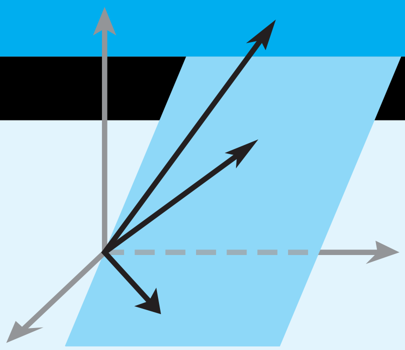


CHAPTER 6

Inner Product Spaces

CHAPTER CONTENTS

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.

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

A real vector space with an inner product is called a ***real inner 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 $\mathbf{u}$ and $\mathbf{v}$ in R^n to be

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n \tag{1}$$

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 $\mathbf{u}$ and $\mathbf{v}$ are vectors in Euclidean n-space, then norm and distance can be expressed in terms of the dot product as

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

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, whose proof is left for the exercises, 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* $\mathbf{v} = \mathbf{0}$.

(*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

$$\langle \mathbf{u}, \mathbf{v} \rangle = w_1 u_1 v_1 + w_2 u_2 v_2 + \cdots + w_n u_n v_n \tag{2}$$

defines an inner product on R^n that we call the ***weighted Euclidean inner product with weights*** $w_1, w_2, \ldots, w_n$.

Note that the standard Euclidean inner product in Formula (1) 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

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

satisfies the four inner product axioms.

Solution

Axiom 1: Interchanging $\mathbf{u}$ and $\mathbf{v}$ in Formula (3) does not change the sum on the right side, so $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$.

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

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

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

Axiom 3: $\langle k\mathbf{u}, \mathbf{v} \rangle = 3(ku_1)v_1 + 2(ku_2)v_2$

$$\begin{aligned}&= k(3u_1 v_1 + 2u_2 v_2) \\ &= k\langle \mathbf{u}, \mathbf{v} \rangle\end{aligned}$$

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

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 is 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 $\bar{x}$) is

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \cdots + f_n x_n}{f_1 + f_2 + \cdots + f_n} = \frac{1}{m}(f_1 x_1 + f_2 x_2 + \cdots + f_n x_n) \tag{4}$$

If we let

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

then (4) can be expressed as the weighted Euclidean inner product

$$\bar{x} = \langle \mathbf{f}, \mathbf{x} \rangle = w_1 f_1 x_1 + w_2 f_2 x_2 + \cdots + w_n f_n x_n$$

▶ EXAMPLE 2 Calculating with 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

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

we have

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

and

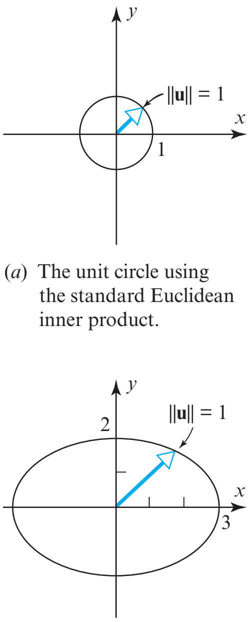
$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \langle (1, -1), (1, -1) \rangle^{1/2}$$
$$= [3(1)(1) + 2(-1)(-1)]^{1/2} = \sqrt{5}$$ ◀

Unit Circles and Spheres in Inner Product Spaces

DEFINITION 3 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.



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

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

▲ Figure 6.1.1

▶ 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 $\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{9}u_1v_1 + \frac{1}{4}u_2v_2$.

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*).

Solution (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*. Though this may seem odd when viewed geometrically, it makes sense algebraically since all points on the ellipse are 1 unit away from the origin relative to the given weighted Euclidean inner product. In short, weighting has the effect of distorting the space that we are used to seeing through "unweighted 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 $\mathbf{u}$ and $\mathbf{v}$ be vectors in R^n that are expressed in *column form*, and let A be an *invertible* $n \times n$ matrix. It can be shown (Exercise 47) that if $\mathbf{u} \cdot \mathbf{v}$ is the Euclidean inner product on R^n, then the formula

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

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 $\mathbf{u}$ and $\mathbf{v}$ are in column form, then $\mathbf{u} \cdot \mathbf{v}$ can be written as $\mathbf{v}^T\mathbf{u}$ from which it follows that (5) can be expressed as

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

or equivalently as

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{v}^T A^T A \mathbf{u} \tag{6}$$

EXAMPLE 4 Matrices Generating Weighted Euclidean Inner Products

The standard Euclidean and weighted Euclidean inner products are special cases 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 (5) yields

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

and the weighted Euclidean inner product

$$\langle \mathbf{u}, \mathbf{v} \rangle = w_1 u_1 v_1 + w_2 u_2 v_2 + \cdots + w_n u_n v_n \tag{7}$$

is generated by the matrix

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

This can be seen by observing that $A^T A$ is the $n \times n$ diagonal matrix whose diagonal entries are the weights $w_1, w_2, \ldots, w_n$.

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 The Standard Inner Product on M_{nn}

If $\mathbf{u} = U$ and $\mathbf{v} = V$ are matrices in the vector space M_{nn}, then the formula

$$\langle \mathbf{u}, \mathbf{v} \rangle = \operatorname{tr}(U^T V) \tag{8}$$

defines an inner product on M_{nn} called the ***standard inner product*** on that space (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 we can see 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

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

which is just the dot product of the corresponding entries in the two matrices. And it follows from this that

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle} = \sqrt{\text{tr}\langle U^T U \rangle} = \sqrt{u_1^2 + u_2^2 + u_3^2 + u_4^2}$$

For example, if

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

then

$$\langle \mathbf{u}, \mathbf{v} \rangle = \text{tr}(U^T V) = 1(-1) + 2(0) + 3(3) + 4(2) = 16$$

and

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle} = \sqrt{\text{tr}(U^T U)} = \sqrt{1^2 + 2^2 + 3^2 + 4^2} = \sqrt{30}$$

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{\text{tr}(V^T V)} = \sqrt{(-1)^2 + 0^2 + 3^2 + 2^2} = \sqrt{14}$$

▶ EXAMPLE 7 The Standard Inner Product on P_n

If

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

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

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

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

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

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

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

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

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

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

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

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

CALCULUS REQUIRED

▶ EXAMPLE 10 An Integral 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 $C[a, b]$:

Axiom 1: $\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b f(x)g(x)\,dx = \int_a^b g(x)f(x)\,dx = \langle \mathbf{g}, \mathbf{f} \rangle$

Axiom 2: $\langle \mathbf{f} + \mathbf{g}, \mathbf{h} \rangle = \int_a^b (f(x) + g(x))h(x)\,dx$

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

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

Axiom 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 in Formula (13) holds if and only if the function f is identically zero on $[a, b]$, that is, if and only if $\mathbf{f} = \mathbf{0}$; 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 that is different from both the standard inner product and the evaluation inner product.

WARNING 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) have different meanings.

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 of the 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

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

Exercise Set 6.1

1. Let R^2 have the weighted Euclidean inner product

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

and let $\mathbf{u} = (1, 1)$, $\mathbf{v} = (3, 2)$, $\mathbf{w} = (0, -1)$, and $k = 3$. Compute the stated quantities.

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

2. Follow the directions of Exercise 1 using the weighted Euclidean inner product

$$\langle \mathbf{u}, \mathbf{v}\rangle = \tfrac{1}{2}u_1v_1 + 5u_2v_2$$

▶ In Exercises 3–4, compute the quantities in parts (a)–(f) of Exercise 1 using the inner product on R^2 generated by A. ◀

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

▶ In Exercises 5–6, find a matrix that generates the stated weighted inner product on R^2. ◀

5. $\langle \mathbf{u}, \mathbf{v}\rangle = 2u_1v_1 + 3u_2v_2$ 6. $\langle \mathbf{u}, \mathbf{v}\rangle = \frac{1}{2}u_1v_1 + 5u_2v_2$

▶ In Exercises 7–8, use the inner product on R^2 generated by the matrix A to find $\langle \mathbf{u}, \mathbf{v}\rangle$ for the vectors $\mathbf{u} = (0, -3)$ and $\mathbf{v} = (6, 2)$. ◀

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

▶ In Exercises 9–10, compute the standard inner product on M_{22} of the given matrices. ◀

9. $U = \begin{bmatrix} 3 & -2 \\ 4 & 8 \end{bmatrix}$, $V = \begin{bmatrix} -1 & 3 \\ 1 & 1 \end{bmatrix}$

10. $U = \begin{bmatrix} 1 & 2 \\ -3 & 5 \end{bmatrix}$, $V = \begin{bmatrix} 4 & 6 \\ 0 & 8 \end{bmatrix}$

▶ In Exercises 11–12, find the standard inner product on P_2 of the given polynomials. ◀

11. $\mathbf{p} = -2 + x + 3x^2$, $\mathbf{q} = 4 - 7x^2$

12. $\mathbf{p} = -5 + 2x + x^2$, $\mathbf{q} = 3 + 2x - 4x^2$

▶ In Exercises 13–14, a weighted Euclidean inner product on R^2 is given for the vectors $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$. Find a matrix that generates it. ◀

13. $\langle \mathbf{u}, \mathbf{v}\rangle = 3u_1v_1 + 5u_2v_2$ 14. $\langle \mathbf{u}, \mathbf{v}\rangle = 4u_1v_1 + 6u_2v_2$

▶ In Exercises 15–16, a sequence of sample points is given. Use the evaluation inner product on P_3 at those sample points to find $\langle \mathbf{p}, \mathbf{q}\rangle$ for the polynomials

$$\mathbf{p} = x + x^3 \quad \text{and} \quad \mathbf{q} = 1 + x^2$$ ◀

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

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

▶ In Exercises 17–18, find $\|\mathbf{u}\|$ and $d(\mathbf{u}, \mathbf{v})$ relative to the weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v}\rangle = 2u_1v_1 + 3u_2v_2$ on R^2. ◀

17. $\mathbf{u} = (-3, 2)$ and $\mathbf{v} = (1, 7)$

18. $\mathbf{u} = (-1, 2)$ and $\mathbf{v} = (2, 5)$

▶ In Exercises 19–20, find $\|\mathbf{p}\|$ and $d(\mathbf{p}, \mathbf{q})$ relative to the standard inner product on P_2. ◀

19. $\mathbf{p} = -2 + x + 3x^2$, $\mathbf{q} = 4 - 7x^2$

20. $\mathbf{p} = -5 + 2x + x^2$, $\mathbf{q} = 3 + 2x - 4x^2$

▶ In Exercises 21–22, find $\|U\|$ and $d(U, V)$ relative to the standard inner product on M_{22}. ◀

21. $U = \begin{bmatrix} 3 & -2 \\ 4 & 8 \end{bmatrix}$, $V = \begin{bmatrix} -1 & 3 \\ 1 & 1 \end{bmatrix}$

22. $U = \begin{bmatrix} 1 & 2 \\ -3 & 5 \end{bmatrix}$, $V = \begin{bmatrix} 4 & 6 \\ 0 & 8 \end{bmatrix}$

▶ In Exercises 23–24, let

$$\mathbf{p} = x + x^3 \quad \text{and} \quad \mathbf{q} = 1 + x^2$$

Find $\|\mathbf{p}\|$ and $d(\mathbf{p}, \mathbf{q})$ relative to the evaluation inner product on P_3 at the stated sample points. ◀

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

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

▶ In Exercises 25–26, find $\|\mathbf{u}\|$ and $d(\mathbf{u}, \mathbf{v})$ for the vectors $\mathbf{u} = (-1, 2)$ and $\mathbf{v} = (2, 5)$ relative to the inner product on R^2 generated by the matrix A. ◀

25. $A = \begin{bmatrix} 4 & 0 \\ 3 & 5 \end{bmatrix}$ 26. $A = \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}$

▶ In Exercises 27–28, suppose that $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in an inner product space such that

$$\langle \mathbf{u}, \mathbf{v}\rangle = 2, \quad \langle \mathbf{v}, \mathbf{w}\rangle = -6, \quad \langle \mathbf{u}, \mathbf{w}\rangle = -3$$
$$\|\mathbf{u}\| = 1, \quad \|\mathbf{v}\| = 2, \quad \|\mathbf{w}\| = 7$$

Evaluate the given expression. ◀

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

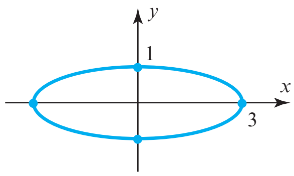
28. (a) $\langle \mathbf{u} - \mathbf{v} - 2\mathbf{w}, 4\mathbf{u} + \mathbf{v}\rangle$ (b) $\|2\mathbf{w} - \mathbf{v}\|$

▶ In Exercises 29–30, sketch the unit circle in R^2 using the given inner product. ◀

29. $\langle \mathbf{u}, \mathbf{v}\rangle = \frac{1}{4}u_1v_1 + \frac{1}{16}u_2v_2$ 30. $\langle \mathbf{u}, \mathbf{v}\rangle = 2u_1v_1 + u_2v_2$

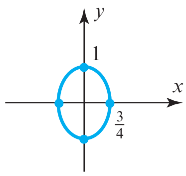
▶ In Exercises 31–32, find a weighted Euclidean inner product on R^2 for which the "unit circle" is the ellipse shown in the accompanying figure. ◀

31.



▲ Figure Ex-31

32.



▲ Figure Ex-31

▶ In Exercises 33–34, let $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$. Show that the expression does *not* define an inner product on R^3, and list all inner product axioms that fail to hold. ◀

33. $\langle \mathbf{u}, \mathbf{v} \rangle = u_1^2 v_1^2 + u_2^2 v_2^2 + u_3^2 v_3^2$

34. $\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 - u_2 v_2 + u_3 v_3$

▶ In Exercises 35–36, suppose that $\mathbf{u}$ and $\mathbf{v}$ are vectors in an inner product space. Rewrite the given expression in terms of $\langle \mathbf{u}, \mathbf{v} \rangle$, $\|\mathbf{u}\|^2$, and $\|\mathbf{v}\|^2$. ◀

35. $\langle 2\mathbf{v} - 4\mathbf{u}, \mathbf{u} - 3\mathbf{v} \rangle$ **36.** $\langle 5\mathbf{u} + 6\mathbf{v}, 4\mathbf{v} - 3\mathbf{u} \rangle$

37. ***(Calculus required)*** Let the vector space P_2 have the inner product

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

Find the following for $\mathbf{p} = 1$ and $\mathbf{q} = x^2$.

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

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

38. ***(Calculus required)*** Let the vector space P_3 have the inner product

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

Find the following for $\mathbf{p} = 2x^3$ and $\mathbf{q} = 1 - x^3$.

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

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

▶ ***(Calculus required)*** In Exercises 39–40, use the inner product

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

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

39. $\mathbf{f} = \cos 2\pi x$, $\mathbf{g} = \sin 2\pi x$ **40.** $\mathbf{f} = x$, $\mathbf{g} = e^x$

Working with Proofs

41. Prove parts (a) and (b) of Theorem 6.1.1.

42. Prove parts (c) and (d) of Theorem 6.1.1.

43. (a) Let $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$. Prove that $\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1v_1 + 5u_2v_2$ defines an inner product on R^2 by showing that the inner product axioms hold.

(b) What conditions must k_1 and k_2 satisfy for $\langle \mathbf{u}, \mathbf{v} \rangle = k_1u_1v_1 + k_2u_2v_2$ to define an inner product on R^2? Justify your answer.

44. Prove that the following identity holds for vectors in any inner product space.

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

45. Prove that the following identity holds for vectors in any inner product space.

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

46. 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 that in 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 $\langle \mathbf{u}, k\mathbf{v} \rangle = \overline{k}\langle \mathbf{u}, \mathbf{v} \rangle$.

47. Prove that Formula (5) defines an inner product on R^n.

48. (a) Prove that if $\mathbf{v}$ is a fixed vector in a real inner product space V, then the mapping $T: V \to R$ defined by $T(\mathbf{x}) = \langle \mathbf{x}, \mathbf{v} \rangle$ is a linear transformation.

(b) Let $V = R^3$ have the Euclidean inner product, and let $\mathbf{v} = (1, 0, 2)$. Compute $T(1, 1, 1)$.

(c) Let $V = P_2$ have the standard inner product, and let $\mathbf{v} = 1 + x$. Compute $T(x + x^2)$.

(d) Let $V = P_2$ have the evaluation inner product at the points $x_0 = 1$, $x_1 = 0$, $x_2 = -1$, and let $\mathbf{v} = 1 + x$. Compute $T(x + x^2)$.

True-False Exercises

TF. 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.

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

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

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

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

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

(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.

Working with Technology

T1. (a) Confirm that the following matrix generates an inner product.

$$A = \begin{bmatrix} 5 & 8 & 6 & -13 \\ 3 & -1 & 0 & -9 \\ 0 & 1 & -1 & 0 \\ 2 & 4 & 3 & -5 \end{bmatrix}$$

(b) For the following vectors, use the inner product in part (a) to compute $\langle \mathbf{u}, \mathbf{v} \rangle$, first by Formula (5) and then by Formula (6).

$$\mathbf{u} = \begin{bmatrix} 1 \\ -2 \\ 0 \\ 3 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 2 \end{bmatrix}$$

T2. Let the vector space P_4 have the evaluation inner product at the points

$$-2, \quad -1, \quad 0, \quad 1, \quad 2$$

and let

$$\mathbf{p} = p(x) = x + x^3 \quad \text{and} \quad \mathbf{q} = q(x) = 1 + x^2 + x^4$$

(a) Compute $\langle \mathbf{p}, \mathbf{q} \rangle$, $\|\mathbf{p}\|$, and $\|\mathbf{q}\|$.

(b) Verify that the identities in Exercises 44 and 45 hold for the vectors $\mathbf{p}$ and $\mathbf{q}$.

T3. Let the vector space M_{33} have the standard inner product and let

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

(a) Use Formula (8) to compute $\langle \mathbf{u}, \mathbf{v} \rangle$, $\|\mathbf{u}\|$, and $\|\mathbf{v}\|$.

(b) Verify that the identities in Exercises 44 and 45 hold for the vectors $\mathbf{u}$ and $\mathbf{v}$.

6.2 Angle and Orthogonality in Inner Product Spaces

In Section 3.2 we defined the notion of "angle" between vectors 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 $\mathbf{u}$ and $\mathbf{v}$ 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 the Cauchy–Schwarz inequality will enable us to define the angle between two vectors in *any* real inner product space.

THEOREM 6.2.1 Cauchy–Schwarz Inequality

If $\mathbf{u}$ *and* $\mathbf{v}$ *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 $\mathbf{u} \ne \mathbf{0}$. 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

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

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 $b^2-4ac \le 0$. Expressing the coefficients a, b, and c in terms of the vectors $\mathbf{u}$ and $\mathbf{v}$ gives $4\langle \mathbf{u},\mathbf{v}\rangle^2 - 4\langle \mathbf{u},\mathbf{u}\rangle\langle \mathbf{v},\mathbf{v}\rangle \le 0$ or, equivalently,

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

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

$$|\langle \mathbf{u},\mathbf{v}\rangle| \le \langle \mathbf{u},\mathbf{u}\rangle^{1/2}\langle \mathbf{v},\mathbf{v}\rangle^{1/2} \quad \text{or equivalently} \quad |\langle \mathbf{u},\mathbf{v}\rangle| \le \|\mathbf{u}\|\,\|\mathbf{v}\|$$

which completes the proof. ◀

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

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

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

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 a first step, we leave it as an exercise for you to use the Cauchy–Schwarz inequality to show that

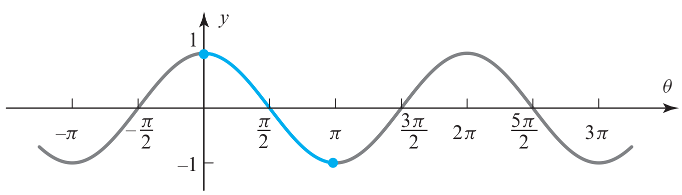
$$-1 \le \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|} \le 1 \tag{6}$$

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

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

(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 the Angle Between Vectors in M_{22}

Let M_{22} have the standard inner product. Find the cosine of the angle between the vectors

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

Solution We showed in Example 6 of the previous section that

$$\langle \mathbf{u}, \mathbf{v} \rangle = 16, \quad \|\mathbf{u}\| = \sqrt{30}, \quad \|\mathbf{v}\| = \sqrt{14}$$

from which it follows that

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|} = \frac{16}{\sqrt{30}\sqrt{14}} \approx 0.78 \blacktriangleleft$$

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 basic geometry theorems remained valid (see Theorems 3.2.5, 3.2.6, and 3.2.7). By making only minor adjustments to the proofs of those theorems, one 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*) $\|\mathbf{u} + \mathbf{v}\| \le \|\mathbf{u}\| + \|\mathbf{v}\|$ **[Triangle inequality for vectors]**

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

Proof (a)

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

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

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 $\mathbf{u}$ and $\mathbf{v}$ 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 a generalization of Definition 1 in Section 3.3 and is applicable even if one or both of the vectors is zero.

DEFINITION 1 Two vectors $\mathbf{u}$ and $\mathbf{v}$ in an inner product space V 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 Theorem 3.3.3 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 a real 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:

$$\begin{aligned} \|\mathbf{p}+\mathbf{q}\|^2 = \langle \mathbf{p}+\mathbf{q}, \mathbf{p}+\mathbf{q} \rangle &= \int_{-1}^{1} (x+x^2)(x+x^2)\,dx \\ &= \int_{-1}^{1} x^2\,dx + 2\int_{-1}^{1} x^3\,dx + \int_{-1}^{1} x^4\,dx = \frac{2}{3} + 0 + \frac{2}{5} = \frac{16}{15} \end{aligned}$$

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 a real 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 4.8.6 we stated three properties of orthogonal complements in R^n. The following theorem generalizes parts (a) and (b) of that theorem to general real inner product spaces.

THEOREM 6.2.4 *If W is a subspace of a real inner product space V, then:*

(a) *$W^\perp$ is a subspace of V.*

(b) $W \cap W^\perp = \{\mathbf{0}\}$.

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 $W^\perp$.

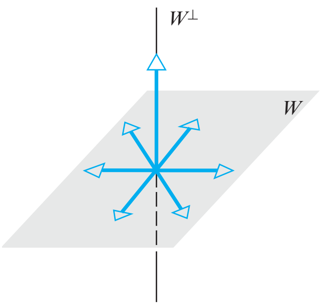
Proof (b) If $\mathbf{v}$ is any vector in both W and $W^{\perp}$, then $\mathbf{v}$ is orthogonal to itself; that is, $\langle \mathbf{v}, \mathbf{v} \rangle = 0$. It follows from the positivity axiom for inner products that $\mathbf{v} = \mathbf{0}$. ◀

The next theorem, which we state without proof, generalizes part (*c*) of Theorem 4.8.6. Note, however, that this theorem applies only to finite-dimensional inner product spaces, whereas Theorem 4.8.6 does not have this restriction.

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 real finite-dimensional inner product space V, then the orthogonal complement of $W^{\perp}$ is W; that is,*

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



▲ 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 4.8.7). 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

$$\mathbf{w}_1 = (1, 3, -2, 0, 2, 0), \quad \mathbf{w}_2 = (2, 6, -5, -2, 4, -3),$$
$$\mathbf{w}_3 = (0, 0, 5, 10, 0, 15), \quad \mathbf{w}_4 = (2, 6, 0, 8, 4, 18)$$

Find a basis for the orthogonal complement of W.

Solution The subspace 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. ◀

Exercise Set 6.2

▶ In Exercises 1–2, find the cosine of the angle between the vectors with respect to the Euclidean inner product. ◀

1. (a) $\mathbf{u} = (1, -3),\ \mathbf{v} = (2, 4)$

(b) $\mathbf{u} = (-1, 5, 2),\ \mathbf{v} = (2, 4, -9)$

(c) $\mathbf{u} = (1, 0, 1, 0),\ \mathbf{v} = (-3, -3, -3, -3)$

2. (a) $\mathbf{u} = (-1, 0),\ \mathbf{v} = (3, 8)$

(b) $\mathbf{u} = (4, 1, 8),\ \mathbf{v} = (1, 0, -3)$

(c) $\mathbf{u} = (2, 1, 7, -1),\ \mathbf{v} = (4, 0, 0, 0)$

▶ In Exercises 3–4, find the cosine of the angle between the vectors with respect to the standard inner product on P_2. ◀

3. $\mathbf{p} = -1 + 5x + 2x^2,\ \mathbf{q} = 2 + 4x - 9x^2$

4. $\mathbf{p} = x - x^2,\ \mathbf{q} = 7 + 3x + 3x^2$

▶ In Exercises 5–6, find the cosine of the angle between A and B with respect to the standard inner product on M_{22}. ◀

5. $A = \begin{bmatrix} 2 & 6 \\ 1 & -3 \end{bmatrix},\ B = \begin{bmatrix} 3 & 2 \\ 1 & 0 \end{bmatrix}$

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

▶ In Exercises 7–8, determine whether the vectors are orthogonal with respect to the Euclidean inner product. ◀

7. (a) $\mathbf{u} = (-1, 3, 2),\ \mathbf{v} = (4, 2, -1)$

(b) $\mathbf{u} = (-2, -2, -2),\ \mathbf{v} = (1, 1, 1)$

(c) $\mathbf{u} = (a, b),\ \mathbf{v} = (-b, a)$

8. (a) $\mathbf{u} = (u_1, u_2, u_3),\ \mathbf{v} = (0, 0, 0)$

(b) $\mathbf{u} = (-4, 6, -10, 1),\ \mathbf{v} = (2, 1, -2, 9)$

(c) $\mathbf{u} = (a, b, c),\ \mathbf{v} = (-c, 0, a)$

▶ In Exercises 9–10, show that the vectors are orthogonal with respect to the standard inner product on P_2. ◀

9. $\mathbf{p} = -1 - x + 2x^2,\ \mathbf{q} = 2x + x^2$

10. $\mathbf{p} = 2 - 3x + x^2,\ \mathbf{q} = 4 + 2x - 2x^2$

▶ In Exercises 11–12, show that the matrices are orthogonal with respect to the standard inner product on M_{22}. ◀

11. $U = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix},\ V = \begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix}$

12. $U = \begin{bmatrix} 5 & -1 \\ 2 & -2 \end{bmatrix},\ V = \begin{bmatrix} 1 & 3 \\ -1 & 0 \end{bmatrix}$

▶ In Exercises 13–14, show that the vectors are not orthogonal with respect to the Euclidean inner product on R^2, and then find a value of k for which the vectors are orthogonal with respect to the weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 2u_1v_1 + ku_2v_2$. ◀

13. $\mathbf{u} = (1, 3),\ \mathbf{v} = (2, -1)$

14. $\mathbf{u} = (2, -4),\ \mathbf{v} = (0, 3)$

15. If the vectors $\mathbf{u} = (1, 2)$ and $\mathbf{v} = (2, -4)$ are orthogonal with respect to the weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = w_1u_1v_1 + w_2u_2v_2$, what must be true of the weights w_1 and w_2?

16. 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 $\mathbf{w} = (3, 2, 5, 4)$.

17. Do there exist scalars k and l such that the vectors

$$\mathbf{p}_1 = 2 + kx + 6x^2, \quad \mathbf{p}_2 = l + 5x + 3x^2, \quad \mathbf{p}_3 = 1 + 2x + 3x^2$$

are mutually orthogonal with respect to the standard inner product on P_2?

18. Show that the vectors

$$\mathbf{u} = \begin{bmatrix} 3 \\ 3 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 5 \\ -8 \end{bmatrix}$$

are orthogonal with respect to the inner product on R^2 that is generated by the matrix

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

[See Formulas (5) and (6) of Section 6.1.]

19. Let P_2 have the evaluation inner product at the points

$$x_0 = -2, \quad x_1 = 0, \quad x_2 = 2$$

Show that the vectors $\mathbf{p} = x$ and $\mathbf{q} = x^2$ are orthogonal with respect to this inner product.

20. Let M_{22} have the standard inner product. Determine whether the matrix A is in the subspace spanned by the matrices U and V.

$$A = \begin{bmatrix} -1 & 1 \\ 0 & 2 \end{bmatrix}, \quad U = \begin{bmatrix} 1 & -1 \\ 3 & 0 \end{bmatrix}, \quad V = \begin{bmatrix} 4 & 0 \\ 9 & 2 \end{bmatrix}$$

▶ In Exercises 21–24, confirm that the Cauchy–Schwarz inequality holds for the given vectors using the stated inner product. ◀

21. $\mathbf{u} = (1, 0, 3)$, $\mathbf{v} = (2, 1, -1)$ using the weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 2u_1v_1 + 3u_2v_2 + u_3v_3$ in R^3.

22. $U = \begin{bmatrix} -1 & 2 \\ 6 & 1 \end{bmatrix}$ and $V = \begin{bmatrix} 1 & 0 \\ 3 & 3 \end{bmatrix}$

using the standard inner product on M_{22}.

23. $\mathbf{p} = -1 + 2x + x^2$ and $\mathbf{q} = 2 - 4x^2$ using the standard inner product on P_2.

24. The vectors

$$\mathbf{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

with respect to the inner product in Exercise 18.

25. 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 $\mathbf{w}_1 = (1, -1, 3, 0)$ and $\mathbf{w}_2 = (4, 0, 9, 2)$.

26. Let P_3 have the standard inner product, and let

$$\mathbf{p} = -1 - x + 2x^2 + 4x^3$$

Determine whether $\mathbf{p}$ is orthogonal to the subspace spanned by the polynomials $\mathbf{w}_1 = 2 - x^2 + x^3$ and $\mathbf{w}_2 = 4x - 2x^2 + 2x^3$.

▶ In Exercises 27–28, find a basis for the orthogonal complement of the subspace of R^n spanned by the vectors. ◀

27. $\mathbf{v}_1 = (1, 4, 5, 2)$, $\mathbf{v}_2 = (2, 1, 3, 0)$, $\mathbf{v}_3 = (-1, 3, 2, 2)$

28. $\mathbf{v}_1 = (1, 4, 5, 6, 9)$, $\mathbf{v}_2 = (3, -2, 1, 4, -1)$, $\mathbf{v}_3 = (-1, 0, -1, -2, -1)$, $\mathbf{v}_4 = (2, 3, 5, 7, 8)$

▶ In Exercises 29–30, assume that R^n has the Euclidean inner product. ◀

29. (a) Let W be the line in R^2 with equation $y = 2x$. Find an equation for $W^\perp$.

(b) Let W be the plane in R^3 with equation $x - 2y - 3z = 0$. Find parametric equations for $W^\perp$.

30. (a) Let W be the y-axis in an xyz-coordinate system in R^3. Describe the subspace $W^\perp$.

(b) Let W be the yz-plane of an xyz-coordinate system in R^3. Describe the subspace $W^\perp$.

31. (***Calculus required***) Let $C[0, 1]$ have the integral inner product

$$\langle \mathbf{p}, \mathbf{q} \rangle = \int_0^1 p(x)q(x)\,dx$$

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

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

(b) Find $\|\mathbf{p}\|$ and $\|\mathbf{q}\|$.

32. (a) Find the cosine of the angle between the vectors $\mathbf{p}$ and $\mathbf{q}$ in Exercise 31.

(b) Find the distance between the vectors $\mathbf{p}$ and $\mathbf{q}$ in Exercise 31.

33. (***Calculus required***) Let $C[-1, 1]$ have the integral inner product

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

and let $\mathbf{p} = p(x) = x^2 - x$ and $\mathbf{q} = q(x) = x + 1$.

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

(b) Find $\|\mathbf{p}\|$ and $\|\mathbf{q}\|$.

34. (a) Find the cosine of the angle between the vectors $\mathbf{p}$ and $\mathbf{q}$ in Exercise 33.

(b) Find the distance between the vectors $\mathbf{p}$ and $\mathbf{q}$ in Exercise 33.

35. (***Calculus required***) Let $C[0, 1]$ have the inner product in Exercise 31.

(a) Show that the vectors

$$\mathbf{p} = p(x) = 1 \quad \text{and} \quad \mathbf{q} = q(x) = \tfrac{1}{2} - x$$

are orthogonal.

(b) Show that the vectors in part (a) satisfy the Theorem of Pythagoras.

36. (***Calculus required***) Let $C[-1, 1]$ have the inner product in Exercise 33.

(a) Show that the vectors

$$\mathbf{p} = p(x) = x \quad \text{and} \quad \mathbf{q} = q(x) = x^2 - 1$$

are orthogonal.

(b) Show that the vectors in part (a) satisfy the Theorem of Pythagoras.

37. Let V be an inner product space. Show that if $\mathbf{u}$ and $\mathbf{v}$ are orthogonal unit vectors in V, then $\|\mathbf{u} - \mathbf{v}\| = \sqrt{2}$.

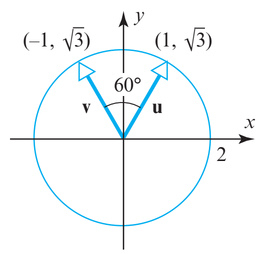
38. Let V be an inner product space. Show that if $\mathbf{w}$ is orthogonal to both $\mathbf{u}_1$ and $\mathbf{u}_2$, then it is orthogonal to $k_1\mathbf{u}_1 + k_2\mathbf{u}_2$ for all scalars k_1 and k_2. Interpret this result geometrically in the case where V is R^3 with the Euclidean inner product.

39. (***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.

40. As illustrated in the accompanying figure, the vectors $\mathbf{u} = (1, \sqrt{3})$ and $\mathbf{v} = (-1, \sqrt{3})$ 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 $\mathbf{u}$ and $\mathbf{v}$ are orthogonal unit vectors.



◀ Figure Ex-40

Working with Proofs

41. Let V be an inner product space. Prove 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$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$.

42. Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ be a basis for an inner product space V. Prove that the zero vector is the only vector in V that is orthogonal to all of the basis vectors.

43. Let $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$ be a basis for a subspace W of V. Prove that $W^\perp$ consists of all vectors in V that are orthogonal to every basis vector.

44. 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

$$\|\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_r\|^2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 + \cdots + \|\mathbf{v}_r\|^2$$

45. Prove: If $\mathbf{u}$ and $\mathbf{v}$ are $n \times 1$ matrices and A is an $n \times n$ matrix, then

$$(\mathbf{v}^T A^T A\mathbf{u})^2 \le (\mathbf{u}^T A^T A\mathbf{u})(\mathbf{v}^T A^T A\mathbf{v})$$

46. Use the Cauchy–Schwarz inequality to prove that for all real values of a, b, and θ,

$$(a\cos\theta + b\sin\theta)^2 \le a^2 + b^2$$

47. 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

$$|w_1u_1v_1 + w_2u_2v_2 + \cdots + w_nu_nv_n|$$
$$\le (w_1u_1^2 + w_2u_2^2 + \cdots + w_nu_n^2)^{1/2}(w_1v_1^2 + w_2v_2^2 + \cdots + w_nv_n^2)^{1/2}$$

48. Prove that equality holds in the Cauchy–Schwarz inequality if and only if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent.

49. ***(Calculus required)*** Let $f(x)$ and $g(x)$ be continuous functions on [0, 1]. Prove:

(a) $$\left[\int_0^1 f(x)g(x)\,dx\right]^2 \le \left[\int_0^1 f^2(x)\,dx\right]\left[\int_0^1 g^2(x)\,dx\right]$$

(b) $$\left[\int_0^1 [f(x)+g(x)]^2\,dx\right]^{1/2} \le \left[\int_0^1 f^2(x)\,dx\right]^{1/2} + \left[\int_0^1 g^2(x)\,dx\right]^{1/2}$$

[*Hint:* Use the Cauchy–Schwarz inequality.]

50. Prove that Formula (4) holds for all nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ in a real inner product space V.

51. Let $T_A\colon R^2 \to R^2$ be multiplication by

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

and let $\mathbf{x} = (1, 1)$.

(a) Assuming that R^2 has the Euclidean inner product, find all vectors $\mathbf{v}$ in R^2 such that $\langle \mathbf{x}, \mathbf{v}\rangle = \langle T_A(\mathbf{x}), T_A(\mathbf{v})\rangle$.

(b) Assuming that R^2 has the weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v}\rangle = 2u_1v_1 + 3u_2v_2$, find all vectors $\mathbf{v}$ in R^2 such that $\langle \mathbf{x}, \mathbf{v}\rangle = \langle T_A(\mathbf{x}), T_A(\mathbf{v})\rangle$.

52. Let $T\colon P_2 \to P_2$ be the linear transformation defined by

$$T(a + bx + cx^2) = 3a - cx^2$$

and let $\mathbf{p} = 1 + x$.

(a) Assuming that P_2 has the standard inner product, find all vectors $\mathbf{q}$ in P_2 such that $\langle \mathbf{p}, \mathbf{q}\rangle = \langle T(\mathbf{p}), T(\mathbf{q})\rangle$.

(b) Assuming that P_2 has the evaluation inner product at the points $x_0 = -1$, $x_1 = 0$, $x_2 = 1$, find all vectors $\mathbf{q}$ in P_2 such that $\langle \mathbf{p}, \mathbf{q}\rangle = \langle T(\mathbf{p}), T(\mathbf{q})\rangle$.

True-False Exercises

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

(a) If $\mathbf{u}$ is orthogonal to every vector of a subspace W, then $\mathbf{u} = \mathbf{0}$.

(b) If $\mathbf{u}$ is a vector in both W and $W^\perp$, then $\mathbf{u} = \mathbf{0}$.

(c) If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $W^\perp$, then $\mathbf{u} + \mathbf{v}$ is in $W^\perp$.

(d) If $\mathbf{u}$ is a vector in $W^\perp$ and k is a real number, then $k\mathbf{u}$ is in $W^\perp$.

(e) If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, then $|\langle \mathbf{u}, \mathbf{v}\rangle| = \|\mathbf{u}\|\|\mathbf{v}\|$.

(f) If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, then $\|\mathbf{u} + \mathbf{v}\| = \|\mathbf{u}\| + \|\mathbf{v}\|$.

Working with Technology

T1. (a) We know that the row space and null space of a matrix are orthogonal complements relative to the Euclidean inner product. Confirm this fact for the matrix

$$A = \begin{bmatrix} 2 & -1 & 3 & 5 \\ 4 & -3 & 1 & 3 \\ 3 & -2 & 3 & 4 \\ 4 & -1 & 15 & 17 \\ 7 & -6 & -7 & 0 \end{bmatrix}$$

(b) Find a basis for the orthogonal complement of the column space of A.

T2. In each part, confirm that the vectors $\mathbf{u}$ and $\mathbf{v}$ satisfy the Cauchy–Schwarz inequality relative to the stated inner product.

(a) M_{44} with the standard inner product.

$$\mathbf{u} = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & -1 & 0 & 1 \\ 3 & 0 & 0 & 2 \\ 0 & 4 & -3 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 2 & 2 & 1 & 3 \\ 3 & -1 & 0 & 1 \\ 1 & 0 & 0 & -2 \\ -3 & 1 & 2 & 0 \end{bmatrix}$$

(b) R^4 with the weighted Euclidean inner product with weights $w_1 = \frac{1}{2}$, $w_2 = \frac{1}{4}$, $w_3 = \frac{1}{8}$, $w_4 = \frac{1}{8}$.

$$\mathbf{u} = (1, -2, 2, 1) \quad \text{and} \quad \mathbf{v} = (0, -3, 3, -2)$$

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 can often be 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 ***orthonormal***.

▶ EXAMPLE 1 An Orthogonal Set in R^3

Let

$$\mathbf{v}_1 = (0, 1, 0), \quad \mathbf{v}_2 = (1, 0, 1), \quad \mathbf{v}_3 = (1, 0, -1)$$

and assume that R^3 has the Euclidean inner product. It follows that the set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is orthogonal since $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, \mathbf{v}_3 \rangle = \langle \mathbf{v}_2, \mathbf{v}_3 \rangle = 0$. ◀

It frequently happens that one has found a set of orthogonal vectors in an inner product space but what is actually needed is a set of *orthonormal* vectors. A simple way to convert an orthogonal set of nonzero vectors into an orthonormal set is to multiply each vector **v** in the orthogonal set by the reciprocal of its length to create a vector of norm 1 (called a ***unit vector***). To see why this works, suppose that **v** is a nonzero vector in an inner product space, and let

$$\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v} \tag{1}$$

Note that Formula (1) is identical to Formula (4) of Section 3.2, but whereas Formula (4) was valid only for vectors in R^n with the Euclidean inner product, Formula (1) is valid in general inner product spaces.

Then it follows from Theorem 6.1.1(*b*) with $k = \|\mathbf{v}\|$ that

$$\|\mathbf{u}\| = \left\| \frac{1}{\|\mathbf{v}\|}\mathbf{v} \right\| = \left| \frac{1}{\|\mathbf{v}\|} \right| \|\mathbf{v}\| = \frac{1}{\|\mathbf{v}\|}\|\mathbf{v}\| = 1$$

This process of multiplying a vector **v** by the reciprocal of its length is called ***normalizing*** **v**. We leave it as an exercise to show that normalizing the vectors in an orthogonal set of nonzero vectors preserves the orthogonality of the vectors and produces an orthonormal set.

▶ EXAMPLE 2 Constructing an Orthonormal Set

The Euclidean norms of the vectors in Example 1 are

$$\|\mathbf{v}_1\| = 1, \quad \|\mathbf{v}_2\| = \sqrt{2}, \quad \|\mathbf{v}_3\| = \sqrt{2}$$

Consequently, normalizing $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ yields

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = (0, 1, 0), \quad \mathbf{u}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right),$$

$$\mathbf{u}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = \left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}\right)$$

We leave it for you to verify that the set $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is orthonormal by showing that

$$\langle \mathbf{u}_1, \mathbf{u}_2 \rangle = \langle \mathbf{u}_1, \mathbf{u}_3 \rangle = \langle \mathbf{u}_2, \mathbf{u}_3 \rangle = 0 \quad \text{and} \quad \|\mathbf{u}_1\| = \|\mathbf{u}_2\| = \|\mathbf{u}_3\| = 1 \blacktriangleleft$$

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

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n = \mathbf{0} \tag{2}$$

To demonstrate that $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is linearly independent, we must prove that $k_1 = k_2 = \cdots = k_n = 0$.

For each $\mathbf{v}_i$ in S, it follows from (2) 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 (2) 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 for P_n

Recall from Example 7 of Section 6.1 that the standard inner product of the polynomials

$$\mathbf{p} = a_0 + a_1x + \cdots + a_nx^n \quad \text{and} \quad \mathbf{q} = b_0 + b_1x + \cdots + b_nx^n$$

is

$$\langle \mathbf{p}, \mathbf{q} \rangle = a_0b_0 + a_1b_1 + \cdots + a_nb_n$$

and the norm of $\mathbf{p}$ relative to this inner product is

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

You should be able to see from these formulas that the standard basis

$$S = \{1, x, x^2, \ldots, x^n\}$$

is orthonormal with respect to this inner product.

EXAMPLE 4 An Orthonormal Basis

In Example 2 we showed that the vectors

$$\mathbf{u}_1 = (0, 1, 0), \quad \mathbf{u}_2 = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right), \quad \text{and} \quad \mathbf{u}_3 = \left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}\right)$$

form an orthonormal set with respect to the Euclidean inner product on R^3. By Theorem 6.3.1, these vectors form a linearly independent set, and since R^3 is three-dimensional, it follows from Theorem 4.5.4 that $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal basis for R^3. ◀

Coordinates Relative to Orthonormal Bases

One way to express a vector $\mathbf{u}$ 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 $\mathbf{u}$ is any vector in V, then*

$$\mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{u}, \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{v}_n\rangle}{\|\mathbf{v}_n\|^2}\mathbf{v}_n \tag{3}$$

(*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 $\mathbf{u}$ is any vector in V, then*

$$\mathbf{u} = \langle \mathbf{u}, \mathbf{v}_1\rangle\mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2\rangle\mathbf{v}_2 + \cdots + \langle \mathbf{u}, \mathbf{v}_n\rangle\mathbf{v}_n \tag{4}$$

Proof (a) Since $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for V, every vector $\mathbf{u}$ 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

$$c_i = \frac{\langle \mathbf{u}, \mathbf{v}_i\rangle}{\|\mathbf{v}_i\|^2} \tag{5}$$

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 (5), which completes the proof.

Proof (b) In this case, $\|\mathbf{v}_1\| = \|\mathbf{v}_2\| = \cdots = \|\mathbf{v}_n\| = 1$, so Formula (3) simplifies to Formula (4). ◀

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 $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is

$$(\mathbf{u})_S = \left(\frac{\langle \mathbf{u}, \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}, \frac{\langle \mathbf{u}, \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}, \dots, \frac{\langle \mathbf{u}, \mathbf{v}_n\rangle}{\|\mathbf{v}_n\|^2}\right) \tag{6}$$

and relative to an orthonormal basis $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is

$$(\mathbf{u})_S = (\langle \mathbf{u}, \mathbf{v}_1\rangle, \langle \mathbf{u}, \mathbf{v}_2\rangle, \dots, \langle \mathbf{u}, \mathbf{v}_n\rangle) \tag{7}$$

▶ EXAMPLE 5 A Coordinate Vector Relative to an Orthonormal Basis

Let

$$\mathbf{v}_1 = (0, 1, 0), \quad \mathbf{v}_2 = \left(-\tfrac{4}{5}, 0, \tfrac{3}{5}\right), \quad \mathbf{v}_3 = \left(\tfrac{3}{5}, 0, \tfrac{4}{5}\right)$$

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

$$\langle \mathbf{u}, \mathbf{v}_1\rangle = 1, \quad \langle \mathbf{u}, \mathbf{v}_2\rangle = -\tfrac{1}{5}, \quad \text{and} \quad \langle \mathbf{u}, \mathbf{v}_3\rangle = \tfrac{7}{5}$$

Therefore, by Theorem 6.3.2 we have

$$\mathbf{u} = \mathbf{v}_1 - \tfrac{1}{5}\mathbf{v}_2 + \tfrac{7}{5}\mathbf{v}_3$$

that is,

$$(1, 1, 1) = (0, 1, 0) - \tfrac{1}{5}\left(-\tfrac{4}{5}, 0, \tfrac{3}{5}\right) + \tfrac{7}{5}\left(\tfrac{3}{5}, 0, \tfrac{4}{5}\right)$$

Thus, the coordinate vector of $\mathbf{u}$ relative to S is

$$(\mathbf{u})_S = (\langle \mathbf{u}, \mathbf{v}_1\rangle, \langle \mathbf{u}, \mathbf{v}_2\rangle, \langle \mathbf{u}, \mathbf{v}_3\rangle) = \left(1, -\tfrac{1}{5}, \tfrac{7}{5}\right)$$

▶ EXAMPLE 6 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

$$\mathbf{v}_1 = \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|} = (0, 1, 0), \quad \mathbf{v}_2 = \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|} = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right),$$

$$\mathbf{v}_3 = \frac{\mathbf{w}_3}{\|\mathbf{w}_3\|} = \left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$$

Solution (b) It follows from Formula (4) 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

$$\langle \mathbf{u}, \mathbf{v}_1 \rangle = (1, 2, 4) \cdot (0, 1, 0) = 2$$
$$\langle \mathbf{u}, \mathbf{v}_2 \rangle = (1, 2, 4) \cdot \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}} \right) = \frac{5}{\sqrt{2}}$$
$$\langle \mathbf{u}, \mathbf{v}_3 \rangle = (1, 2, 4) \cdot \left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}} \right) = \frac{3}{\sqrt{2}}$$

and hence that

$$(1, 2, 4) = 2(0, 1, 0) + \frac{5}{\sqrt{2}} \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}} \right) + \frac{3}{\sqrt{2}} \left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}} \right) \blacktriangleleft$$

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 *Projection Theorem* (see Theorem 3.3.2) that dealt with the problem of decomposing a vector $\mathbf{u}$ 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 $\mathbf{u}$ on some nonzero vector $\mathbf{a}$ 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, which we will state without proof.

THEOREM 6.3.3 Projection Theorem

If W is a finite-dimensional subspace of an inner product space V, then every vector $\mathbf{u}$ in V can be expressed in exactly one way as

$$\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2 \tag{8}$$

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 (8) are commonly denoted by

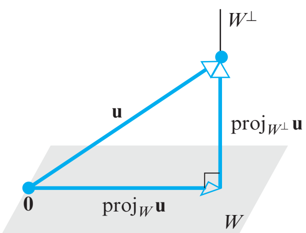
$$\mathbf{w}_1 = \text{proj}_W \mathbf{u} \quad \text{and} \quad \mathbf{w}_2 = \text{proj}_{W^\perp} \mathbf{u} \tag{9}$$

These are called the ***orthogonal projection of* $\mathbf{u}$ *on* W** and the ***orthogonal projection of* $\mathbf{u}$ *on* $W^\perp$**, respectively. The vector $\mathbf{w}_2$ is also called the ***component of* $\mathbf{u}$ *orthogonal to* W**. Using the notation in (9), Formula (8) can be expressed as

$$\mathbf{u} = \text{proj}_W \mathbf{u} + \text{proj}_{W^\perp} \mathbf{u} \tag{10}$$

(Figure 6.3.1). Moreover, since $\text{proj}_{W^\perp}\mathbf{u} = \mathbf{u} - \text{proj}_W \mathbf{u}$, we can also express Formula (10) as

$$\mathbf{u} = \text{proj}_W \mathbf{u} + (\mathbf{u} - \text{proj}_W \mathbf{u}) \tag{11}$$



▲ Figure 6.3.1

The following theorem provides formulas for calculating orthogonal projections.

> Although Formulas (12) and (13) are expressed in terms of orthogonal and orthonormal basis vectors, the resulting vector $\text{proj}_W \mathbf{u}$ does not depend on the basis vectors that are used.

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*

$$\text{proj}_W \mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{u}, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{v}_r \rangle}{\|\mathbf{v}_r\|^2}\mathbf{v}_r \tag{12}$$

(*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*

$$\text{proj}_W \mathbf{u} = \langle \mathbf{u}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2 \rangle \mathbf{v}_2 + \cdots + \langle \mathbf{u}, \mathbf{v}_r \rangle \mathbf{v}_r \tag{13}$$

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

$$\text{proj}_W \mathbf{u} = \mathbf{w}_1 = \frac{\langle \mathbf{w}_1, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{w}_1, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{w}_1, \mathbf{v}_r \rangle}{\|\mathbf{v}_r\|^2}\mathbf{v}_r \tag{14}$$

Since $\mathbf{w}_2$ is orthogonal to W, it follows that

$$\langle \mathbf{w}_2, \mathbf{v}_1 \rangle = \langle \mathbf{w}_2, \mathbf{v}_2 \rangle = \cdots = \langle \mathbf{w}_2, \mathbf{v}_r \rangle = 0$$

so we can rewrite (14) 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 (b) In this case, $\|\mathbf{v}_1\| = \|\mathbf{v}_2\| = \cdots = \|\mathbf{v}_r\| = 1$, so Formula (14) simplifies to Formula (13). ◀

▶ EXAMPLE 7 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 (13) the orthogonal projection of $\mathbf{u} = (1, 1, 1)$ on W is

$$\begin{aligned} \text{proj}_W \mathbf{u} &= \langle \mathbf{u}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2 \rangle \mathbf{v}_2 \\ &= (1)(0, 1, 0) + \left(-\tfrac{1}{5}\right)\left(-\tfrac{4}{5}, 0, \tfrac{3}{5}\right) \\ &= \left(\tfrac{4}{25}, 1, -\tfrac{3}{25}\right) \end{aligned}$$

The component of $\mathbf{u}$ orthogonal to W is

$$\text{proj}_{W^\perp} \mathbf{u} = \mathbf{u} - \text{proj}_W \mathbf{u} = (1, 1, 1) - \left(\tfrac{4}{25}, 1, -\tfrac{3}{25}\right) = \left(\tfrac{21}{25}, 0, \tfrac{28}{25}\right)$$

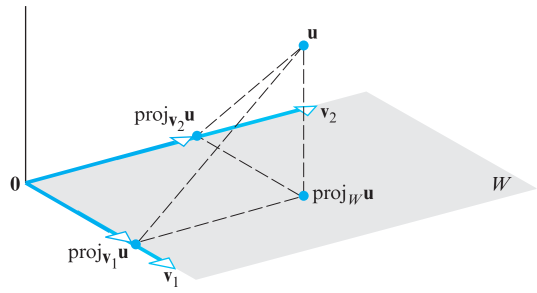
Observe that $\text{proj}_{W^\perp} \mathbf{u}$ is orthogonal to both $\mathbf{v}_1$ and $\mathbf{v}_2$, so this vector is orthogonal to each vector in the space W spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$, as it should be. ◀

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 (12) 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 $\mathbf{u}$ along $\mathbf{a}$. This suggests that we can think of (12) 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

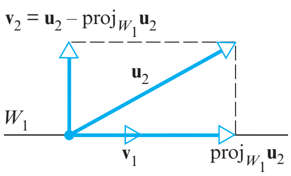
The Gram–Schmidt Process

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 $\mathbf{v}_1 = \mathbf{u}_1$.



▲ Figure 6.3.3

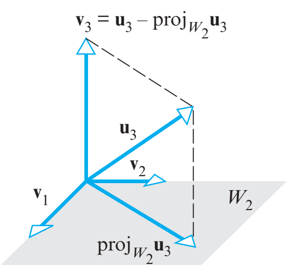
Step 2. As illustrated in Figure 6.3.3, we can obtain a vector $\mathbf{v}_2$ that is orthogonal to $\mathbf{v}_1$ by computing the component of $\mathbf{u}_2$ that is orthogonal to the space W_1 spanned by $\mathbf{v}_1$. Using Formula (12) to perform this computation, we obtain

$$\mathbf{v}_2 = \mathbf{u}_2 - \text{proj}_{W_1}\mathbf{u}_2 = \mathbf{u}_2 - \frac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1$$

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 preceding 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 $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$.



▲ Figure 6.3.4

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 (12) 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 $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ ensures that $\mathbf{v}_3 \neq \mathbf{0}$. We leave the details for you.

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 (12),

$$\mathbf{v}_4 = \mathbf{u}_4 - \text{proj}_{W_3}\mathbf{u}_4 = \mathbf{u}_4 - \frac{\langle \mathbf{u}_4, \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\langle \mathbf{u}_4, \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 - \frac{\langle \mathbf{u}_4, \mathbf{v}_3\rangle}{\|\mathbf{v}_3\|^2}\mathbf{v}_3$$

Continuing in this way we will produce after r steps an orthogonal set of nonzero vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$. Since such sets are linearly independent, we will have produced 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 $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ into an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$, 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 $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_r\}$, normalize the orthogonal basis vectors.

Jorgen Pederson Gram (1850–1916)

Historical Note Erhardt Schmidt (1875–1959) was a 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 structure 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.

Historical Note Gram was a Danish actuary whose early education was at village schools supplemented by 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: http://www-history.mcs.st-and.ac.uk/PictDisplay/Gram.html*]

▶ EXAMPLE 8 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. $$\mathbf{v}_2 = \mathbf{u}_2 - \operatorname{proj}_{W_1} \mathbf{u}_2 = \mathbf{u}_2 - \frac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1$$

$$= (0, 1, 1) - \frac{2}{3}(1, 1, 1) = \left(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

Step 3. $$\mathbf{v}_3 = \mathbf{u}_3 - \operatorname{proj}_{W_2} \mathbf{u}_3 = \mathbf{u}_3 - \frac{\langle \mathbf{u}_3, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\langle \mathbf{u}_3, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2$$

$$= (0, 0, 1) - \frac{1}{3}(1, 1, 1) - \frac{1/3}{2/3}\left(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

$$= \left(0, -\frac{1}{2}, \frac{1}{2}\right)$$

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.

CALCULUS REQUIRED

▶ EXAMPLE 9 Legendre Polynomials

Let the vector space P_2 have the inner product

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

Apply the Gram–Schmidt process to transform the standard basis $\{1, x, x^2\}$ 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

$$\langle \mathbf{u}_3, \mathbf{v}_1 \rangle = \int_{-1}^{1} x^2\,dx = \frac{x^3}{3}\bigg]_{-1}^{1} = \frac{2}{3}$$

$$\langle \mathbf{u}_3, \mathbf{v}_2 \rangle = \int_{-1}^{1} x^3\,dx = \frac{x^4}{4}\bigg]_{-1}^{1} = 0$$

$$\|\mathbf{v}_1\|^2 = \langle \mathbf{v}_1, \mathbf{v}_1 \rangle = \int_{-1}^{1} 1\,dx = x\bigg]_{-1}^{1} = 2$$

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 last example are often scaled so all three functions have a value of 1 at $x = 1$. The resulting polynomials

$$1, \quad x, \quad \frac{1}{2}(3x^2 - 1)$$

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 that Q has orthonormal column vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$. Thus, A and Q can be written in partitioned form as

$$A = [\mathbf{u}_1 \mid \mathbf{u}_2 \mid \cdots \mid \mathbf{u}_n] \quad \text{and} \quad Q = [\mathbf{q}_1 \mid \mathbf{q}_2 \mid \cdots \mid \mathbf{q}_n]$$

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

$$\begin{aligned}
\mathbf{u}_1 &= \langle \mathbf{u}_1, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{u}_1, \mathbf{q}_2 \rangle \mathbf{q}_2 + \cdots + \langle \mathbf{u}_1, \mathbf{q}_n \rangle \mathbf{q}_n \\
\mathbf{u}_2 &= \langle \mathbf{u}_2, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{u}_2, \mathbf{q}_2 \rangle \mathbf{q}_2 + \cdots + \langle \mathbf{u}_2, \mathbf{q}_n \rangle \mathbf{q}_n \\
&\vdots \\
\mathbf{u}_n &= \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \mathbf{q}_2 + \cdots + \langle \mathbf{u}_n, \mathbf{q}_n \rangle \mathbf{q}_n
\end{aligned}$$

Recalling from Section 1.3 (Example 9) that the jth column vector of a matrix product is a linear combination of the column vectors of the first factor with coefficients coming from the jth column of the second factor, it follows that these relationships can be expressed in matrix form as

$$[\mathbf{u}_1 \mid \mathbf{u}_2 \mid \cdots \mid \mathbf{u}_n] = [\mathbf{q}_1 \mid \mathbf{q}_2 \mid \cdots \mid \mathbf{q}_n] \begin{bmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \\ \langle \mathbf{u}_1, \mathbf{q}_2 \rangle & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \\ \vdots & \vdots & & \vdots \\ \langle \mathbf{u}_1, \mathbf{q}_n \rangle & \langle \mathbf{u}_2, \mathbf{q}_n \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \end{bmatrix}$$

or more briefly as

$$A = QR \tag{15}$$

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

$$R = \begin{bmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \\ 0 & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \end{bmatrix} \tag{16}$$

We leave it for you to show that R is invertible by showing that its diagonal entries are nonzero. Thus, Equation (15) 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 (15) a ***QR*-decomposition of *A***. In summary, we have the following theorem.

> It is common in numerical linear algebra to say that a matrix with linearly independent columns has ***full column rank***.

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.

Recall from Theorem 5.1.5 (the Equivalence Theorem) that a *square* matrix has linearly independent column vectors if and only if it is invertible. Thus, it follows from Theorem 6.3.7 that *every invertible matrix has a QR-decomposition.*

EXAMPLE 10 *QR*-Decomposition of a 3 × 3 Matrix

Find a 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 8)

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

Thus, it follows from Formula (16) that R is

$$R = \begin{bmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \langle \mathbf{u}_3, \mathbf{q}_1 \rangle \\ 0 & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \langle \mathbf{u}_3, \mathbf{q}_2 \rangle \\ 0 & 0 & \langle \mathbf{u}_3, \mathbf{q}_3 \rangle \end{bmatrix} = \begin{bmatrix} \frac{3}{\sqrt{3}} & \frac{2}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

from which it follows that a QR-decomposition of A is

$$\underset{A}{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}} = \underset{Q}{\begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{bmatrix}} \underset{R}{\begin{bmatrix} \frac{3}{\sqrt{3}} & \frac{2}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}} \blacktriangleleft$$

Exercise Set 6.3

1. In each part, determine whether the set of vectors is orthogonal and whether it is orthonormal 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) $(0, 0),\ (0, 1)$

2. In each part, determine whether the set of vectors is orthogonal and whether it is orthonormal 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) $\left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}\right),\ \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0\right)$

3. In each part, determine whether the set of vectors is orthogonal with respect to the standard inner product on P_2 (see 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) $p_1(x) = 1,\ \ p_2(x) = \frac{1}{\sqrt{2}}x + \frac{1}{\sqrt{2}}x^2,\ \ p_3(x) = x^2$

4. In each part, determine whether the set of vectors is orthogonal with respect to the standard inner product on M_{22} (see 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) $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},\ \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},\ \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix},\ \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}$

In Exercises 5–6, show that the column vectors of A form an orthogonal basis for the column space of A with respect to the Euclidean inner product, and then find an orthonormal basis for that column space.

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

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

7. Verify that the vectors

$$\mathbf{v}_1 = \left(-\tfrac{3}{5}, \tfrac{4}{5}, 0\right),\quad \mathbf{v}_2 = \left(\tfrac{4}{5}, \tfrac{3}{5}, 0\right),\quad \mathbf{v}_3 = (0, 0, 1)$$

form an orthonormal basis for R^3 with respect to the Euclidean inner product, and then use Theorem 6.3.2(*b*) to express the vector $\mathbf{u} = (1, -2, 2)$ as a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$.

8. Use Theorem 6.3.2(*b*) to express the vector $\mathbf{u} = (3, -7, 4)$ as a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$ in Exercise 7.

9. Verify that the vectors

$$\mathbf{v}_1 = (2, -2, 1),\quad \mathbf{v}_2 = (2, 1, -2),\quad \mathbf{v}_3 = (1, 2, 2)$$

form an orthogonal basis for R^3 with respect to the Euclidean inner product, and then use Theorem 6.3.2(*a*) to express the vector $\mathbf{u} = (-1, 0, 2)$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$.

10. Verify that the vectors

$$\mathbf{v}_1 = (1, -1, 2, -1),\quad \mathbf{v}_2 = (-2, 2, 3, 2),$$
$$\mathbf{v}_3 = (1, 2, 0, -1),\quad \mathbf{v}_4 = (1, 0, 0, 1)$$

form an orthogonal basis for R^4 with respect to the Euclidean inner product, and then use Theorem 6.3.2(*a*) to express the vector $\mathbf{u} = (1, 1, 1, 1)$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, and $\mathbf{v}_4$.

In Exercises 11–14, find the coordinate vector $(\mathbf{u})_S$ for the vector $\mathbf{u}$ and the basis S that were given in the stated exercise.

11. Exercise 7

12. Exercise 8

13. Exercise 9

14. Exercise 10

In Exercises 15–18, let R^2 have the Euclidean inner product.

(a) Find the orthogonal projection of $\mathbf{u}$ onto the line spanned by the vector $\mathbf{v}$.

(b) Find the component of $\mathbf{u}$ orthogonal to the line spanned by the vector $\mathbf{v}$, and confirm that this component is orthogonal to the line.

15. $\mathbf{u} = (-1, 6);\ \ \mathbf{v} = \left(\frac{3}{5}, \frac{4}{5}\right)$

16. $\mathbf{u} = (2, 3);\ \ \mathbf{v} = \left(\frac{5}{13}, \frac{12}{13}\right)$

17. $\mathbf{u} = (2, 3);\ \ \mathbf{v} = (1, 1)$

18. $\mathbf{u} = (3, -1);\ \ \mathbf{v} = (3, 4)$

In Exercises 19–22, let R^3 have the Euclidean inner product.

(a) Find the orthogonal projection of $\mathbf{u}$ onto the plane spanned by the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$.

(b) Find the component of $\mathbf{u}$ orthogonal to the plane spanned by the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, and confirm that this component is orthogonal to the plane.

19. $\mathbf{u} = (4, 2, 1);\ \ \mathbf{v}_1 = \left(\frac{1}{3}, \frac{2}{3}, -\frac{2}{3}\right),\ \ \mathbf{v}_2 = \left(\frac{2}{3}, \frac{1}{3}, \frac{2}{3}\right)$

20. $\mathbf{u} = (3, -1, 2);\ \ \mathbf{v}_1 = \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}\right),\ \ \mathbf{v}_2 = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$

21. $\mathbf{u} = (1, 0, 3);\ \ \mathbf{v}_1 = (1, -2, 1),\ \ \mathbf{v}_2 = (2, 1, 0)$

22. $\mathbf{u} = (1, 0, 2);\ \ \mathbf{v}_1 = (3, 1, 2),\ \ \mathbf{v}_2 = (-1, 1, 1)$

In Exercises 23–24, the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal with respect to the Euclidean inner product on R^4. Find the orthogonal projection of $\mathbf{b} = (1, 2, 0, -2)$ on the subspace W spanned by these vectors.

23. $\mathbf{v}_1 = (1, 1, 1, 1),\ \ \mathbf{v}_2 = (1, 1, -1, -1)$

24. $\mathbf{v}_1 = (0, 1, -4, -1),\ \ \mathbf{v}_2 = (3, 5, 1, 1)$

▶ In Exercises 25–26, the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are orthonormal with respect to the Euclidean inner product on R^4. Find the orthogonal projection of $\mathbf{b} = (1, 2, 0, -1)$ onto the subspace W spanned by these vectors. ◀

25. $\mathbf{v}_1 = \left(0, \frac{1}{\sqrt{18}}, -\frac{4}{\sqrt{18}}, -\frac{1}{\sqrt{18}}\right)$, $\mathbf{v}_2 = \left(\frac{1}{2}, \frac{5}{6}, \frac{1}{6}, \frac{1}{6}\right)$,
$\mathbf{v}_3 = \left(\frac{1}{\sqrt{18}}, 0, \frac{1}{\sqrt{18}}, -\frac{4}{\sqrt{18}}\right)$

26. $\mathbf{v}_1 = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$, $\mathbf{v}_2 = \left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}\right)$,
$\mathbf{v}_3 = \left(\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right)$

▶ In Exercises 27–28, let R^2 have the Euclidean inner product and 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. ◀

27. $\mathbf{u}_1 = (1, -3)$, $\mathbf{u}_2 = (2, 2)$ **28.** $\mathbf{u}_1 = (1, 0)$, $\mathbf{u}_2 = (3, -5)$

▶ In Exercises 29–30, let R^3 have the Euclidean inner product and use the Gram–Schmidt process to transform the basis $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ into an orthonormal basis. ◀

29. $\mathbf{u}_1 = (1, 1, 1)$, $\mathbf{u}_2 = (-1, 1, 0)$, $\mathbf{u}_3 = (1, 2, 1)$

30. $\mathbf{u}_1 = (1, 0, 0)$, $\mathbf{u}_2 = (3, 7, -2)$, $\mathbf{u}_3 = (0, 4, 1)$

31. 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.

$$\mathbf{u}_1 = (0, 2, 1, 0), \quad \mathbf{u}_2 = (1, -1, 0, 0),$$
$$\mathbf{u}_3 = (1, 2, 0, -1), \quad \mathbf{u}_4 = (1, 0, 0, 1)$$

32. Let R^3 have the Euclidean inner product. Find an orthonormal basis for the subspace spanned by $(0, 1, 2)$, $(-1, 0, 1)$, $(-1, 1, 3)$.

33. Let $\mathbf{b}$ and W be as in Exercise 23. Find vectors $\mathbf{w}_1$ in W and $\mathbf{w}_2$ in $W^\perp$ such that $\mathbf{b} = \mathbf{w}_1 + \mathbf{w}_2$.

34. Let $\mathbf{b}$ and W be as in Exercise 25. Find vectors $\mathbf{w}_1$ in W and $\mathbf{w}_2$ in $W^\perp$ such that $\mathbf{b} = \mathbf{w}_1 + \mathbf{w}_2$.

35. Let R^3 have the Euclidean inner product. The subspace of R^3 spanned by the vectors $\mathbf{u}_1 = (1, 1, 1)$ and $\mathbf{u}_2 = (2, 0, -1)$ is a plane 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.

36. 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.

37. 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.

38. Verify that the set of vectors $\{(1, 0), (0, 1)\}$ is orthogonal with respect to the inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 4u_1 v_1 + u_2 v_2$ on R^2; then convert it to an orthonormal set by normalizing the vectors.

39. Find vectors $\mathbf{x}$ and $\mathbf{y}$ in R^2 that are orthonormal with respect to the inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1 v_1 + 2u_2 v_2$ but are not orthonormal with respect to the Euclidean inner product.

40. In Example 3 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 positive x-axis. Solve that same problem using Theorem 6.3.4.

41. This exercise illustrates that the orthogonal projection resulting from Formula (12) in Theorem 6.3.4 does not depend on which orthogonal basis vectors are used.

(a) Let R^3 have the Euclidean inner product, and let W be the subspace of R^3 spanned by the orthogonal vectors

$$\mathbf{v}_1 = (1, 0, 1) \quad \text{and} \quad \mathbf{v}_2 = (0, 1, 0)$$

Show that the orthogonal vectors

$$\mathbf{v}_1' = (1, 1, 1) \quad \text{and} \quad \mathbf{v}_2' = (1, -2, 1)$$

span the same subspace W.

(b) Let $\mathbf{u} = (-3, 1, 7)$ and show that the same vector $\text{proj}_W \mathbf{u}$ results regardless of which of the bases in part (a) is used for its computation.

42. (***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 9).

(a) $1 + x + 4x^2$ (b) $2 - 7x^2$ (c) $4 + 3x$

43. (***Calculus required***) Let P_2 have the inner product

$$\langle \mathbf{p}, \mathbf{q} \rangle = \int_0^1 p(x)q(x)\,dx$$

Apply the Gram–Schmidt process to transform the standard basis $S = \{1, x, x^2\}$ into an orthonormal basis.

44. Find an orthogonal basis for the column space of the matrix

$$A = \begin{bmatrix} 6 & 1 & -5 \\ 2 & 1 & 1 \\ -2 & -2 & 5 \\ 6 & 8 & -7 \end{bmatrix}$$

▶ In Exercises 45–48, we obtained the column vectors of Q by applying the Gram–Schmidt process to the column vectors of A. Find a QR-decomposition of the matrix A. ◀

45. $A = \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}$, $Q = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}$

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

47. $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 1 & 2 & 0 \end{bmatrix}, \quad Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \end{bmatrix}$

48. $A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 0 & 3 & 1 \end{bmatrix}, \quad Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{\sqrt{2}}{2\sqrt{19}} & -\frac{3}{\sqrt{19}} \\ \frac{1}{\sqrt{2}} & -\frac{\sqrt{2}}{2\sqrt{19}} & \frac{3}{\sqrt{19}} \\ 0 & \frac{3\sqrt{2}}{\sqrt{19}} & \frac{1}{\sqrt{19}} \end{bmatrix}$

49. Find a QR-decomposition of the matrix

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

50. In the Remark following Example 8 we discussed two alternative ways to perform the calculations in the Gram–Schmidt process: normalizing each orthogonal basis vector as soon as it is calculated and scaling the orthogonal basis vectors at each step to eliminate fractions. Try these methods in Example 8.

Working with Proofs

51. Prove part (a) of Theorem 6.3.6.

52. 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.

53. Prove that the diagonal entries of R in Formula (16) are nonzero.

54. Show that matrix Q in Example 10 has the property $QQ^T = I_3$, and prove that every $m \times n$ matrix Q with orthonormal column vectors has the property $QQ^T = I_m$.

55. (a) Prove that if W is a subspace of a finite-dimensional vector space V, then the mapping $T: V \to W$ defined by $T(\mathbf{v}) = \text{proj}_W \mathbf{v}$ is a linear transformation.

(b) What are the range and kernel of the transformation in part (a)?

True-False Exercises

TF. 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.

(b) Every orthogonal set of vectors in an inner product space is linearly independent.

(c) Every nontrivial subspace of R^3 has an orthonormal basis with respect to the Euclidean inner product.

(d) Every nonzero finite-dimensional inner product space has an orthonormal basis.

(e) $\text{proj}_W \mathbf{x}$ is orthogonal to every vector of W.

(f) If A is an $n \times n$ matrix with a nonzero determinant, then A has a QR-decomposition.

Working with Technology

T1. (a) Use the Gram–Schmidt process to find an orthonormal basis relative to the Euclidean inner product for the column space of

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

(b) Use the method of Example 9 to find a QR-decomposition of A.

T2. Let P_4 have the evaluation inner product at the points $-2, -1, 0, 1, 2$. Find an orthogonal basis for P_4 relative to this inner product by applying the Gram–Schmidt process to the vectors

$$\mathbf{p}_0 = 1, \quad \mathbf{p}_1 = x, \quad \mathbf{p}_2 = x^2, \quad \mathbf{p}_3 = x^3, \quad \mathbf{p}_4 = x^4$$

6.4 Best Approximation; Least Squares

There are many applications in which some linear system $A\mathbf{x} = \mathbf{b}$ of m equations in n unknowns should be consistent on physical grounds but fails to be so because of measurement errors in the entries of A or $\mathbf{b}$. In such cases one looks for vectors that come as close as possible to being solutions in the sense that they minimize $\|\mathbf{b} - A\mathbf{x}\|$ with respect to the Euclidean inner product on R^m. In this section we will discuss methods for finding such minimizing vectors.

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 errors in the entries of A or $\mathbf{b}$. 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.

If a linear system is consistent, then its exact solutions are the same as its least squares solutions, in which case the least squares error is zero.

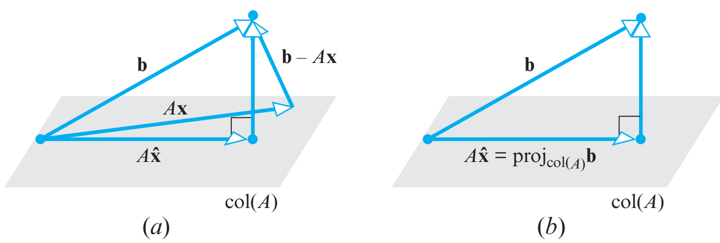
Least Squares Problem Given a linear system $A\mathbf{x}=\mathbf{b}$ of m equations in n unknowns, find a vector $\mathbf{x}$ in R^n that minimizes $\|\mathbf{b}-A\mathbf{x}\|$ with respect to the Euclidean inner product on R^m. We call such a vector, if it exists, a ***least squares solution*** of $A\mathbf{x}=\mathbf{b}$, 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 explain the terminology in this problem, suppose that the column 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 has the effect of minimizing $\|\mathbf{b}-A\mathbf{x}\|^2 = e_1^2+e_2^2+\cdots+e_m^2$.

What is important to keep in mind about the least squares problem is that for every vector $\mathbf{x}$ in R^n, the product $A\mathbf{x}$ is in the column space of A because it is a linear combination of the column vectors of A. That being the case, to find a least squares solution of $A\mathbf{x}=\mathbf{b}$ is equivalent to finding a vector $A\hat{\mathbf{x}}$ in the column space of A that is closest to $\mathbf{b}$ in the sense that it minimizes the length of the vector $\mathbf{b}-A\mathbf{x}$. This is illustrated in Figure 6.4.1*a*, which also suggests that $A\hat{\mathbf{x}}$ is the orthogonal projection of $\mathbf{b}$ on the column space of A, that is, $A\hat{\mathbf{x}}=\text{proj}_{\text{col}(A)}\mathbf{b}$ (Figure 6.4.1*b*). The next theorem will confirm this conjecture.



▶ Figure 6.4.1

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

$$\|\mathbf{b} - \mathbf{w}\|^2 = \|\mathbf{b} - \text{proj}_W \mathbf{b}\|^2 + \|\text{proj}_W \mathbf{b} - \mathbf{w}\|^2$$

If $\mathbf{w} \neq \text{proj}_W \mathbf{b}$, 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

$$\|\mathbf{b} - \text{proj}_W \mathbf{b}\| < \|\mathbf{b} - \mathbf{w}\| \quad ◀$$

It follows from Theorem 6.4.1 that if $V = R^n$ and $W = \text{col}(A)$, then the best approximation to $\mathbf{b}$ from $\text{col}(A)$ is $\text{proj}_{\text{col}(A)}\mathbf{b}$. But every vector in the column space of A is expressible in the form $A\mathbf{x}$ for some vector $\mathbf{x}$, so there is at least one vector $\hat{\mathbf{x}}$ in $\text{col}(A)$ for which $A\hat{\mathbf{x}} = \text{proj}_{\text{col}(A)}\mathbf{b}$. Each such vector is a least squares solution of $A\mathbf{x} = \mathbf{b}$. Note, however, that although there may be more than one least squares solution of $A\mathbf{x} = \mathbf{b}$, each such solution $\hat{\mathbf{x}}$ has the same error vector $\mathbf{b} - A\hat{\mathbf{x}}$.

Finding Least Squares Solutions

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 A and then solve the equation

$$A\mathbf{x} = \text{proj}_W \mathbf{b} \tag{2}$$

However, we can avoid calculating 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 4.8.7(b) that this vector lies in the null space of A^T, and hence that

$$A^T(\mathbf{b} - \text{proj}_W \mathbf{b}) = \mathbf{0}$$

Thus, (3) simplifies to

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

which we can rewrite as

$$A^TA\mathbf{x} = A^T\mathbf{b} \tag{4}$$

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 squares solutions 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*

$$\text{proj}_W \mathbf{b} = A\mathbf{x} \tag{6}$$

EXAMPLE 1 Unique Least Squares Solution

Find the least squares solution, the least squares error vector, and the least squares error of the linear system

$$\begin{aligned} x_1 - x_2 &= 4 \\ 3x_1 + 2x_2 &= 1 \\ -2x_1 + 4x_2 &= 3 \end{aligned}$$

Solution It will be convenient to express the system in the matrix form $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} \tag{7}$$

It follows that

$$A^T A = \begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 14 & -3 \\ -3 & 21 \end{bmatrix} \tag{8}$$

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 10 \end{bmatrix}$$

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,

$$x_1 = \tfrac{17}{95}, \quad x_2 = \tfrac{143}{285}$$

The least squares 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 least squares error is

$$\|\mathbf{b} - A\mathbf{x}\| \approx 4.556 \quad ◀$$

The computations in the next example are a little tedious for hand computation, so in absence of a calculating utility you may want to just read through it for its ideas and logical flow.

EXAMPLE 2 Infinitely Many Least Squares Solutions

Find the least squares solutions, the least squares error vector, and the least squares error of the linear system

$$\begin{aligned} 3x_1 + 2x_2 - x_3 &= 2 \\ x_1 - 4x_2 + 3x_3 &= -2 \\ x_1 + 10x_2 - 7x_3 &= 1 \end{aligned}$$

Solution The matrix form of the system is $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 3 & 2 & -1 \\ 1 & -4 & 3 \\ 1 & 10 & -7 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$$

It follows that

$$A^TA = \begin{bmatrix} 11 & 12 & -7 \\ 12 & 120 & -84 \\ -7 & -84 & 59 \end{bmatrix} \quad \text{and} \quad A^T\mathbf{b} = \begin{bmatrix} 5 \\ 22 \\ -15 \end{bmatrix}$$

so the augmented matrix for the normal system $A^TA\mathbf{x} = A^T\mathbf{b}$ is

$$\left[\begin{array}{ccc|c} 11 & 12 & -7 & 5 \\ 12 & 120 & -84 & 22 \\ -7 & -84 & 59 & -15 \end{array}\right]$$

The reduced row echelon form of this matrix is

$$\left[\begin{array}{ccc|c} 1 & 0 & \frac{1}{7} & \frac{2}{7} \\ 0 & 1 & -\frac{5}{7} & \frac{13}{84} \\ 0 & 0 & 0 & 0 \end{array}\right]$$

from which it follows that there are infinitely many least squares solutions, and that they are given by the parametric equations

$$\begin{aligned} x_1 &= \tfrac{2}{7} - \tfrac{1}{7}t \\ x_2 &= \tfrac{13}{84} + \tfrac{5}{7}t \\ x_3 &= t \end{aligned}$$

As a check, let us verify that all least squares solutions produce the same least squares error vector and the same least squares error. To see that this is so, we first compute

$$\mathbf{b} - A\mathbf{x} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 & 2 & -1 \\ 1 & -4 & 3 \\ 1 & 10 & -7 \end{bmatrix} \begin{bmatrix} \frac{2}{7} - \frac{1}{7}t \\ \frac{13}{84} + \frac{5}{7}t \\ t \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{7}{6} \\ -\frac{1}{3} \\ \frac{11}{6} \end{bmatrix} = \begin{bmatrix} \frac{5}{6} \\ -\frac{5}{3} \\ -\frac{5}{6} \end{bmatrix}$$

Since $\mathbf{b} - A\mathbf{x}$ does not depend on t, all least squares solutions produce the same error vector, namely

$$\|\mathbf{b} - A\mathbf{x}\| = \sqrt{\left(\tfrac{5}{6}\right)^2 + \left(-\tfrac{5}{3}\right)^2 + \left(-\tfrac{5}{6}\right)^2} = \tfrac{5}{6}\sqrt{6}$$ ◀

Conditions for Uniqueness of Least Squares Solutions

We know from Theorem 6.4.2 that the system $A^TA\mathbf{x} = A^T\mathbf{b}$ of normal equations that is associated with the system $A\mathbf{x} = \mathbf{b}$ is consistent. Thus, it follows from Theorem 1.6.1 that every linear system $A\mathbf{x} = \mathbf{b}$ has either one least squares solution (as in Example 1) or infinitely many least squares solutions (as in Example 2). Since A^TA is a square matrix, the former occurs if A^TA is invertible and the latter if it is not. The next two theorems are concerned with this idea.

THEOREM 6.4.3 *If A is an $m \times n$ matrix, then the following are equivalent.*

(*a*) *The column vectors of A are linearly independent.*

(*b*) *A^TA 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 the column vectors of A are linearly independent. The matrix A^TA has size $n \times n$, so we can prove that this matrix is invertible by showing that the linear system $A^TA\mathbf{x} = \mathbf{0}$ 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 4.8.7(*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. ◀

The next theorem, which follows directly from Theorems 6.4.2 and 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 linear system $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution. This solution is given by*

$$\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b} \tag{9}$$

Moreover, if W is the column space of A, then the orthogonal projection of $\mathbf{b}$ on W is

$$\text{proj}_W \mathbf{b} = A\mathbf{x} = A(A^TA)^{-1}A^T\mathbf{b} \tag{10}$$

▶ EXAMPLE 3 A Formula Solution to Example 1

Use Formula (9) and the matrices in Formulas (7) and (8) to find the least squares solution of the linear system in Example 1.

Solution We leave it for you to verify that

$$\begin{aligned}\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b} &= \begin{bmatrix} 14 & -3 \\ -3 & 21 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} \\ &= \frac{1}{285}\begin{bmatrix} 21 & 3 \\ 3 & 14 \end{bmatrix} \begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{17}{95} \\ \frac{143}{285} \end{bmatrix}\end{aligned}$$

which agrees with the result obtained in Example 1. ◀

It follows from Formula (10) that the standard matrix for the orthogonal projection on the column space of a matrix A is

$$P = A(A^TA)^{-1}A^T \tag{11}$$

We will use this result in the next example.

▶ EXAMPLE 4 Orthogonal Projection on a Column Space

We showed in Formula (4) of Section 4.9 that the standard matrix for the orthogonal projection onto the line W through the origin of R^2 that makes an angle θ with the positive x-axis is

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

Derive this result using Formula (11).

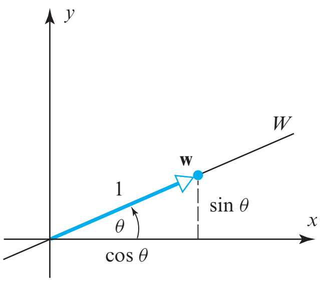


Figure 6.4.2

Solution To apply Formula (11) we must find a matrix A for which the line W is the column space. Since the line is one-dimensional and consists of all scalar multiples of the vector $\mathbf{w} = (\cos\theta, \sin\theta)$ (see Figure 6.4.2), we can take A to be

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

Since A^TA is the 1×1 identity matrix (verify), it follows that

$$A(A^TA)^{-1}A^T = AA^T = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} [\cos\theta \quad \sin\theta]$$

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

More on the Equivalence Theorem

As our final result in the main part of this section we will add one additional part to Theorem 5.1.5.

THEOREM 6.4.5 Equivalent Statements

If A is an $n \times n$ matrix, then the following statements are equivalent.

(a) *A is invertible.*

(b) *$A\mathbf{x} = \mathbf{0}$ 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 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 kernel of T_A is $\{\mathbf{0}\}$.*

(s) *The range of T_A is R^n.*

(t) *T_A is one-to-one.*

(u) *$\lambda = 0$ is not an eigenvalue of A.*

(v) *A^TA is invertible.*

The proof of part (*v*) follows from part (*h*) of this theorem and Theorem 6.4.3 applied to square matrices.

OPTIONAL

Another View of Least Squares

Recall from Theorem 4.8.7 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

$$\mathbf{x} = \mathbf{x}_{\text{row}(A)} + \mathbf{x}_{\text{null}(A)} \quad \text{and} \quad \mathbf{b} = \mathbf{b}_{\text{null}(A^T)} + \mathbf{b}_{\text{col}(A)}$$

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 $\mathbf{b}_{\text{null}(A^T)}$ 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 $A\mathbf{x} = \mathbf{b}_{\text{col}(A)}$.

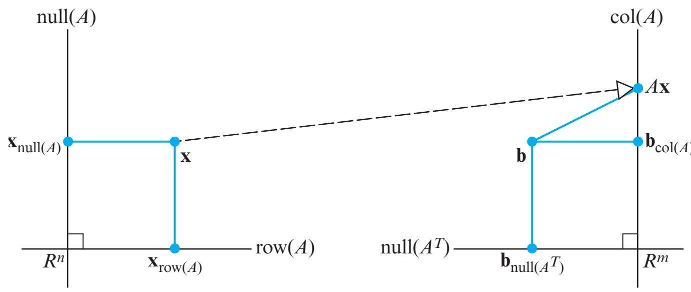


Figure 6.4.3

OPTIONAL

The Role of QR-Decomposition in Least Squares Problems

Formulas (9) and (10) 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 6.4.6 *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*

$$\mathbf{x} = R^{-1}Q^T\mathbf{b} \tag{12}$$

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 (12) by making the substitution $A = QR$ in (9) and using the fact that $Q^TQ = I$ to obtain

$$\begin{aligned}\mathbf{x} &= \big((QR)^T(QR)\big)^{-1}(QR)^T\mathbf{b}\\ &= (R^TQ^TQR)^{-1}(QR)^T\mathbf{b}\\ &= R^{-1}(R^T)^{-1}R^TQ^T\mathbf{b}\\ &= R^{-1}Q^T\mathbf{b}\end{aligned}$$

Exercise Set 6.4

▶ In Exercises 1–2, find the associated normal equation. ◀

1. $\begin{bmatrix} 1 & -1 \\ 2 & 3 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}$

2. $\begin{bmatrix} 2 & -1 & 0 \\ 3 & 1 & 2 \\ -1 & 4 & 5 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}$

▶ In Exercises 3–6, find the least squares solution of the equation $A\mathbf{x} = \mathbf{b}$. ◀

3. $A = \begin{bmatrix} 1 & -1 \\ 2 & 3 \\ 4 & 5 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}$

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

5. $A = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & -2 \\ 1 & 1 & 0 \\ 1 & 1 & -1 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 6 \\ 0 \\ 9 \\ 3 \end{bmatrix}$

6. $A = \begin{bmatrix} 2 & 0 & -1 \\ 1 & -2 & 2 \\ 2 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 0 \\ 6 \\ 0 \\ 6 \end{bmatrix}$

▶ In Exercises 7–10, find the least squares error vector and least squares error of the stated equation. Verify that the least squares error vector is orthogonal to the column space of A. ◀

7. The equation in Exercise 3.

8. The equation in Exercise 4.

9. The equation in Exercise 5.

10. The equation in Exercise 6.

▶ In Exercises 11–14, find parametric equations for all least squares solutions of $A\mathbf{x} = \mathbf{b}$, and confirm that all of the solutions have the same error vector. ◀

11. $A = \begin{bmatrix} 2 & 1 \\ 4 & 2 \\ -2 & -1 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$

12. $A = \begin{bmatrix} 1 & 3 \\ -2 & -6 \\ 3 & 9 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

13. $A = \begin{bmatrix} -1 & 3 & 2 \\ 2 & 1 & 3 \\ 0 & 1 & 1 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 7 \\ 0 \\ -7 \end{bmatrix}$

14. $A = \begin{bmatrix} 3 & 2 & -1 \\ 1 & -4 & 3 \\ 1 & 10 & -7 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$

▶ In Exercises 15–16, use Theorem 6.4.2 to find the orthogonal projection of $\mathbf{b}$ on the column space of A, and check your result using Theorem 6.4.4. ◀

15. $A = \begin{bmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix}$

16. $A = \begin{bmatrix} 5 & 1 \\ 1 & 3 \\ 4 & -2 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} -4 \\ 2 \\ 3 \end{bmatrix}$

17. Find the orthogonal projection of $\mathbf{u}$ on the subspace of R^3 spanned by the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$.

$$\mathbf{u} = (1, -6, 1); \quad \mathbf{v}_1 = (-1, 2, 1), \quad \mathbf{v}_2 = (2, 2, 4)$$

18. Find the orthogonal projection of $\mathbf{u}$ on the subspace of R^4 spanned by the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$.

$$\mathbf{u} = (6, 3, 9, 6); \quad \mathbf{v}_1 = (2, 1, 1, 1), \quad \mathbf{v}_2 = (1, 0, 1, 1),$$
$$\mathbf{v}_3 = (-2, -1, 0, -1)$$

▶ In Exercises 19–20, use the method of Example 3 to find the standard matrix for the orthogonal projection on the stated subspace of R^2. Compare your result to that in Table 3 of Section 4.9. ◀

19. the x-axis

20. the y-axis

▶ In Exercises 21–22, use the method of Example 3 to find the standard matrix for the orthogonal projection on the stated subspace of R^3. Compare your result to that in Table 4 of Section 4.9. ◀

21. the xz-plane

22. the yz-plane

▶ In Exercises 23–24, a QR-factorization of A is given. Use it to find the least squares solution of $A\mathbf{x} = \mathbf{b}$. ◀

23. $A = \begin{bmatrix} 3 & 1 \\ -4 & 1 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 5 & -\frac{1}{5} \\ 0 & \frac{7}{5} \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

24. $A = \begin{bmatrix} 3 & -6 \\ 4 & -8 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & 0 \\ \frac{4}{5} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 5 & -10 \\ 0 & 1 \end{bmatrix}$; $\mathbf{b} = \begin{bmatrix} -1 \\ 7 \\ 2 \end{bmatrix}$

25. Let W be the plane with equation $5x - 3y + z = 0$.

(a) Find a basis for W.

(b) Find the standard matrix for the orthogonal projection onto W.

26. Let W be the line with parametric equations

$$x = 2t, \quad y = -t, \quad z = 4t$$

(a) Find a basis for W.

(b) Find the standard matrix for the orthogonal projection on W.

27. Find the orthogonal projection of $\mathbf{u} = (5, 6, 7, 2)$ on the solution space of the homogeneous linear system

$$\begin{aligned} x_1 + \quad x_2 + x_3 \qquad &= 0 \\ 2x_2 + x_3 + x_4 &= 0 \end{aligned}$$

28. Show that if $\mathbf{w} = (a, b, c)$ is a nonzero vector, then the standard matrix for the orthogonal projection of R^3 onto the line span$\{\mathbf{w}\}$ is

$$P = \frac{1}{a^2 + b^2 + c^2} \begin{bmatrix} a^2 & ab & ac \\ ab & b^2 & bc \\ ac & bc & c^2 \end{bmatrix}$$

29. 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.

Working with Proofs

30. 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.

31. 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 $\mathbf{x} = \mathbf{0}$.

32. Prove the implication $(b) \Rightarrow (a)$ of Theorem 6.4.3.

True-False Exercises

TF. 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.

(b) If A^TA is invertible, then A is invertible.

(c) If A is invertible, then A^TA is invertible.

(d) If $A\mathbf{x} = \mathbf{b}$ is a consistent linear system, then $A^TA\mathbf{x} = A^T\mathbf{b}$ is also consistent.

(e) If $A\mathbf{x} = \mathbf{b}$ is an inconsistent linear system, then $A^TA\mathbf{x} = A^T\mathbf{b}$ is also inconsistent.

(f) Every linear system has a least squares solution.

(g) Every linear system has a unique least squares solution.

(h) If A is an $m \times n$ matrix with linearly independent columns and $\mathbf{b}$ is in R^m, then $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution.

Working with Technology

T1. (a) Use Theorem 6.4.4 to show that the following linear system has a unique least squares solution, and use the method of Example 1 to find it.

$$\begin{aligned} x_1 + \quad x_2 + x_3 &= \ \ 1 \\ 4x_1 + 2x_2 + x_3 &= 10 \\ 9x_1 + 3x_2 + x_3 &= \ \ 9 \\ 16x_1 + 4x_2 + x_3 &= 16 \end{aligned}$$

(b) Check your result in part (a) using Formula (9).

T2. Use your technology utility to perform the computations and confirm the results obtained in Example 2.

6.5 Mathematical Modeling Using Least Squares

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

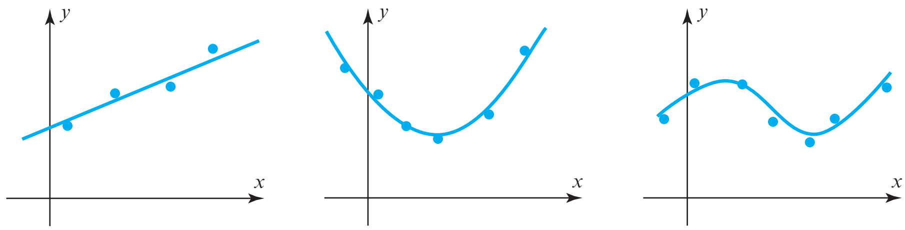
$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

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. This curve is called a ***mathematical model*** of the data. Some examples 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: $y = a + bx + cx^2 + dx^3$



▶ Figure 6.5.1 (a) $y = a + bx$ (b) $y = a + bx + cx^2$ (c) $y = a + bx + cx^2 + dx^3$

Least Squares Fit of a Straight Line

When data points are obtained experimentally, there is generally some measurement "error," 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 case: fitting a straight line to data points.

Suppose we want to fit a straight line $y = a + bx$ to the experimentally determined points

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

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

$$\begin{aligned} y_1 &= a + bx_1 \\ y_2 &= a + bx_2 \\ &\vdots \\ y_n &= a + bx_n \end{aligned} \tag{1}$$

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

$$M\mathbf{v} = \mathbf{y} \tag{2}$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad M = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix} \tag{3}$$

If there are measurement errors in the data, then the data points will typically not lie on a line, and (1) will be inconsistent. In this case we look for a least squares approximation to the values of a and b by solving the normal system

$$M^T M\mathbf{v} = M^T\mathbf{y}$$

For simplicity, let us assume that the x-coordinates of the data points are not all the same, so M has linearly independent column vectors (Exericse 14) and the normal system has the unique solution

$$\mathbf{v}^* = \begin{bmatrix} a^* \\ b^* \end{bmatrix} = (M^T M)^{-1} M^T \mathbf{y}$$

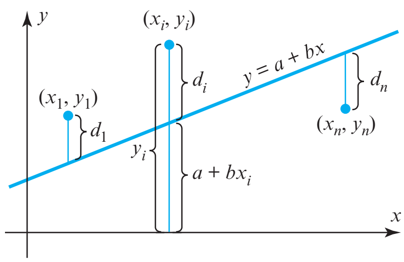
[see Formula (9) of Theorem 6.4.4]. The line $y = a^* + b^* x$ that results from this solution is called the ***least squares line of best fit*** or the ***regression line***. It follows from (2) and (3) that this line minimizes

$$\|\mathbf{y} - M\mathbf{v}\|^2 = [y_1 - (a + bx_1)]^2 + [y_2 - (a + bx_2)]^2 + \cdots + [y_n - (a + bx_n)]^2$$

The quantities

$$d_1 = |y_1 - (a + bx_1)|, \quad d_2 = |y_2 - (a + bx_2)|, \ldots, \quad d_n = |y_n - (a + bx_n)|$$

are called ***residuals***. Since the residual d_i is the distance between the data point (x_i, y_i) and the regression line (Figure 6.5.2), we can interpret its value as the "error" in y_i at the point x_i. If we assume that the value of each x_i is exact, then all the errors are in the y_i so the regression line can be described as *the line that minimizes the sum of the squares of the data errors*—hence the name, "least squares line of best fit." In summary, we have the following theorem.



▶ Figure 6.5.2 d_i measures the vertical error.

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

$$M = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{4}$$

Then there is a unique least squares straight line fit

$$y = a^* + b^* x \tag{5}$$

to the data points. Moreover,

$$\mathbf{v}^* = \begin{bmatrix} a^* \\ b^* \end{bmatrix} \tag{6}$$

is given by the formula

$$\mathbf{v}^* = (M^T M)^{-1} M^T \mathbf{y} \tag{7}$$

which expresses the fact that $\mathbf{v} = \mathbf{v}^*$ is the unique solution of the normal equation

$$M^T M \mathbf{v} = M^T \mathbf{y} \tag{8}$$

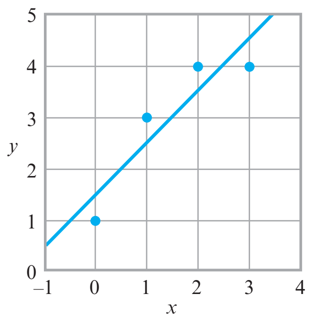


Figure 6.5.3

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.)

Solution We have

$$M = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \quad M^TM = \begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix}, \quad \text{and} \quad (M^TM)^{-1} = \frac{1}{10}\begin{bmatrix} 7 & -3 \\ -3 & 2 \end{bmatrix}$$

$$\mathbf{v}^* = (M^TM)^{-1}M^T\mathbf{y} = \frac{1}{10}\begin{bmatrix} 7 & -3 \\ -3 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 1 \end{bmatrix}$$

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 $y = 0$). Suppose further that, as illustrated in Figure 6.5.4, various weights are attached to the end of the spring and the following table of resulting spring lengths is recorded. Find the least squares straight line fit to the data and use it to approximate the spring constant.

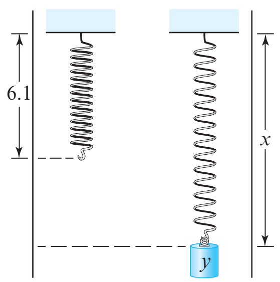


Figure 6.5.4

Weight y **(lb)**	0	2	4	6
Length x **(in)**	6.1	7.6	8.7	10.4

Solution The mathematical problem is to fit a line $y = a + bx$ to the four data points

$$(6.1, 0), \quad (7.6, 2), \quad (8.7, 4), \quad (10.4, 6)$$

For these data the matrices M and $\mathbf{y}$ in (4) are

$$M = \begin{bmatrix} 1 & 6.1 \\ 1 & 7.6 \\ 1 & 8.7 \\ 1 & 10.4 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 0 \\ 2 \\ 4 \\ 6 \end{bmatrix}$$

so

$$\mathbf{v}^* = \begin{bmatrix} a^* \\ b^* \end{bmatrix} = (M^TM)^{-1}M^T\mathbf{y} \approx \begin{bmatrix} -8.6 \\ 1.4 \end{bmatrix}$$

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. ◀

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

$$y = a_0 + a_1x + \cdots + a_mx^m \tag{9}$$

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 (9) yields the n equations

$$\begin{aligned} y_1 &= a_0 + a_1x_1 + \cdots + a_mx_1^m \\ y_2 &= a_0 + a_1x_2 + \cdots + a_mx_2^m \\ &\vdots \\ y_n &= a_0 + a_1x_n + \cdots + a_mx_n^m \end{aligned}$$

or in matrix form,

$$\mathbf{y} = M\mathbf{v} \tag{10}$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad M = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} \tag{11}$$

As before, the solutions of the normal equations

$$M^TM\mathbf{v} = M^T\mathbf{y}$$

determine the coefficients of the polynomial, and the vector $\mathbf{v}$ minimizes

$$\|\mathbf{y} - M\mathbf{v}\|$$

Conditions that guarantee the invertibility of M^TM are discussed in the exercises (Exercise 16). If M^TM is invertible, then the normal equations have a unique solution $\mathbf{v} = \mathbf{v}^*$, which is given by

$$\mathbf{v}^* = (M^TM)^{-1}M^T\mathbf{y} \tag{12}$$

▶ 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 in accordance with the equation

$$s = s_0 + v_0t + \tfrac{1}{2}gt^2 \tag{13}$$

where

$$\begin{aligned} s &= \text{vertical displacement downward relative to some reference point} \\ s_0 &= \text{displacement from the reference point at time } t = 0 \\ v_0 &= \text{velocity at time } t = 0 \\ g &= \text{acceleration of gravity at the Earth's surface} \end{aligned}$$

Suppose that a laboratory experiment is performed to approximate g by measuring the displacement s relative to a fixed reference point of a falling weight at various times. Use the experimental results shown in the following table to approximate g.

Time t (sec)	.1	.2	.3	.4	.5
Displacement s (ft)	−0.18	0.31	1.03	2.48	3.73

Solution For notational simplicity, let $a_0 = s_0$, $a_1 = v_0$, and $a_2 = \frac{1}{2}g$ in (13), so our mathematical problem is to fit a quadratic curve

$$s = a_0 + a_1 t + a_2 t^2 \tag{14}$$

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 $\mathbf{y}$ in (11) 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 (12),

$$\mathbf{v}^* = \begin{bmatrix} a_0^* \\ a_1^* \\ a_2^* \end{bmatrix} = (M^TM)^{-1}M^T\mathbf{y} \approx \begin{bmatrix} -0.40 \\ 0.35 \\ 16.1 \end{bmatrix}$$

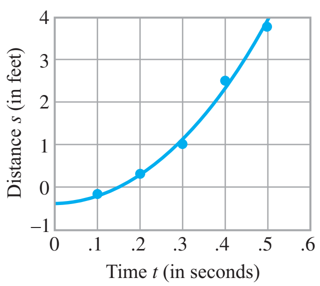
so the least squares quadratic fit is

$$s = -0.40 + 0.35t + 16.1t^2$$

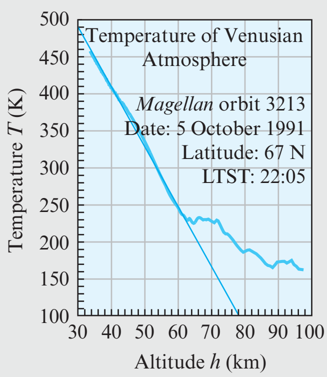
From this equation we estimate that $\frac{1}{2}g = 16.1$ and hence that $g = 32.2$ ft/sec^2. Note that this equation also provides the following estimates of the initial displacement and velocity of the weight:

$$s_0 = a_0^* = -0.40 \text{ ft}$$
$$v_0 = a_1^* = 0.35 \text{ ft/sec}$$

In Figure 6.5.5 we have plotted the data points and the approximating polynomial. ◀



▲ Figure 6.5.5



Source: NASA

Historical Note On October 5, 1991 the *Magellan* spacecraft entered the atmosphere of Venus and transmitted the temperature T in kelvins (K) versus the altitude h in kilometers (km) until its signal was lost at an altitude of about 34 km. Discounting the initial 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. The accuracy of this result has been confirmed by more recent flybys of Venus.

Exercise Set 6.5

▶ In Exercises 1–2, find the least squares straight line fit $y = ax + b$ to the data points, and show that the result is reasonable by graphing the fitted line and plotting the data in the same coordinate system. ◀

1. $(0, 0),\ (1, 2),\ (2, 7)$

2. $(0, 1),\ (2, 0),\ (3, 1),\ (3, 2)$

▶ In Exercises 3–4, find the least squares quadratic fit $y = a_0 + a_1x + a_2x^2$ to the data points, and show that the result is reasonable by graphing the fitted curve and plotting the data in the same coordinate system. ◀

3. $(2, 0),\ (3, -10),\ (5, -48),\ (6, -76)$

4. $(1, -2),\ (0, -1),\ (1, 0),\ (2, 4)$

5. 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 $X = 1/x$.

6. Find a curve of the form $y = a + b\sqrt{x}$ that best fits the data points $(3, 1.5)$, $(7, 2.5)$, $(10, 3)$ by making the substitution $X = \sqrt{x}$. Show that the result is reasonable by graphing the fitted curve and plotting the data in the same coordinate system.

Working with Proofs

7. Prove that the matrix M in Equation (3) has linearly independent columns if and only if at least two of the numbers $x_1, x_2, \ldots, x_n$ are distinct.

8. Prove that the columns of the $n \times (m + 1)$ matrix M in Equation (11) 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.]

9. Let M be the matrix in Equation (11). Using Exercise 8, show that a sufficient condition for the matrix M^TM 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.

True-False Exercises

TF. 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.

(b) If the data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ are not collinear, then (2) is an inconsistent system.

(c) If the data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ do not lie on a vertical line, then the expression

$$|y_1 - (a + bx_1)|^2 + |y_2 - (a + bx_2)^2| + \cdots + |y_n - (a + bx_n)|^2$$

is minimized by taking a and b to be the coefficients in the least squares line $y = a + bx$ of best fit to the data.

(d) If the data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ do not lie on a vertical line, then the expression

$$|y_1 - (a + bx_1)| + |y_2 - (a + bx_2)| + \cdots + |y_n - (a + bx_n)|$$

is minimized by taking a and b to be the coefficients in the least squares line $y = a + bx$ of best fit to the data.

Working with Technology

▶ In Exercises T1–T7, find the normal system for the least squares cubic fit $y = a_0 + a_1x + a_2x^2 + a_3x^3$ to the data points. Solve the system and show that the result is reasonable by graphing the fitted curve and plotting the data in the same coordinate system. ◀

T1. $(-1, -14),\ (0, -5),\ (1, -4),\ (2, 1),\ (3, 22)$

T2. $(0, -10),\ (1, -1),\ (2, 0),\ (3, 5),\ (4, 26)$

T3. 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.

T4. *Pathfinder* is an experimental, lightweight, remotely piloted, solar-powered aircraft that was used in a series of experiments by NASA to determine the feasibility of applying solar power for long-duration, high-altitude flights. In August 1997 *Pathfinder* recorded the 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 the least squares line $H = H_0 + kT$ of best fit.

Table Ex-T4

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

▶ Three important models in applications are

exponential models $(y = ae^{bx})$
power function models $(y = ax^b)$
logarithmic models $(y = a + b\ln x)$

where a and b are to be determined to fit experimental data as closely as possible. Exercises T5–T7 are concerned with a procedure, called ***linearization***, by which the data are transformed to a form in which a least squares straight line fit can be used to approximate the constants. Calculus is required for these exercises. ◀

T5. (a) Show that making the substitution $Y = \ln y$ in the equation $y = ae^{bx}$ produces the equation $Y = bx + \ln a$ whose graph in the xY-plane is a line of slope b and Y-intercept $\ln a$.

(b) Part (a) suggests that a curve of the form $y = ae^{bx}$ can be fitted to n data points (x_i, y_i) by letting $Y_i = \ln y_i$, then fitting a straight line to the transformed data points (x_i, Y_i) by least squares to find b and $\ln a$, and then computing a from $\ln a$. Use this method to fit an exponential model to the following data, and graph the curve and data in the same coordinate system.

x	0	1	2	3	4	5	6	7
y	3.9	5.3	7.2	9.6	12	17	23	31

T6. (a) Show that making the substitutions

$$X = \ln x \quad \text{and} \quad Y = \ln y$$

in the equation $y = ax^b$ produces the equation $Y = bX + \ln a$ whose graph in the XY-plane is a line of slope b and Y-intercept $\ln a$.

(b) Part (a) suggest that a curve of the form $y = ax^b$ can be fitted to n data points (x_i, y_i) by letting $X_i = \ln x_i$ and $Y_i = \ln y_i$, then fitting a straight line to the transformed data points (X_i, Y_i) by least squares to find b and $\ln a$, and then computing a from $\ln a$. Use this method to fit a power function model to the following data, and graph the curve and data in the same coordinate system.

x	2	3	4	5	6	7	8	9
y	1.75	1.91	2.03	2.13	2.22	2.30	2.37	2.43

T7. (a) Show that making the substitution $X = \ln x$ in the equation $y = a + b \ln x$ produces the equation $y = a + bX$ whose graph in the Xy-plane is a line of slope b and y-intercept a.

(b) Part (a) suggests that a curve of the form $y = a + b \ln x$ can be fitted to n data points (x_i, y_i) by letting $X_i = \ln x_i$ and then fitting a straight line to the transformed data points (X_i, y_i) by least squares to find b and a. Use this method to fit a logarithmic model to the following data, and graph the curve and data in the same coordinate system.

x	2	3	4	5	6	7	8	9
y	4.07	5.30	6.21	6.79	7.32	7.91	8.23	8.51

6.6 Function Approximation; Fourier Series

In this section we will show how orthogonal projections can be used to approximate certain types of functions by simpler functions. 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:

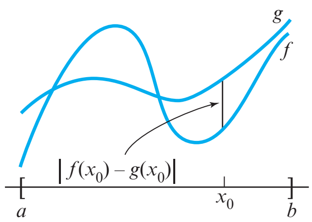
(a) Find the best possible approximation to e^x over $[0, 1]$ by a polynomial of the form $a_0 + a_1x + a_2x^2$.

(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



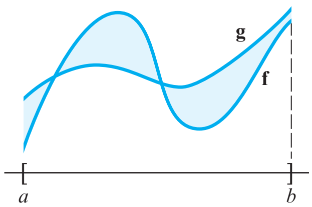
▲ Figure 6.6.1 The deviation between f and g at x_0.

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 difficulty 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

$$\text{error} = \int_a^b |f(x) - g(x)|\,dx \tag{1}$$



▲ 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 $[a, b]$.

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.

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

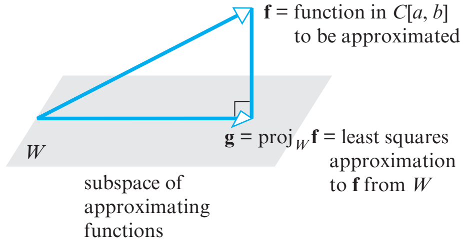
Least Squares Approximation Problem Let **f** be a function that is continuous on an interval $[a, b]$, let $C[a, b]$ have the inner product

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

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 $\mathbf{g}$, this problem is equivalent to looking for a function $\mathbf{g}$ in W that is closest to $\mathbf{f}$. But we know from Theorem 6.4.1 that $\mathbf{g} = \text{proj}_W \mathbf{f}$ is such a function (Figure 6.6.3). Thus, we have the following result.



▶ Figure 6.6.3

THEOREM 6.6.1 *If* $\mathbf{f}$ *is a continuous function on* $[a, b]$, *and* W *is a finite-dimensional subspace of* $C[a, b]$, *then the function* $\mathbf{g}$ *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* ***least squares approximation*** *to* $\mathbf{f}$ *from* W.

Fourier Series

A function of the form

$$\begin{aligned} T(x) = c_0 + c_1 \cos x + c_2 \cos 2x + \cdots + c_n \cos nx \\ + d_1 \sin x + d_2 \sin 2x + \cdots + d_n \sin nx \end{aligned} \tag{2}$$

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

$$c_0 = 2,\quad c_1 = 1,\quad c_2 = -3,\quad c_3 = 0,\quad c_4 = 0,\quad d_1 = 0,\quad d_2 = 0,\quad d_3 = 0,\quad d_4 = 7$$

It is evident from (2) that the trigonometric polynomials of order n or less are the various possible linear combinations of

$$1,\ \cos x,\ \cos 2x, \ldots,\ \cos nx, \qquad \sin x,\ \sin 2x, \ldots,\ \sin nx \tag{3}$$

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 $\mathbf{f}$ from W is the orthogonal projection of $\mathbf{f}$ 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

$$\begin{gathered} \mathbf{g}_0 = \frac{1}{\sqrt{2\pi}}, \quad \mathbf{g}_1 = \frac{1}{\sqrt{\pi}}\cos x, \ldots, \quad \mathbf{g}_n = \frac{1}{\sqrt{\pi}}\cos nx, \\ \mathbf{g}_{n+1} = \frac{1}{\sqrt{\pi}}\sin x, \ldots, \quad \mathbf{g}_{2n} = \frac{1}{\sqrt{\pi}}\sin nx \end{gathered} \tag{5}$$

(see Exercise 6). If we introduce the notation

$$\begin{gathered} a_0 = \frac{2}{\sqrt{2\pi}}\langle \mathbf{f}, \mathbf{g}_0\rangle, \quad a_1 = \frac{1}{\sqrt{\pi}}\langle \mathbf{f}, \mathbf{g}_1\rangle, \ldots, \quad a_n = \frac{1}{\sqrt{\pi}}\langle \mathbf{f}, \mathbf{g}_n\rangle \\ b_1 = \frac{1}{\sqrt{\pi}}\langle \mathbf{f}, \mathbf{g}_{n+1}\rangle, \ldots, \quad b_n = \frac{1}{\sqrt{\pi}}\langle \mathbf{f}, \mathbf{g}_{2n}\rangle \end{gathered} \tag{6}$$

then on substituting (5) in (4), we obtain

$$\operatorname{proj}_W \mathbf{f} = \frac{a_0}{2} + [a_1\cos x + \cdots + a_n \cos nx] + [b_1 \sin x + \cdots + b_n \sin nx] \tag{7}$$

where

$$\begin{aligned}
a_0 &= \frac{2}{\sqrt{2\pi}}\langle \mathbf{f}, \mathbf{g}_0\rangle = \frac{2}{\sqrt{2\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{2\pi}}\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\,dx \\
a_1 &= \frac{1}{\sqrt{\pi}}\langle \mathbf{f}, \mathbf{g}_1\rangle = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{\pi}}\cos x\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos x\,dx \\
&\;\;\vdots \\
a_n &= \frac{1}{\sqrt{\pi}}\langle \mathbf{f}, \mathbf{g}_n\rangle = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{\pi}}\cos nx\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos nx\,dx \\
b_1 &= \frac{1}{\sqrt{\pi}}\langle \mathbf{f}, \mathbf{g}_{n+1}\rangle = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{\pi}}\sin x\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin x\,dx \\
&\;\;\vdots \\
b_n &= \frac{1}{\sqrt{\pi}}\langle \mathbf{f}, \mathbf{g}_{2n}\rangle = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(x)\frac{1}{\sqrt{\pi}}\sin nx\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin nx\,dx
\end{aligned}$$

In short,

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx\,dx, \quad b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin kx\,dx \tag{8}$$

The numbers $a_0, a_1, \ldots, a_n, b_1, \ldots, b_n$ are called the ***Fourier coefficients*** of $\mathbf{f}$.

▶ 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),

$$x \approx \pi - 2\sin x - \sin 2x$$

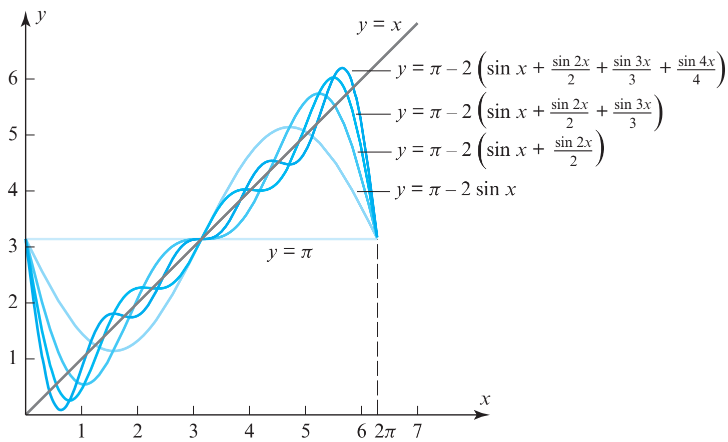
Solution (b) The least squares approximation to x on $[0, 2\pi]$ by a trigonometric polynomial of order n or less is

$$x \approx \frac{a_0}{2} + [a_1\cos x + \cdots + a_n\cos nx] + [b_1\sin x + \cdots + b_n\sin nx]$$

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

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 who made him a baron.

[*Image: Hulton Archive/Getty Images*]

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. ◀

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.

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.

4. (a) Find the least squares approximation of e^x over the interval $[0, 1]$ by a polynomial of the form $a_0 + a_1x$.

(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 $a_0 + a_1x + a_2x^2$.

(b) Find the mean square error of the approximation.

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 $[0, 2\pi]$.

10. What is the Fourier series of $\sin(3x)$?

True-False Exercises

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

(a) If a function $\mathbf{f}$ in $C[a, b]$ is approximated by the function $\mathbf{g}$, then the mean square error is the same as the area between the graphs of $f(x)$ and $g(x)$ over the interval $[a, b]$.

(b) Given a finite-dimensional subspace W of $C[a, b]$, the function $\mathbf{g} = \text{proj}_W \mathbf{f}$ minimizes the mean square error.

(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 $\langle \mathbf{f}, \mathbf{g} \rangle = \int_0^{2\pi} f(x)g(x)\,dx$.

(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 $\langle \mathbf{f}, \mathbf{g} \rangle = \int_0^{2\pi} f(x)g(x)\,dx$.

(e) $\{1, \cos x, \sin x, \cos 2x, \sin 2x\}$ is a linearly independent subset of $C[0, 2\pi]$.

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 $\mathbf{u}_3$.

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 $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v} = \mathbf{v}^T\mathbf{u}$.]

3. Let M_{22} have the inner product $\langle U, V \rangle = \text{tr}(U^TV) = \text{tr}(V^TU)$ 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.

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

$$(a_1 + a_2 + \cdots + a_n)\left(\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n}\right) \ge n^2$$

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 $\mathbf{u}_3 = (-1, 0, 1)$.

8. Find a weighted Euclidean inner product on R^n such that the vectors

$$\begin{aligned} \mathbf{v}_1 &= (1, 0, 0, \ldots, 0) \\ \mathbf{v}_2 &= (0, \sqrt{2}, 0, \ldots, 0) \\ \mathbf{v}_3 &= (0, 0, \sqrt{3}, \ldots, 0) \\ &\vdots \\ \mathbf{v}_n &= (0, 0, 0, \ldots, \sqrt{n}\,) \end{aligned}$$

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.

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

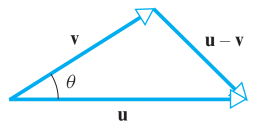


Figure Ex-10

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

$$(k, 0, 0, \ldots, 0), \quad (0, k, 0, \ldots, 0), \ldots, \quad (0, 0, 0, \ldots, k)$$

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 in part (a) as the dimension of R^n approaches ∞?

12. Let $\mathbf{u}$ and $\mathbf{v}$ 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 $\mathbf{u}$ 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 $\mathbf{u}$ 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 $\mathbf{b}$ is orthogonal to the column space of A, then the least squares solution of $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x} = \mathbf{0}$.

17. Is there any value of s for which $x_1 = 1$ and $x_2 = 2$ is the least squares solution of the following linear system?

$$\begin{aligned} x_1 - x_2 &= 1 \\ 2x_1 + 3x_2 &= 1 \\ 4x_1 + 5x_2 &= s \end{aligned}$$

Explain your reasoning.

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

$$\langle \mathbf{f}, \mathbf{g}\rangle = \int_0^{2\pi} f(x)g(x)\,dx$$

20. Let W be the intersection of the planes

$$x + y + z = 0 \quad \text{and} \quad x - y + z = 0$$

in R^3. Find an equation for $W^\perp$.

21. Prove that if $ad - bc \neq 0$, then the matrix

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

has a unique QR-decomposition $A = QR$, where

$$Q = \frac{1}{\sqrt{a^2 + c^2}} \begin{bmatrix} a & -c \\ c & a \end{bmatrix}$$

$$R = \frac{1}{\sqrt{a^2 + c^2}} \begin{bmatrix} a^2 + c^2 & ab + cd \\ 0 & ad - bc \end{bmatrix}$$