & 0 & 9 \end{bmatrix}$$

Answer:

(c) $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}$

4. Prove: If A is a square matrix, then A and A^T have the same characteristic polynomial.

5. Prove: If A is a square matrix and $p(\lambda) = \det(\lambda I - A)$ is the characteristic polynomial of A, then the coefficient of λ^{n-1} in $p(\lambda)$ is the negative of the trace of A.

6. Prove: If $b \neq 0$, then

$$A = \begin{bmatrix} a & b \\ 0 & a \end{bmatrix}$$

is not diagonalizable.

7. In advanced linear algebra, one proves the ***Cayley—Hamilton Theorem***, which states that a square matrix

A satisfies its characteristic equation; that is, if

$$c_0 + c_1\lambda + c_2\lambda^2 + \ldots + c_{n-1}\lambda^{n-1} + \lambda^n = 0$$

is the characteristic equation of A, then

$$c_0 I + c_1 A + c_2 A^2 + \ldots + c_{n-1}A^{n-1} + A^n = 0$$

Verify this result for

$$\text{(a)}\quad A = \begin{bmatrix} 3 & 6 \\ 1 & 2 \end{bmatrix} \qquad \text{(b)}\quad A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -3 & 3 \end{bmatrix}$$

In Exercises 8–10, use the Cayley—Hamilton Theorem, stated in Exercise 7.

8. (a) Use Exercise 18 of Section 5.1 to prove the Cayley—Hamilton Theorem for 2×2 matrices.

(b) Prove the Cayley—Hamilton Theorem for $n \times n$ diagonalizable matrices.

9. The Cayley—Hamilton Theorem provides a method for calculating powers of a matrix. For example, if A is a 2×2 matrix with characteristic equation

$$c_0 + c_1\lambda + \lambda^2 = 0$$

then $c_0 I + c_1 A + A^2 = 0$, so

$$A^2 = -c_1 A - c_0 I$$

Multiplying through by A yields $A^3 = -c_1 A^2 - c_0 A$, which expresses A^3 in terms of A^2 and A, and multiplying through by A^2 yields $A^4 = -c_1 A^3 - c_0 A^2$, which expresses A^4 in terms of A^3 and A^2. Continuing in this way, we can calculate successive powers of A by expressing them in terms of lower powers. Use this procedure to calculate A^2, A^3, A^4, and A^5 for

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

Answer:

$$A^2 = \begin{bmatrix} 15 & 30 \\ 5 & 10 \end{bmatrix},\ A^3 = \begin{bmatrix} 75 & 150 \\ 25 & 50 \end{bmatrix},\ A^4 = \begin{bmatrix} 375 & 750 \\ 125 & 250 \end{bmatrix},\ A^5 = \begin{bmatrix} 1875 & 3750 \\ 625 & 1250 \end{bmatrix}$$

10. Use the method of the preceding exercise to calculate A^3 and A^4 for

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

11. Find the eigenvalues of the matrix

$$A = \begin{bmatrix} c_1 & c_2 & \ldots & c_n \\ c_1 & c_2 & \ldots & c_n \\ \vdots & \vdots & & \vdots \\ c_1 & c_2 & \ldots & c_n \end{bmatrix}$$

Answer:

$0,\ \mathrm{tr}(A)$

12. (a) It was shown in Exercise 17 of Section 5.1 that if A is an $n \times n$ matrix, then the coefficient of λ^n in the characteristic polynomial of A is 1. (A polynomial with this property is called ***monic***.) Show that the matrix

$$\begin{bmatrix} 0 & 0 & 0 & \dots & 0 & -c_0 \\ 1 & 0 & 0 & \dots & 0 & -c_1 \\ 0 & 1 & 0 & \dots & 0 & -c_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -c_{n-1} \end{bmatrix}$$

has characteristic polynomial

$$p(\lambda) = c_0 + c_1\lambda + \dots + c_{n-1}\lambda^{n-1} + \lambda^n$$

This shows that every monic polynomial is the characteristic polynomial of some matrix. The matrix in this example is called the ***companion matrix*** of $p(\lambda)$. [*Hint:* Evaluate all determinants in the problem by adding a multiple of the second row to the first to introduce a zero at the top of the first column, and then expanding by cofactors along the first column.]

(b) Find a matrix with characteristic polynomial

$$p(\lambda) = 1 - 2\lambda + \lambda^2 + 3\lambda^3 + \lambda^4$$

13. A square matrix A is called ***nilpotent*** if $A^n = 0$ for some positive integer n. What can you say about the eigenvalues of a nilpotent matrix?

Answer:

They are all 0.

14. Prove: If A is an $n \times n$ matrix and n is odd, then A has at least one real eigenvalue.

15. Find a 3×3 matrix A that has eigenvalues $\lambda = 0$, 1, and -1 with corresponding eigenvectors

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

respectively.

Answer:

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

16. Suppose that a 4×4 matrix A has eigenvalues $\lambda_1 = 1, \lambda_2 = -2, \lambda_3 = 3$, and $\lambda_4 = -3$.

(a) Use the method of Exercise 16 of Section 5.1 to find $\det(A)$.

(b) Use Exercise 5 above to find $\mathrm{tr}(A)$.

17. Let A be a square matrix such that $A^3 = A$. What can you say about the eigenvalues of A?

Answer:

They are all 0, 1, or -1.

18. (a) Solve the system

$$\begin{aligned} y_1' &= y_1 + 3y_2 \\ y_2' &= 2y_1 + 4y_2 \end{aligned}$$

(b) Find the solution satisfying the initial conditions $y_1(0) = 5$ and $y_2(0) = 6$.

Copyright © 2010 John Wiley & Sons, Inc. All rights reserved.

CHAPTER

6 Inner Product Spaces

CHAPTER CONTENTS

6.1. Inner Products
6.2. Angle and Orthogonality in Inner Product Spaces
6.3. Gram–Schmidt Process; QR-Decomposition
6.4. Best Approximation; Least Squares
6.5. Least Squares Fitting to Data
6.6. Function Approximation; Fourier Series

INTRODUCTION

In Chapter 3 we defined the dot product of vectors in R^n, and we used that concept to define notions of length, angle, distance, and orthogonality. In this chapter we will generalize those ideas so they are applicable in any vector space, not just R^n. We will also discuss various applications of these ideas.

Copyright © 2010 John Wiley & Sons, Inc. All rights reserved.

6.1 Inner Products

In this section we will use the most important properties of the dot product on R^n as axioms, which, if satisfied by the vectors in a vector space V, will enable us to extend the notions of length, distance, angle, and perpendicularity to general vector spaces.

General Inner Products

In Definition 4 of Section 3.2 we defined the dot product of two vectors in R^n, and in Theorem 3.2.2 we listed four fundamental properties of such products. Our first goal in this section is to extend the notion of a dot product to general real vector spaces by using those four properties as axioms. We make the following definition.

Note that Definition 1 applies only to *real* vector spaces. A definition of inner products on *complex* vector spaces is given in the exercises. Since we will have little need for complex vector spaces from this point on, you can assume that all vector spaces under discussion are real, even though some of the theorems are also valid in complex vector spaces.

DEFINITION 1

An ***inner product*** on a real vector space V is a function that associates a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ with each pair of vectors in V in such a way that the following axioms are satisfied for all vectors **u**, **v**, and **w** in V and all scalars k.

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$ [**Symmetry axiom**]
2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$ [**Additivity axiom**]
3. $\langle k\mathbf{u}, \mathbf{v} \rangle = \mathbf{k}\langle \mathbf{u}, \mathbf{v} \rangle$ [**Homogeneity axiom**]
4. $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = 0$ [**Positivity axiom**]

A real vector space with an inner product is called a ***real product space***.

Because the axioms for a real inner product space are based on properties of the dot product, these inner product space axioms will be satisfied automatically if we define the inner product of two vectors **u** and **v** in R^n to be

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v} = \mathbf{u}_1\mathbf{v}_1 + \mathbf{u}_2\mathbf{v}_2 + \cdots + \mathbf{u}_n\mathbf{v}_n$$

This inner product is commonly called the ***Euclidean inner product*** (or the ***standard inner product***) on R^n to distinguish it from other possible inner products that might be defined on R^n. We call R^n with the Euclidean inner product ***Euclidean n-space***.

Inner products can be used to define notions of norm and distance in a general inner product space just as we did with dot products in R^n. Recall from Formulas 11 and 19 of Section 3.2 that if **u** and **v** are vectors in Euclidean n-space, then norm and distance can be expressed in terms of the dot product as

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} \quad \text{and} \quad \mathbf{d}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{(\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v})}$$

Motivated by these formulas we make the following definition.

DEFINITION 2

If V is a real inner product space, then the ***norm*** (or ***length***) of a vector $\mathbf{v}$ in V is denoted by $\|\mathbf{v}\|$ and is defined by

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

and the ***distance*** between two vectors is denoted by $d(\mathbf{u}, \mathbf{v})$ and is defined by

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{\langle \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle}$$

A vector of norm 1 is called a ***unit vector***.

The following theorem, which we state without proof, shows that norms and distances in real inner product spaces have many of the properties that you might expect.

THEOREM 6.1.1

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in a real inner product space V, and if k is a scalar, then:

(a) $\|\mathbf{v}\| \geq 0$ with equality if and only if $\mathbf{v} = 0$.

(b) $\|k\mathbf{v}\| = |k|\|\mathbf{v}\|$.

(c) $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$.

(d) $d(\mathbf{u}, \mathbf{v}) \geq 0$ with equality if and only if $\mathbf{u} = \mathbf{v}$.

Although the Euclidean inner product is the most important inner product on R^n, there are various applications in which it is desirable to modify it by *weighting* each term differently. More precisely, if

$$w_1, w_2, \ldots, w_n$$

are *positive* real numbers, which we will call ***weights***, and if $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ are vectors in R^n, then it can be shown that the formula

$$\langle \mathbf{u}, \mathbf{v} \rangle = w_1 u_1 v_1 + w_2 u_2 v_2 + \cdots + w_n u_n v_n \tag{1}$$

defines an inner product on R^n that we call the ***weighted Euclidean inner product with weights*** $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n$.

Note that the standard Euclidean inner product is the special case of the weighted Euclidean inner product in which all the weights are 1.

EXAMPLE 1 Weighted Euclidean Inner Product

Let $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$ be vectors in R^2. Verify that the weighted Euclidean inner product

$$\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1 v_1 + 2u_2 v_2 \tag{2}$$

satisfies the four inner product axioms.

Solution Axiom 1: Interchanging $\mathbf{u}$ and $\mathbf{v}$ in Formula 2 does not change the sum on the right side, so

$\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$.

Axiom 2: If $\mathbf{w} = (w_1, w_2)$, then

$$\begin{aligned} \langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle &= 3(u_1 + v_1)w_1 + 2(u_2 + v_2)w_2 \\ &= 3(u_1w_1 + v_1w_1) + 2(u_2w_2 + v_2w_2) \\ &= (3u_1w_1 + 2u_2w_2) + (3v_1w_1 + 2v_2w_2) \\ &= \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle \end{aligned}$$

Axiom 3:

$$\begin{aligned} \langle k\mathbf{u}, \mathbf{v} \rangle &= 3(ku_1)v_1 + 2(ku_2)v_2 \\ &= k(3u_1v_1 + 2u_2v_2) \\ &= k\langle \mathbf{u}, \mathbf{v} \rangle \end{aligned}$$

Axiom 4: $\langle \mathbf{v}, \mathbf{v} \rangle = 3(v_1v_1) + 2(v_2v_2) = 3v_1^2 + 2v_2^2 \geq 0$ with equality if and only if $v_1 = v_2 = 0$; that is, if and only if $\mathbf{v} = \mathbf{0}$.

In Example 1, we are using subscripted w's to denote the components of thevector $\mathbf{w}$, not the weights. The weights are the numbers 3 and 2 in Formula 2.

An Application of Weighted Euclidean Inner Products

To illustrate one way in which a weighted Euclidean inner product can arise, suppose that some physical experiment has n possible numerical outcomes

$$x_1, x_2, \ldots, x_n$$

and that a series of m repetitions of the experiment yields these values with various frequencies. Specifically, suppose that x_1 occurs f_1 times, x_2 occurs f_2 times, and so forth. Since there are a total of m repetitions of the experiment, it follows that

$$f_1 + f_2 + \cdots + f_n = m$$

Thus, the ***arithmetic average*** of the observed numerical values (denoted by $\overline{x}$) is

$$\overline{x} = \frac{f_1x_1 + f_2x_2 + \cdots + f_nx_n}{f_1 + f_2 + \cdots + f_n} = \frac{1}{m}(f_1x_1 + f_2x_2 + \cdots + f_nx_n) \tag{3}$$

If we let

$$\begin{aligned} \mathbf{f} &= (f_1, f_2, \ldots, f_n) \\ \mathbf{x} &= (x_1, x_2, \ldots, x_n) \\ w_1 &= w_2 = \ldots = w_n = 1/m \end{aligned}$$

then 3 can be expressed as the weighted Euclidean inner product

$$\overline{x} = \langle \mathbf{f}, \mathbf{x} \rangle = w_1f_1x_1 + w_2f_2x_2 + \cdots + w_nf_nx_n$$

EXAMPLE 2 Using a Weighted Euclidean Inner Product ◀

It is important to keep in mind that norm and distance depend on the inner product being used. If the inner product is changed, then the norms and distances between vectors also change. For example, for the vectors $\mathbf{u} = (1, 0)$ and $\mathbf{v} = (0, 1)$ in R^2 with the Euclidean inner product we have

$$\|\mathbf{u}\| = \sqrt{1^2 + 0^2} = 1$$

and

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \|(1, -1)\| = \sqrt{1^2 + (-1)^2} = \sqrt{2}$$

but if we change to the weighted Euclidean inner product

$$\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1v_1 + 2u_2v_2$$

we have

$$\|\mathbf{u}\| = \langle \mathbf{u}, \mathbf{u} \rangle^{1/2} = [3(1)(1) + 2(0)(0)]^{1/2} = \sqrt{3}$$

and

$$\begin{aligned} d(\mathbf{u}, \mathbf{v}) &= \|\mathbf{u} - \mathbf{v}\| = \langle (1, -1), (1, -1) \rangle^{1/2} \\ &= [3(1)(1) + 2(-1)(-1)]^{1/2} = \sqrt{5} \end{aligned}$$

Unit Circles and Spheres in Inner Product Spaces

If V is an inner product space, then the set of points in V that satisfy

$$\|\mathbf{u}\| = 1$$

is called the ***unit sphere*** or sometimes the ***unit circle*** in V.

EXAMPLE 3 Unusual Unit Circles in R^2

(a) Sketch the unit circle in an xy-coordinate system in R^2 using the Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = u_1v_1 + u_2v_2$.

(b) Sketch the unit circle in an xy-coordinate system in R^2 using the weighted Euclidean inner product $\left\langle \mathbf{u}, \mathbf{v} \right\rangle = \frac{1}{9}u_1v_1 + \frac{1}{4}u_2v_2$.

Solution

(a) If $\mathbf{u} = (x, y)$, then $\|\mathbf{u}\| = \langle \mathbf{u}, \mathbf{u} \rangle^{1/2} = \sqrt{x^2 + y^2}$, so the equation of the unit circle is $\sqrt{x^2 + y^2} = 1$, or, on squaring both sides,

$$x^2 + y^2 = 1$$

As expected, the graph of this equation is a circle of radius 1 centered at the origin (Figure 6.1.1 *a*).

(b) If $\mathbf{u} = (x, y)$, then $\|\mathbf{u}\| = \langle \mathbf{u}, \mathbf{u} \rangle^{1/2} = \sqrt{\frac{1}{9}x^2 + \frac{1}{4}y^2}$, so the equation of the unit circle is $\sqrt{\frac{1}{9}x^2 + \frac{1}{4}y^2} = 1$, or, on squaring both sides,

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

The graph of this equation is the ellipse shown in Figure 6.1.1*b*.

(*a*) The unit circle using the standard Euclidean inner product.

(*b*) The unit circle using a weighted Euclidean inner product.

Figure 6.1.1

Remark It may seem odd that the "unit circle" in the second part of the last example turned out to have an elliptical shape. This will make more sense if you think of circles and spheres in general vector spaces *algebraically* $(\|\mathbf{u}\| = 1)$ rather than geometrically. The change in geometry occurs because the norm, not being Euclidean, has the effect of distorting the space that we are used to seeing through "Euclidean eyes."

Inner Products Generated by Matrices

The Euclidean inner product and the weighted Euclidean inner products are special cases of a general class of inner products on R^n called ***matrix inner products***. To define this class of inner products, let **u** and **v** be vectors in R^n that are expressed in *column form*, and let A be an *nvertible* $n \times n$ matrix. It can be shown (Exercise 31) that if $\mathbf{u} \cdot \mathbf{v}$ is the Euclidean inner product on R^n, then the formula

$$\langle \mathbf{u}, \mathbf{v} \rangle = A\mathbf{u} \cdot A\mathbf{v} \tag{4}$$

also defines an inner product; it is called the ***inner product on R^n generated by A***.

Recall from Table 1 of Section 3.2 that if **u** and **v** are in column form, then $\mathbf{u} \cdot \mathbf{v}$ can be written as $\mathbf{v}^T\mathbf{u}$ from which it follows that 4 can be expressed as

$$\langle \mathbf{u}, \mathbf{v} \rangle = (A\mathbf{v})^T A\mathbf{u}$$

or, equivalently as

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{v}^T A^T A \mathbf{u} \tag{5}$$

EXAMPLE 4 Matrices Generating Weighted Euclidean Inner Products

The standard Euclidean and weighted Euclidean inner products are examples of matrix inner products. The standard Euclidean inner product on R^n is generated by the $n \times n$ identity matrix, since setting $A = I$ in Formula 4 yields

$$\langle \mathbf{u}, \mathbf{v} \rangle = I\mathbf{u} \cdot I\mathbf{v} = \mathbf{u} \cdot v$$

and the weighted Euclidean inner product

$$\langle \mathbf{u}, \mathbf{v} \rangle = w_1 u_1 v_1 + w_2 u_2 v_2 + \cdots + w_n u_n v_n \tag{6}$$

is generated by the matrix

$$A = \begin{bmatrix} \sqrt{w_1} & 0 & 0 & \dots & 0 \\ 0 & \sqrt{w_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \sqrt{w_n} \end{bmatrix} \tag{7}$$

This can be seen by first observing that $A^T A$ is the $n \times n$ diagonal matrix whose diagonal entries are the weights $w_1, w_2, \ldots, w_n$ and then observing that 5 simplifies to 6 when A is the matrix in Formula 7.

EXAMPLE 5 Example 1 Revisited

Every diagonal matrix with positive diagonal entries generates a weighted inner product. Why?

The weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1 v_1 + 2u_2 v_2$ discussed in Example 1 is the inner product on R^2 generated by

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

Other Examples of Inner Products

So far, we have only considered examples of inner products on R^n. We will now consider examples of inner products on some of the other kinds of vector spaces that we discussed earlier.

EXAMPLE 6 An Inner Product on M_{nn}

If U and V are $n \times n$ matrices, then the formula

$$\langle U, V\rangle = \operatorname{tr}\left(U^T V\right) \tag{8}$$

defines an inner product on the vector space M_{nn} (see Definition 8 of Section 1.3 for a definition of trace). This can be proved by confirming that the four inner product space axioms are satisfied, but you can visualize why this is so by computing 8 for the 2×2 matrices

$$U = \begin{bmatrix} u_1 & u_2 \\ u_3 & u_4 \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} v_1 & v_2 \\ v_3 & v_4 \end{bmatrix}$$

This yields

$$\langle U, V\rangle = \operatorname{tr}\left(U^T V\right) = u_1 v_1 + u_2 v_2 + u_3 v_3 + u_4 v_4$$

which is just the dot product of the corresponding entries in the two matrices. For example, if

$$U = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \text{ and } V = \begin{bmatrix} -1 & 0 \\ 3 & 2 \end{bmatrix}$$

then

$$\langle U, V\rangle = 1(-1) + 2(0) + 3(3) + 4(2) = 16$$

The norm of a matrix U relative to this inner product is

$$\|U\| = \langle U, U\rangle^{1/2} = \sqrt{u_1^2 + u_2^2 + u_3^2 + u_4^2}$$

EXAMPLE 7 The Standard Inner Product on P_n

If

$$\mathbf{p} = a_0 + a_1 x + \cdots + a_n x^n \text{ and } \mathbf{q} = b_0 + b_1 x + \cdots + b_n x^n$$

are polynomials in P_n, then the following formula defines an inner product on P_n (verify) that we will call the ***standard inner product*** on this space:

$$\langle \mathbf{p}, \mathbf{q}\rangle = a_0 b_0 + a_1 b_1 + \cdots + a_n b_n \tag{9}$$

The norm of a polynomial $\mathbf{p}$ relative to this inner product is

$$\|\mathbf{p}\| = \sqrt{\langle \mathbf{p}, \mathbf{p}\rangle} = \sqrt{a_0^2 + a_1^2 + \cdots + a_n^2}$$

EXAMPLE 8 The Evaluation Inner Product on P_n

If

$$\mathbf{p} = p(x) = a_0 + a_1 x + \cdots + a_n x^n \text{ and } \mathbf{q} = \mathrm{q}(\mathrm{x}) = \mathrm{b}_0 + \mathrm{b}_1 \mathrm{x} + \cdots + \mathrm{b}_n \mathrm{x}^n$$

are polynomials in P_n, and if $x_0, x_1, \ldots, x_n$ are distinct real numbers (called ***sample points***), then the formula

$$\langle \mathbf{p}, \mathbf{q}\rangle = p(x_0)q(x_0) + p(x_1)q(x_1) + \cdots + p(x_n)q(x_n) \tag{10}$$

defines an inner product on P_n called the ***evaluation inner product*** at $x_0, x_1, \ldots, x_n$. Algebraically, this can be viewed as the dot product in R^n of the n-tuples

$$(p(x_0), p(x_1), \ldots, p(x_n)) \quad \text{and} \quad (q(x_0), q(x_1), \ldots, q(x_n))$$

and hence the first three inner product axioms follow from properties of the dot product. The fourth inner product axiom follows from the fact that

$$\langle \mathbf{p}, \mathbf{p} \rangle = [p(x_0)]^2 + [p(x_1)]^2 + \cdots + [p(x_n)]^2 \geq 0$$

with equality holding if and only if

$$p(x_0) = p(x_1) = \ldots = p(x_n) = 0$$

But a nonzero polynomial of degree n or less can have at most n distinct roots, so it must be that $\mathbf{p} = \mathbf{0}$, which proves that the fourth inner product axiom holds.

The norm of a polynomial $\mathbf{p}$ relative to the evaluation inner product is

$$\|\mathbf{p}\| = \sqrt{\langle \mathbf{p}, \mathbf{p} \rangle} = \sqrt{[p(x_0)]^2 + [p(x_1)]^2 + \cdots + [p(x_n)]^2} \tag{11}$$

EXAMPLE 9 Working with the Evaluation Inner Product

Let P_2 have the evaluation inner product at the points

$$x_0 = -2, \ x_1 = 0, \ \text{and } x_2 = 2$$

Compute $\langle \mathbf{p}, \mathbf{q} \rangle$ and $\|\mathbf{p}\|$ for the polynomials $\mathbf{p} = p(x) = x^2$ and $\mathbf{q} = q(x) = 1 + x$.

Solution It follows from 10 and 11 that

$$\langle \mathbf{p}, \mathbf{q} \rangle = p(-2)q(-2) + p(0)q(0) + p(2)q(2) = (4)(-1) + (0)(1) + (4)(3) = 8$$

$$\begin{aligned} \|\mathbf{p}\| &= \sqrt{[p(x_0)]^2 + [p(x_1)]^2 + [p(x_2)]^2} = \sqrt{[p(-2)]^2 + [p(0)]^2 + [p(2)]^2} \\ &= \sqrt{4^2 + 0^2 + 4^2} = \sqrt{32} = 4\sqrt{2} \end{aligned}$$

CALCULUS REQUIRED

EXAMPLE 10 An Inner Product on $C[a, b]$

Let $\mathbf{f} = f(x)$ and $\mathbf{g} = g(x)$ be two functions in $C[a, b]$ and define

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b f(x)g(x)\,dx \tag{12}$$

We will show that this formula defines an inner product on $C[a, b]$ by verifying the four inner product axioms for functions $\mathbf{f} = f(x)$, $\mathbf{g} = g(x)$, and $\mathbf{h} = h(x)$ in $C[a, b]$:

1.

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b f(x)g(x)\,dx = \int_a^b g(x)f(x)\,dx = \langle \mathbf{g}, \mathbf{f} \rangle$$

which proves that Axiom 1 holds.

2.

$$\begin{aligned} \langle \mathbf{f} + \mathbf{g}, \mathbf{h} \rangle &= \int_a^b (f(x) + g(x))h(x)\,dx \\ &= \int_a^b f(x)h(x)\,dx + \int_a^b g(x)h(x)\,dx \\ &= \langle \mathbf{f}, \mathbf{h} \rangle + \langle \mathbf{g}, \mathbf{h} \rangle \end{aligned}$$

which proves that Axiom 2 holds.

3. $$\langle k\mathbf{f}, \mathbf{g}\rangle = \int_a^b kf(x)g(x)\,dx = k\int_a^b f(x)g(x)\,dx = k\langle \mathbf{f}, \mathbf{g}\rangle$$

which proves that Axiom 3 holds.

4. If $\mathbf{f} = f(x)$ is any function in $C[a, b]$, then

$$\langle \mathbf{f}, \mathbf{f}\rangle = \int_a^b f^2(x)\,dx \geq 0 \tag{13}$$

since $f^2(x) \geq 0$ for all x in the interval $[a, b]$. Moreover because f is continuous on $[a, b]$, the equality holds in Formula 13 if and only if the function f is identically zero on $[a, b]$, that is, if and only if $\mathbf{f} = 0$; and this proves that Axiom 4 holds.

CALCULUS REQUIRED

EXAMPLE 11 Norm of a Vector in $C[a, b]$

If $C[a, b]$ has the inner product that was defined in Example 10, then the norm of a function $\mathbf{f} = f(x)$ relative to this inner product is

$$\|\mathbf{f}\| = \langle \mathbf{f}, \mathbf{f}\rangle^{1/2} = \sqrt{\int_a^b f^2(x)\,dx} \tag{14}$$

and the unit sphere in this space consists of all functions $\mathbf{f}$ in $C[a, b]$ that satisfy the equation

$$\int_a^b f^2(x)\,dx = 1$$

Remark Note that the vector space P_n is a subspace of $C[a, b]$ because polynomials are continuous functions. Thus, Formula 12 defines an inner product on P_n.

Remark Recall from calculus that the arc length of a curve $y = f(x)$ over an interval $[a, b]$ is given by the formula

$$L = \int_a^b \sqrt{1 + [f'(x)]^2}\,dx \tag{15}$$

Do not confuse this concept of arc length with $\|\mathbf{f}\|$, which is the length (norm) of $\mathbf{f}$ when $\mathbf{f}$ is viewed as a vector in $C[a, b]$. Formulas 14 and 15 are quite different.

Algebraic Properties of Inner Products

The following theorem lists some of the algebraic properties of inner products that follow from the inner product axioms. This result is a generalization of Theorem 3.2.3, which applied only to the dot product on R^n.

THEOREM 6.1.2

If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in a real inner product space V, and if k is a scalar, then

(a) $\langle \mathbf{0}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{0} \rangle = 0$
(b) $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$
(c) $\langle \mathbf{u}, \mathbf{v} - \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle - \langle \mathbf{u}, \mathbf{w} \rangle$
(d) $\langle \mathbf{u} - \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle - \langle \mathbf{v}, \mathbf{w} \rangle$
(e) $k\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, k\mathbf{v} \rangle$

Proof We will prove part (*b*) and leave the proofs ofthe remaining parts as exercises.

$$\begin{aligned}
\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle &= \langle \mathbf{v} + \mathbf{w}, \mathbf{u} \rangle && \text{[By symmetry]} \\
&= \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{w}, \mathbf{u} \rangle && \text{[By additivity]} \\
&= \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle && \text{[By symmetry]}
\end{aligned}$$

The following example illustrates how Theorem 6.1.2 and the defining properties of inner products can be used to perform algebraic computations with inner products. As you read through the example, you will find it instructive to justify the steps.

EXAMPLE 12 Calculating with Inner Products

$$\begin{aligned}
\langle \mathbf{u} - 2\mathbf{v}, 3\mathbf{u} + 4\mathbf{v} \rangle &= \langle \mathbf{u}, 3\mathbf{u} + 4\mathbf{v} \rangle - \langle 2\mathbf{v}, 3\mathbf{u} + 4\mathbf{v} \rangle \\
&= \langle \mathbf{u}, 3\mathbf{u} \rangle + \langle \mathbf{u}, 4\mathbf{v} \rangle - \langle 2\mathbf{v}, 3\mathbf{u} \rangle - \langle 2\mathbf{v}, 4\mathbf{v} \rangle \\
&= 3\langle \mathbf{u}, \mathbf{u} \rangle + 4\langle \mathbf{u}, \mathbf{v} \rangle - 6\langle \mathbf{v}, \mathbf{u} \rangle - 8\langle \mathbf{v}, \mathbf{v} \rangle \\
&= 3\|\mathbf{u}\|^2 + 4\langle \mathbf{u}, \mathbf{v} \rangle - 6\langle \mathbf{u}, \mathbf{v} \rangle - 8\|\mathbf{v}\|^2 \\
&= 3\|\mathbf{u}\|^2 - 2\langle \mathbf{u}, \mathbf{v} \rangle - 8\|\mathbf{v}\|^2
\end{aligned}$$

Concept Review

- Inner product axioms
- Euclidean inner product
- Euclidean *n*-space
- Weighted Euclidean inner product
- Unit circle (sphere)
- Matrix inner product
- Norm in an inner product space
- Distance between two vectors in an inner product space
- Examples of inner products
- Properties of inner products

Skills

- Compute the inner product of two vectors.
- Find the norm of a vector.
- Find the distance between two vectors.

- Show that a given formula defines an inner product.
- Show that a given formula does not define an inner product by demonstrating that at least one of the inner product space axioms fails.

Exercise Set 6.1

1. Let $\langle \mathbf{u}, \mathbf{v} \rangle$ be the Euclidean inner product on R^2, and let $\mathbf{u} = (1, 1)$, $\mathbf{v} = (3, 2)$, $\mathbf{w} = (0, -1)$, and $k = 3$. Compute the following.

(a) $\langle \mathbf{u}, \mathbf{v} \rangle$

(b) $\langle k\mathbf{v}, \mathbf{w} \rangle$

(c) $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle$

(d) $\|\mathbf{v}\|$

(e) $d(\mathbf{u}, \mathbf{v})$

(f) $\|\mathbf{u} - k\mathbf{v}\|$

Answer:

(a) 5

(b) -6

(c) -3

(d) $\sqrt{13}$

(e) $\sqrt{5}$

(f) $\sqrt{89}$

2. Repeat Exercise 1 for the weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 2u_1v_1 + 3u_2v_2$.

3. Let $\langle \mathbf{u}, \mathbf{v} \rangle$ be the Euclidean inner product on R^2, and let $\mathbf{u} = (3, -2)$, $\mathbf{v} = (4, 5)$, $\mathbf{w} = (-1, 6)$, and $k = -4$. Verify the following.

(a) $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$

(b) $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$

(c) $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$

(d) $\langle k\mathbf{u}, \mathbf{v} \rangle = k\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, k\mathbf{v} \rangle$

(e) $\langle \mathbf{0}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{0} \rangle = 0$

Answer:

(a) 2

(b) 11

(c) -13

(d) -8

(e) 0

4. Repeat Exercise 3 for the weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 4u_1v_1 + 5u_2v_2$.

5. Let $\langle \mathbf{u}, \mathbf{v} \rangle$ be the inner product on R^2 generated by $\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$, and let $\mathbf{u} = (2, 1)$, $\mathbf{v} = (-1, 1)$, $\mathbf{w} = (0, -1)$. Compute the following.

(a) $\langle \mathbf{u}, \mathbf{v} \rangle$
(b) $\langle \mathbf{v}, \mathbf{w} \rangle$
(c) $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle$
(d) $\|\mathbf{v}\|$
(e) $d(\mathbf{v}, \mathbf{w})$
(f) $\|\mathbf{v} - \mathbf{w}\|^2$

Answer:

(a) -5
(b) 1
(c) -7
(d) 1
(e) 1
(f) 1

6. Repeat Exercise 5 for the inner product on R^2 generated by $\begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}$.

7. Compute $\langle \mathbf{u}, \mathbf{v} \rangle$ using the inner product in Example 6.

(a) $\mathbf{u} = \begin{bmatrix} 3 & -2 \\ 4 & 8 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -1 & 3 \\ 1 & 1 \end{bmatrix}$
(b) $\mathbf{u} = \begin{bmatrix} 1 & 2 \\ -3 & 5 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 4 & 6 \\ 0 & 8 \end{bmatrix}$

Answer:

(a) 3
(b) 56

8. Compute $\langle \mathbf{p}, \mathbf{q} \rangle$ using the inner product in Example 7.

(a) $\mathbf{p} = -2 + x + 3x^2$, $\mathbf{q} = 4 - 7x^2$
(b) $\mathbf{p} = -5 + 2x + x^2$, $\mathbf{q} = 3 + 2x - 4x^2$

9. (a) Use Formula 4 to show that $\langle \mathbf{u}, \mathbf{v} \rangle = 9u_1v_1 + 4u_2v_2$ is the inner product on R^2 generated by

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

(b) Use the inner product in part (a) to compute $\langle \mathbf{u}, \mathbf{v} \rangle$ if $\mathbf{u} = (-3, 2)$ and $\mathbf{v} = (1, 7)$.

Answer:

(b) 29

10. (a) Use Formula 4 to show that

$$\langle \mathbf{u}, \mathbf{v} \rangle = 5u_1v_1 - u_1v_2 - u_2v_1 + 10u_2v_2$$

is the inner product on R^2 generated by

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

(b) Use the inner product in part (a) to compute $\langle \mathbf{u}, \mathbf{v} \rangle$ if $\mathbf{u} = (0, -3)$ and $\mathbf{v} = (6, 2)$.

11. Let $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$. In each part, the given expression is an inner product on R^2. Find a matrix that generates it.

(a) $\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1v_1 + 5u_2v_2$

(b) $\langle \mathbf{u}, \mathbf{v} \rangle = 4u_1v_1 + 6u_2v_2$

Answer:

(a) $\begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{5} \end{bmatrix}$

(b) $\begin{bmatrix} 2 & 0 \\ 0 & \sqrt{6} \end{bmatrix}$

12. Let P_2 have the inner product in Example 7. In each part, find $\|\mathbf{p}\|$.

(a) $\mathbf{p} = -2 + 3x + 2x^2$

(b) $\mathbf{p} = 4 - 3x^2$

13. Let M_{22} have the inner product in Example 6. In each part, find $\|A\|$.

(a) $A = \begin{bmatrix} -2 & 5 \\ 3 & 6 \end{bmatrix}$

(b) $A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

Answer:

(a) $\sqrt{74}$

(b) 0

14. Let P_2 have the inner product in Example 7. Find $d(\mathbf{p}, \mathbf{q})$.

$$\mathbf{p} = 3 - x + x^2, \quad \mathbf{q} = 2 + 5x^2$$

15. Let M_{22} have the inner product in Example 6. Find $d(A, B)$.

(a) $A = \begin{bmatrix} 2 & 6 \\ 9 & 4 \end{bmatrix}, B = \begin{bmatrix} -4 & 7 \\ 1 & 6 \end{bmatrix}$

(b) $A = \begin{bmatrix} -2 & 4 \\ 1 & 0 \end{bmatrix}, B = \begin{bmatrix} -5 & 1 \\ 6 & 2 \end{bmatrix}$

Answer:

(a) $\sqrt{105}$

(b) $\sqrt{47}$

16. Let P_2 have the inner product of Example 9, and let $\mathbf{p} = 1 + x + x^2$ and $\mathbf{q} = 1 - 2x^2$. Compute the following.

(a) $\langle \mathbf{p}, \mathbf{q} \rangle$

(b) $\|\mathbf{p}\|$

(c) $d(\mathbf{p}, \mathbf{q})$

17. Let P_3 have the evaluation inner product at the sample points

$$x_0 = -1, \ x_1 = 0, \ x_2 = 1, \ x_3 = 2$$

Find $\langle \mathbf{p}, \mathbf{q} \rangle$ and $\|\mathbf{p}\|$ for $\mathbf{p} = x + x^3$ and $\mathbf{q} = 1 + x^2$.

Answer:

$\langle \mathbf{p}, \mathbf{q} \rangle = 50, \ \|\mathbf{p}\| = 6\sqrt{3}$

18. In each part, use the given inner product on R^2 to find $\|\mathbf{w}\|$, where $\mathbf{w} = (-1, 3)$.

(a) the Euclidean inner product

(b) the weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1v_1 + 2u_2v_2$, where $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$

(c) the inner product generated by the matrix

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

19. Use the inner products in Exercise 18 to find $d(\mathbf{u}, \mathbf{v})$ for $\mathbf{u} = (-1, 2)$ and $\mathbf{v} = (2, 5)$.

Answer:

(a) $3\sqrt{2}$

(b) $3\sqrt{5}$

(c) $3\sqrt{13}$

20. Suppose that $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors such that

$$\langle \mathbf{u}, \mathbf{v} \rangle = 2, \ \langle \mathbf{v}, \mathbf{w} \rangle = -3, \ \langle \mathbf{u}, \mathbf{w} \rangle = 5$$
$$\|\mathbf{u}\| = 1, \ \|\mathbf{v}\| = 2, \ \|\mathbf{w}\| = 7$$

Evaluate the given expression.

(a) $\langle \mathbf{u} + \mathbf{v}, \mathbf{v} + \mathbf{w} \rangle$

(b) $\langle 2\mathbf{v} - \mathbf{w}, 3\mathbf{u} + 2\mathbf{w} \rangle$

(c) $\langle \mathbf{u} - \mathbf{v} - 2\mathbf{w}, 4\mathbf{u} + \mathbf{v} \rangle$

(d) $\|\mathbf{u} + \mathbf{v}\|$

(e) $\|2\mathbf{w} - \mathbf{v}\|$

(f) $\|\mathbf{u} - 2\mathbf{v} + 4\mathbf{w}\|$

21. Sketch the unit circle in R^2 using the given inner product.

(a) $\left\langle \mathbf{u}, \mathbf{v} \right\rangle = \frac{1}{4}u_1v_1 + \frac{1}{16}u_2v_2$

(b) $\langle \mathbf{u}, \mathbf{v} \rangle = 2u_1v_1 + u_2v_2$

Answer:

(a)

(b)

22. Find a weighted Euclidean inner product on R^2 for which the unit circle is the ellipse shown in the accompanying figure.

Figure Ex-22

23. Let $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$. Show that the following are inner products on R^2 by verifying that the inner product axioms hold.

(a) $\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1v_1 + 5u_2v_2$

(b) $\langle \mathbf{u}, \mathbf{v} \rangle = 4u_1v_1 + u_2v_1 + u_1v_2 + 4u_2v_2$

Answer:

For $V = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, then $\langle V, V \rangle = -2 < 0$, so Axiom 4 fails.

24. Let $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$. Determine which of the following are inner products on R^3. For those that are not, list the axioms that do not hold.

(a) $\langle \mathbf{u}, \mathbf{v} \rangle = u_1v_1 + u_3v_3$

(b) $\left\langle \mathbf{u}, \mathbf{v} \right\rangle = u_1^2v_1^2 + u_2^2v_2^2 + u_3^2v_3^2$

(c) $\langle \mathbf{u}, \mathbf{v} \rangle = 2u_1v_1 + u_2v_2 + 4u_3v_3$

(d) $\langle \mathbf{u}, \mathbf{v} \rangle = u_1v_1 - u_2v_2 + u_3v_3$

25. Show that the following identity holds for vectors in any inner product space.

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

Answer:

(a) $-\frac{28}{15}$

(b) 0

26. Show that the following identity holds for vectors in any inner product space.

$$\left\langle \mathbf{u}, \mathbf{v} \right\rangle = \frac{1}{4}\|\mathbf{u} + \mathbf{v}\|^2 - \frac{1}{4}\|\mathbf{u} - \mathbf{v}\|^2$$

27. Let $U = \begin{bmatrix} u_1 & u_2 \\ u_3 & u_4 \end{bmatrix}$ and $V = \begin{bmatrix} v_1 & v_2 \\ v_3 & v_4 \end{bmatrix}$. Show that $\langle U, V \rangle = u_1v_1 + u_2v_3 + u_3v_2 + u_4v_4$ is *not* an inner product on M_{22}.

28. ***Calculus required*** Let the vector space P_2 have the inner product

$$\left\langle \mathbf{p}, \mathbf{q} \right\rangle = \int_{-1}^{1} p(x)q(x)\, dx$$

(a) Find $\|\mathbf{p}\|$ for $\mathbf{p} = 1$, $\mathbf{p} = x$, and $\mathbf{p} = x^2$.

(b) Find $d(\mathbf{p}, \mathbf{q})$ if $\mathbf{p} = 1$ and $\mathbf{q} = x$.

29. ***Calculus required*** Use the inner product

$$\langle \mathbf{p}, \mathbf{q} \rangle = \int_{-1}^{1} p(x)q(x)\,dx$$

on P_3, to compute $\langle \mathbf{p}, \mathbf{q} \rangle$.

(a) $\mathbf{p} = 1 - x + x^2 + 5x^3$, $\mathbf{q} = x - 3x^2$

(b) $\mathbf{p} = x - 5x^3$, $\mathbf{q} = 2 + 8x^2$

30. ***Calculus required*** In each part, use the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_{0}^{1} f(x)g(x)\,dx$$

on $C[0, 1]$ to compute $\langle \mathbf{f}, \mathbf{g} \rangle$.

(a) $\mathbf{f} = \cos 2\pi x$, $\mathbf{g} = \sin 2\pi x$

(b) $\mathbf{f} = x$, $\mathbf{g} = e^x$

(c) $\mathbf{f} = \tan\frac{\pi}{4}x$, $\mathbf{g} = 1$

31. Prove that Formula 4 defines an inner product on R^n.

32. The definition of a complex vector space was given in the first margin note in Section 4.1. The definition of a ***complex inner product*** on a complex vector space V is identical to Definition 1 except that scalars are allowed to be complex numbers, and Axiom 1 is replaced by $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$. The remaining axioms are unchanged. A complex vector space with a complex inner product is called a ***complex inner product space***. Prove that if V is a complex inner product space then $\langle \mathbf{u}, k\mathbf{v} \rangle = \bar{k}\langle \mathbf{u}, \mathbf{v} \rangle$.

True-False Exercises

In parts (a)–(g) determine whether the statement is true or false, and justify your answer.

(a) The dot product on R^2 is an example of a weighted inner product.

Answer:

True

(b) The inner product of two vectors cannot be a negative real number.

Answer:

False

(c) $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{w}, \mathbf{u} \rangle$.

Answer:

True

(d) $\langle k\mathbf{u}, k\mathbf{v} \rangle = k^2 \langle \mathbf{u}, \mathbf{v} \rangle$.

Answer:

True

(e) If $\langle \mathbf{u}, \mathbf{v} \rangle = 0$, then $\mathbf{u} = 0$ or $\mathbf{v} = 0$.

Answer:

False

(f) If $\|\mathbf{v}\|^2 = 0$, then $\mathbf{v} = 0$.

Answer:

True

(g) If A is an $n \times n$ matrix, then $\langle \mathbf{u}, \mathbf{v} \rangle = A\mathbf{u} \cdot A\mathbf{v}$ defines an inner product on R^n.

Answer:

False

Copyright © 2010 John Wiley & Sons, Inc. All rights reserved.

6.2 Angle and Orthogonality in Inner Product Spaces

In Section 3.2 we defined the notion of "angle" between vector in R^n. In this section we will extend this idea to general vector spaces. This will enable us to extend the notion of orthogonality as well, thereby setting the groundwork for a variety of new applications.

Cauchy–Schwarz Inequality

Recall from Formula 20 of Section 3.2 that the angle θ between two vectors **u** and **v** in R^n is

$$\theta = \cos^{-1}\left(\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|}\right) \tag{1}$$

We were assured that this formula was valid because it followed from the Cauchy–Schwarz inequality (Theorem 3.2.4) that

$$-1 \le \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|} \le 1 \tag{2}$$

as required for the inverse cosine to be defined. The following generalization of Theorem 3.2.4 will enable us to define the angle between two vectors in *any* real inner product space.

THEOREM 6.2.1 Cauchy–Schwarz Inequality

If **u** and **v** are vectors in a real inner product space V, then

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \le \|\mathbf{u}\|\,\|\mathbf{v}\| \tag{3}$$

Proof We warn you in advance that the proof presented here depends on a clever trick that is not easy to motivate.

In the case where $\mathbf{u} = \mathbf{0}$ the two sides of 3 are equal since $\langle \mathbf{u}, \mathbf{v}\rangle$ and $\|\mathbf{u}\|$ are both zero. Thus, we need only consider the case where $\mathbf{u} \neq 0$. Making this assumption, let

$$a = \langle \mathbf{u}, \mathbf{u}\rangle, \quad b = 2\langle \mathbf{u}, \mathbf{v}\rangle, \quad c = \langle \mathbf{v}, \mathbf{v}\rangle$$

and let t be any real number. Since the positivity axiom states that the inner product of any vector with itself is nonnegative, it follows that

$$\begin{aligned} 0 \le \langle t\mathbf{u}+\mathbf{v},\ t\mathbf{u}+\mathbf{v}\rangle &= \langle \mathbf{u},\ \mathbf{u}\rangle t^2 + 2\langle \mathbf{u},\ \mathbf{v}\rangle t + \langle \mathbf{v},\ \mathbf{v}\rangle \\ &= at^2 + bt + c \end{aligned}$$

This inequality implies that the quadratic polynomial $at^2 + bt + c$ has either no real roots or a repeated real root. Therefore, its discriminant must satisfy the inequality $b^2 - 4ac \leq 0$. Expressing the coefficients a, b, and c in terms of the vectors **u** and **v** gives $4\langle \mathbf{u}, \mathbf{v}\rangle^2 - 4\langle \mathbf{u}, \mathbf{u}\rangle\langle \mathbf{v}, \mathbf{v}\rangle \leq 0$ or, equivalently,

$$\langle \mathbf{u}, \mathbf{v}\rangle^2 \leq \langle \mathbf{u}, \mathbf{u}\rangle\langle \mathbf{v}, \mathbf{v}\rangle$$

Taking square roots of both sides and using the fact that $\langle \mathbf{u}, \mathbf{u}\rangle$ and $\langle \mathbf{v}, \mathbf{v}\rangle$ are nonnegative yields

$$\left|\langle \mathbf{u}, \mathbf{v}\rangle\right| \leq \langle \mathbf{u}, \mathbf{u}\rangle^{1/2}\langle \mathbf{v}, \mathbf{v}\rangle^{1/2} \quad \textit{or equivalently} \quad \left|\langle \mathbf{u}, \mathbf{v}\rangle\right| \leq \|\mathbf{u}\|\|\mathbf{v}\|$$

which completes the proof.

The following two alternative forms of the Cauchy–Schwarz inequality are useful to know:

$$\langle \mathbf{u}, \mathbf{v}\rangle^2 \leq \langle \mathbf{u}, \mathbf{u}\rangle\langle \mathbf{v}, \mathbf{v}\rangle \tag{4}$$

$$\langle \mathbf{u}, \mathbf{v}\rangle^2 \leq \|\mathbf{u}\|^2\|\mathbf{v}\|^2 \tag{5}$$

The first of these formulas was obtained in the proof of Theorem 6.2.1, and the second is a variation of the first.

Angle Between Vectors

Our next goal is to define what is meant by the "angle" between vectors in a real inner product space. As the first step, we leave it for you to use the Cauchy–Schwarz inequality to show that

$$-1 \leq \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\|\mathbf{v}\|} \leq 1 \tag{6}$$

This being the case, there is a unique angle θ in radian measure for which

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\|\mathbf{v}\|} \quad \text{and} \quad 0 \leq \theta \leq \pi \tag{7}$$

(Figure 6.2.1). This enables us to *define* the ***angle θ between*** **u** and **v** to be

$$\theta = \cos^{-1}\left(\frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\|\mathbf{v}\|}\right) \tag{8}$$

Figure 6.2.1

EXAMPLE 1 Cosine of an Angle Between Two Vectors in R^4

Let R^4 have the Euclidean inner product. Find the cosine of the angle θ between the vectors $\mathbf{u} = (4, 3, 1, -2)$ and $\mathbf{v} = (-2, 1, 2, 3)$.

Solution We leave it for you to verify that

$$\|\mathbf{u}\| = \sqrt{30}, \quad \|\mathbf{v}\| = \sqrt{18}, \quad \text{and} \quad \langle \mathbf{u}, \mathbf{v} \rangle = -9$$

from which it follows that

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\|\mathbf{v}\|} = -\frac{9}{\sqrt{30}\sqrt{18}} = -\frac{3}{2\sqrt{15}}$$

Properties of Length and Distance in General Inner Product Spaces

In Section 3.2 we used the dot product to extend the notions of length and distance to R^n, and we showed that various familiar theorems remained valid (see Theorem 3.2.5, Theorem 3.2.6, and Theorem 3.2.7). By making only minor adjustments to the proofs of those theorems, we can show that they remain valid in any real inner product space. For example, here is the generalization of Theorem 3.2.5 (the triangle inequalities).

THEOREM 6.2.2

If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in a real inner product space V, and if k is any scalar, then:

(a) $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ [**Triangle inequality for vectors**]

(b) $d(\mathbf{u}, \mathbf{v}) \leq d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v})$ [**Triangle inequality for distances**]

Proof (a)

$$\begin{aligned}
\|\mathbf{u} + \mathbf{v}\|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle \\
&= \langle \mathbf{u}, \mathbf{u} \rangle + 2\langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\
&\leq \langle \mathbf{u}, \mathbf{u} \rangle + |\langle \mathbf{u}, \mathbf{v} \rangle| + \langle \mathbf{v}, \mathbf{v} \rangle && \text{[Property of absolute value]} \\
&\leq \langle \mathbf{u}, \mathbf{u} \rangle + 2\|\mathbf{u}\|\|\mathbf{v}\| + \langle \mathbf{v}, \mathbf{v} \rangle && \text{[By (3)]} \\
&= \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| + \|\mathbf{v}\|^2 \\
&= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2
\end{aligned}$$

Taking square roots gives $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

Proof (b) Identical to the proof of part (*b*) of Theorem 3.2.5.

Orthogonality

Although Example 1 is a useful mathematical exercise, there is only an occasional need to compute angles in vector spaces other than R^2 and R^3. A problem of more interest in general vector spaces is ascertaining whether the angle between vectors is $\pi / 2$. You should be able to see from Formula 8 that if **u** and **v** are *nonzero* vectors, then the angle between them is $\theta = \pi / 2$ if and only if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. Accordingly, we make the following definition (which is applicable even if one or both of the vectors is zero).

DEFINITION 1

Two vectors **u** and **v** in an inner product space are called ***orthogonal*** if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

As the following example shows, orthogonality depends on the inner product in the sense that for different inner products two vectors can be orthogonal with respect to one but not the other.

EXAMPLE 2 Orthogonality Depends on the Inner Product

The vectors $\mathbf{u} = (1, 1)$ and $\mathbf{v} = (1, -1)$ are orthogonal with respect to the Euclidean inner product on R^2, since

$$\mathbf{u} \cdot \mathbf{v} = (1)(1) + (1)(-1) = 0$$

However, they are not orthogonal with respect to the weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1v_1 + 2u_2v_2$, since

$$\langle \mathbf{u}, \mathbf{v} \rangle = 3(1)(1) + 2(1)(-1) = 1 \neq 0$$

EXAMPLE 3 Orthogonal Vectors in M_{22}

If M_{22} has the inner product of Example 6 in the preceding section, then the matrices

$$U = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}$$

are orthogonal, since

$$\langle U, V \rangle = 1(0) + 0(2) + 1(0) + 1(0) = 0$$

CALCULUS REQUIRED

EXAMPLE 4 Orthogonal Vectors in P_2

Let P_2 have the inner product

$$\langle \mathbf{p}, \mathbf{q} \rangle = \int_{-1}^{1} p(x)q(x)\,dx$$

and let $\mathbf{p} = x$ and $\mathbf{q} = x^2$. Then

$$\|\mathbf{p}\| = \langle \mathbf{p}, \mathbf{p} \rangle^{1/2} = \left[\int_{-1}^{1} xx\,dx\right]^{1/2} = \left[\int_{-1}^{1} x^2\,dx\right]^{1/2} = \sqrt{\frac{2}{3}}$$

$$\|\mathbf{q}\| = \langle \mathbf{q}, \mathbf{q} \rangle^{1/2} = \left[\int_{-1}^{1} x^2x^2\,dx\right]^{1/2} = \left[\int_{-1}^{1} x^4\,dx\right]^{1/2} = \sqrt{\frac{2}{5}}$$

$$\langle \mathbf{p}, \mathbf{q} \rangle = \int_{-1}^{1} xx^2\,dx = \int_{-1}^{1} x^3\,dx = 0$$

Because $\langle \mathbf{p}, \mathbf{q} \rangle = 0$, the vectors $\mathbf{p} = x$ and $\mathbf{q} = x^2$ are orthogonal relative to the given inner product.

In Section 3.3 we proved the Theorem of Pythagoras for vectors in Euclidean n-space. The following theorem extends this result to vectors in any real inner product space.

THEOREM 6.2.3 Generalized Theorem of Pythagoras

If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal vectors in an inner product space, then

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

Proof The orthogonality of $\mathbf{u}$ and $\mathbf{v}$ implies that $\langle \mathbf{u}, \mathbf{v} \rangle = 0$, so

$$\begin{aligned} \|\mathbf{u} + \mathbf{v}\|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle = \|\mathbf{u}\|^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2 \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \end{aligned}$$

CALCULUS REQUIRED

EXAMPLE 5 Theorem of Pythagoras in P_2

In Example 4 we showed that $\mathbf{p} = x$ and $\mathbf{q} = x^2$ are orthogonal with respect to the inner product

$$\langle \mathbf{p}, \mathbf{q} \rangle = \int_{-1}^{1} p(x)q(x)\, dx$$

on P_2. It follows from Theorem 6.2.3 that

$$\|\mathbf{p} + \mathbf{q}\|^2 = \|\mathbf{p}\|^2 + \|\mathbf{q}\|^2$$

Thus, from the computations in Example 4, we have

$$\|\mathbf{p} + \mathbf{q}\|^2 = \left(\sqrt{\frac{2}{3}}\right)^2 + \left(\sqrt{\frac{2}{5}}\right)^2 = \frac{2}{3} + \frac{2}{5} = \frac{16}{15}$$

We can check this result by direct integration:

$$\|\mathbf{p} + \mathbf{q}\|^2 = \langle \mathbf{p} + \mathbf{q}, \mathbf{p} + \mathbf{q} \rangle = \int_{-1}^{1} \left(x + x^2\right)\left(x + x^2\right) dx$$

$$= \int_{-1}^{1} x^2\, dx + 2\int_{-1}^{1} x^3\, dx + \int_{-1}^{1} x^4\, dx = \frac{2}{3} + 0 + \frac{2}{5} = \frac{16}{15}$$

Orthogonal Complements

In Section 4.8 we defined the notion of an *orthogonal complement* for subspaces of R^n, and we used that definition to establish a geometric link between the fundamental spaces of a matrix. The following definition extends that idea to general inner product spaces.

DEFINITION 2

If W is a subspace of an inner product space V, then the set of all vectors in V that are orthogonal to every vector in W is called the ***orthogonal complement*** of W and is denoted by the symbol $W^{\perp}$.

In Theorem 4.8.8 we stated three properties of orthogonal complements in R^n. The following theorem generalizes parts (a) and (b) of that theorem to general inner product spaces.

THEOREM 6.2.4

If W is a subspace of an inner product space V, then:

(a) $W^{\perp}$ is a subspace of V.

(b) $W \cap W^{\perp} = \{0\}$.

Proof (a) The set $W^{\perp}$ contains at least the zero vector, since $\langle \mathbf{0}, \mathbf{w} \rangle = 0$ for every vector $\mathbf{w}$ in W. Thus, it remains to show that $W^{\perp}$ is closed under addition and scalar multiplication. To do this, suppose that $\mathbf{u}$ and $\mathbf{v}$ are vectors in $W^{\perp}$, so that for every vector $\mathbf{w}$ in W we have $\langle \mathbf{u}, \mathbf{w} \rangle = 0$ and $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. It follows from the additivity and homogeneity axioms of inner products that

$$\begin{aligned} \langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle &= \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle = 0 + 0 = 0 \\ \langle k\mathbf{u}, \mathbf{w} \rangle &= k\langle \mathbf{u}, \mathbf{w} \rangle = k(0) = 0 \end{aligned}$$

which proves that $\mathbf{u} + \mathbf{v}$ and $k\mathbf{u}$ are in $W^{\perp}$.

Proof (b) If $\mathbf{v}$ is any vector in both W and $W^{\perp}$, then $\mathbf{v}$ is orthogonal to itself; that is, $\langle \mathbf{v}, \mathbf{v} \rangle = 0$. It follows from the positivity axiom for inner products that $\mathbf{v} = \mathbf{0}$.

The next theorem, which we state without proof, generalizes part (*c*) of Theorem 4.8.8. Note, however, that this theorem applies only to finite-dimensional inner product spaces, whereas Theorem 6.2.5 does not have this restriction.

THEOREM 6.2.5

Theorem 6.2.5 implies that in a finite-dimensional inner product space orthogonal complements occur in pairs, each being orthogonal to the other (Figure 6.2.2).

Theorem 6.2.5 If W is a subspace of a finite-dimensional inner product space V, then the orthogonal complement of $W^{\perp}$ is W; that is,

$$\left(W^{\perp}\right)^{\perp} = W$$

Figure 6.2.2 Each vector in W is orthogonal to each vector in $W^{\perp}$ and conversely

In our study of the fundamental spaces of a matrix in Section 4.8 we showed that the row space and null space of a matrix are orthogonal complements with respect to the Euclidean inner product on R^n (Theorem 4.8.9). The following example takes advantage of that fact.

EXAMPLE 6 Basis for an Orthogonal Complement

Let W be the subspace of R^6 spanned by the vectors

$$\mathbf{w}_1 = (1, 3, -2, 0, 2, 0), \qquad \mathbf{w}_2 = (2, 6, -5, -2, 4, -3),$$
$$\mathbf{w}_3 = (0, 0, 5, 10, 0, 15), \qquad \mathbf{w}_4 = (2, 6, 0, 8, 4, 18)$$

Find a basis for the orthogonal complement of W.

Solution The space W is the same as the row space of the matrix

$$A = \begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 \\ 2 & 6 & -5 & -2 & 4 & -3 \\ 0 & 0 & 5 & 10 & 0 & 15 \\ 2 & 6 & 0 & 8 & 4 & 18 \end{bmatrix}$$

Since the row space and null space of A are orthogonal complements, our problem reduces to finding a basis for the null space of this matrix. In Example 4 of Section 4.7 we showed that

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

form a basis for this null space. Expressing these vectors in comma-delimited form (to match that of $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$, and $\mathbf{w}_4$), we obtain the basis vectors

$$\mathbf{v}_1 = (-3, 1, 0, 0, 0, 0), \quad \mathbf{v}_2 = (-4, 0, -2, 1, 0, 0), \quad \mathbf{v}_3 = (-2, 0, 0, 0, 1, 0)$$

You may want to check that these vectors are orthogonal to $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$, and $\mathbf{w}_4$ by computing the necessary dot products.

Concept Review

- Cauchy–Schwarz inequality
- Angle between vectors
- Orthogonal vectors
- Orthogonal complement

Skills

- Find the angle between two vectors in an inner product space.
- Determine whether two vectors in an inner product space are orthogonal.
- Find a basis for the orthogonal complement of a subspace of an inner product space.

Exercise Set 6.2

1. Let R^2, R^3, and R^4 have the Euclidean inner product. In each part, find the cosine of the angle between **u** and **v**.

(a) $\mathbf{u} = (1, -3),\ \mathbf{v} = (2, 4)$

(b) $\mathbf{u} = (-1, 0),\ \mathbf{v} = (3, 8)$

(c) $\mathbf{u} = (-1, 5, 2),\ \mathbf{v} = (2, 4, -9)$

(d) $\mathbf{u} = (4, 1, 8),\ \mathbf{v} = (1, 0, -3)$

(e) $\mathbf{u} = (1, 0, 1, 0),\ \mathbf{v} = (-3, -3, -3, -3)$

(f) $\mathbf{u} = (2, 1, 7, -1),\ \mathbf{v} = (4, 0, 0, 0)$

Answer:

(a) $-\frac{1}{\sqrt{2}}$

(b) $-\frac{3}{\sqrt{73}}$

(c) 0

(d) $-\frac{20}{9\sqrt{10}}$

(e) $-\frac{1}{\sqrt{2}}$

(f) $\frac{2}{\sqrt{55}}$

2. Let P_2 have the inner product in Example 7 of Section 6.1 . Find the cosine of the angle between **p**and **q**.

(a) $\mathbf{p} = -1 + 5x + 2x^2$, $\mathbf{q} = 2 + 4x - 9x^2$

(b) $\mathbf{p} = x - x^2$, $\mathbf{q} = 7 + 3x + 3x^2$

3. Let M_{22} have the inner product in Example 6 of Section 6.1 . Find the cosine of the angle between A and B.

(a) $A = \begin{bmatrix} 2 & 6 \\ 1 & -3 \end{bmatrix}, B = \begin{bmatrix} 3 & 2 \\ 1 & 0 \end{bmatrix}$

(b) $A = \begin{bmatrix} 2 & 4 \\ -1 & 3 \end{bmatrix}, B = \begin{bmatrix} -3 & 1 \\ 4 & 2 \end{bmatrix}$

Answer:

(a) $\dfrac{19}{10\sqrt{7}}$

(b) 0

4. In each part, determine whether the given vectors are orthogonal withrespect to the Euclidean inner product.

(a) $\mathbf{u} = (-1, 3, 2),\ \mathbf{v} = (4, 2, -1)$

(b) $\mathbf{u} = (-2, -2, -2),\ \mathbf{v} = (1, 1, 1)$

(c) $\mathbf{u} = (u_1, u_2, u_3),\ \mathbf{v} = (0, 0, 0)$

(d) $\mathbf{u} = (-4, 6, -10, 1),\ \mathbf{v} = (2, 1, -2, 9)$

(e) $\mathbf{u} = (0, 3, -2, 1),\ \mathbf{v} = (5, 2, -1, 0)$

(f) $\mathbf{u} = (a, b),\ \mathbf{v} = (-b, a)$

5. Show that $\mathbf{p} = 1 - x + 2x^2$ and $\mathbf{q} = 2x + x^2$ are orthogonal with respect to the inner product in Exercise 2.

6. Let

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

Which of the following matrices are orthogonal to A with respect to the inner product in Exercise 3?

(a) $\begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}$

(c) $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} 2 & 1 \\ 5 & 2 \end{bmatrix}$

7. Do there exist scalars k and l such that the vectors $\mathbf{u} = (2, k, 6)$, $\mathbf{v} = (l, 5, 3)$, and $\mathbf{w} = (1, 2, 3)$ are mutually orthogonal with respect to the Euclidean inner product?

Answer:

No

8. Let R^3 have the Euclidean inner product, and suppose that $\mathbf{u} = (1, 1, -1)$ and $\mathbf{v} = (6, 7, -15)$. Find a value of k for which $\|k\mathbf{u} + \mathbf{v}\| = 13$.

9. Let R^3 have the Euclidean inner product. For which values of k are $\mathbf{u}$ and $\mathbf{v}$ orthogonal?

(a) $\mathbf{u} = (2, 1, 3),\ \mathbf{v} = (1, 7, k)$

(b) $\mathbf{u} = (k, k, 1),\ \mathbf{v} = (k, 5, 6)$

Answer:

(a) $k = -3$

(b) $k = -2, -3$

10. Let R^4 have the Euclidean inner product. Find two unit vectors that are orthogonal to all three of the vectors $\mathbf{u} = (2, 1, -4, 0)$, $\mathbf{v} = (-1, -1, 2, 2)$, and $\mathbf{w} = (3, 2, 5, 4)$.

11. In each part, verify that the Cauchy–Schwarz inequality holds for the given vectors using the Euclidean inner product.

(a) $\mathbf{u} = (3, 2),\ \mathbf{v} = (4, -1)$

(b) $\mathbf{u} = (-3, 1, 0),\ \mathbf{v} = (2, -1, 3)$

(c) $\mathbf{u} = (-4, 2, 1),\ \mathbf{v} = (8, -4, -2)$

(d) $\mathbf{u} = (0, -2, 2, 1),\ \mathbf{v} = (-1, -1, 1, 1)$

12. In each part, verify that the Cauchy–Schwarz inequality holds for the given vectors.

(a) $\mathbf{u} = (-2, 1)$ and $\mathbf{v} = (1, 0)$ using the inner product of Example 1 of Section 6.1 .

(b) $U = \begin{bmatrix} -1 & 2 \\ 6 & 1 \end{bmatrix}$ and $V = \begin{bmatrix} 1 & 0 \\ 3 & 3 \end{bmatrix}$ using the inner product in Example 6 of Section 6.1 .

(c) $\mathbf{p} = -1 + 2x + x^2$ and $\mathbf{q} = 2 - 4x^2$ using the inner product given in Example 7 of Section 6.1 .

13. Let R^4 have the Euclidean inner product, and let $\mathbf{u} = (-1, 1, 0, 2)$. Determine whether the vector $\mathbf{u}$ is orthogonal to the subspace spanned by the vectors $\mathbf{w}_1 = (0, 0, 0, 0)$, $\mathbf{w}_2 = (1, -1, 3, 0)$, and $\mathbf{w}_3 = (4, 0, 9, 2)$.

Answer:

No

In Exercises 14–15, assume that R^n has the Euclidean inner product.

14. Let W be the line in R^2 with equation $y = 2x$. Find an equation for $W^\perp$.

15. (a) Let W be the plane in R^3 with equation $x - 2y - 3z = 0$. Find parametric equations for $W^\perp$.

(b) Let W be the line in R^3 with parametric equations

$$x = 2t, \quad y = -5t, \quad z = 4t$$

Find an equation for $W^\perp$.

(c) Let W be the intersection of the two planes

$$x + y + z = 0 \quad \text{and} \quad x - y + z = 0$$

in R^3. Find an equation for $W^\perp$.

Answer:

(a) $x = t,\ y = -2t,\ z = -3t$

(b) $2x - 5y + 4z = 0$

(c) $x - z = 0$

16. Find a basis for the orthogonal complement of the subspace of R^n spanned by the vectors.

(a) $\mathbf{v}_1 = (1, -1, 3), \mathbf{v}_2 = (5, -4, -4), \mathbf{v}_3 = (7, -6, 2)$

(b) $\mathbf{v}_1 = (2, 0, -1), \mathbf{v}_2 = (4, 0, -2)$

(c) $\mathbf{v}_1 = (1, 4, 5, 2), \mathbf{v}_2 = (2, 1, 3, 0), \mathbf{v}_3 = (-1, 3, 2, 2)$

(d) $\mathbf{v}_1 = (1, 4, 5, 6, 9), \mathbf{v}_2 = (3, -2, 1, 4, -1), \mathbf{v}_3 = (-1, 0, -1, -2, -1), \mathbf{v}_4 = (2, 3, 5, 7, 8)$

17. Let V be an inner product space. Show that if $\mathbf{u}$ and $\mathbf{v}$ are orthogonal unit vectors in V, then $\|\mathbf{u} - \mathbf{v}\| = \sqrt{2}$.

18. Let V be an inner product space. Show that if $\mathbf{w}$ is orthogonal to both $\mathbf{u}_1$ and $\mathbf{u}_2$, then it is orthogonal to $k_1\mathbf{u}_1 + k_2\mathbf{u}_2$ for all scalars k_1 and k_2. Interpret this result geometrically in the case where V is R^3 with the Euclidean inner product.

19. Let V be an inner product space. Show that if $\mathbf{w}$ is orthogonal to each of the vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r$, then it is orthogonal to every vector in span $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$.

20. Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ be a basis for an inner product space V. Show that the zero vector is the only vector in V that is orthogonal to all of the basis vectors.

21. Let $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$ be a basis for a subspace W of V. Show that $W^\perp$ consists of all vectors in V that are orthogonal to every basis vector.

22. Prove the following generalization of Theorem 6.2.3: If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ are pairwise orthogonal vectors in an inner product space V, then

$$\|\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_r\|^2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 + \cdots + \|\mathbf{v}_r\|^2$$

23. Prove: If $\mathbf{u}$ and $\mathbf{v}$ are $n \times 1$ matrices and A is an $n \times n$ matrix, then

$$\left(\mathbf{v}^T A^T A\mathbf{u}\right)^2 \leq \left(\mathbf{u}^T A^T A\mathbf{u}\right)\left(\mathbf{v}^T A^T A\mathbf{v}\right)$$

24. Use the Cauchy–Schwarz inequality to prove that for all real values of a, b, and θ,

$$(a\cos\theta + b\sin\theta)^2 \leq a^2 + b^2$$

25. Prove: If $w_1, w_2, \ldots, w_n$ are positive real numbers, and if $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ are any two vectors in R^n, then

$$\begin{aligned}&|w_1u_1v_1 + w_2u_2v_2 + \cdots + w_nu_nv_n| \\ &\leq \left(w_1u_1^2 + w_2u_2^2 + \cdots + w_nu_n^2\right)^{1/2}\left(w_1v_1^2 + w_2v_2^2 + \cdots + w_nv_n^2\right)^{1/2}\end{aligned}$$

26. Show that equality holds in the Cauchy–Schwarz inequality if and only if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent.

27. Use vector methods to prove that a triangle that is inscribed in a circle so that it has a diameter for a side must be a right triangle. [*Hint:* Express the vectors $\overrightarrow{AB}$ and $\overrightarrow{BC}$ in the accompanying figure in terms of $\mathbf{u}$ and$\mathbf{v}$.]

Figure Ex-27

28. As illustrated in the accompanying figure, the vectors $\mathbf{u} = \left(1, \sqrt{3}\right)$ and $\mathbf{v} = \left(-1, \sqrt{3}\right)$ have norm 2 and an angle of 60° between them relative to the Euclidean inner product. Find a weighted Euclidean inner product with respect to which **u** and **v** are orthogonal unit vectors.

Figure Ex-28

29. ***Calculus required*** Let $f(x)$ and $g(x)$ be continuous functions on $[0, 1]$. Prove:

(a) $\left[\int_0^1 f(x)g(x)\,dx\right]^2 \leq \left[\int_0^1 f^2(x)\,dx\right]\left[\int_0^1 g^2(x)\,dx\right]$

(b) $\left[\int_0^1 [f(x)+g(x)]^2\,dx\right]^{1/2} \leq \left[\int_0^1 f^2(x)\,dx\right]^{1/2} + \left[\int_0^1 g^2(x)\,dx\right]^{1/2}$

[*Hint:* Use the Cauchy–Schwarz inequality.]

30. ***Calculus required*** Let $C[0, \pi]$ have the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_0^{\pi} f(x)g(x)\,dx$$

and let $\mathbf{f}_n = \cos nx$ $(n = 0, 1, 2, \ldots)$. Show that if $k \neq l$, then $\mathbf{f}_k$ and $\mathbf{f}_l$ are orthogonal vectors.

31. (a) Let W be the line $y = x$ in an xy-coordinate system in R^2. Describe the subspace $W^{\perp}$.

(b) Let W be the y-axis in an xyz-coordinate system in R^3. Describe the subspace $W^{\perp}$.

(c) Let W be the yz-plane of an xyz-coordinate system in R^3. Describe the subspace $W^{\perp}$.

Answer:

(a) The line $y = -x$

(b) The xz-plane

(c) The x-axis

32. Prove that Formula 4 holds for all nonzero vectors **u** and **v** in an inner product space V.

True-False Exercises

In parts (a)–(f) determine whether the statement is true or false, and justify your answer.

(a) If **u** is orthogonal to every vector of a subspace W, then $\mathbf{u} = \mathbf{0}$.

Answer:

False

(b) If **u** is a vector in both W and $W^{\perp}$, then $\mathbf{u} = 0$.

Answer:

True

(c) If **u** and **v** are vectors in $W^{\perp}$, then $\mathbf{u} + \mathbf{v}$ is in $W^{\perp}$.

Answer:

True

(d) If **u** is a vector in $W^{\perp}$ and k is a real number, then $k\mathbf{u}$ is in $W^{\perp}$.

Answer:

True

(e) If **u** and **v** are orthogonal, then $|\langle \mathbf{u}, \mathbf{v} \rangle| = \|\mathbf{u}\| \|\mathbf{v}\|$.

Answer:

False

(f) If **u** and **v** are orthogonal, then $\|\mathbf{u} + \mathbf{v}\| = \|\mathbf{u}\| + \|\mathbf{v}\|$.

Answer:

False

Copyright © 2010 John Wiley & Sons, Inc. All rights reserved.

6.3 Gram–Schmidt Process; *QR*-Decomposition

In many problems involving vector spaces, the problem solver is free to choose any basis for the vector space that seems appropriate. In inner product spaces, the solution of a problem is often greatly simplified by choosing a basis in which the vectors are orthogonal to one another. In this section we will show how such bases can be obtained.

Orthogonal and Orthonormal Sets

Recall from Section 6.2 that two vectors in an inner product space are said to be *orthogonal* if their inner product is zero. The following definition extends the notion of orthogonality to sets of vectors in an inner product space.

DEFINITION 1

A set of two or more vectors in a real inner product space is said to be ***orthogonal*** if all pairs of distinct vectors in the set are orthogonal. An orthogonal set in which each vector has norm 1 is said to be ***orthogonal***.

EXAMPLE 1 An Orthogonal Set in R^3

Let

$$\mathbf{u}_1 = (0, 1, 0), \quad \mathbf{u}_2 = (1, 0, 1), \quad \mathbf{u}_3 = (1, 0, -1)$$

and assume that R^3 has the Euclidean inner product. It follows that the set of vectors $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is orthogonal since $\langle \mathbf{u}_1, \mathbf{u}_2 \rangle = \langle \mathbf{u}_1, \mathbf{u}_3 \rangle = \langle \mathbf{u}_2, \mathbf{u}_3 \rangle = 0$.

If $\mathbf{v}$ is a nonzero vector in an inner product space, then it follows from Theorem 6.1.1*b* with $k = \|\mathbf{v}\|$ that

$$\left\| \frac{1}{\|\mathbf{v}\|}\mathbf{v} \right\| = \left| \frac{1}{\|\mathbf{v}\|} \right| \|\mathbf{v}\| = \frac{1}{\|\mathbf{v}\|}\|\mathbf{v}\| = 1$$

from which we see that multiplying a nonzero vector by the reciprocal of its norm produces a vector of norm 1. This process is called ***normalizing*** $\mathbf{v}$. It follows that *any orthogonal set of nonzero vectors can be converted to an orthonormal set by normalizing each of its vectors*.

EXAMPLE 2 Constructing an Orthonormal Set

The Euclidean norms of the vectors in Example 1 are

$$\|\mathbf{u}_1\| = 1, \quad \|\mathbf{u}_2\| = \sqrt{2}, \quad \|\mathbf{u}_3\| = \sqrt{2}$$

Consequently, normalizing $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ yields

$$\mathbf{v}_1 = \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} = (0, 1, 0), \quad \mathbf{v}_2 = \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|} = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right),$$

$$\mathbf{v}_3 = \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|} = \left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}\right)$$

We leave it for you to verify that the set $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is orthonormal by showing that

$$\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, \mathbf{v}_3 \rangle = \langle \mathbf{v}_2, \mathbf{v}_3 \rangle = 0 \quad \text{and} \quad \|\mathbf{v}_1\| = \|\mathbf{v}_2\| = \|\mathbf{v}_3\| = 1$$

In R^2 any two nonzero perpendicular vectors are linearly independent because neither is a scalar multiple of the other; and in R^3 any three nonzero mutually perpendicular vectors are linearly independent because no one lies in the plane of the other two (and hence is not expressible as a linear combination of the other two). The following theorem generalizes these observations.

THEOREM 6.3.1

If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthogonal set of nonzero vectors in an inner product space, then S is linearly independent.

Proof Assume that

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n = 0 \tag{1}$$

To demonstrate that $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is linearly independent, we must prove that $k_1 = k_2 = \ldots = k_n = 0$.

For each $\mathbf{v}_i$ in S, it follows from 1 that

$$\langle k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n, \mathbf{v}_i \rangle = \langle \mathbf{0}, \mathbf{v_i} \rangle = 0$$

or, equivalently,

$$k_1\langle \mathbf{v}_1, \mathbf{v}_i \rangle + k_2\langle \mathbf{v}_2, \mathbf{v}_i \rangle + \cdots + k_n\langle \mathbf{v}_n, \mathbf{v}_i \rangle = 0$$

From the orthogonality of S it follows that $\langle \mathbf{v}_j, \mathbf{v}_i \rangle = 0$ when $j \neq i$, so this equation reduces to

$$k_i\langle \mathbf{v}_i, \mathbf{v}_i \rangle = 0$$

Since the vectors in S are assumed to be nonzero, it follows from the positivity axiom for inner products that $\langle \mathbf{v}_i, \mathbf{v}_i \rangle \neq 0$. Thus, the preceding equation implies that each k_i in Equation 1 is zero, which is what we wanted to prove.

Since an orthonormal set is orthogonal, and since its vectors are nonzero (norm 1), it follows from Theorem 6.3.1 that every *orthonormal* set is linearly independent.

In an inner product space, a basis consisting of orthonormal vectors is called an ***orthonormal basis***, and a basis

consisting of orthogonal vectors is called an ***orthogonal basis***. A familiar example of an orthonormal basis is the standard basis for R^n with the Euclidean inner product:

$$\mathbf{e}_1 = (1, 0, 0, \ldots, 0), \quad \mathbf{e}_2 = (0, 1, 0, \ldots, 0), \ldots, \quad \mathbf{e}_n = (0, 0, 0, \ldots, 1)$$

EXAMPLE 3 An Orthonormal Basis

In Example 2 we showed that the vectors

$$\mathbf{v}_1 = (0, 1, 0), \quad \mathbf{v}_2 = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right), \quad \text{and} \quad \mathbf{v}_3 = \left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}\right)$$

form an orthonormal set with respect to the Euclidean inner product on R^3. By Theorem 6.3.1, these vectors form a linearlyindependent set, and since R^3 is three-dimensional, it follows from Theorem 4.5.4 that $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis for R^3.

Coordinates Relative to Orthonormal Bases

One way to express a vector **u** as a linear combination of basis vectors

$$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$$

is to convert the vector equation

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

to a linear system and solve for the coefficients $c_1, c_2, \ldots, c_n$. However, if the basis happens to be orthogonal or orthonormal, then the following theorem shows that the coefficients can be obtained more simply by computing appropriate inner products.

THEOREM 6.3.2

(a) If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthogonal basis for an inner product space *V*, and if **u** is any vector in *V*, then

$$\mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{u}, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{v}_n \rangle}{\|\mathbf{v}_n\|^2}\mathbf{v}_n \tag{2}$$

(b) If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthonormal basis for an inner product space *V*, and if **u** is any vector in *V*, then

$$\mathbf{u} = \langle \mathbf{u}, \mathbf{v}_1 \rangle\mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2 \rangle\mathbf{v}_2 + \cdots + \langle \mathbf{u}, \mathbf{v}_n \rangle\mathbf{v}_n \tag{3}$$

Proof (a) Since $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for *V*, every vector **u** in *V* can be expressed in the form

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

We will complete the proof by showing that

$$c_i = \frac{\langle \mathbf{u}, \mathbf{v}_i \rangle}{\|\mathbf{v}_i\|^2} \tag{4}$$

for $i = 1, 2, \ldots, n$. To do this, observe first that

$$\begin{aligned} \langle \mathbf{u}, \mathbf{v}_i \rangle &= \langle c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n, \mathbf{v}_i \rangle \\ &= c_1\langle \mathbf{v}_1, \mathbf{v}_i \rangle + c_2\langle \mathbf{v}_2, \mathbf{v}_i \rangle + \cdots + c_n\langle \mathbf{v}_n, \mathbf{v}_i \rangle \end{aligned}$$

Since S is an orthogonal set, all of the inner products in the last equality are zero except the ith, so we have

$$\langle \mathbf{u}, \mathbf{v}_i \rangle = c_i \langle \mathbf{v}_i, \mathbf{v}_i \rangle = c_i \|\mathbf{v}_i\|^2$$

Solving this equation for c_i yields 4, which completes the proof.

Proof (b) In this case, $\|\mathbf{v}_1\| = \|\mathbf{v}_2\| = \ldots = \|\mathbf{v}_n\| = 1$, so Formula 2 simplifies to Formula 3.

Using the terminology and notation from Definition 2 of Section 4.4, it follows from Theorem 6.3.2 that the coordinate vector of a vector $\mathbf{u}$ in V relative to an orthogonal basis $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is

$$(\mathbf{u})_S = \left(\frac{\langle \mathbf{u}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}, \frac{\langle \mathbf{u}, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}, \ldots, \frac{\langle \mathbf{u}, \mathbf{v}_n \rangle}{\|\mathbf{v}_n\|^2} \right) \tag{5}$$

and relative to an orthonormal basis $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is

$$(\mathbf{u})_S = (\langle \mathbf{u}, \mathbf{v}_1 \rangle, \langle \mathbf{u}, \mathbf{v}_2 \rangle, \ldots, \langle \mathbf{u}, \mathbf{v}_n \rangle) \tag{6}$$

EXAMPLE 4 A Coordinate Vector Relative to an Orthonormal Basis

Let

$$\mathbf{v}_1 = (0, 1, 0), \quad \mathbf{v}_2 = \left(-\frac{4}{5}, 0, \frac{3}{5}\right), \quad \mathbf{v}_3 = \left(\frac{3}{5}, 0, \frac{4}{5}\right)$$

It is easy to check that $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis for R^3 with the Euclidean inner product. Express the vector $\mathbf{u} = (1, 1, 1)$ as a linear combination of the vectors in S, and find the coordinate vector $(\mathbf{u})_S$.

Solution We leave it for you to verify that

$$\langle \mathbf{u}, \mathbf{v}_1 \rangle = 1, \quad \langle \mathbf{u}, \mathbf{v}_2 \rangle = -\frac{1}{5}, \quad \text{and} \quad \langle \mathbf{u}, \mathbf{v}_3 \rangle = \frac{7}{5}$$

Therefore, by Theorem 6.3.2 we have

$$\mathbf{u} = \mathbf{v}_1 - \frac{1}{5}\mathbf{v}_2 + \frac{7}{5}\mathbf{v}_3$$

that is,

$$(1, 1, 1) = (0, 1, 0) - \frac{1}{5}\left(-\frac{4}{5}, 0, \frac{3}{5}\right) + \frac{7}{5}\left(\frac{3}{5}, 0, \frac{4}{5}\right)$$

Thus, the coordinate vector of $\mathbf{u}$ relative to S is

$$(\mathbf{u})_S = (\langle \mathbf{u}, \mathbf{v}_1 \rangle, \langle \mathbf{u}, \mathbf{v}_2 \rangle, \langle \mathbf{u}, \mathbf{v}_3 \rangle) = \left(1, -\frac{1}{5}, \frac{7}{5}\right)$$

EXAMPLE 5 An Orthonormal Basis from an Orthogonal Basis

(a) Show that the vectors

$$\mathbf{w}_1 = (0, 2, 0), \quad \mathbf{w}_2 = (3, 0, 3), \quad \mathbf{w}_3 = (-4, 0, 4)$$

form an orthogonal basis for R^3 with the Euclidean inner product, and use that basis to find an orthonormal basis by normalizing each vector.

(b) Express the vector $\mathbf{u} = (1, 2, 4)$ as a linear combination of the orthonormal basis vectors obtained in part (a).

Solution

(a) The given vectors form an orthogonal set since

$$\langle \mathbf{w}_1, \mathbf{w}_2 \rangle = 0, \quad \langle \mathbf{w}_1, \mathbf{w}_3 \rangle = 0, \quad \langle \mathbf{w}_2, \mathbf{w}_3 \rangle = 0$$

It follows from Theorem 6.3.1 that these vectors are linearly independent and hence form a basis for R^3 by Theorem 4.5.4. We leave it for you to calculate the norms of $\mathbf{w}_1$, $\mathbf{w}_2$, and $\mathbf{w}_3$ and then obtain the orthonormal basis

$$\mathbf{v}_1 = \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|} = (0, 1, 0), \quad \mathbf{v}_2 = \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|} = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right),$$

$$\mathbf{v}_3 = \frac{\mathbf{w}_3}{\|\mathbf{w}_3\|} = \left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$$

(b) It follows from Formula 3 that

$$\mathbf{u} = \langle \mathbf{u}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2 \rangle \mathbf{v}_2 + \langle \mathbf{u}, \mathbf{v}_3 \rangle \mathbf{v}_3$$

We leave it for you to confirm that

$$\begin{aligned}
\langle \mathbf{u}, \mathbf{v}_1 \rangle &= (1, 2, 4) \cdot (0, 1, 0) = 2 \\
\langle \mathbf{u}, \mathbf{v}_2 \rangle &= (1, 2, 4) \cdot \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right) = \frac{5}{\sqrt{2}} \\
\langle \mathbf{u}, \mathbf{v}_3 \rangle &= (1, 2, 4) \cdot \left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right) = \frac{3}{\sqrt{2}}
\end{aligned}$$

and hence that

$$(1, 2, 4) = 2(0, 1, 0) + \frac{5}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right) + \frac{3}{\sqrt{2}}\left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$$

Orthogonal Projections

Many applied problems are best solved by working with orthogonal or orthonormal basis vectors. Such bases are typically found by starting with some simple basis (say a standard basis) and then converting that basis into an

orthogonal or orthonormal basis. To explain exactly how that is done will require some preliminary ideas about orthogonal projections.

In Section 3.3 we proved a result called the *Prohection Theorem* (see Theorem 3.3.2) which dealt with the problem of decomposing a vector **u** in R^n into a sum of two terms, $\mathbf{w}_1$ and $\mathbf{w}_2$, in which $\mathbf{w}_1$ is the orthogonal projection of **u** on some nonzero vector **a** and $\mathbf{w}_2$ is orthogonal to $\mathbf{w}_1$ (Figure 3.3.2). That result is a special case of the following more general theorem.

THEOREM 6.3.3 Projection Theorem

If W is a finite-dimensional subspace of an inner product space V,then every vector **u** in V can be expressed in exactly oneway as

$$\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2 \tag{7}$$

where $\mathbf{w}_1$ is in W and $\mathbf{w}_2$ is in $W^{\perp}$.

The vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ in Formula 7 are commonly denoted by

$$\mathbf{w}_1 = \text{proj}_W \mathbf{u} \quad \text{and} \quad \mathbf{w}_2 = \text{proj}_{W^{\perp}} \mathbf{u} \tag{8}$$

They are called the ***orthogonal projection of*** **u** ***on W*** and the ***orthogonal projection of*** **u** ***on*** $W^{\perp}$, respectively. The vector $\mathbf{w}_2$ is also called the ***component of*** **u** ***orthogonal to W***. Using the notation in 8, Formula 7 can be expressed as

$$\mathbf{u} = \text{proj}_W \mathbf{u} + \text{proj}_{W^{\perp}} \mathbf{u} \tag{9}$$

(Figure 6.3.1). Moreover, since $\text{proj}_{W^{\perp}}\mathbf{u} = \mathbf{u} - \text{proj}_W \mathbf{u}$, we can also express Formula 9 as

$$\mathbf{u} = \text{proj}_W \mathbf{u} + (\mathbf{u} - \text{proj}_W \mathbf{u}) \tag{10}$$

Figure 6.3.1

The following theorem provides formulas for calculating orthogonal projections.

THEOREM 6.3.4

Let W be a finite-dimensional subspace of an inner product space V.

(a) If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is an orthogonal basis for W, and $\mathbf{u}$ is any vector in V, then

$$\text{proj}_W \mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{u}, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{v}_r \rangle}{\|\mathbf{v}_r\|^2}\mathbf{v}_r \tag{11}$$

(b) If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is an orthonormal basis for W, and $\mathbf{u}$ is any vector in V, then

$$\text{proj}_W \mathbf{u} = \langle \mathbf{u}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2 \rangle \mathbf{v}_2 + \cdots + \langle \mathbf{u}, \mathbf{v}_r \rangle \mathbf{v}_r \tag{12}$$

Proof (a) It follows from Theorem 6.3.3 that the vector $\mathbf{u}$ can be expressed in the form $\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1 = \text{proj}_W \mathbf{u}$ is in W and $\mathbf{w}_2$ is in $W^\perp$; and it follows from Theorem 6.3.2 that the component $\text{proj}_W \mathbf{u} = \mathbf{w}_1$ can be expressed in terms of the basis vectors for W as

$$\text{proj}_W \mathbf{u} = \mathbf{w}_1 = \frac{\langle \mathbf{w}_1, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{w}_1, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{w}_1, \mathbf{v}_r \rangle}{\|\mathbf{v}_r\|^2}\mathbf{v}_r \tag{13}$$

Since $\mathbf{w}_2$ is orthogonal to W, it follows that

$$\langle \mathbf{w}_2, \mathbf{v}_1 \rangle = \langle \mathbf{w}_2, \mathbf{v}_2 \rangle = \ldots = \langle \mathbf{w}_2, \mathbf{v}_r \rangle = 0$$

so we can rewrite 13 as

$$\text{proj}_W \mathbf{u} = \mathbf{w}_1 = \frac{\langle \mathbf{w}_1 + \mathbf{w}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{w}_1 + \mathbf{w}_2, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{w}_1 + \mathbf{w}_2, \mathbf{v}_r \rangle}{\|\mathbf{v}_r\|^2}\mathbf{v}_r$$

or, equivalently, as

$$\text{proj}_W \mathbf{u} = \mathbf{w}_1 = \frac{\langle \mathbf{u}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{u}, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{v}_r \rangle}{\|\mathbf{v}_r\|^2}\mathbf{v}_r$$

Proof (a) In this case, $\|\mathbf{v}_1\| = \|\mathbf{v}_2\| = \ldots = \|\mathbf{v}_r\| = 1$, so Formula 13 simplifies to Formula 12.

EXAMPLE 6 Calculating Projections ◀

Let R^3 have the Euclidean inner product, and let W be the subspace spanned by the orthonormal vectors $\mathbf{v}_1 = (0, 1, 0)$ and $\mathbf{v}_2 = \left(-\frac{4}{5}, 0, \frac{3}{5}\right)$. From Formula 12 the orthogonal projection of $\mathbf{u} = (1, 1, 1)$ on W is

$$\begin{aligned}
\text{proj}_W \mathbf{u} &= \langle \mathbf{u}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2 \rangle \mathbf{v}_2 \\
&= (1)(0, 1, 0) + \left(-\frac{1}{5}\right)\left(-\frac{4}{5}, 0, \frac{3}{5}\right) \\
&= \left(\frac{4}{25}, 1, -\frac{3}{25}\right)
\end{aligned}$$

The component of $\mathbf{u}$ orthogonal to W is

$$\text{proj}_{W^\perp}\mathbf{u} = \mathbf{u} - \text{proj}_W\mathbf{u} = (1, 1, 1) - \left(\frac{4}{25}, 1, -\frac{3}{25}\right) = \left(\frac{21}{25}, 0, \frac{28}{25}\right)$$

Observe that $\text{proj}_{W^\perp}\mathbf{u}$ is orthogonal to both $\mathbf{v}_1$ and $\mathbf{v}_2$, so this vector is orthogonal to each vector in the space W spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$, as it should be.

A Geometric Interpretation of Orthogonal Projections

If W is a one-dimensional subspace of an inner product space V, say span $\{\mathbf{a}\}$, then Formula 11 has only the one term

$$\text{proj}_W\mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{a}\rangle}{\|\mathbf{a}\|^2}\mathbf{a}$$

In the special case where V is R^3 with the Euclidean inner product, this is exactly Formula 10 of Section 3.3 for the orthogonal projection of **u** along **a**. This suggests that we can think of 11 as the sum of orthogonal projections on "axes" determined by the basis vectors for the subspace W (Figure 6.3.2).

Figure 6.3.2

The Gram–Schmidt Process

We have seen that orthonormal bases exhibit a variety of useful properties. Our next theorem, which is the main result in this section, shows that every nonzero finite-dimensional vector space has an orthonormal basis. The proof of this result is extremely important, since it provides an algorithm, or method, for converting an arbitrary basis into an orthonormal basis.

THEOREM 6.3.5

Every nonzero finite-dimensional inner product space has an orthonormal basis.

Proof Let W be any nonzero finite-dimensional subspace of an inner product space, and suppose that $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ is any basis for W. It suffices to show that W has an orthogonal basis, since the vectors in that basis can be normalized to obtain an orthonormal basis. The following sequence of steps will produce an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ for W:

Step 1. Let $\mathbf{v}_1 = \mathbf{u}_1$.

Step 2. As illustrated in Figure 6.3.3, we can obtain a vector $\mathbf{v}_2$ that is orthogonal to $\mathbf{v}_1$ by computing the component of $\mathbf{u}_2$ that is orthogonal to the space W_1 spanned by $\mathbf{v}_1$. Using Formula 11 to perform this computation we obtain

$$\mathbf{v}_2 = \mathbf{u}_2 - \text{proj}_{W_1}\mathbf{u}_2 = \mathbf{u}_2 - \frac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1$$

Of course, if $\mathbf{v}_2 = \mathbf{0}$, then $\mathbf{v}_2$ is not a basis vector. But this cannot happen, since it would then follow from the above formula for $\mathbf{v}_2$ that

$$\mathbf{u}_2 = \frac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 = \frac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{u}_1\|^2}\mathbf{u}_1$$

which implies that $\mathbf{u}_2$ is a multiple of $\mathbf{u}_1$, contradicting the linear independence of the basis $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$.

Figure 6.3.3

Step 3. To construct a vector $\mathbf{v}_3$ that is orthogonal to both $\mathbf{v}_1$ and $\mathbf{v}_2$, we compute the component of $\mathbf{u}_3$ orthogonal to the space W_2 spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$ (Figure 6.3.4). Using Formula 11 to perform this computation we obtain

$$\mathbf{v}_3 = \mathbf{u}_3 - \text{proj}_{W_2}\mathbf{u}_3 = \mathbf{u}_3 - \frac{\langle \mathbf{u}_3, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\langle \mathbf{u}_3, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2$$

As in Step 2, the linear independence of $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ ensures that $\mathbf{v}_3 \neq \mathbf{0}$. We leave the details for you.

Figure 6.3.4

Step 4. To determine a vector $\mathbf{v}_4$ that is orthogonal to $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$, we compute the component of $\mathbf{u}_4$ orthogonal to the space W_3 spanned by $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. From 11,

$$\mathbf{v}_4 = \mathbf{u}_4 - \text{proj}_{W_3}\mathbf{u}_4 = \mathbf{u}_4 - \frac{\langle \mathbf{u}_4, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\langle \mathbf{u}_4, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 - \frac{\langle \mathbf{u}_4, \mathbf{v}_3 \rangle}{\|\mathbf{v}_3\|^2}\mathbf{v}_3$$

Continuing in this way we will produce an orthogonal set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ after r steps. Since orthogonal sets are linearly independent, this set will be an orthogonal basis for the r-dimensional space W. By normalizing these basis vectors we can obtain an orthonormal basis.

The step-by-step construction of an orthogonal (or orthonormal) basis given in the foregoing proof is called the ***Gram–Schmidt process***. For reference, we provide the following summary of the steps.

The Gram–Schmidt Process

To convert a basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ into an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$, perform the following computations:

Step 1. $\mathbf{v}_1 = \mathbf{u}_1$

Step 2. $\mathbf{v}_2 = \mathbf{u}_2 - \dfrac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1$

Step 3. $\mathbf{v}_3 = \mathbf{u}_3 - \dfrac{\langle \mathbf{u}_3, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \dfrac{\langle \mathbf{u}_3, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2$

Step 4. $\mathbf{v}_4 = \mathbf{u}_4 - \dfrac{\langle \mathbf{u}_4, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \dfrac{\langle \mathbf{u}_4, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 - \dfrac{\langle \mathbf{u}_4, \mathbf{v}_3 \rangle}{\|\mathbf{v}_3\|^2}\mathbf{v}_3$

$\vdots$

(continue for *r* steps)

Optional Step. To convert the orthogonal basis into an orthonormal basis $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_r\}$, normalize the orthogonal basis vectors.

EXAMPLE 7 Using the Gram–Schmidt Process

Assume that the vector space R^3 has the Euclidean inner product. Apply the Gram–Schmidt process to transform the basis vectors

$$\mathbf{u}_1 = (1, 1, 1), \quad \mathbf{u}_2 = (0, 1, 1), \quad \mathbf{u}_3 = (0, 0, 1)$$

into an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, and then normalize the orthogonal basis vectors to obtain an orthonormal basis $\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$.

Solution

Step 1. $\mathbf{v}_1 = \mathbf{u}_1 = (1, 1, 1)$

Step 2.
$$\begin{aligned} \mathbf{v}_2 &= \mathbf{u}_2 - \operatorname{proj}_{W_1}\mathbf{u}_2 = \mathbf{u}_2 - \frac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 \\ &= (0, 1, 1) - \tfrac{2}{3}(1, 1, 1) = \left(-\tfrac{2}{3}, \tfrac{1}{3}, \tfrac{1}{3}\right) \end{aligned}$$

Step 3.

$$\begin{aligned}\mathbf{v}_3 &= \mathbf{u}_3 - \text{proj}_{W_2}\mathbf{u}_3 = \mathbf{u}_3 - \frac{\langle \mathbf{u}_3, \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\langle \mathbf{u}_3, \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 \\ &= (0, 0, 1) - \frac{1}{3}(1, 1, 1) - \frac{1/3}{2/3}\left(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right) \\ &= \left(0, -\frac{1}{2}, \frac{1}{2}\right)\end{aligned}$$

Thus,

$$\mathbf{v}_1 = (1, 1, 1), \quad \mathbf{v}_2 = \left(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right), \quad \mathbf{v}_3 = \left(0, -\frac{1}{2}, \frac{1}{2}\right)$$

form an orthogonal basis for R^3. The norms of these vectors are

$$\|\mathbf{v}_1\| = \sqrt{3}, \quad \|\mathbf{v}_2\| = \frac{\sqrt{6}}{3}, \quad \|\mathbf{v}_3\| = \frac{1}{\sqrt{2}}$$

so an orthonormal basis for R^3 is

$$\mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \quad \mathbf{q}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \left(-\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right),$$

$$\mathbf{q}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = \left(0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

Remark In the last example we normalized at the end to convert the orthogonal basis into an orthonormal basis. Alternatively, we could have normalized each orthogonal basis vector as soon as it was obtained, thereby producing an orthonormal basis step by step. However, that procedure generally has the disadvantage in hand calculation of producing more square roots to manipulate. A more useful variation is to "scale" the orthogonal basis vectors at each step to eliminate some of the fractions. For example, after Step 2 above, we could have multiplied by 3 to produce $(-2, 1, 1)$ as the second orthogonal basis vector, thereby simplifying the calculations in Step 3.

Erhardt Schmidt (1875–1959)

Historical Note Schmidt wasa German mathematician who studied for his doctoral degree at Göttingen University under David Hilbert, one of the giants of modern mathematics. For most of his life he taught at Berlin University where, in addition to making important contributions to many branches of mathematics, he fashioned some of Hilbert's ideas into a general concept, called a *Hilbert space*—a fundamental idea in

the study of infinite-dimensional vector spaces.He first described the process that bears his name in a paper on integral equations that he published in 1907.
[*Image: Archives of the Mathematisches Forschungsinst*]

Jorgen Pederson Germ (1850–1916)

Historical Note Gram was a Danish actuary whose early education was at village schools supplementedby private tutoring. He obtained a doctorate degree in mathematics while working for the Hafnia Life Insurance Company, where he specialized in the mathematics of accident insurance.It was in his dissertation that his contributions to the Gram–Schmidt process were formulated. He eventually became interested in abstract mathematics and received a gold medal from the Royal Danish Society of Sciences and Letters in recognition of his work. His lifelong interest in applied mathematics never wavered, however, and he produced a variety of treatises on Danish forest management.
[*Image: wikipedia*]

CALCULUS REQUIRED

EXAMPLE 8 Legendre Polynomials

Let the vector space P_2 have the inner product

$$\left\langle \mathbf{p}, \mathbf{q} \right\rangle = \int_{-1}^{1} p(x)q(x)\,dx$$

Apply the Gram–Schmidt process to transform the standard basis $\left\{1, x, x^2\right\}$ for P_2 into an orthogonal basis $\{\phi_1(x), \phi_2(x), \phi_3(x)\}$.

Solution Take $\mathbf{u}_1 = 1$, $\mathbf{u}_2 = x$, and $\mathbf{u}_3 = x^2$.

Step 1. $\mathbf{v}_1 = \mathbf{u}_1 = 1$

Step 2. We have

$$\langle \mathbf{u}_2, \mathbf{v}_1 \rangle = \int_{-1}^{1} x\,dx = 0$$

so

$$\mathbf{v}_2 = \mathbf{u}_2 - \frac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 = \mathbf{u}_2 = x$$

Step 3. We have

$$\langle \mathbf{u}_3, \mathbf{v}_1 \rangle = \int_{-1}^{1} x^2\,dx = \frac{x^3}{3}\Big]_{-1}^{1} = \frac{2}{3}$$

$$\langle \mathbf{u}_3, \mathbf{v}_2 \rangle = \int_{-1}^{1} x^3\,dx = \frac{x^4}{4}\Big]_{-1}^{1} = 0$$

$$\|\mathbf{v}_1\|^2 = \langle \mathbf{v}_1, \mathbf{v}_1 \rangle = \int_{-1}^{1} 1\,dx = x\Big]_{-1}^{1} = 2$$

so

$$\mathbf{v}_3 = \mathbf{u}_3 - \frac{\langle \mathbf{u}_3, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\langle \mathbf{u}_3, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 = x^2 - \frac{1}{3}$$

Thus, we have obtained the orthogonal basis $\{\phi_1(x), \phi_2(x), \phi_3(x)\}$ in which

$$\phi_1(x) = 1, \quad \phi_2(x) = x, \quad \phi_3(x) = x^2 - \frac{1}{3}$$

Remark The orthogonal basis vectors in the foregoing example are often scaled so all three functions have a value of 1 at $x = 1$. The resulting polynomials

$$1, \quad x, \quad \frac{1}{2}\left(3x^2 - 1\right)$$

which are known as the first three ***Legendre polynomials***, play an important role in a variety of applications. The scaling does not affect the orthogonality.

Extending Orthonormal Sets to Orthonormal Bases

Recall from part (*b*) of Theorem 4.5.5 that a linearly independent set in a finite-dimensional vector space can be enlarged to a basis by adding appropriate vectors. The following theorem is an analog of that result for orthogonal and orthonormal sets in finite-dimensional inner product spaces.

THEOREM 6.3.6

If W is a finite-dimensional inner product space, then:

(a) Every orthogonal set of nonzero vectors in W can be enlarged to an orthogonal basis for W.

(b) Every orthonormal set in W can be enlarged to an orthonormal basis for W.

We will prove part (*b*) and leave part (*a*) as an exercise.

Proof (b) Suppose that $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\}$ is an orthonormal set of vectors in *W*. Part (*b*) of Theorem 4.5.5 tells us that we can enlarge *S* to some basis

$$S'' = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s, \mathbf{v}_{s+1}, \ldots, \mathbf{v}_k\}$$

for *W*. If we now apply the Gram–Schmidt process to the set S', then the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s$, will not be affected since they are already orthonormal, and the resulting set

$$S'' = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s, \mathbf{v}_{s+1}, \ldots, \mathbf{v}_k\}$$

will be an orthonormal basis for *W*.

OPTIONAL

QR-Decomposition

In recent years a numerical algorithm based on the Gram–Schmidt process, and known as ***QR-decomposition***, has assumed growing importance as the mathematical foundation for a wide variety of numerical algorithms, including those for computing eigenvalues of large matrices. The technical aspects of such algorithms are discussed in textbooks that specialize in the numerical aspects of linear algebra. However, we will discuss some of the underlying ideas here. We begin by posing the following problem.

Problem

If *A* is an $m \times n$ matrix with linearly independent column vectors, and if *Q* is the matrix that results by applying the Gram–Schmidt process to the column vectors of *A*, what relationship, if any, exists between *A* and *Q*?

To solve this problem, suppose that the column vectors of *A* are $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ and the orthonormal column vectors of *Q* are $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$. Thus, *A* and *Q* can be written in partitioned form as

$$A = [\mathbf{u}_1 | \mathbf{u}_2 | \cdots | \mathbf{u}_n] \quad \text{and } Q = [\mathbf{q}_1 | \mathbf{q}_2 | \cdots | \mathbf{q}_n]$$

It follows from Theorem 6.3.2*b* that $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ are expressible in terms of the vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$ as

$$\begin{array}{rcccccc}
\mathbf{u}_1 & = & \langle \mathbf{u}_1, \mathbf{q}_1 \rangle \mathbf{q}_1 & + & \langle \mathbf{u}_1, \mathbf{q}_2 \rangle \mathbf{q}_2 & + \cdots + & \langle \mathbf{u}_1, \mathbf{q}_n \rangle \mathbf{q}_n \\
\mathbf{u}_2 & = & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle \mathbf{q}_1 & + & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle \mathbf{q}_2 & + \cdots + & \langle \mathbf{u}_2, \mathbf{q}_n \rangle \mathbf{q}_n \\
\vdots & & \vdots & & \vdots & & \vdots \\
\mathbf{u}_n & = & \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \mathbf{q}_1 & + & \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \mathbf{q}_2 & + \cdots + & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \mathbf{q}_n
\end{array}$$

Recalling from Section 1.3 (Example 9) that the *j*th column vector of amatrix product is a linear combination of the column vectors of the first factor with coefficients coming from the *j*th column of the second factor, it follows that these relationships can be expressed in matrix form as

$$[\mathbf{u}_1 | \mathbf{u}_2 | \cdots | \mathbf{u}_n] = [\mathbf{q}_1 | \mathbf{q}_2 | \cdots | \mathbf{q}_n] \begin{bmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \\ \langle \mathbf{u}_1, \mathbf{q}_2 \rangle & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \\ \vdots & \vdots & & \vdots \\ \langle \mathbf{u}_1, \mathbf{q}_n \rangle & \langle \mathbf{u}_2, \mathbf{q}_n \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \end{bmatrix}$$

or more briefly as

$$A = QR \tag{14}$$

where R is the second factor in the product. However, it is a property of the Gram–Schmidt process that for $j \geq 2$, the vector $\mathbf{q}_j$ is orthogonal to $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{j-1}$. Thus, all entries below the main diagonal of R are zero, and R has the form

$$R = \begin{bmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \\ 0 & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \end{bmatrix} \tag{15}$$

We leave it for you to show that R is invertible by showing that its diagonal entries are nonzero. Thus, Equation 14 is a factorization of A into the product of a matrix Q with orthonormal column vectors and an invertible upper triangular matrix R. We call Equation 14 the ***QR-decomposition of A***. In summary, we have the following theorem.

THEOREM 6.3.7 ***QR*-Decomposition**

If A is an $m \times n$ matrix with linearly independent column vectors, then A can be factored as

$$A = QR$$

where Q is an $m \times n$ matrix with orthonormal column vectors, and R is an $n \times n$ invertible upper triangular matrix.

It is common in numerical linear algebra to say that a matrix with linearly independent columns has ***full column rank***.

Recall from Theorem 5.1.6 (the Equivalence Theorem) that a *square* matrix has linearly independent column vectors if and only if it is invertible. Thus, it follows from the foregoing theorem that *every invertible matrix has a QR-decomposition*.

EXAMPLE 9 *QR*-Decomposition of a 3 × 3 Matrix

Find the *QR*-decomposition of

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

Solution The column vectors of A are

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

Applying the Gram–Schmidt process with normalization to these column vectors yields the

orthonormal vectors (see Example 7)

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

Thus, it follows from Formula 15 that R is

$$R = \begin{bmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \langle \mathbf{u}_3, \mathbf{q}_1 \rangle \\ 0 & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \langle \mathbf{u}_3, \mathbf{q}_2 \rangle \\ 0 & 0 & \langle \mathbf{u}_3, \mathbf{q}_3 \rangle \end{bmatrix} = \begin{bmatrix} \frac{3}{\sqrt{3}} & \frac{2}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Show that the matrix Q in Example 9 has the property $QQ^T = I$, and show that *every* $m \times n$ matrix with orthonormal column vectors has this property.

from which it follows that the QR-decomposition of A is

$$\begin{array}{ccccc} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} & = & \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{bmatrix} & \begin{bmatrix} \frac{3}{\sqrt{3}} & \frac{2}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \\ A & = & Q & R \end{array}$$

Concept Review

- Orthogonal and orthonormal sets
- Normalizing a vector
- Orthogonal projections
- Gram–Schmidt process
- QR-decomposition

Skills

- Determine whether a set of vectors is orthogonal (or orthonormal).
- Compute the coordinates of a vector with respect to an orthogonal (or orthonormal) basis.
- Find the orthogonal projection of a vector onto a subspace.
- Use the Gram–Schmidt process to construct an orthogonal (or orthonormal) basis for an inner product space.
- Find the QR-decomposition of an invertible matrix.

Exercise Set 6.3

1. Which of the following sets of vectors are orthogonal with respect to the Euclidean inner product on R^2?

(a) $(0, 1),\ (2, 0)$

(b) $\left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

(c) $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right), \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

(d) $(0, 0), (0, 1)$

Answer:

(a), (b), (d)

2. Which of the sets in Exercise 1 are orthonormal with respect to the Euclidean inner product on R^2?

3. Which of the following sets of vectors are orthogonal with respect to the Euclidean inner product on R^3?

(a) $\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right), \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}\right), \left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$

(b) $\left(\frac{2}{3}, -\frac{2}{3}, \frac{1}{3}\right), \left(\frac{2}{3}, \frac{1}{3}, -\frac{2}{3}\right), \left(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}\right)$

(c) $(1, 0, 0), \left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right),\ (0, 0, 1)$

(d) $\left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}\right), \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0\right)$

Answer:

(b), (d)

4. Which of the sets in Exercise 3 are orthonormal with respect to the Euclidean inner product on R^3?

5. Which of the following sets of polynomials are orthonormal with respect to the inner product on P_2 discussed in Example 7 of Section 6.1 ?

(a) $p_1(x) = \frac{2}{3} - \frac{2}{3}x + \frac{1}{3}x^2,\ p_2(x) = \frac{2}{3} + \frac{1}{3}x - \frac{2}{3}x^2,\ p_3(x) = \frac{1}{3} + \frac{2}{3}x + \frac{2}{3}x^2$

(b) $p_1(x) = 1,\ p_2(x) = \frac{1}{\sqrt{2}}x + \frac{1}{\sqrt{2}}x^2,\ p_3(x) = x^2$

Answer:

(a)

6. Which of the following sets of matrices are orthonormal with respect to the inner product on M_{22} discussed in Example 6 of Section 6.1 ?

(a) $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & \frac{2}{3} \\ \frac{1}{3} & -\frac{2}{3} \end{bmatrix}, \begin{bmatrix} 0 & \frac{2}{3} \\ -\frac{2}{3} & \frac{1}{3} \end{bmatrix}, \begin{bmatrix} 0 & \frac{1}{3} \\ \frac{2}{3} & \frac{2}{3} \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}$

7. Verify that the given vectors form an orthogonal set with respect to the Euclidean inner product; then convert it to an orthonormal set by normalizing the vectors.

(a) $(-1, 2), (6, 3)$

(b) $(1, 0, -1), (2, 0, 2), (0, 5, 0)$

(c) $\left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right), \left(-\frac{1}{2}, \frac{1}{2}, 0\right), \left(\frac{1}{3}, \frac{1}{3}, -\frac{2}{3}\right)$

Answer:

(a) $\left(-\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right), \left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right)$

(b) $\left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}\right), \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right), (0, 1, 0)$

(c) $\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right), \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}\right)$

8. Verify that the set of vectors $\{(1, 0), (0, 1)\}$ is orthogonal with respect to the inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 4u_1v_1 + u_2v_2$ on R^2; then convert it to an orthonormal set by normalizing the vectors.

9. Verify that the vectors

$$\mathbf{v}_1 = \left(-\frac{3}{5}, \frac{4}{5}, 0\right), \mathbf{v}_2 = \left(\frac{4}{5}, \frac{3}{5}, 0\right), \mathbf{v}_3 = (0, 0, 1)$$

form an orthonormal basis for R^3 with the Euclidean inner product; then use Theorem 6.3.2*b* to express each of the following as linear combinations of $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$.

(a) $(1, -1, 2)$

(b) $(3, -7, 4)$

(c) $\left(\frac{1}{7}, -\frac{3}{7}, \frac{5}{7}\right)$

Answer:

(a) $-\frac{7}{5}\mathbf{v}_1 + \frac{1}{5}\mathbf{v}_2 + 2\mathbf{v}_3$

(b) $-\frac{37}{5}\mathbf{v}_1 - \frac{9}{5}\mathbf{v}_2 + 4\mathbf{v}_3$

(c) $-\frac{3}{7}\mathbf{v}_1 - \frac{1}{7}\mathbf{v}_2 + \frac{5}{7}\mathbf{v}_3$

10. Verify that the vectors

$$\mathbf{v}_1 = (1, -1, 2, -1), \qquad \mathbf{v}_2 = (-2, 2, 3, 2),$$
$$\mathbf{v}_3 = (1, 2, 0, -1), \qquad \mathbf{v}_4 = (1, 0, 0, 1)$$

form an orthogonal basis for R^4 with the Euclidean inner product; then use Theorem 6.3.2*a* to express each of the following as linear combinations of $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, and $\mathbf{v}_4$.

(a) $(1, 1, 1, 1)$

(b) $\left(\sqrt{2}, -3\sqrt{2}, 5\sqrt{2}, -\sqrt{2}\right)$

(c) $\left(-\frac{1}{3}, \frac{2}{3}, -\frac{1}{3}, \frac{4}{3}\right)$

11. (a) Show that the vectors

$$\begin{aligned} \mathbf{v}_1 &= (1, -2, 3, -4), & \mathbf{v}_2 &= (2, 1, -4, -3), \\ \mathbf{v}_3 &= (-3, 4, 1, -2), & \mathbf{v}_4 &= (4, 3, 2, 1) \end{aligned}$$

form an orthogonal basis for R^4 with the Euclidean inner product.

(b) Use Theorem 6.3.2*a* to express $\mathbf{u} = (-1, 2, 3, 7)$ as a linear combination of the vectors in part (a).

Answer:

(b) $\mathbf{u} = -\frac{4}{5}\mathbf{v}_1 - \frac{11}{10}\mathbf{v}_2 + 0\mathbf{v}_3 + \frac{1}{2}\mathbf{v}_4$

In Exercises 12–13, an orthonormal basis with respect to the Euclidean inner product is given. Use Theorem 6.3.2*b* to find the coordinate vector of **w** with respect to that basis.

12. (a) $\mathbf{w} = (3, 7)$; $\mathbf{u}_1 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$, $\mathbf{u}_2 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

(b) $\mathbf{w} = (-1, 0, 2)$; $\mathbf{u}_1 = \left(\frac{2}{3}, -\frac{2}{3}, \frac{1}{3}\right)$, $\mathbf{u}_2 = \left(\frac{2}{3}, \frac{1}{3}, -\frac{2}{3}\right)$, $\mathbf{u}_3 = \left(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}\right)$

13. (a) $\mathbf{w} = (2, 0, 5)$; $\mathbf{u}_1 = \left(\frac{2}{3}, \frac{1}{3}, \frac{2}{3}\right)$, $\mathbf{u}_2 = \left(\frac{1}{3}, \frac{2}{3}, -\frac{2}{3}\right)$, $\mathbf{u}_3 = \left(\frac{2}{3}, -\frac{2}{3}, -\frac{1}{3}\right)$

(b) $\mathbf{w} = (-1, 1, 2)$; $\mathbf{u}_1 = \left(\frac{3}{\sqrt{11}}, \frac{1}{\sqrt{11}}, \frac{1}{\sqrt{11}}\right)$, $\mathbf{u}_2 = \left(-\frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right)$,
$\mathbf{u}_3 = \left(-\frac{1}{\sqrt{66}}, -\frac{4}{\sqrt{66}}, \frac{7}{\sqrt{66}}\right)$

Answer:

(a) $\mathbf{w} = \frac{14}{3}\mathbf{u}_1 - \frac{8}{3}\mathbf{u}_2 - \frac{1}{3}\mathbf{u}_3$

(b) $\mathbf{w} = \frac{5}{\sqrt{6}}\mathbf{u}_2 + \frac{11}{\sqrt{66}}\mathbf{u}_3$

In Exercises 14–15, the given vectors are orthogonal with respect to the Euclidean inner product. Find $\text{proj}_W \mathbf{x}$, where $\mathbf{x} = (1, 2, 0, -2)$ and W is the subspace of R^4 spanned by the vectors.

14. (a) $\mathbf{v}_1 = (1, 1, 1, 1), \mathbf{v}_2 = (1, 1, -1, -1)$

(b) $\mathbf{v}_1 = (0, 1, -4, -1), \mathbf{v}_2 = (3, 5, 1, 1)$

15. (a) $\mathbf{v}_1 = (1, 1, 1, 1), \mathbf{v}_2 = (1, 1, -1, -1), \mathbf{v}_3 = (1, -1, 1, -1)$

(b) $\mathbf{v}_1 = (0, 1, -4, -1), \mathbf{v}_2 = (3, 5, 1, 1), \mathbf{v}_3 = (1, 0, 1, -4)$

Answer:

(a) $\left(\frac{7}{4}, \frac{5}{4}, -\frac{3}{4}, -\frac{5}{4}\right)$

(b) $\left(\frac{17}{12}, \frac{7}{4}, -\frac{1}{12}, -\frac{23}{12}\right)$

In Exercises 16–17, the given vectors are orthonormal with respect to the Euclidean inner product. Use Theorem 6.3.4*b* to find $\text{proj}_W \mathbf{x}$, where $\mathbf{x} = (1, 2, 0, -1)$ and W is the subspace of R^4 spanned by the vectors.

16. (a) $\mathbf{v}_1 = \left(0, \frac{1}{\sqrt{18}}, -\frac{4}{\sqrt{18}}, -\frac{1}{\sqrt{18}}\right), \mathbf{v}_2 = \left(\frac{1}{2}, \frac{5}{6}, \frac{1}{6}, \frac{1}{6}\right)$

(b) $\mathbf{v}_1 = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right), \mathbf{v}_2 = \left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}\right)$

17. (a) $\mathbf{v}_1 = \left(0, \frac{1}{\sqrt{18}}, -\frac{4}{\sqrt{18}}, -\frac{1}{\sqrt{18}}\right), \mathbf{v}_2 = \left(\frac{1}{2}, \frac{5}{6}, \frac{1}{6}, \frac{1}{6}\right), \mathbf{v}_3 = \left(\frac{1}{\sqrt{18}}, 0, \frac{1}{\sqrt{18}}, -\frac{4}{\sqrt{18}}\right)$

(b) $\mathbf{v}_1 = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right), \mathbf{v}_2 = \left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}\right), \mathbf{v}_3 = \left(\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right)$

Answer:

(a) $\left(\frac{23}{18}, \frac{11}{6}, -\frac{1}{18}, -\frac{17}{18}\right)$

(b) $\left(\frac{3}{2}, \frac{3}{2}, -\frac{1}{2}, -\frac{1}{2}\right)$

18. In Example 6 of Section 4.9 we found the orthogonal projection of the vector $\mathbf{x} = (1, 5)$ onto the line through the origin making an angle of $\pi / 6$ radians with the *x*-axis. Solve that same problem using Theorem 6.3.4.

19. Find the vectors $\mathbf{w}_1$ in W and $\mathbf{w}_2$ in $W^\perp$ such that $\mathbf{x} = \mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{x}$ and W are as given in

(a) Exercise 14(a).

(b) Exercise 15(a).

Answer:

(a) $\mathbf{w}_1 = \left(\frac{3}{2}, \frac{3}{2}, -1, -1\right)$, $\mathbf{w}_2 = \left(-\frac{1}{2}, \frac{1}{2}, 1, -1\right)$

(b) $\mathbf{w}_1 = \left(\frac{7}{4}, \frac{5}{4}, -\frac{3}{4}, -\frac{5}{4}\right)$, $\mathbf{w}_2 = \left(-\frac{3}{4}, \frac{3}{4}, \frac{3}{4}, -\frac{3}{4}\right)$

20. Find the vectors $\mathbf{w}_1$ in W and $\mathbf{w}_2$ in $W^{\perp}$ such that $\mathbf{x} = \mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{x}$ and W are as given in

(a) Exercise 16(a).

(b) Exercise 17(a).

21. Let R^2 have the Euclidean inner product. Use the Gram–Schmidt process to transform the basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ into an orthonormal basis. Draw both sets of basis vectors in the *xy*-plane.

(a) $\mathbf{u}_1 = (1, -3)$, $\mathbf{u}_2 = (2, 2)$

(b) $\mathbf{u}_1 = (1, 0)$, $\mathbf{u}_2 = (3, -5)$

Answer:

(a) $\mathbf{v}_1 = \left(\frac{1}{\sqrt{10}}, -\frac{3}{\sqrt{10}}\right)$, $\mathbf{v}_2 = \left(\frac{3}{\sqrt{10}}, \frac{1}{\sqrt{10}}\right)$

(b) $\mathbf{v}_1 = (1, 0)$, $\mathbf{v}_2 = (0, -1)$

22. Let R^3 have the Euclidean inner product. Use theGram–Schmidt process to transform the basis $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ into an orthonormal basis.

(a) $\mathbf{u}_1 = (1, 1, 1), \mathbf{u}_2 = (-1, 1, 0), \mathbf{u}_3 = (1, 2, 1)$

(b) $\mathbf{u}_1 = (1, 0, 0), \mathbf{u}_2 = (3, 7, -2), \mathbf{u}_3 = (0, 4, 1)$

23. Let R^4 have the Euclidean inner product. Use the Gram–Schmidt process to transform the basis $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ into an orthonormal basis.

$$\mathbf{u}_1 = (0, 2, 1, 0), \qquad \mathbf{u}_2 = (1, -1, 0, 0),$$
$$\mathbf{u}_3 = (1, 2, 0, -1), \quad \mathbf{u}_4 = (1, 0, 0, 1)$$

Answer:

$$\mathbf{v}_1 = \left(0, \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}, 0\right), \ \mathbf{v}_2 = \left(\frac{5}{\sqrt{30}}, -\frac{1}{\sqrt{30}}, \frac{2}{\sqrt{30}}, 0\right),$$
$$\mathbf{v}_3 = \left(\frac{1}{\sqrt{10}}, \frac{1}{\sqrt{10}}, -\frac{2}{\sqrt{10}}, -\frac{2}{\sqrt{10}}\right), \ \mathbf{v}_4 = \left(\frac{1}{\sqrt{15}}, \frac{1}{\sqrt{15}}, -\frac{2}{\sqrt{15}}, \frac{3}{\sqrt{15}}\right)$$

24. Let R^3 have the Euclidean inner product. Find an orthonormal basis for the subspace spanned by $(0, 1, 2)$, $(-1, 0, 1)$, $(-1, 1, 3)$.

25. Let R^3 have the inner product

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + 2u_2 v_2 + 3u_3 v_3$$

Use the Gram–Schmidt process to transform $\mathbf{u}_1 = (1, 1, 1)$, $\mathbf{u}_2 = (1, 1, 0)$, $\mathbf{u}_3 = (1, 0, 0)$ into an orthonormal basis.

Answer:

$$\mathbf{v}_1 = \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right), \ \mathbf{v}_2 = \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}\right), \ \mathbf{v}_3 = \left(\frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, 0\right)$$

26. Let R^3 have the Euclidean inner product. The subspace of R^3 spanned by the vectors $\mathbf{u}_1 = \left(\frac{4}{5}, 0, -\frac{3}{5}\right)$ and $\mathbf{u}_2 = (0, 1, 0)$ is aplane passing through the origin. Express $\mathbf{w} = (1, 2, 3)$ in the form $\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1$ lies in the plane and $\mathbf{w}_2$ is perpendicular to the plane.

27. Repeat Exercise 26 with $\mathbf{u}_1 = (1, 1, 1)$ and $\mathbf{u}_2 = (2, 0, -1)$.

Answer:

$$\mathbf{w}_1 = \left(\frac{13}{14}, \frac{31}{14}, \frac{40}{14}\right), \ \mathbf{w}_2 = \left(\frac{1}{14}, -\frac{3}{14}, \frac{2}{14}\right)$$

28. Let R^4 have the Euclidean inner product. Express the vector $\mathbf{w} = (-1, 2, 6, 0)$ in the form $\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1$ is in the space W spanned by $\mathbf{u}_1 = (-1, 0, 1, 2)$ and $\mathbf{u}_2 = (0, 1, 0, 1)$, and $\mathbf{w}_2$ is orthogonal to W.

29. Find the QR-decomposition of the matrix, where possible.

(a) $\begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 4 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 1 \\ -2 & 1 \\ 2 & 1 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 1 & 2 & 0 \end{bmatrix}$

(e) $\begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 0 & 3 & 1 \end{bmatrix}$

(f) $\begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 1 \\ 1 & 0 & 1 \\ -1 & 1 & 1 \end{bmatrix}$

Answer:

(a) $\begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \sqrt{5} & \sqrt{5} \\ 0 & \sqrt{5} \end{bmatrix}$

(b) $\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 3\sqrt{2} \\ 0 & \sqrt{3} \end{bmatrix}$

(c) $\begin{bmatrix} \frac{1}{3} & \frac{8}{\sqrt{234}} \\ -\frac{2}{3} & \frac{11}{\sqrt{234}} \\ \frac{2}{3} & \frac{7}{\sqrt{234}} \end{bmatrix} \begin{bmatrix} 3 & \frac{1}{3} \\ 0 & \frac{\sqrt{26}}{3} \end{bmatrix}$

(d) $\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{2} \\ 0 & \sqrt{3} & -\frac{1}{\sqrt{3}} \\ 0 & 0 & \frac{4}{\sqrt{6}} \end{bmatrix}$

(e) $\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{\sqrt{2}}{2\sqrt{19}} & -\frac{3}{\sqrt{19}} \\ \frac{1}{\sqrt{2}} & -\frac{\sqrt{2}}{2\sqrt{19}} & \frac{3}{\sqrt{19}} \\ 0 & \frac{3\sqrt{2}}{\sqrt{19}} & \frac{1}{\sqrt{19}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \frac{3}{\sqrt{2}} & \sqrt{2} \\ 0 & \frac{\sqrt{19}}{\sqrt{2}} & \frac{3\sqrt{2}}{\sqrt{19}} \\ 0 & 0 & \frac{1}{\sqrt{19}} \end{bmatrix}$

(f) Columns not linearly independent

30. In Step 3 of the proof of Theorem 6.3.5, it was stated that "the linear independence of $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ ensures that $\mathbf{v}_3 \neq \mathbf{0}$." Prove this statement.

31. Prove that the diagonal entries of R in Formula 15 are nonzero.

32. ***Calculus required*** Use Theorem 6.3.2*a* to express the following polynomials as linear combinations of the first three Legendre polynomials (see the Remark following Example 8).

(a) $1 + x + 4x^2$

(b) $2 - 7x^2$

(c) $4 + 3x$

33. ***Calculus required*** Let P_2 have the inner product

$$\left\langle \mathbf{p}, \mathbf{q} \right\rangle = \int_0^1 p(x)q(x)\,dx$$

Apply the Gram–Schmidt process to transform the standard basis $S = \left\{1, x, x^2\right\}$ into an orthonormal basis.

Answer:

$\mathbf{v}_1 = 1,\ \mathbf{v}_2 = \sqrt{3}(2x - 1),\ \mathbf{v}_3 = \sqrt{5}(6x^2 - 6x + 1)$

34. Find vectors **x** and **y** in R^2 that are orthonormal with respect to the inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1v_1 + 2u_2v_2$ but are not orthonormal with respect to the Euclidean inner product.

True-False Exercises

In parts (a)–(f) determine whether the statement is true or false, and justify your answer.

(a) Every linearly independent set of vectors in an inner product space is orthogonal.

Answer:

False

(b) Every orthogonal set of vectors in an inner product space is linearly independent.

Answer:

False

(c) Every nontrivial subspace of R^3 has an orthonormal basis with respect to the Euclidean inner product.

Answer:

True

(d) Every nonzero finite-dimensional inner product space has an orthonormal basis.

Answer:

True

(e) $\text{proj}_W \mathbf{x}$ is orthogonal to every vector of W.

Answer:

False

(f) If A is an $n \times n$ matrix with a nonzero determinant, then A has a QR-decomposition.

Answer:

True

Copyright © 2010 John Wiley & Sons, Inc. All rights reserved.

6.4 Best Approximation; Least Squares

In this section we will be concerned with linear systems that cannot be solved exactly and for which an approximate solution is needed. Such systems commonly occur in applications where measurement errors "perturb" the coefficients of a consistent system sufficiently to produce inconsistency.

Least Squares Solutions of Linear Systems

Suppose that $A\mathbf{x} = \mathbf{b}$ is an *inconsistent* linear system of m equations in n unknowns in which we suspect the inconsistency to be caused by measurement errors in the coefficients of A. Since no exact solution is possible, we will look for a vector $\mathbf{x}$ that comes as "close as possible" to being a solution in the sense that it minimizes $\|\mathbf{b} - A\mathbf{x}\|$ with respect to the Euclidean inner product on R^m. You can think of $A\mathbf{x}$ as an approximation to $\mathbf{b}$ and $\|\mathbf{b} - A\mathbf{x}\|$ as the *error* in that approximation—the smaller the error, the better the approximation. This leads to the following problem.

Least Squares Problem

Given a linear system $A\mathbf{x} = \mathbf{b}$ of m equations in n unknowns, find a vector $\mathbf{x}$ that minimizes $\|\mathbf{b} - A\mathbf{x}\|$ with respect to the Euclidean inner product on R^m. We call such an $\mathbf{x}$ a ***least squares solution*** of the system, we call $\mathbf{b} - A\mathbf{x}$ the ***least squares error vector***, and we call $\|\mathbf{b} - A\mathbf{x}\|$ the ***least squares error***.

To clarify the above terminology, suppose that the matrix form of $\mathbf{b} - A\mathbf{x}$ is

$$\mathbf{b} - A\mathbf{x} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix}$$

The term "least squares solution" results from the fact that minimizing $\|\mathbf{b} - A\mathbf{x}\|$ also minimizes $\|\mathbf{b} - A\mathbf{x}\|^2 = e_1^2 + e_2^2 + \ldots + e_m^2$.

Best Approximation

Suppose that $\mathbf{b}$ is a fixed vector in R^3 that we would like to approximate by a vector $\mathbf{w}$ that is required to lie in some subspace W of R^3. Unless $\mathbf{b}$ happens to be in W, then any such approximation will result in an "error vector" $\mathbf{b} - \mathbf{w}$ that cannot be made equal to $\mathbf{0}$ no matter how $\mathbf{w}$ is chosen (Figure 6.4.1*a*). However, by choosing

$$\mathbf{w} = \text{proj}_W\, \mathbf{b}$$

we can make the length of the error vector

$$\|\mathbf{b} - \mathbf{w}\| = \|\mathbf{b} - \text{proj}_W\, \mathbf{b}\|$$

as small as possible (Figure 6.4.1*b*).

Figure 6.4.1

These geometric ideas suggest the following general theorem.

THEOREM 6.4.1 Best Approximation Theorem

If W is a finite-dimensional subspace of an inner product space V, and if $\mathbf{b}$ is a vector in V, then $\text{proj}_W\ \mathbf{b}$ is the ***best approximation*** to $\mathbf{b}$ from W in the sense that

$$\|\mathbf{b} - \text{proj}_W\ \mathbf{b}\| < \|\mathbf{b} - \mathbf{w}\|$$

for every vector $\mathbf{w}$ in W that is different from $\text{proj}_W\ \mathbf{b}$.

Proof For every vector $\mathbf{w}$ in W, we can write

$$\mathbf{b} - \mathbf{w} = (\mathbf{b} - \text{proj}_W\ \mathbf{b}) + (\text{proj}_W\ \mathbf{b} - \mathbf{w}) \tag{1}$$

But $\text{proj}_W\ \mathbf{b} - \mathbf{w}$ being a difference of vectors in W is itself in W; and since $\mathbf{b} - \text{proj}_W\ \mathbf{b}$ is orthogonal to W, the two terms on the right side of 1 are orthogonal. Thus, it follows from the Theorem of Pythagoras (Theorem 6.2.3) that

$$\|\mathbf{b} - \mathbf{w}\|^2 = \|\mathbf{b} - \text{proj}_W\ \mathbf{b}\|^2 + \|\text{proj}_W\ \mathbf{b} - \mathbf{w}\|^2$$

Since $\mathbf{w} \neq \text{proj}_W\ \mathbf{b}$, it follows that the second term in this sum is positive, and hence that

$$\|\mathbf{b} - \text{proj}_W\ \mathbf{b}\|^2 < \|\mathbf{b} - \mathbf{w}\|^2$$

Since norms are nonnegative, it follows (from a property of inequalities) that

$$\|\mathbf{b} - \text{proj}_W\ \mathbf{b}\| < \|\mathbf{b} - \mathbf{w}\|$$

Least Squares Solutions of Linear Systems

One way to find a least squares solution of $A\mathbf{x} = \mathbf{b}$ is to calculate the orthogonal projection $\text{proj}_W\ \mathbf{b}$ on the column space W of the matrix A and then solve the equation

$$A\mathbf{x} = \text{proj}_W\ \mathbf{b} \tag{2}$$

However, we can avoid the need to calculate the projection by rewriting 2 as

$$\mathbf{b} - A\mathbf{x} = \mathbf{b} - \text{proj}_W\ \mathbf{b}$$

and then multiplying both sides of this equation by A^T to obtain

$$A^T(\mathbf{b} - A\mathbf{x}) = A^T(\mathbf{b} - \text{proj}_W\ \mathbf{b}) \tag{3}$$

Since $\mathbf{b} - \text{proj}_W\ \mathbf{b}$ is the component of $\mathbf{b}$ that is orthogonal to the column space of A, it follows from Theorem 4.8.9*b* that this vector lies in the null space of A^T, and hence that

$$A^T(\mathbf{b} - \text{proj}_W\ \mathbf{b}) = 0$$

Thus, 3 simplifies to

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

which we can rewrite as

$$A^T A\mathbf{x} = A^T\mathbf{b} \tag{4}$$

This is called the ***normal equation*** or the ***normal system*** associated with $A\mathbf{x} = \mathbf{b}$. When viewed as a linear system, the individual equations are called the ***normal equations*** associated with $A\mathbf{x} = \mathbf{b}$.

In summary, we have established the following result.

THEOREM 6.4.2

For every linear system $A\mathbf{x} = \mathbf{b}$, the associated normal system

$$A^T A\mathbf{x} = A^T\mathbf{b} \tag{5}$$

is consistent, and all solutions of 5 are least squaressolutions of $A\mathbf{x} = \mathbf{b}$. Moreover, if W is the column space of A, and $\mathbf{x}$ is any least squares solution of $A\mathbf{x} = \mathbf{b}$, then the orthogonal projection of $\mathbf{b}$ on W is

$$\text{proj}_W\, \mathbf{b} = A\mathbf{x} \tag{6}$$

If a linear system is consistent, then its exact solutions are the same as its least squares solutions, in which case the error is zero.

EXAMPLE 1 Least Squares Solution

(a) Find all least squares solutions of the linear system

$$\begin{array}{rcrcr} x_1 & - & x_2 & = & 4 \\ 3x_1 & + & 2x_2 & = & 1 \\ -2x_1 & + & 4x_2 & = & 3 \end{array}$$

(b) Find the error vector and the error.

Solution

(a) It will be convenient to express the system in the matrix form $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix}$$

It follows that

$$A^T A = \begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 14 & -3 \\ -3 & 21 \end{bmatrix}$$

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 10 \end{bmatrix}$$

so the normal system $A^T A\mathbf{x} = A^T\mathbf{b}$ is

$$\begin{bmatrix} 14 & -3 \\ -3 & 21 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 10 \end{bmatrix}$$

Solving this system yields a unique least squares solution, namely,

$$x_1 = \frac{17}{95}, \quad x_2 = \frac{143}{285}$$

(b) The error vector is

$$\mathbf{b}-A\mathbf{x}=\begin{bmatrix}4\\1\\3\end{bmatrix}-\begin{bmatrix}1&-1\\3&2\\-2&4\end{bmatrix}\begin{bmatrix}\frac{17}{95}\\\frac{143}{285}\end{bmatrix}=\begin{bmatrix}4\\1\\3\end{bmatrix}-\begin{bmatrix}-\frac{92}{285}\\\frac{439}{285}\\\frac{95}{57}\end{bmatrix}=\begin{bmatrix}\frac{1232}{285}\\-\frac{154}{285}\\\frac{4}{3}\end{bmatrix}$$

and the error is

$$\|\mathbf{b}-A\mathbf{x}\|\approx 4.556$$

EXAMPLE 2 Orthogonal Projection on a Subspace

Find the orthogonal projection of the vector $\mathbf{u}=(-3,\ -3,8,9)$ on the subspace of R^4 spanned by the vectors

$$\mathbf{u}_1=(3,1,0,1),\quad \mathbf{u}_2=(1,2,1,1),\quad \mathbf{u}_3=(-1,0,2,\ -1)$$

Solution We could solve this problem by first using the Gram–Schmidt process to convert $\{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3\}$ into an orthonormal basis and then applying the method used in Example 6 of Section 6.3 . However, the following method is more efficient.

The subspace W of R^4 spanned by $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ is the column space of the matrix

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

Thus, if $\mathbf{u}$ is expressed as a column vector, we can find the orthogonal projection of $\mathbf{u}$ on W by finding a least squares solution of the system $A\mathbf{x}=\mathbf{u}$ and then calculating $\text{proj}_W\ \mathbf{u}=A\mathbf{x}$ from the least squares solution. The computations are as follows: The system $A\mathbf{x}=\mathbf{u}$ is

$$\begin{bmatrix}3&1&-1\\1&2&0\\0&1&2\\1&1&-1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}-3\\-3\\8\\9\end{bmatrix}$$

so

$$A^TA=\begin{bmatrix}3&1&0&1\\1&2&1&1\\-1&0&2&-1\end{bmatrix}\begin{bmatrix}3&1&-1\\1&2&0\\0&1&2\\1&1&-1\end{bmatrix}=\begin{bmatrix}11&6&-4\\6&7&0\\-4&0&6\end{bmatrix}$$

$$A^T\mathbf{u}=\begin{bmatrix}3&1&0&1\\1&2&1&1\\-1&0&2&-1\end{bmatrix}\begin{bmatrix}-3\\-3\\8\\9\end{bmatrix}=\begin{bmatrix}-3\\8\\10\end{bmatrix}$$

The normal system $A^TA\mathbf{x}=A^T\mathbf{u}$ in this case is

$$\begin{bmatrix}11&6&-4\\6&7&0\\-4&0&6\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}-3\\8\\10\end{bmatrix}$$

Solving this system yields

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$$

as the least squares solution of $A\mathbf{x} = \mathbf{u}$ (verify), so

$$\text{proj}_W\ \mathbf{u} = A\mathbf{x} = \begin{bmatrix} 3 & 1 & -1 \\ 1 & 2 & 0 \\ 0 & 1 & 2 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \\ 4 \\ 0 \end{bmatrix}$$

or, in comma-delimited notation, $\text{proj}_W\ \mathbf{u} = (-2, 3, 4, 0)$.

Uniqueness of Least Squares Solutions

In general, least squares solutions of linear systems are not unique. Although the linear system in Example 1 turned out to have a unique least squares solution, that occurred only because the coefficient matrix of the system happened to satisfy certain conditions that guarantee uniqueness. Our next theorem will show what those conditions are.

THEOREM 6.4.3

If A is an $m \times n$ matrix, then the following are equivalent.

(a) A has linearly independent column vectors.

(b) $A^T A$ is invertible.

Proof We will prove that $(a) \Rightarrow (b)$ and leave the proof that $(b) \Rightarrow (a)$ as an exercise.

(a) ⇒ (b) Assume that A has linearly independent column vectors. The matrix $A^T A$ has size $n \times n$, so we can prove that this matrix is invertible by showing that the linear system $A^T A\mathbf{x} = \mathbf{0}$ has only the trivial solution. But if $\mathbf{x}$ is any solution of this system, then $A\mathbf{x}$ is in the null space of A^T and also in the column space of A. By Theorem 4.8.9*b* these spaces are orthogonal complements, so part (*b*) of Theorem 6.2.4 implies that $A\mathbf{x} = \mathbf{0}$. But A is assumed to have linearly independent column vectors, so $\mathbf{x} = \mathbf{0}$ by Theorem 1.3.1.

As an exercise, try using Formula 7 to solve the problem in part (a) of Example 1.

The next theorem, which follows directly from Theorem 6.4.2 and Theorem 6.4.3, gives an explicit formula for the least squares solution of a linear system in which the coefficient matrix has linearly independent column vectors.

THEOREM 6.4.4

If A is an $m \times n$ matrix with linearly independent column vectors, then for every $m \times 1$ matrix $\mathbf{b}$, the linearsystem $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution. This solution is given by

$$\mathbf{x} = \left(A^T A\right)^{-1} A^T \mathbf{b} \tag{7}$$

Moreover, if W is the column space of A, then the orthogonalprojection of $\mathbf{b}$ on W is

$$\operatorname{proj}_W \mathbf{b} = A\mathbf{x} = A\left(A^T A\right)^{-1} A^T \mathbf{b} \tag{8}$$

OPTIONAL

The Role of QR-Decomposition in Least Squares Problems

Formulas 7 and 8 have theoretical use but are not well suited for numerical computation. In practice, least squares solutions of $A\mathbf{x} = \mathbf{b}$ are typically found by using some variation of Gaussian elimination to solve the normal equations or by using QR-decomposition and the following theorem.

THEOREM 6.4.5

If A is an $m \times n$ matrix with linearly independent column vectors, and if $A = QR$ is a QR-decomposition of A (see Theorem 6.3.7), then for each $\mathbf{b}$ in R^m the system $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution given by

$$\mathbf{x} = R^{-1} Q^T \mathbf{b} \tag{9}$$

A proof of this theorem and a discussion of its use can be found in many books on numerical methods of linear algebra. However, you can obtain Formula 9 by making the substitution $A = QR$ in 7 and using the fact that $Q^T Q = I$ to obtain

$$\begin{aligned}
\mathbf{x} &= \left((QR)^T (QR)\right)^{-1} (QR)^T \mathbf{b} \\
&= \left(R^T Q^T QR\right)^{-1} (QR)^T \mathbf{b} \\
&= R^{-1}\left(R^T\right)^{-1} R^T Q^T \mathbf{b} \\
&= R^{-1} Q^T \mathbf{b}
\end{aligned}$$

Orthogonal Projections on Subspaces of R^m

In Section 4.8 we showed how to compute orthogonal projections on the coordinate axes of a rectangular coordinate system in R^3 and more generally on lines through the origin of R^3. We will now consider the problem of finding orthogonal projections on *subspaces* of R^m. We begin with the following definition.

DEFINITION 1

If W is a subspace of R^m, then the linear transformation $P: R^m \to W$ that maps each vector $\mathbf{x}$ in R^m into its orthogonal projection $\operatorname{proj}_W \mathbf{x}$ in W is called the ***orthogonal projection of R^m*** on ***W***

It follows from Formula 7 that the standard matrix for the transformation P is

$$[P] = A\left(A^T A\right)^{-1} A^T \tag{10}$$

where A is constructed using any basis for W as its column vectors.

EXAMPLE 3 The Standard Matrix for an Orthogonal Projection on a Line

We showed in Formula 16 of Section 4.9 that

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

is the standard matrix for the orthogonal projection on the line W through the origin of R^2 that makes an angle θ with the positive x-axis. Derive this result using Formula 10.

Solution The column vectors of A can be formed from any basis for W. Since W is one-dimensional, we can take $\mathbf{w} = (\cos\theta, \sin\theta)$ as the basis vector (Figure 6.4.2), so

$$A = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$$

We leave it for you to show that $A^T A$ is the 1×1 identity matrix. Thus, Formula 10 simplifies to

$$\begin{aligned} [P] &= A\left(A^T A\right)^{-1} A^T = AA^T = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} [\cos\theta \quad \sin\theta] \\ &= \begin{bmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{bmatrix} = P_\theta \end{aligned}$$

Figure 6.4.2

Another View of Least Squares

Recall from Theorem 4.8.9 that the null space and row space of an $m \times n$ matrix A are orthogonal complements, as are the null space of A^T and the column space of A. Thus, given a linear system $A\mathbf{x} = \mathbf{b}$ in which A is an $m \times n$ matrix, the Projection Theorem (6.3.3) tells us that the vectors $\mathbf{x}$ and $\mathbf{b}$ can each be decomposed into sums of orthogonal terms as

$$\mathbf{x} = \mathbf{x}_{\text{row}(A)} + \mathbf{x}_{\text{null}(A)} \quad \text{and} \quad \mathbf{b} = \mathbf{b}_{\text{null}\left(A^T\right)} + \mathbf{b}_{\text{col}(A)}$$

where $\mathbf{x}_{\text{row}(A)}$ and $\mathbf{x}_{\text{null}(A)}$ are the orthogonal projections of $\mathbf{x}$ on the row space of A and the null space of A, and the vectors $\mathbf{b}_{\text{null}\left(A^T\right)}$ and $\mathbf{b}_{\text{col}(A)}$ are the orthogonal projections of $\mathbf{b}$ on the null space of A^T and the column space of A.

In Figure 6.4.3 we have represented the fundamental spaces of A by perpendicular lines in R^n and R^m on which we indicated the orthogonal projections of $\mathbf{x}$ and $\mathbf{b}$. (This, of course, is only pictorial since the fundamental spaces need not be one-dimensional.) The figure shows $A\mathbf{x}$ as a point in the column space of A and conveys that $\mathbf{b}_{\text{col}(A)}$ is the point in col(A) that is closest to $\mathbf{b}$. This

illustrates that the least squares solutions of $A\mathbf{x} = \mathbf{b}$ are the exact solutions of the equation $A\mathbf{x} = \mathbf{b}_{\text{col}(A)}$.

Figure 6.4.3

More on the Equivalence Theorem

As our final result in the main part of this section we will add one additional part to Theorem 5.1.6.

THEOREM 6.4.6 Equivalent Statements

If A is an $n \times n$ matrix, then the following statements are equivalent.

(a) A is invertible.

(b) $A\mathbf{x} = 0$ has only the trivial solution.

(c) The reduced row echelon form of A is I_n.

(d) A is expressible as a product of elementary matrices.

(e) $A\mathbf{x} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.

(f) $A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.

(g) $\det(A) \neq 0$.

(h) The column vectors of A are linearly independent.

(i) The row vectors of A are linearly independent.

(j) The column vectors of A span R^n.

(k) The row vectors of A span R^n.

(l) The column vectors of A form a basis for R^n.

(m) The row vectors of A form a basis for R^n.

(n) A has $\text{rank} n$.

(o) A has nullity 0.

(p) The orthogonal complement of the null space of A is R^n.

(q) The orthogonal complement of the row space of A is $\{\mathbf{0}\}$.

(r) The range of T_A is R^n.

(s) T_A is one-to-one.

(t) $\lambda = 0$ is not an eigenvalue of A.

(u) $A^T A$ is invertible.

The proof of part (*u*) follows from part (*h*) of this theorem and Theorem 6.4.3 applied to square matrices.

OPTIONAL

We now have all the ingredients needed to prove Theorem 6.3.3 in the special case where V is the vector space R^m.

Proof of Theorem 6.3.3 We will leave the case where $W = \{0\}$ as an exercise, so assume that $W \neq \{0\}$. Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be any basis for W, and form the $m \times k$ matrix M that has these basis vectors as successive columns. This makes W the column space of M and hence $W^\perp$ the null space of M^T. We will complete the proof by showing that every vector $\mathbf{u}$ in R^m can be written in exactly one way as

$$\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$$

where $\mathbf{w}_1$ is in the column space of M and $M^T\mathbf{w}_2 = \mathbf{0}$. However, to say that $\mathbf{w}_1$ is in the column space of M is equivalent to saying $\mathbf{w}_1 = M\mathbf{x}$ for some vector $\mathbf{x}$ in R^m, and to say that $M^T\mathbf{w}_2 = \mathbf{0}$ is equivalent to saying that $M^T(\mathbf{u} - \mathbf{w}_1) = \mathbf{0}$. Thus, if we can show that the equation

$$M^T(\mathbf{u} - M\mathbf{x}) = 0 \tag{11}$$

has a unique solution for $\mathbf{x}$, then $\mathbf{w}_1 = M\mathbf{x}$ and $\mathbf{w}_2 = \mathbf{x} - \mathbf{w}_1$ will be uniquely determined vectors with the required properties. To do this, let us rewrite 11 as

$$M^T M\mathbf{x} = M^T\mathbf{u}$$

Since the matrix M has linearly independent column vectors, the matrix $M^T M$ is invertible by Theorem 6.4.6 and hence the equation has a unique solution as required to complete the proof.

Concept Review

- Least squares problem
- Least squares solution
- Least squares error vector
- Least squares error
- Best approximation
- Normal equation
- Orthogonal projection

Skills

- Find the least squares solution of a linear system.
- Find the error and error vector associated with a least squares solution to a linear system.
- Use the techniques developed in this section to compute orthogonal projections.
- Find the standard matrix of an orthogonal projection.

Exercise Set 6.4

1. Find the normal system associated with the given linear system.

(a) $\begin{bmatrix} 1 & -1 \\ 2 & 3 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}$

(b) $\begin{bmatrix} 2 & -1 & 0 \\ 3 & 1 & 2 \\ -1 & 4 & 5 \\ 1 & 2 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}$

Answer:

(a) $\begin{bmatrix} 21 & 25 \\ 25 & 35 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 20 \\ 20 \end{bmatrix}$

(b) $\begin{bmatrix} 15 & -1 & 5 \\ -1 & 22 & 30 \\ 5 & 30 & 45 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 9 \\ 13 \end{bmatrix}$

In Exercises 2–4, find the least squares solution of the linear equation $A\mathbf{x} = \mathbf{b}$.

2. (a) $A = \begin{bmatrix} 1 & -1 \\ 2 & 3 \\ 4 & 5 \end{bmatrix}; \mathbf{b} = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}$

(b) $A = \begin{bmatrix} 2 & -2 \\ 1 & 1 \\ 3 & 1 \end{bmatrix}; \mathbf{b} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$

3. (a) $A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ -1 & 2 \end{bmatrix}; \mathbf{b} = \begin{bmatrix} 7 \\ 0 \\ -7 \end{bmatrix}$

(b) $A = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & -2 \\ 1 & 1 & 0 \\ 1 & 1 & -1 \end{bmatrix}; \mathbf{b} = \begin{bmatrix} 6 \\ 0 \\ 9 \\ 3 \end{bmatrix}$

Answer:

(a) $x_1 = 5,\ x_2 = \frac{1}{2}$

(b) $x_1 = 12,\ x_2 = -3,\ x_3 = 9$

4. (a) $A = \begin{bmatrix} 3 & 2 & -1 \\ 1 & -4 & 3 \\ 1 & 10 & -7 \end{bmatrix}; \mathbf{b} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$

(b) $A = \begin{bmatrix} 2 & 0 & -1 \\ 1 & -2 & 2 \\ 2 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}; \mathbf{b} = \begin{bmatrix} 0 \\ 6 \\ 0 \\ 6 \end{bmatrix}$

In Exercises 5–6, find the least squares error vector $\mathbf{e} = \mathbf{b} - A\mathbf{x}$ resulting from the least squares solution $\mathbf{x}$ and verify that it is orthogonal to the column space of A.

5. (a) A and $\mathbf{b}$ are as in Exercise 3(a).

(b) A and $\mathbf{b}$ are as in Exercise 3(b).

Answer:

(a) $\mathbf{e} = \begin{bmatrix} \frac{3}{2} \\ \frac{9}{2} \\ -3 \end{bmatrix}$

(b) $\mathbf{e} = \begin{bmatrix} 3 \\ -3 \\ 0 \\ 3 \end{bmatrix}$

6. (a) A and $\mathbf{b}$ are as in Exercise 4(a).

(b) A and $\mathbf{b}$ are as in Exercise 4(b).

7. Find all least squares solutions of $A\mathbf{x} = \mathbf{b}$ andconfirm that all of the solutions have the same error vector. Compute the least squares error.

(a) $A = \begin{bmatrix} 2 & 1 \\ 4 & 2 \\ -2 & 1 \end{bmatrix}; \mathbf{b} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$

(b) $A = \begin{bmatrix} 1 & 3 \\ -2 & -6 \\ 3 & 9 \end{bmatrix}; \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

(c) $A = \begin{bmatrix} -1 & 3 & 2 \\ 2 & 1 & 3 \\ 0 & 1 & 1 \end{bmatrix}; \mathbf{b} = \begin{bmatrix} 7 \\ 0 \\ -7 \end{bmatrix}$

Answer:

(a) Solution: $\mathbf{x} = \left(\frac{1}{10}, \frac{6}{5}\right)$; least squares error: $\frac{4}{5}\sqrt{5}$

(b) Solution: $\mathbf{x} = \left(\frac{2}{7}, 0\right) + t(-3, 1)$ (t a real number); least squares error: $\frac{1}{7}\sqrt{42}$

(c) Solution: $\mathbf{x} = \left(-\frac{7}{6}, \frac{7}{6}, 0\right) + t(-1, -1, 1)$ (t a real number); least squares error: $\frac{1}{2}\sqrt{294}$

8. Find the orthogonal projection of $\mathbf{u}$ on the subspace of R^3 spanned by the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$.

(a) $\mathbf{u} = (2, 1, 3); \quad \mathbf{v}_1 = (1, 1, 0), \quad \mathbf{v}_2 = (1, 2, 1)$

(b) $\mathbf{u} = (1, -6, 1); \quad \mathbf{v}_1 = (-1, 2, 1), \quad \mathbf{v}_2 = (2, 2, 4)$

9. Find the orthogonal projection of $\mathbf{u}$ on the subspace of R^4 spanned by the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$.

(a) $\mathbf{u} = (6, 3, 9, 6); \mathbf{v}_1 = (2, 1, 1, 1), \mathbf{v}_2 = (1, 0, 1, 1), \mathbf{v}_3 = (-2, -1, 0, -1)$

(b) $\mathbf{u} = (-2, 0, 2, 4); \mathbf{v}_1 = (1, 1, 3, 0), \mathbf{v}_2 = (-2, -1, -2, 1), \mathbf{v}_3 = (-3, -1, 1, 3)$

Answer:

(a) $(7, 2, 9, 5)$

(b) $\left(-\frac{12}{5}, -\frac{4}{5}, \frac{12}{5}, \frac{16}{5}\right)$

10. Find the orthogonal projection of $\mathbf{u} = (5, 6, 7, 2)$ on the solution space of the homogeneous linear system

$$\begin{aligned} x_1 + x_2 + x_3 \quad &= 0 \\ 2x_2 + x_3 + x_4 &= 0 \end{aligned}$$

11. In each part, find $\det\left(A^T A\right)$, and apply Theorem 6.4.3 to determine whether A has linearly independent column vectors.

(a) $A = \begin{bmatrix} -1 & 3 & 2 \\ 2 & 1 & 3 \\ 0 & 1 & 1 \end{bmatrix}$

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

Answer:

(a) $\det(A^T A) = 0$; A does not have linearly independent column vectors.

(b) $\det(A^T A) = 0$; A does not have linearly independent column vectors.

12. Use Formula 10 and the method of Example 3 to find the standard matrix for the orthogonal projection $P: R^2 \to R^2$ onto

(a) the x-axis.

(b) the y-axis.

[*Note:* Compare your results to Table 3 of Section 4.9.]

13. Use Formula 10 and the method of Example 3 to find the standard matrix for the orthogonal projection $P: R^3 \to R^3$ onto

(a) the xz-plane.

(b) the yz-plane.

[*Note:* Compare your results to Table 4 of Section 4.9.]

Answer:

(a) $[P] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(b) $[P] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

14. Show that if $\mathbf{w} = (a, b, c)$ is a nonzero vector, then the standard matrix for the orthogonal projection of R^3 on the line $\operatorname{span}\{\mathbf{w}\}$ is

$$P = \frac{1}{a^2 + b^2 + c^2} \begin{bmatrix} a^2 & ab & ac \\ ab & b^2 & bc \\ ac & bc & c^2 \end{bmatrix}$$

15. Let W be the plane with equation $5x - 3y + z = 0$.

(a) Find a basis for W.

(b) Use Formula 10 to find the standard matrix for the orthogonal projection on W.

(c) Use the matrix obtained in part (b) to find the orthogonal projection of a point $P_0(x_0, y_0, z_0)$ on W.

(d) Find the distance between the point $P_0(1, -2, 4)$ and the plane W, and check your result using Theorem 3.3.4.

Answer:

(a) $(1, 0, -5)$, $(0, 1, 3)$

(b) $[P] = \frac{1}{35} \begin{bmatrix} 10 & 15 & -5 \\ 15 & 26 & 3 \\ -5 & 3 & 34 \end{bmatrix}$

(c) $\left(\frac{2x_0+3y_0-z_0}{7}, \frac{15x_0+26y_0+3z_0}{35}, \frac{-5x_0+3y_0+34z_0}{35}\right)$

(d) $\frac{3\sqrt{35}}{7}$

16. Let W be the line with parametric equations

$$x=2t, \quad y=-t, \quad z=4t$$

(a) Find a basis for W.

(b) Use Formula 10 to find the standard matrix for the orthogonal projection on W.

(c) Use the matrix obtained in part (b) to find the orthogonalprojection of a point $P_0(x_0, y_0, z_0)$ on W.

(d) Find the distance between the point $P_0(2, 1, -3)$ and the line W.

17. In R^3, consider the line l given by the equations

$$x=t, \quad y=t, \quad z=t$$

and the line m given by the equations

$$x=s, \quad y=2s-1, \quad z=1$$

Let P be a point on l, and let Q be a point on m. Find the values of t and s that minimize the distance between the lines by minimizing the squared distance $\|P-Q\|^2$.

Answer:

$s=t=1$

18. Prove: If A has linearly independent column vectors, and if $A\mathbf{x}=\mathbf{b}$ is consistent, then the least squares solution of $A\mathbf{x}=\mathbf{b}$ and the exact solution of $A\mathbf{x}=\mathbf{b}$ are the same.

19. Prove: If A has linearly independent column vectors, and if $\mathbf{b}$ is orthogonal to the column space of A, then the least squares solution of $A\mathbf{x}=\mathbf{b}$ is $\mathbf{x}=\mathbf{0}$.

20. Let $P: R^m \to W$ be the orthogonal projection of R^m onto a subspace W.

(a) Prove that $[P]^2=[P]$.

(b) What does the result in part (a) imply about the composition $P \circ P$?

(c) Show that $[P]$ is symmetric.

21. Let A be an $m \times n$ matrix with linearly independent row vectors. Find a standard matrix for the orthogonal projection of R^n onto the row space of A. [*Hint:* Start with Formula 10.]

Answer:

$[P]=A^T(AA^T)^{-1}A$

22. Prove the implication $(b) \Rightarrow (a)$ of Theorem 6.4.3.

True-False Exercises

In parts (a)–(h) determine whether the statement is true or false, and justify your answer.

(a) If A is an $m \times n$ matrix, then A^TA is a square matrix.

Answer:

True

(b) If A^TA is invertible, then A is invertible.

Answer:

False

(c) If A is invertible, then $A^T A$ is invertible.

Answer:

True

(d) If $A\mathbf{x} = \mathbf{b}$ is a consistent linear system, then $A^T A\mathbf{x} = A^T\mathbf{b}$ is also consistent.

Answer:

True

(e) If $A\mathbf{x} = \mathbf{b}$ is an inconsistent linear system, then $A^T A\mathbf{x} = A^T\mathbf{b}$ is also inconsistent.

Answer:

False

(f) Every linear system has a least squares solution.

Answer:

True

(g) Every linear system has a unique least squares solution.

Answer:

False

(h) If A is an $m \times n$ matrix with linearly independent columns and **b** is in R^m, then $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution.

Answer:

True

Copyright © 2010 John Wiley & Sons, Inc. All rights reserved.

6.5 Least Squares Fitting to Data

In this section we will use results about orthogonal projections in inner product spaces to obtain a technique for fitting a line or other polynomial curve to a set of experimentally determined points in the plane.

Fitting a Curve to Data

A common problem in experimental work is to obtain a mathematical relationship $y = f(x)$ between two variables x and y by "fitting" a curve to points in the plane corresponding to various experimentally determined values of x and y, say

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

On the basis of theoretical considerations or simply by observing the pattern of the points, the experimenter decides on the general form of the curve $y = f(x)$ to be fitted. Some possibilities are (Figure 6.5.1)

(a) A straight line: $y = a + bx$

(b) A quadratic polynomial: $y = a + bx + cx^2$

(c) A cubic polynomial: $y = a + bx + cx^2 + dx^3$

Because the points are obtained experimentally, there is often some measurement "error" in the data, making it impossible to find a curve of the desired form that passes through all the points. Thus, the idea is to choose the curve (by determining its coefficients) that "best" fits the data. We begin with the simplest and most common case: fitting a straight line to data points.

Figure 6.5.1

Least Squares Fit of a Straight Line

Suppose we want to fit a straight line $y = a + bx$ to the experimentally determined points

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

If the data points were collinear, the line would pass through all n points, and the unknown coefficients a and b would satisfy the equations

$$\begin{aligned} y_1 &= a + bx_1 \\ y_2 &= a + bx_2 \\ &\vdots \\ y_n &= a + bx_n \end{aligned}$$

We can write this system in matrix form as

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

or more compactly as

$$M\mathbf{v} = \mathbf{y} \tag{1}$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad M = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix} \tag{2}$$

If the data points are not collinear, then it is impossible to find coefficients *a* and *b* that satisfy system 1 exactly; that is, the system is inconsistent. In this case we will look for a least squares solution

$$\mathbf{v} = \mathbf{v}^* = \begin{bmatrix} a^* \\ b^* \end{bmatrix}$$

We call a line $y = a^* + b^* x$ whose coefficients come from a least squares solution a ***regression line*** or a ***least squares straight line fit*** to the data. To explain this terminology, recall that a least squares solution of 1 minimizes

$$\|\mathbf{y} - M\mathbf{v}\| \tag{3}$$

If we express the square of 3 in terms of components, we obtain

$$\|\mathbf{y} - M\mathbf{v}\|^2 = (y_1 - a - bx_1)^2 + (y_2 - a - bx_2)^2 + \ldots + (y_n - a - bx_n)^2 \tag{4}$$

If we now let

$$d_1 = |y_1 - a - bx_1|, \quad d_2 = |y_2 - a - bx_2|, \ldots, \quad d_n = |y_n - a - bx_n|$$

then 4 can be written as

$$\|\mathbf{y} - M\mathbf{v}\|^2 = d_1^2 + d_2^2 + \ldots + d_n^2 \tag{5}$$

As illustrated in Figure 6.5.2, the number d_i can be interpreted as the vertical distance between the line $y = a + bx$ and the data point (x_i, y_i). This distance is a measure of the "error" at the point (x_i, y_i)

resulting from the inexact fit of $y = a + bx$ to the data points, the assumption being that the x_i are known exactly and that all the error is in the measurement of the y_i. Since 3 and 5 are minimized by the same vector $\mathbf{v}^*$, the least squares straight line fit minimizes the sum of the squares of the estimated errors d_i, hence the name *least squares straight line fit*.

Figure 6.5.2 d_i measures the vertical error in the least squares straight line.

Normal Equations

Recall from Theorem 6.4.2 that the least squares solutions of 1 can be obtained by solving the associated normal system

$$M^T M \mathbf{v} = M^T \mathbf{y}$$

the equations of which are called the *normal equations*.

In the exercises it will be shown that the column vectors of M are linearly independent if and only if the n data points do not lie on a vertical line in the xy-plane. In this case it follows from Theorem 6.4.4 that the least squares solution is unique and is given by

$$\mathbf{v}^* = \left(M^T M\right)^{-1} M^T \mathbf{y}$$

In summary, we have the following theorem.

THEOREM 6.5.1 Uniqueness of the Least Squares Solution

Let $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be a set of two or more data points, not all lying on a vertical line, and let

$$M = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Then there is a unique least squares straight line fit

$$y = a^* + b^* x$$

to the data points. Moreover,

$$\mathbf{v}^* = \begin{bmatrix} a^* \\ b^* \end{bmatrix}$$

is given by the formula

$$\mathbf{v}^* = \left(M^T M\right)^{-1} M^T \mathbf{y} \tag{6}$$

which expresses the fact that $\mathbf{v} = \mathbf{v}^*$ is the unique solution of the normal equations

$$M^T M \mathbf{v} = M^T \mathbf{y} \tag{7}$$

EXAMPLE 1 Least Squares Straight Line Fit

Find the least squares straight line fit to the four points $(0, 1)$, $(1, 3)$, $(2, 4)$, and $(3, 4)$. (See Figure 6.5.3.)

Figure 6.5.3

Solution We have

$$M = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \quad M^T M = \begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix}, \quad \text{and } \left(M^T M\right)^{-1} = \frac{1}{10}\begin{bmatrix} 7 & -3 \\ -3 & 2 \end{bmatrix}$$

$$v^* = \left(M^T M\right)^{-1} M^T y = \frac{1}{10}\begin{bmatrix} 7 & -3 \\ -3 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 1 \end{bmatrix}$$

so the desired line is $y = 1.5 + x$.

EXAMPLE 2 Spring Constant

Hooke's law in physics states that the length x of a uniform spring is a linear function of the force y applied to it. If we express this relationship as $y = a + bx$, then the coefficient b is called the ***spring constant***. Suppose a particular unstretched spring has a measured length of 6.1 inches (i.e., $x = 6.1$ when $y = 0$). Forces of 2 pounds, 4 pounds, and 6 pounds are then applied to the spring, and the corresponding lengths are found to be 7.6 inches, 8.7 inches, and 10.4 inches (see Figure 6.5.4). Find the spring constant.

x_i	y_i
6.1	0
7.6	2
8.7	4
10.4	6

Figure 6.5.4

Solution We have

$$M = \begin{bmatrix} 1 & 6.1 \\ 1 & 7.6 \\ 1 & 8.7 \\ 1 & 10.4 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 0 \\ 2 \\ 4 \\ 6 \end{bmatrix},$$

and

$$\mathbf{v}^* = \begin{bmatrix} a^* \\ b^* \end{bmatrix} = \left(M^T M\right)^{-1} M^T \mathbf{y} \approx \begin{bmatrix} -8.6 \\ 1.4 \end{bmatrix}$$

where the numerical values have been rounded to one decimal place. Thus, the estimated value of the spring constant is $b^* \approx 1.4$ pounds/inch.

Source: NASA

Historical Note On October 5, 1991 the *Magellan* spacecraft entered the atmosphere of Venus and transmitted thetemperature T in kelvins (K) versus the altitude h in kilometers (km) until its signal was lost at an altitude of about 34 km. Discounting theinitial erratic signal, the data strongly suggested a linear relationship, so a least squares straight line fit was used on the linear part of the data to obtain the equation

$$T = 737.5 - 8.125h$$

By setting $h = 0$ in this equation, the surface temperature of Venus was estimated at $T \approx 737.5$K.

Least Squares Fit of a Polynomial

The technique described for fitting a straight line to data points can be generalized to fitting a polynomial of specified degree to data points. Let us attempt to fit a polynomial of fixed degree m

$$y = a_0 + a_1 x + \ldots + a_m x^m \tag{8}$$

to n points

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

Substituting these n values of x and y into 8 yields the n equations

$$\begin{array}{ccccccc} y_1 & = & a_0 & + & a_1 x_1 & + \ldots + & a_m x_1^m \\ y_2 & = & a_0 & + & a_1 x_2 & + \ldots + & a_m x_2^m \\ \vdots & & \vdots & & \vdots & & \vdots \\ y_n & = & a_0 & + & a_1 x_n & + \ldots + & a_m x_n^m \end{array}$$

or, in matrix form,

$$\mathbf{y} = M\mathbf{v} \tag{9}$$

where

$$\mathbf{y}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad M=\begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^m \\ 1 & x_2 & x_2^2 & \dots & x_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^m \end{bmatrix}, \quad \mathbf{v}=\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} \tag{10}$$

As before, the solutions of the normal equations

$$M^T M\mathbf{v}=M^T\mathbf{y}$$

determine the coefficients of the polynomial, and the vector **v** minimizes

$$\|\mathbf{y}-M\mathbf{v}\|$$

Conditions that guarantee the invertibility of $M^T M$ are discussed in the exercises (Exercise 7). If $M^T M$ is invertible, then the normal equations have a unique solution $\mathbf{v}=\mathbf{v}^*$, which is given by

$$\mathbf{v}^*=\left(M^T M\right)^{-1} M^T\mathbf{y} \tag{11}$$

EXAMPLE 3 Fitting a Quadratic Curve to Data

According to Newton's second law of motion, a body near the Earth's surface falls vertically downward according to the equation

$$s=s_0+v_0 t+\frac{1}{2}gt^2 \tag{12}$$

where

s = vertical displacement downward relative to some fixed point
s_0 = initial displacement at time $t=0$
v_0 = initial velocity at time $t=0$
g = acceleration of gravity at the Earth's surface

from Equation 12 by releasing a weight with unknown initial displacement and velocity and measuring the distance it has fallen at certain times relative to a fixed reference point. Suppose that a laboratory experiment is performed to evaluate g. Suppose it is found that at times $t=.1, .2, .3, .4$, and .5 seconds the weight has fallen $s=-0.18, 0.31, 1.03, 2.48$, and 3.73 feet, respectively, from the reference point. Find an approximate value of g using these data.

Solution The mathematical problem is to fit a quadratic curve

$$s=a_0+a_1 t+a_2 t^2 \tag{13}$$

to the five data points:

$$(.1, -0.18), \quad (.2, 0.31), \quad (.3, 1.03), \quad (.4, 2.48), \quad (.5, 3.73)$$

With the appropriate adjustments in notation, the matrices M and **y** in 10 are

$$M = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \\ 1 & t_4 & t_4^2 \\ 1 & t_5 & t_5^2 \end{bmatrix} = \begin{bmatrix} 1 & .1 & .01 \\ 1 & .2 & .04 \\ 1 & .3 & .09 \\ 1 & .4 & .16 \\ 1 & .5 & .25 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{bmatrix} = \begin{bmatrix} -0.18 \\ 0.31 \\ 1.03 \\ 2.48 \\ 3.73 \end{bmatrix}$$

Thus, from 11,

$$\mathbf{v}^* = \begin{bmatrix} a_0^* \\ a_1^* \\ a_2^* \end{bmatrix} = \left(M^T M\right)^{-1} M^T \mathbf{y} \doteq \begin{bmatrix} -0.40 \\ 0.35 \\ 16.1 \end{bmatrix}$$

From 12 and 13, we have $a_2 = \frac{1}{2}g$, so the estimated value of g is

$$g = 2a_2^* = 2(16.1) = 32.2 \text{ feet / second}^2$$

If desired, we can also estimate the initial displacement and initial velocity of the weight:

$$\begin{aligned} s_0 &= a_0^* = -0.40 \text{ feet} \\ v_0 &= a_1^* = 0.35 \text{ feet / second} \end{aligned}$$

In Figure 6.5.5 we have plotted the five data points and the approximating polynomial.

Figure 6.5.5

Concept Review

- Least squares straight line fit
- Regression line
- Least squares polynomial fit

Skills

- Find the least squares straight line fit to a set of data points.
- Find the least squares polynomial fit to a set of data points.
- Use the techniques of this section to solve applied problems.

Exercise Set 6.5

1. Find the least squares straight line fit to the three points $(0, 0)$, $(1, 2)$, and $(2, 7)$.

Answer:

$$y = -\frac{1}{2} + \frac{7}{2}x$$

2. Find the least squares straight line fit to the four points $(0, 1)$, $(2, 0)$, $(3, 1)$, and $(3, 2)$.

3. Find the quadratic polynomial that best fits the four points $(2, 0)$, $(3, -10)$, $(5, -48)$, and $(6, -76)$.

Answer:

$$y = 2 + 5x - 3x^2$$

4. Find the cubic polynomial that best fits the five points $(-1, -14)$, $(0, -5)$, $(1, -4)$, $(2, 1)$, and $(3, 22)$.

5. Show that the matrix M in Equation 2 has linearly independent columns if and only if at least two of the numbers $x_1, x_2, \ldots, x_n$ are distinct.

6. Show that the columns of the $n \times (m+1)$ matrix M in Equation 10 are linearly independent if $n > m$ and at least $m+1$ of the numbers $x_1, x_2, \ldots, x_n$ are distinct. [*Hint:* A nonzero polynomial of degree m has at most m distinct roots.]

7. Let M be the matrix in Equation 10. Using Exercise 6, show that a sufficient condition for the matrix $M^T M$ to be invertible is that $n > m$ and that at least $m+1$ of the numbers $x_1, x_2, \ldots, x_n$ are distinct.

8. The owner of a rapidly expanding business finds that for the first five months of the year the sales (in thousands) are \$4.0, \$4.4, \$5.2, \$6.4, and \$8.0. The owner plots these figures on a graph and conjectures that for the rest of the year, the sales curve can be approximated by a quadratic polynomial. Find the least squares quadratic polynomial fit to the sales curve, and use it to project the sales for the twelfth month of the year.

9. A corporation obtains the following data relating the number of sales representatives on its staff to annual sales:

Number of Sales Representatives	5	10	15	20	25	30
Annual Sales (millions)	3.4	4.3	5.2	6.1	7.2	8.3

Explain how you might use least squares methods to estimate the annual sales with 45 representatives, and discuss the assumptions that you are making. (You need not perform the actual computations.)

10. *Pathfinder* is an experimental, lightweight,remotely piloted,solar-powered aircraft that was used in aseries of experiments by NASA to determine the feasibilityof applyingsolar power for long-duration,high-altitude flight. In August 1997 *Pathfinder* recordedthe data in the accompanying table relating altitude H and temperature T. Show that a linear model is reasonable by plotting the data, and then find theleast squares line $H = H_0 + kT$ of best fit.

Table Ex-10

Altitude H (thousands of feet)	15	20	25	30	35	40	45
Temperature T (°C)	4.5	−5.9	−16.1	−27.6	−39.8	−50.2	−62.9

11. Find a curve of the form $y = a + (b / x)$ that best fits the data points $(1, 7)$, $(3, 3)$, $(6, 1)$ by making the substitution $X = 1 / x$. Draw the curve and plot the data points in the same coordinate system.

Answer:

$$y = \frac{5}{21} + \frac{48}{7x}$$

True-False Exercises

In parts (a)–(d) determine whether the statement is true or false, and justify your answer.

(a) Every set of data points has a unique least squares straight line fit.

Answer:

False

(b) If the data points (x_1, y_1), (x_2, y_2), …, (x_n, y_n) are not collinear, then 1 is an inconsistent system.

Answer:

True

(c) If $y = a + bx$ is the least squares line fit to the data points (x_1, y_1), (x_2, y_2), …, (x_n, y_n), then $d_i = |y_i - (a + bx_i)|$ is minimal for every $1 \le i \le n$.

Answer:

False

(d) If $y = a + bx$ is the least squares line fit to the data points (x_1, y_1), (x_2, y_2), ..., (x_n, y_n), then $\sum_{i=1}^{n} |y_i - (a + bx_i)|^2$ is minimal.

Answer:

True

Copyright © 2010 John Wiley & Sons, Inc. All rights reserved.

6.6 Function Approximation; Fourier Series

In this section we will show orthogonal projections can be used to approximate certain types of functions by simpler functions that are easier to work with. The ideas explained here have important applications in engineering and science. Calculus is required.

Best Approximations

All of the problems that we will study in this section will be special cases of the following general problem.

APPROXIMATION PROBLEM

Given a function *f* that is continuous on an interval $[a, b]$, find the "best possible approximation" to *f* using only functions from a specified subspace *W* of $C[a, b]$.

Here are some examples of such problems:

(a) Find the best possible approximation to e^x over $[0, 1]$ by a polynomial of the form $a_0 + a_1x + a_2x^2$.

(b) Find the best possible approximation to $\sin \pi x$ over $[-1, 1]$ by a function of the form $a_0 + a_1e^x + a_2e^{2x} + a_3e^{3x}$.

(c) Find the best possible approximation to *x* over $[0, 2\pi]$ by a function of the form $a_0 + a_1 \sin x + a_2 \sin 2x + b_1 \cos x + b_2 \cos 2x$.

In the first example *W* is the subspace of $C[0, 1]$ spanned by 1, x, and x^2; in the second example *W* is the subspace of $C[-1, 1]$ spanned by 1, e^x, e^{2x}, and e^{3x}; and in the third example *W* is the subspace of $C[0, 2\pi]$ spanned by 1, $\sin x$, $\sin 2x$, $\cos x$, and $\cos 2x$.

Measurements of Error

To solve approximation problems of the preceding types, we first need to make the phrase "best approximation over $[a, b]$" mathematically precise. To do this we will need some way of quantifying the error that results when one continuous function is approximated by another over an interval $[a, b]$. If we were to approximate $f(x)$ by $g(x)$, and if we were concerned only with the error in that approximation at a *single point* x_0, then it would be natural to define the error to be

$$\text{error} = |f(x_0) - g(x_0)|$$

sometimes called the ***deviation*** between *f* and *g* at x_0 (Figure 6.6.1). However, we are not concerned simply with measuring the error at a single point but rather with measuring it over the *entire* interval $[a, b]$. The problem is that an approximation may have small deviations in one part of the interval and large deviations in another. One possible way of accounting for this is to integrate the deviation $|f(x) - g(x)|$ over the interval $[a, b]$ and define the error over the interval to be

$$\text{error} = \int_a^b \left| f(x) - g(x) \right| dx \tag{1}$$

Geometrically, 1 is the area between the graphs of $f(x)$ and $g(x)$ over the interval $[a, b]$ (Figure 6.6.2); the greater the area, the greater the overall error.

Figure 6.6.1 The deviation between f and g x_0

Figure 6.6.2 The area between the graphs of **f** and **g** over $[a, b]$ measures the error in approximating f by g over $[a, b]$

Although 1 is natural and appealing geometrically, most mathematicians and scientists generally favor the following alternative measure of error, called the ***mean square error***:

$$\text{mean square error} = \int_a^b [f(x) - g(x)]^2 \, dx$$

Mean square error emphasizes the effect of larger errors because of the squaring and has the added advantage that it allows us to bring to bear the theory of inner product spaces. To see how, suppose that **f** is a continuous function on $[a, b]$ that we want to approximate by a function **g** from a subspace W of $C[a, b]$, and suppose that $C[a, b]$ is given the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b f(x) g(x) \, dx$$

It follows that

$$\|\mathbf{f} - \mathbf{g}\|^2 = \langle \mathbf{f} - \mathbf{g}, \mathbf{f} - \mathbf{g} \rangle = \int_a^b [f(x) - g(x)]^2 \, dx = \text{mean square error}$$

so minimizing the mean square error is the same as minimizing $\|\mathbf{f} - \mathbf{g}\|^2$. Thus the approximation problem posed informally at the beginning of this section can be restated more precisely as follows.

Least Squares Approximation

LEAST SQUARES APPROXIMATION PROBLEM

Let **f** be a function that is continuous on an interval $[a, b]$, let $C[a, b]$ have the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b f(x)g(x)\,dx$$

and let W be a finite-dimensional subspace of $C[a, b]$. Find a function **g** in W that minimizes

$$\|\mathbf{f} - \mathbf{g}\|^2 = \int_a^b [f(x) - g(x)]^2\,dx$$

Since $\|\mathbf{f} - \mathbf{g}\|^2$ and $\|\mathbf{f} - \mathbf{g}\|$ are minimized by the same function **g**, this problem is equivalent to looking for a function **g** in W that is closest to **f**. But we know from Theorem 6.4.1 that $\mathbf{g} = \text{proj}_W \mathbf{f}$ is such a function (Figure 6.6.3).

Figure 6.6.3

Thus, we have the following result.

THEOREM 6.6.1

If **f** is a continuous function on $[a, b]$, and W is a finite-dimensional subspace of $C[a, b]$, then the function **g** in W that minimizes the mean square error

$$\int_a^b [f(x) - g(x)]^2\,dx$$

is $\mathbf{g} = \text{proj}_W \mathbf{f}$, where the orthogonal projection is relative to the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b f(x)g(x)\,dx$$

The function $\mathbf{g} = \text{proj}_W \mathbf{f}$ is called the ***last squares approximation*** to **f** from W.

Fourier Series

A function of the form

$$T(x) = c_0 + c_1 \cos x + c_2 \cos 2x + \cdots + c_n \cos nx + d_1 \sin x + d_2 \sin 2x + \cdots + d_n \sin nx \tag{2}$$

is called a ***trigonometric polynomial***; if c_n and d_n are not both zero, then $T(x)$ is said to have ***order n***. For example,

$$T(x) = 2 + \cos x - 3 \cos 2x + 7 \sin 4x$$

is a trigonometric polynomial of order 4 with

$$c_0 = 2, c_1 = 1, c_2 = -3, c_3 = 0, c_4 = 0, \ d_1 = 0, d_2 = 0, d_3 = 0, d_4 = 7$$

It is evident from 2 that the trigonometric polynomials of order *n* or less are the various possible linear combinations of

$$1, \ \cos x, \ \cos 2x, \ldots, \ \cos nx, \ \sin x, \ \sin 2x, \ldots, \ \sin nx \tag{3}$$

It can be shown that these $2n + 1$ functions are linearly independent and thus form a basis for a $(2n + 1)$ -dimensional subspace of $C[a, b]$.

Let us now consider the problem of finding the least squares approximation of a continuous function $f(x)$ over the interval $[0, 2\pi]$ by a trigonometric polynomial of order *n* or less. As noted above, the least squares approximation to **f** from *W* is the orthogonal projection of **f** on *W*. To find this orthogonal projection, we must find an orthonormal basis $\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_{2n}$ for *W*, after which we can compute the orthogonal projection on *W* from the formula

$$\text{proj}_W \mathbf{f} = \langle \mathbf{f}, \mathbf{g}_0 \rangle \mathbf{g}_0 + \langle \mathbf{f}, \mathbf{g}_1 \rangle \mathbf{g}_1 + \cdots + \langle \mathbf{f}, \mathbf{g}_{2n} \rangle \mathbf{g}_{2n} \tag{4}$$

(see Theorem 6.3.4*b*). An orthonormal basis for *W* can be obtained by applying the Gram–Schmidt process to the basis vectors in 3 using the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_0^{2\pi} f(x) g(x) \, dx$$

This yields the orthonormal basis

$$\begin{gathered} \mathbf{g}_0 = \frac{1}{\sqrt{2\pi}}, \ \mathbf{g}_1 = \frac{1}{\sqrt{\pi}} \cos x, \ldots, \ \mathbf{g}_n = \frac{1}{\sqrt{\pi}} \cos nx, \\ \mathbf{g}_{n+1} = \frac{1}{\sqrt{\pi}} \sin x, \ldots, \ \mathbf{g}_{2n} = \frac{1}{\sqrt{\pi}} \sin nx \end{gathered} \tag{5}$$

(see Exercise 6). If we introduce the notation

$$a_0 = \frac{2}{\sqrt{2\pi}}\left\langle \mathbf{f}, \mathbf{g}_0 \right\rangle,\ a_1 = \frac{1}{\sqrt{\pi}}\left\langle \mathbf{f}, \mathbf{g}_1 \right\rangle, \ldots,\ a_n = \frac{1}{\sqrt{\pi}}\left\langle \mathbf{f}, \mathbf{g}_n \right\rangle$$
$$b_1 = \frac{1}{\sqrt{\pi}}\left\langle \mathbf{f}, \mathbf{g}_{n+1} \right\rangle, \ldots,\ b_n = \frac{1}{\sqrt{\pi}}\left\langle \mathbf{f}, \mathbf{g}_{2n} \right\rangle \tag{6}$$

then on substituting 5 in 4, we obtain

$$\text{proj}_W \mathbf{f} = \frac{a_0}{2} + [a_1 \cos x + \cdots + a_n \cos nx] + [b_1 \sin x + \cdots + b_n \sin nx] \tag{7}$$

where

$$a_0 = \frac{2}{\sqrt{2\pi}}\left\langle \mathbf{f}, \mathbf{g}_0 \right\rangle = \frac{2}{\sqrt{2\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{2\pi}}\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\,dx$$
$$a_1 = \frac{1}{\sqrt{\pi}}\left\langle \mathbf{f}, \mathbf{g}_1 \right\rangle = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{\pi}}\cos x\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos x\,dx$$
$$\vdots$$
$$a_n = \frac{1}{\sqrt{\pi}}\left\langle \mathbf{f}, \mathbf{g}_n \right\rangle = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{\pi}}\cos nx\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos nx\,dx$$
$$b_1 = \frac{1}{\sqrt{\pi}}\left\langle \mathbf{f}, \mathbf{g}_{n+1} \right\rangle = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{\pi}}\sin x\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin x\,dx$$
$$\vdots$$
$$b_n = \frac{1}{\sqrt{\pi}}\left\langle \mathbf{f}, \mathbf{g}_{2n} \right\rangle = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{\pi}}\sin nx\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin nx\,dx$$

In short,

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx\,dx,\ \ b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin kx\,dx \tag{8}$$

The numbers $a_0, a_1, \ldots, a_n, b_1, \ldots, b_n$ are called the ***Fourier coefficients*** of **f**.

EXAMPLE 1 Least Squares Approximations

Find the least squares approximation of $f(x) = x$ on $[0, 2\pi]$ by

(a) a trigonometric polynomial of order 2 or less;

(b) a trigonometric polynomial of order *n* or less.

Solution

(a)

$$a_0 = \frac{1}{\pi}\int_0^{2\pi} f(x)\,dx = \frac{1}{\pi}\int_0^{2\pi} x\,dx = 2\pi \tag{9a}$$

For $k = 1, 2, \ldots$, integration by parts yields (verify)

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx\,dx = \frac{1}{\pi}\int_0^{2\pi} x\cos kx\,dx = 0 \tag{9b}$$

$$b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin kx\,dx = \frac{1}{\pi}\int_0^{2\pi} x\sin kx\,dx = -\frac{2}{k} \tag{9c}$$

Thus, the least squares approximation to x on $[0, 2\pi]$ by a trigonometric polynomial of order 2 or less is

$$x \approx \frac{a_0}{2} + a_1\cos x + a_2\cos 2x + b_1\sin x + b_2\sin 2x$$

or, from (9a), (9b), and (9c),

$$x \approx \pi - 2\sin x - \sin 2x$$

(b) The least squares approximation to x on $[0, 2\pi]$ by a trigonometric polynomial of order n or less is

$$x \approx \frac{a_0}{2} + [a_1\cos x + \cdots + a_n\cos nx] + [b_1\sin x + \cdots + b_n\sin nx]$$

or, from (9a), (9b), and (9c),

$$x \approx \pi - 2\left(\sin x + \frac{\sin 2x}{2} + \frac{\sin 3x}{3} + \cdots + \frac{\sin nx}{n}\right)$$

The graphs of $y = x$ and some of these approximations are shown in Figure 6.6.4.

Figure 6.6.4

It is natural to expect that the mean square error will diminish as the number of terms in the least squares approximation

$$f(x) \approx \frac{a_0}{2} + \sum_{k=1}^{n}(a_k\cos kx + b_k\sin kx)$$

increases. It can be proved that for functions f in $C[0, 2\pi]$, the mean square error approaches zero as $n \to +\infty$; this is denoted by writing

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k\cos kx + b_k\sin kx)$$

The right side of this equation is called the ***Fourier series*** for f over the interval $[0, 2\pi]$. Such series are of major importance in engineering, science, and mathematics.

Jean Baptiste Fourier (1768–1830)

Historical Note Fourier was a French mathematician and physicist who discovered the Fourier series and related ideas while working on problems of heat diffusion. This discovery was one of the most influential in the history of mathematics; it is the cornerstone of many fields of mathematical research and a basic tool in many branches of engineering. Fourier, a political activist during the French revolution, spent time in jail for his defense of many victims during the Terror. He later became a favorite of Napoleon and was named a baron.
[*Image: The Granger Collection, New York*]

Concept Review

- Approximation of functions
- Mean square error
- Least squares approximation
- Trigonometric polynomial
- Fourier coefficients
- Fourier series

Skills

- Find the least squares approximation of a function.
- Find the mean square error of the least squares approximation of a function.
- Compute the Fourier series of a function.

Exercise Set 6.6

1. Find the least squares approximation of $f(x) = 1 + x$ over the interval $[0, 2\pi]$ by

(a) a trigonometric polynomial of order 2 or less.

(b) a trigonometric polynomial of order n or less.

Answer:

(a) $(1+\pi) - 2\sin x - \sin 2x$

(b) $(1+\pi) - 2\left[\sin x + \frac{\sin 2x}{2} + \frac{\sin 3x}{3} + \ldots + \frac{\sin nx}{n}\right]$

2. Find the least squares approximation of $f(x) = x^2$ over the interval $[0, 2\pi]$ by

(a) a trigonometric polynomial of order 3 or less.

(b) a trigonometric polynomial of order n or less.

3. (a) Find the least squares approximation of x over the interval $[0, 1]$ by a function of the form $a + be^x$.

(b) Find the mean square error of the approximation.

Answer:

(a) $-\frac{1}{2} + \frac{1}{e-1}e^x$

(b) $\frac{13}{12} + \frac{1+e}{2(1-e)}$

4. (a) Find the least squares approximation of e^x over the interval $[0, 1]$ by a polynomial of the form $a_0 + a_1 x$.

(b) Find the mean square error of the approximation.

5. (a) Find the least squares approximation of $\sin \pi x$ over the interval $[-1, 1]$ by a polynomial of the form $a_0 + a_1 x + a_2 x^2$.

(b) Find the mean square error of the approximation.

Answer:

(a) $\frac{3}{\pi}x$

(b) $1 - \frac{6}{\pi^2}$

6. Use the Gram–Schmidt process to obtain the orthonormal basis 5 from the basis 3.

7. Carry out the integrations indicated in Formulas 9a, 9b, and 9c.

8. Find the Fourier series of $f(x) = \pi - x$ over the interval $[0, 2\pi]$.

9. Find the Fourier series of $f(x) = 1,\ 0 < x < \pi$ and $f(x) = 0,\ \pi \le x \le 2\pi$ over the interval $[0, 2\pi]$.

Answer:

$$\frac{1}{2} + \sum_{k=1}^{\infty} \frac{1}{k\pi}\left[1 - (-1)^k\right] \sin kx$$

10. What is the Fourier series of $\sin(3x)$?

True-False Exercises

In parts (a)–(e) determine whether the statement is true or false, and justify your answer.

(a) If a function **f** in $C[a, b]$ is approximated by the function **g**, then the mean square error is the same as the area between the graphs of $f(x)$ and $g(x)$ over the interval $[a, b]$.

Answer:

False

(b) Given a finite-dimensional subspace W of $C[a, b]$, the function **g** = proj_W **f** minimizes the mean square error.

Answer:

True

(c) $\{1, \cos x, \sin x, \cos 2x, \sin 2x\}$ is an orthogonal subset of the vector space $C[0, 2\pi]$ with respect to the inner product $\langle \mathbf{f}, \mathbf{g} \rangle = \int_0^{2\pi} f(x)g(x)\,dx$.

Answer:

True

(d) $\{1, \cos x, \sin x, \cos 2x, \sin 2x\}$ is an orthonormal subset of the vector space $C[0, 2\pi]$ with respect to the inner product $\langle \mathbf{f}, \mathbf{g} \rangle = \int_0^{2\pi} f(x)g(x)\,dx$.

Answer:

False

(e) $\{1, \cos x, \sin x, \cos 2x, \sin 2x\}$ is a linearly independent subset of $C[0, 2\pi]$.

Answer:

True

Copyright © 2010 John Wiley & Sons, Inc. All rights reserved.

Chapter 6 Supplementary Exercises

1. Let R^4 have the Euclidean inner product.

(a) Find a vector in R^4 that is orthogonal to $\mathbf{u}_1 = (1, 0, 0, 0)$ and $\mathbf{u}_4 = (0, 0, 0, 1)$ and makes equal angles with $\mathbf{u}_2 = (0, 1, 0, 0)$ and $\mathbf{u}_3 = (0, 0, 1, 0)$.

(b) Find a vector $\mathbf{x} = (x_1, x_2, x_3, x_4)$ of length 1 that is orthogonal to $\mathbf{u}_1$ and $\mathbf{u}_4$ above and such that the cosine of the angle between $\mathbf{x}$ and $\mathbf{u}_2$ is twice the cosine of the angle between $\mathbf{x}$ and $\mathbf{u}_3$.

Answer:

(a) $(0, a, a, 0)$ with $a \neq 0$

(b) $\pm\left(0, \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}, 0\right)$

2. Prove: If $\langle \mathbf{u}, \mathbf{v} \rangle$ is the Euclidean inner product on R^n, and if A is an $n \times n$ matrix, then

$$\langle \mathbf{u}, A\mathbf{v} \rangle = \langle A^T\mathbf{u}, \mathbf{v} \rangle$$

[*Hint:* Use the fact that $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v} = \mathbf{v}^{\mathbf{T}}\mathbf{u}$.]

3. Let M_{22} have the inner product $\langle U, V \rangle = tr(U^T V) = tr(V^T U)$ that was defined in Example 6 of Section 6.1 . Describe the orthogonal complement of

(a) the subspace of all diagonal matrices.

(b) the subspace of symmetric matrices.

Answer:

(a) The subspace of all matrices in M_{22} with only zeros on the diagonal.

(b) The subspace of all skew-symmetric matrices in M_{22}.

4. Let $A\mathbf{x} = \mathbf{0}$ be a system of m equations in n unknowns. Show that

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

is a solution of this system if and only if the vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is orthogonal to every row vector of A with respect to the Euclidean inner product on R^n.

5. Use the Cauchy–Schwarz inequality to show that if $a_1, a_2, \ldots, a_n$ are positive real numbers, then

$$(a_1 + a_2 + \cdots + a_n)\left(\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n}\right) \geq n^2$$

6. Show that if $\mathbf{x}$ and $\mathbf{y}$ are vectors in an inner product space and c is any scalar, then

$$\|c\mathbf{x}+\mathbf{y}\|^2 = c^2\|\mathbf{x}\|^2 + 2c\langle \mathbf{x}, \mathbf{y}\rangle + \|\mathbf{y}\|^2$$

7. Let R^3 have the Euclidean inner product. Find two vectors of length 1 that are orthogonal to all three of the vectors $\mathbf{u}_1 = (1, 1, -1)$, $\mathbf{u}_2 = (-2, -1, 2)$, and $\mathbf{u}_3 = (-1, 0, 1)$.

Answer:

$$\pm\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$$

8. Find a weighted Euclidean inner product on R^n such that the vectors

$$\begin{aligned}
\mathbf{v}_1 &= (1, 0, 0, \ldots, 0) \\
\mathbf{v}_2 &= \left(0, \sqrt{2}, 0, \ldots, 0\right) \\
\mathbf{v}_3 &= \left(0, 0, \sqrt{3}, \ldots, 0\right) \\
&\vdots \\
\mathbf{v}_n &= \left(0, 0, 0, \ldots, \sqrt{n}\right)
\end{aligned}$$

form an orthonormal set.

9. Is there a weighted Euclidean inner product on R^2 for which the vectors $(1, 2)$ and $(3, -1)$ form an orthonormal set? Justify your answer.

Answer:

No

10. If **u** and **v** are vectors in an inner product space V, then **u**, **v**, and $\mathbf{u}-\mathbf{v}$ can be regarded as sides of a "triangle" in V (see the accompanying figure). Prove that the law of cosines holds for any such triangle; that is,

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

where θ is the angle between **u** and **v**.

Figure Ex-10

11. (a) As shown in Figure 3.2.6, the vectors $(k, 0, 0)$, $(0, k, 0)$, and $(0, 0, k)$ form the edges of a cube in R^3 with diagonal (k, k, k). Similarly, the vectors

$$(k, 0, 0, \ldots, 0),\ (0, k, 0, \ldots, 0), \ldots,\ (0, 0, 0, \ldots, k)$$

can be regarded as edges of a "cube" in R^n with diagonal $(k, k, k, \ldots, k)$. Show that each of the above edges makes an angle of θ with the diagonal, where $\cos\theta = 1/\sqrt{n}$.

(b) **Calculus required** What happens to the angle θ inpart (a) as the dimension of R^n approaches $+\infty$?

Answer:

(b) θ approaches $\frac{\pi}{2}$

12. Let **u** and **v** be vectors in an inner product space.

(a) Prove that $\|\mathbf{u}\| = \|\mathbf{v}\|$ if and only if $\mathbf{u} + \mathbf{v}$ and $\mathbf{u} - \mathbf{v}$ are orthogonal.

(b) Give a geometric interpretation of this result in R^2 with the Euclidean inner product.

13. Let **u** be a vector in an inner product space *V*, and let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be an orthonormal basis for *V*. Show that if α_i is the angle between **u** and $\mathbf{v}_i$, then

$$\cos^2\alpha_1 + \cos^2\alpha_2 + \cdots + \cos^2\alpha_n = 1$$

14. Prove: If $\langle \mathbf{u}, \mathbf{v}\rangle_1$ and $\langle \mathbf{u}, \mathbf{v}\rangle_2$ are two inner products on a vector space *V*, then the quantity $\langle \mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{v}\rangle_1 + \langle \mathbf{u}, \mathbf{v}\rangle_2$ is also an inner product.

15. Prove Theorem 6.2.5.

16. Prove: If *A* has linearly independent column vectors, and if **b** is orthogonal to the column space of *A*,then the least squares solution of $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = 0$.

17. Is there any value of *s* for which $x_1 = 1$ and $x_2 = 2$ is the leastsquares solution of the following linear system?

$$\begin{array}{rcrcl} x_1 & - & x_2 & = & 1 \\ 2x_1 & + & 3x_2 & = & 1 \\ 4x_1 & + & 5x_2 & = & s \end{array}$$

Explain your reasoning.

Answer:

No

18. Show that if *p* and *q* are distinct positive integers, then the functions $f(x) = \sin px$ and $g(x) = \sin qx$ are orthogonal with respect to the inner product

$$\langle \mathbf{f}, \mathbf{g}\rangle = \int_0^{2\pi} f(x)g(x)\,dx$$

19. Show that if *p* and *q* are positive integers, then the functions $f(x) = \cos px$ and $g(x) = \sin qx$ are orthogonal with respect to the inner product

$$\langle \mathbf{f}, \mathbf{g}\rangle = \int_0^{2\pi} f(x)g(x)\,dx$$

Copyright © 2010 John Wiley & Sons, Inc. All rights reserved.

CHAPTER

7 Diagonalization and Quadratic Forms

CHAPTER CONTENTS

7.1. Orthogonal Matrices
7.2. Orthogonal Diagonalization
7.3. Quadratic Forms
7.4. Optimization Using Quadratic Forms
7.5. Hermitian, Unitary, and Normal Matrices

INTRODUCTION

In Section 5.2 we found conditions that guaranteed the diagonalizability of an $n \times n$ matrix, but we did not consider what class or classes of matrices might actually satisfy those conditions. In this chapter we will show that every symmetric matrix is diagonalizable. This is an extremely important result because many applications utilize it in some essential way.

Copyright © 2010 John Wiley & Sons, Inc. All rights reserved.

7.1 Orthogonal Matrices

In this section we will discuss the class of matrices whose inverses can be obtained by transposition. Such matrices occur in a variety of applications and arise as well as transition matrices when one orthonormal basis is changed to another.

Orthogonal Matrices

We begin with the following definition.

DEFINITION 1

A square matrix A is said to be ***orthogonal*** if its transpose is the same as its inverse, that is, if

$$A^{-1} = A^T$$

or, equivalently, if

$$AA^T = A^T A = I \tag{1}$$

Recall from Theorem 1.6.3 that if either product in 1 holds, then so does the other. Thus, A is orthogonal if *either* $AA^T = I$ or $A^T A = I$.

EXAMPLE 1 A 3 × 3 Orthogonal Matrix

The matrix

$$A = \begin{bmatrix} \frac{3}{7} & \frac{2}{7} & \frac{6}{7} \\ -\frac{6}{7} & \frac{3}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{6}{7} & -\frac{3}{7} \end{bmatrix}$$

is orthogonal since

$$A^T A = \begin{bmatrix} \frac{3}{7} & -\frac{6}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{3}{7} & \frac{6}{7} \\ \frac{6}{7} & \frac{2}{7} & -\frac{3}{7} \end{bmatrix} \begin{bmatrix} \frac{3}{7} & \frac{2}{7} & \frac{6}{7} \\ -\frac{6}{7} & \frac{3}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{6}{7} & -\frac{3}{7} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

EXAMPLE 2 Rotation and Reflection Matrices are Orthogonal

Recall from Table Table 5 of Section 4.9 that the standard matrix for the counterclockwise rotation of R^2 through an angle θ is

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

This matrix is orthogonal for all choices of θ since

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

We leave it for you to verify that the reflection matrices in Tables Table 1 and Table 2 and the rotation matrices in Table Table 6 of Section 4.9 are all orthogonal.

Observe that for the orthogonal matrices in Example 1 and Example 2, both the row vectors and the column vectors form orthonormal sets with respect to the Euclidean inner product. This is a consequence of the following theorem.

THEOREM 7.1.1

The following are equivalent for an $n \times n$ matrix A.

(a) A is orthogonal

(b) The row vectors of A form an orthonormal set in R^n with the Euclidean inner product.

(c) The column vectors of A form an orthonormal set in R^n with the Euclidean inner product.

Proof We will prove the equivalence of (*a*) and (*b*) and leave the equivalence of (*a*) and (*c*) as an exercise.

(a) ⇔ (b) The entry in the *i*th row and *j*th column of the matrix product AA^T is the dot product of the *i*th row vector of A and the *j*th column vector of A^T (see Formula 5 of Section 1.3). But except for a difference in form, the *j*th column vector of A^T is the *j*th row vector of A. Thus, if the row vectors of A are $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n$, then the matrix product AA^T can be expressed as

$$AA^T = \begin{bmatrix} \mathbf{r}_1 \cdot \mathbf{r}_1 & \mathbf{r}_1 \cdot \mathbf{r}_2 & \ldots & \mathbf{r}_1 \cdot \mathbf{r}_n \\ \mathbf{r}_2 \cdot \mathbf{r}_1 & \mathbf{r}_2 \cdot \mathbf{r}_2 & \ldots & \mathbf{r}_2 \cdot \mathbf{r}_n \\ \vdots & \vdots & & \vdots \\ \mathbf{r}_n \cdot \mathbf{r}_1 & \mathbf{r}_n \cdot \mathbf{r}_2 & \ldots & \mathbf{r}_n \cdot \mathbf{r}_n \end{bmatrix}$$

[see Formula 28 of Section 3.2]. Thus, it follows that $AA^T = I$ if and only if

$$\mathbf{r}_1 \cdot \mathbf{r}_1 = \mathbf{r}_2 \cdot \mathbf{r}_2 = \ldots = \mathbf{r}_n \cdot \mathbf{r}_n = 1$$

and

$$\mathbf{r}_i \cdot \mathbf{r}_j = 0 \text{ when } i \neq j$$

which are true if and only if $\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n\}$ is an orthonormal set in R^n.

WARNING

Note that an orthogonal matrix is one with *orthonormal* rows and columns—not simply orthogonal rows and columns.

The following theorem lists three more fundamental properties of orthogonal matrices. The proofs are all straightforward and are left as exercises.

THEOREM 7.1.2

(a) The inverse of an orthogonal matrix is orthogonal.

(b) A product of orthogonal matrices is orthogonal.

(c) If A is orthogonal, then $\det(A) = 1$ or $\det(A) = -1$.

EXAMPLE 3 det(*A*) = ±1 for an Orthogonal Matrix *A*

The matrix

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

is orthogonal since its row (and column) vectors form orthonormal sets in R^2 with the Euclidean inner product. We leave it for you to verify that $\det(A) = 1$ and that interchanging the rows produces an orthogonal matrix whose determinant is -1.

Orthogonal Matrices as Linear Operators

We observed in Example 2 that the standard matrices for the basic reflection and rotation operators on R^2 and R^3 are orthogonal. The next theorem will explain why this is so.

THEOREM 7.1.3

If A is an $n \times n$ matrix, then the following are equivalent.

(a) A is orthogonal.

(b) $\|Ax\| = \|x\|$ for all x in R^n.

(c) $A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}$ and $\mathbf{y}$ in R^n.

Proof We will prove the sequence of implications $(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (a)$.

(a) ⇒ (b) Assume that A is orthogonal, so that $A^T A = I$. It follows from Formula 26 of Section 3.2 that

$$\|A\mathbf{x}\| = (A\mathbf{x} \cdot A\mathbf{x})^{1/2} = \left(\mathbf{x} \cdot A^T A\mathbf{x}\right)^{1/2} = (\mathbf{x} \cdot \mathbf{x})^{1/2} = \|\mathbf{x}\|$$

(b) ⇒ (c) Assume that $\|A\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$ in R^n. From Theorem 3.2.7 we have

$$\begin{aligned} A\mathbf{x} \cdot A\mathbf{y} &= \frac{1}{4}\|A\mathbf{x} + A\mathbf{y}\|^2 - \frac{1}{4}\|A\mathbf{x} - A\mathbf{y}\|^2 = \frac{1}{4}\|A\left(\mathbf{x} + \mathbf{y}\right)\|^2 - \frac{1}{4}\|A\left(\mathbf{x} - \mathbf{y}\right)\|^2 \\ &= \frac{1}{4}\|\mathbf{x} + \mathbf{y}\|^2 - \frac{1}{4}\|\mathbf{x} - \mathbf{y}\|^2 = \mathbf{x} \cdot \mathbf{y} \end{aligned}$$

(c) ⇒ (a) Assume that $A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}$ and $\mathbf{y}$ in R^n. It follows from Formula 26 of Section 3.2 that

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot A^T A\mathbf{y}$$

which can be rewritten as $\mathbf{x} \cdot \left(A^T A\mathbf{y} - \mathbf{y}\right) = 0$ or as

$$\mathbf{x} \cdot \left(A^T A - I\right)\mathbf{y} = 0$$

Since this equation holds for all $\mathbf{x}$ in R^n, it holds in particular if $\mathbf{x} = \left(A^T A - I\right)\mathbf{y}$, so

$$\left(A^T A - I\right)\mathbf{y} \cdot \left(A^T A - I\right)\mathbf{y} = 0$$

Thus, it follows from the positivity axiom for inner products that

$$\left(A^T A - I\right)\mathbf{y} = \mathbf{0}$$

Since this equation is satisfied by every vector $\mathbf{y}$ in R^n, it must be that $A^T A - I$ is the zero matrix (why?) and hence that $A^T A = I$. Thus, A is orthogonal.

Theorem 7.1.3 has a useful geometric interpretation when considered from the viewpoint of matrix transformations: If A is an orthogonal matrix and $T_A: R^n \to R^n$ is multiplication by A, then we will call T_A an ***orthogonal operator*** on R^n. It follows from parts (a) and (b) of Theorem 7.1.3 that the orthogonal operators on R^n are precisely those operators that leave the lengths of all vectors unchanged. This explains why, in Example 2, we found the standard matrices for the basic reflections and rotations of R^2 and R^3 to be orthogonal.

Parts (a) and (c) of Theorem 7.1.3 imply that orthogonal operators leave the *angle* between two vectors unchanged. Why?