Chapter 6

Inner Product Spaces

Most applications of mathematics are involved with the concept of measurement and hence of the magnitude or relative size of various quantities. So it is not surprising that the fields of real and complex numbers, which have a built-in notion of distance, should play a special role. Except for Section 6.7 we assume that all vector spaces are over the field F, where F denotes either R or C. (See Appendix D for properties of complex numbers.)

We introduce the idea of distance or length into vector spaces, obtaining a much richer structure, the so-called *inner product space* structure. This added structure provides applications to geometry (Sections 6.5 and 6.10), physics (Section 6.8), conditioning in systems of linear equations (Section 6.9), least squares applications (Section 6.3), and quadratic forms (Section 6.7).

6.1 INNER PRODUCTS AND NORMS

Many of the geometric notions such as angle, length, and perpendicularity in R^2 and R^3 may be extended to more general real and complex vector spaces. All of these ideas are related to the concept of *inner product.*

Definition. *Let* V *be a vector space over* F. *An* **inner product** *on* V *is a function that assigns to every ordered pair of vectors* x *and* y *in* V *a scalar in* F, *denoted* $\langle x,y\rangle$, *such that for all* x, y, *and* z *in* V *and all* c *in* F *we have*

(a) $\langle x+z,y\rangle = \langle x,y\rangle + \langle z,y\rangle$.
(b) $\langle cx,y\rangle = c\langle x,y\rangle$.
(c) $\overline{\langle x,y\rangle} = \langle y,x\rangle$, *where the bar denotes complex conjugation.*
(d) $\langle x,x\rangle > 0$ *if* $x \neq 0$.

Note that (c) requires that $\langle x,y\rangle = \langle y,x\rangle$ if $F=R$. Conditions (a) and (b) simply require that the inner product be linear in the first component.

It is easily shown that if $a_1, a_2, \ldots, a_n \in F$ and $y, v_1, v_2, \ldots, v_n \in \mathsf{V}$, then

$$\left\langle \sum_{i=1}^{n} a_i v_i, y \right\rangle = \sum_{i=1}^{n} a_i \langle v_i, y\rangle .$$

Example 1

For $x=(a_1,\ldots,a_n)$ and $y=(b_1,\ldots,b_n)$ in F^n, define

$$\langle x,y\rangle = \sum_{i=1}^{n} a_i\overline{b_i}.$$

The verification that $\langle\cdot,\cdot\rangle$ satisfies conditions (a) through (d) is easy. For example, if $z=(c_1,\ldots,c_n)$, we have for (a)

$$\begin{aligned}\langle x+z,y\rangle &= \sum_{i=1}^{n}(a_i+c_i)\overline{b_i} = \sum_{i=1}^{n} a_i\overline{b_i} + \sum_{i=1}^{n} c_i\overline{b_i}\\ &= \langle x,y\rangle + \langle z,y\rangle .\end{aligned}$$

Thus for $x=(1+i,4)$ and $y=(2-3i,4+5i)$ in C^2 we have

$$\langle x,y\rangle = (1+i)(2+3i)+4(4-5i) = 15-15i. \quad ∎$$

The inner product in Example 1 is called the **standard inner product** on F^n. (In elementary courses in linear algebra, this inner product is called the **dot product**).

Example 2

If $\langle x,y\rangle$ is any inner product on a vector space V and $r>0$, we may define another inner product by the rule $\langle x,y\rangle' = r\langle x,y\rangle$. If $r<0$, then (d) would not hold. ∎

Example 3

Let $\mathsf{V} = \mathsf{C}([0,1])$, the vector space of real-valued continuous functions on $[0,1]$. For $f,g\in\mathsf{V}$, define $\langle f,g\rangle = \int_0^1 f(t)g(t)\,dt$. Since the integral above is linear in f, (a) and (b) are immediate, and (c) is trivial. If $f\neq 0$, then f^2 is bounded away from zero on some subinterval of $[0,1]$ (continuity is used here), and hence $\langle f,f\rangle = \int_0^1 [f(t)]^2\,dt > 0$. ∎

Definition. *Let* $A\in\mathsf{M}_{m\times n}(F)$. *We define the* **conjugate transpose** *or* **adjoint** *of* A *to be the* $n\times m$ *matrix* A^* *such that* $(A^*)_{ij} = \overline{A_{ji}}$ *for all* i,j.

Example 4

Let

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

Then

$$A^* = \begin{pmatrix} -i & 2\\ 1-2i & 3-4i\end{pmatrix}. \quad ∎$$

Notice that if x and y are viewed as column vectors, then $\langle x, y\rangle = y^*x$.

The conjugate transpose of a matrix plays a very important role in the remainder of this chapter. Note that if A has real entries, then A^* is simply the transpose of A.

Example 5

Let $\mathsf{V} = \mathsf{M}_{n\times n}(F)$, and define $\langle A, B\rangle = \operatorname{tr}(B^*A)$ for $A, B \in \mathsf{V}$. (Recall that the trace of a matrix A is defined by $\operatorname{tr} A = \sum_{i=1}^{n} A_{ii}$). We verify that (a) and (d) of the definition of inner product hold and leave (b) and (c) to the reader. For this purpose, let $A, B, C \in \mathsf{V}$. Then (using Exercise 6 of Section 1.3)

$$\begin{aligned}\langle A+B, C\rangle &= \operatorname{tr}(C^*(A+B)) = \operatorname{tr}(C^*A + C^*B)\\ &= \operatorname{tr}(C^*A) + \operatorname{tr}(C^*B) = \langle A, C\rangle + \langle B, C\rangle.\end{aligned}$$

Also

$$\begin{aligned}\langle A, A\rangle = \operatorname{tr}(A^*A) &= \sum_{i=1}^{n}(A^*A)_{ii} = \sum_{i=1}^{n}\sum_{k=1}^{n}(A^*)_{ik}A_{ki}\\ &= \sum_{i=1}^{n}\sum_{k=1}^{n}\overline{A_{ki}}A_{ki} = \sum_{i=1}^{n}\sum_{k=1}^{n}|A_{ki}|^2.\end{aligned}$$

Now if $A \neq O$, then $A_{ki} \neq 0$ for some k and i. So $\langle A, A\rangle > 0$. ∎

A vector space V over F endowed with a specific inner product is called an **inner product space**. If $F = C$, we call V a **complex inner product space**, whereas if $F = R$, we call V a **real inner product space**.

Thus Examples 1, 3, and 5 also provide examples of inner product spaces. *For the remainder of this chapter* F^n *denotes the inner product space with the standard inner product as defined in Example 1.* The reader is cautioned that two distinct inner products on a given vector space yield two distinct inner product spaces.

A very important inner product space that resembles $\mathsf{C}([0,1])$ is the space H of continuous complex-valued functions defined on the interval $[0, 2\pi]$ with the inner product

$$\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(t)\overline{g(t)}\,dt.$$

The reason for the constant $1/2\pi$ will become evident later. This inner product space, which arises often in the context of physical situations, is examined more closely in later sections.

At this point we mention a few facts about integration of complex-valued functions. First, the imaginary number i can be treated as a constant under the integration sign. Second, every complex-valued function f may be written

as $f = f_1 + if_2$, where f_1 and f_2 are real-valued functions. Thus we have

$$\int f = \int f_1 + i\int f_2 \quad \text{and} \quad \overline{\int f} = \int \overline{f}.$$

From these properties, as well as the assumption of continuity, it follows that H is an inner product space (see Exercise 16(a)).

Some properties that follow easily from the definition of an inner product are contained in the next theorem.

Theorem 6.1. *Let* V *be an inner product space. Then for* $x, y, z \in \mathsf{V}$ *and* $c \in F$

(a) $\langle x, y+z\rangle = \langle x, y\rangle + \langle x, z\rangle$
(b) $\langle x, cy\rangle = \bar{c}\,\langle x, y\rangle$
(c) $\langle x, x\rangle = 0$ *if and only if* $x = 0$
(d) *if* $\langle x, y\rangle = \langle x, z\rangle$ *for all* $x \in \mathsf{V}$, *then* $y = z$.

Proof. (a)

$$\begin{aligned}\langle x, y+z\rangle &= \overline{\langle y+z, x\rangle} = \overline{\langle y, x\rangle + \langle z, x\rangle}\\ &= \overline{\langle y, x\rangle} + \overline{\langle z, x\rangle} = \langle x, y\rangle + \langle x, z\rangle.\end{aligned}$$

The proofs of (b), (c), and (d) are left as exercises. ∎

The reader should observe that (a) and (b) of Theorem 6.1 show that the inner product is **conjugate linear** in the second component.

In order to generalize the notion of length in R^3 to arbitrary inner product spaces, we need only observe that the length of $x = (a, b, c) \in \mathsf{R}^3$ is given by $\sqrt{a^2+b^2+c^2} = \sqrt{\langle x, x\rangle}$. This leads to the following definition.

Definition. *Let* V *be an inner product space. For* $x \in \mathsf{V}$ *we define the* ***norm*** *or* ***length*** *of* x *by* $\|x\| = \sqrt{\langle x, x\rangle}$.

Example 6

Let $\mathsf{V} = \mathsf{F}^n$. Then

$$\|(a_1, \ldots, a_n)\| = \left[\sum_{i=1}^{n}|a_i|^2\right]^{1/2}$$

is the Euclidean definition of length. Note that if $n = 1$, we have $\|a\| = |a|$. ∎

As we might expect, the well-known properties of length in R^3 hold in general, as shown below.

Theorem 6.2. *Let* V *be an inner product space over* F. *Then for all* $x, y \in \mathsf{V}$ *and* $c \in F$ *the following are true.*

(a) $\|cx\| = |c| \cdot \|x\|$
(b) $\|x\| = 0$ *if and only if* $x = 0$. *In any case,* $\|x\| \geq 0$
(c) *(Cauchy-Schwarz Inequality)* $|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$
(d) *(Triangle Inequality)* $\|x + y\| \leq \|x\| + \|y\|$.

Proof. We leave the proofs of (a) and (b) as exercises.

(c) If $y = 0$, then the result is immediate. So assume that $y \neq 0$. Then for any $c \in F$, we have

$$\begin{aligned} 0 \leq \|x - cy\|^2 &= \langle x - cy, x - cy\rangle = \langle x, x - cy\rangle - c\langle y, x - cy\rangle \\ &= \langle x, x\rangle - \bar{c}\langle x, y\rangle - c\langle y, x\rangle + c\bar{c}\langle y, y\rangle. \end{aligned}$$

Setting

$$c = \frac{\langle x, y\rangle}{\langle y, y\rangle},$$

the inequality above becomes

$$0 \leq \langle x, x\rangle - \frac{|\langle x, y\rangle|^2}{\langle y, y\rangle} = \|x\|^2 - \frac{|\langle x, y\rangle|^2}{\|y\|^2},$$

from which (c) follows.

$$\begin{aligned} \text{(d) } \|x + y\|^2 &= \langle x + y, x + y\rangle = \langle x, x\rangle + \langle y, x\rangle + \langle x, y\rangle + \langle y, y\rangle \\ &= \|x\|^2 + 2\Re\langle x, y\rangle + \|y\|^2 \\ &\leq \|x\|^2 + 2|\langle x, y\rangle| + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\| \cdot \|y\| + \|y\|^2 \\ &= (\|x\| + \|y\|)^2, \end{aligned}$$

where $\Re\langle x, y\rangle$ denotes the real part of the complex number $\langle x, y\rangle$. Note that we used (c) to prove (d). ∎

The case when equality results in (c) and (d) is considered in Exercise 15.

Example 7

For F^n we may apply (c) and (d) of Theorem 6.2 to the standard inner product to obtain the following well-known inequalities:

$$\left|\sum_{i=1}^{n} a_i \bar{b}_i\right| \leq \left[\sum_{i=1}^{n} |a_i|^2\right]^{1/2} \left[\sum_{i=1}^{n} |b_i|^2\right]^{1/2}$$

and

$$\left[\sum_{i=1}^{n} |a_i + b_i|^2\right]^{1/2} \leq \left[\sum_{i=1}^{n} |a_i|^2\right]^{1/2} + \left[\sum_{i=1}^{n} |b_i|^2\right]^{1/2}. \quad ∎$$

The reader may recall from earlier courses that for R^3 or R^2 we have that $\langle x, y\rangle = \|x\| \cdot \|y\| \cos\theta$, where θ denotes the angle $(0 \leq \theta \leq \pi)$ between x and y. This equation implies (c) immediately since $|\cos\theta| \leq 1$. Notice also that x and y are perpendicular if and only if $\cos\theta = 0$, that is, if and only if $\langle x, y\rangle = 0$.

We are now at the point where we can generalize the notion of perpendicularity to arbitrary inner product spaces.

Definitions. *Let* V *be an inner product space. Vectors x and y in* V *are* ***orthogonal (perpendicular)*** *if $\langle x, y\rangle = 0$. A subset S of* V *is* ***orthogonal*** *if any two distinct elements of S are orthogonal. A vector x in* V *is a* ***unit vector*** *if $\|x\| = 1$. Finally, a subset S of* V *is* ***orthonormal*** *if S is orthogonal and consists entirely of unit vectors.*

Note that if $S = \{v_1, v_2, \ldots\}$, then S is orthonormal if and only if $\langle v_i, v_j\rangle = \delta_{ij}$, where δ_{ij} denotes the Kronecker delta. Also, observe that dividing vectors by nonzero scalars does not affect their orthogonality and that if x is any nonzero vector, then $(1/\|x\|)x$ is a unit vector.

Example 8

In F^3 the set $\{(1, 1, 0), (1, -1, 1), (-1, 1, 2)\}$ is orthogonal but not orthonormal; however, if we divide each vector by its length, we obtain the orthonormal set

$$\left\{\frac{1}{\sqrt{2}}(1, 1, 0), \frac{1}{\sqrt{3}}(1, -1, 1), \frac{1}{\sqrt{6}}(-1, 1, 2)\right\}. \quad ∎$$

Our next example is of an infinite orthonormal set that is important in analysis. This set is used in later examples in this chapter.

Example 9

Recall the inner product space H (defined on page 318). We introduce an important example of an orthonormal subset S of H that is used frequently in analysis. For what follows i is the imaginary number $\sqrt{-1}$. For any integer j let $f_j(t) = e^{ijt}$, where $0 \leq t \leq 2\pi$. (Recall that $e^{ijt} = \cos jt + i \sin jt$.) Now define $S = \{f_j : j \text{ is an integer}\}$. Clearly S is a subset of H. Using the property that $\overline{e^{it}} = e^{-it}$ for every real number t, we have for $j \neq k$ that

$$\begin{aligned} \langle f_j, f_k\rangle &= \frac{1}{2\pi}\int_0^{2\pi} e^{ijt}\overline{e^{ikt}}\, dt = \frac{1}{2\pi}\int_0^{2\pi} e^{i(j-k)t}\, dt \\ &= \frac{1}{2\pi(j-k)} e^{i(j-k)t}\Big|_0^{2\pi} = 0. \end{aligned}$$

Also,

$$\langle f_j, f_j\rangle = \frac{1}{2\pi}\int_0^{2\pi} e^{i(j-j)t}\, dt = \frac{1}{2\pi}\int_0^{2\pi} 1\, dt = 1.$$

In other words, $\langle f_j, f_k\rangle = \delta_{jk}$. ∎

EXERCISES

1. Label the following statements as being true or false.
   (a) An inner product is a scalar-valued function on the set of ordered pairs of vectors.
   (b) An inner product space must be over the field of real or complex numbers.
   (c) An inner product is linear in both components.
   (d) There is exactly one inner product on the vector space R^n.
   (e) The triangle inequality only holds in finite-dimensional inner product spaces.
   (f) Only square matrices have a conjugate-transpose.
   (g) If x, y, and z are vectors in an inner product space such that $\langle x, y\rangle = \langle x, z\rangle$, then $y = z$.
   (h) If $\langle x, y\rangle = 0$ for all x in an inner product space, then $y = 0$.

2. Let $V = C^3$ with the standard inner product. Let $x = (2, 1+i, i)$ and $y = (2-i, 2, 1+2i)$. Compute $\langle x, y\rangle$, $\|x\|$, $\|y\|$, and $\|x+y\|$. Then verify both Cauchy's inequality and the triangle inequality.

3. In $C([0,1])$, let $f(t) = t$ and $g(t) = e^t$. Compute $\langle f, g\rangle$ (as defined in Example 3), $\|f\|$, $\|g\|$, and $\|f+g\|$. Then verify both Cauchy's inequality and the triangle inequality.

4. Let $V = M_{n\times n}(F)$ with $\langle A, B\rangle = \text{tr}(B^*A)$. Complete the proof in Example 5 that $\langle \cdot, \cdot\rangle$ is an inner product on V. If $n = 2$ and

$$A = \begin{pmatrix} 1 & 2+i \\ 3 & i \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1+i & 0 \\ i & -i \end{pmatrix},$$

compute $\|A\|$, $\|B\|$, and $\langle A, B\rangle$.

5. On C^2, show that $\langle x, y\rangle = xAy^*$ is an inner product, where

$$A = \begin{pmatrix} 1 & i \\ -i & 2 \end{pmatrix}.$$

Compute $\langle x, y\rangle$ for $x = (1-i, 2+3i)$ and $y = (2+i, 3-2i)$.

6. Complete the proof of Theorem 6.1.

7. Complete the proof of Theorem 6.2.

8. Provide reasons why each of the following is not an inner product on the given vector spaces.
   (a) $\langle (a,b), (c,d)\rangle = ac - bd$ on R^2.

   (b) $\langle A, B\rangle = \text{tr}(A+B)$ on $M_{2\times 2}(R)$.
   (c) $\langle f, g\rangle = \int_0^1 f'(t)g(t)\,dt$ on $P(R)$, where $'$ denotes differentiation.

9. Let β be a basis for a finite-dimensional inner product space. Prove that if $\langle x, y\rangle = 0$ for all $x \in \beta$, then $y = 0$.

10.† Let V be an inner product space, and suppose that x and y are orthogonal elements of V. Prove that $\|x+y\|^2 = \|x\|^2 + \|y\|^2$. Deduce the Pythagorean theorem in R^2.

11. Prove the *parallelogram law* on an inner product space V; that is, show that

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2 \quad \text{for all } x, y \in V.$$

What does this equation state about parallelograms in R^2?

12.† Let $\{v_1, \ldots, v_k\}$ be an orthogonal set in V, and let $a_1, \ldots, a_k$ be scalars. Prove that

$$\left\| \sum_{i=1}^{k} a_i v_i \right\|^2 = \sum_{i=1}^{k} |a_i|^2 \|v_i\|^2.$$

13. Suppose that $\langle \cdot, \cdot\rangle_1$ and $\langle \cdot, \cdot\rangle_2$ are two inner products on a vector space V. Prove that $\langle \cdot, \cdot\rangle = \langle \cdot, \cdot\rangle_1 + \langle \cdot, \cdot\rangle_2$ is another inner product on V.

14. Let A and B be $n \times n$ matrices, and let c be a scalar. Prove that $(A + cB)^* = A^* + \bar{c}B^*$.

15. (a) Prove that if V is an inner product space, then $|\langle x, y\rangle| = \|x\| \cdot \|y\|$ if and only if one of the vectors x or y is a multiple of the other. *Hint:* If $y \neq 0$, let

$$a = \frac{\langle x, y\rangle}{\|y\|^2}.$$

Then $x = ay + z$, where $\langle y, z\rangle = 0$. By assumption

$$|a| = \frac{\|x\|}{\|y\|}.$$

Apply Exercise 10 to $\|x\|^2 = \|ay + z\|^2$ and obtain $\|z\| = 0$.
   (b) Derive a similar result for the equality $\|x+y\| = \|x\| + \|y\|$, and generalize it to the case of n vectors.

16. (a) Show that the vector space H defined in this section is an inner product space.

(b) Let $\mathsf{V} = \mathsf{C}([0,1])$, and define

$$\langle f, g\rangle = \int_0^{1/2} f(t)g(t)\,dt.$$

Is this an inner product on V?

17. Let T be a linear operator on an inner product space V, and suppose that $\|\mathsf{T}(x)\| = \|x\|$ for all x. Prove that T is one-to-one.

18. Let V be a vector space over F, where $F = R$ or $F = C$, and let W be an inner product space over F with inner product $\langle\cdot,\cdot\rangle$. If $\mathsf{T}\colon \mathsf{V} \to \mathsf{W}$ is linear, prove that $\langle x, y\rangle' = \langle \mathsf{T}(x), \mathsf{T}(y)\rangle$ defines an inner product on V if and only if T is one-to-one.

19. Let V be an inner product space. Prove that

(a) $\|x \pm y\|^2 = \|x\|^2 \pm 2\Re\langle x, y\rangle + \|y\|^2$ for all $x, y \in \mathsf{V}$, where $\Re\langle x, y\rangle$ denotes the real part of the complex number $\langle x, y\rangle$.

(b) $|\,\|x\| - \|y\|\,| \le \|x - y\|$ for all $x, y \in \mathsf{V}$.

20. Let V be an inner product space over F. Prove the *polar identities:* For all $x, y \in \mathsf{V}$

(a) $\langle x, y\rangle = \frac{1}{4}\|x+y\|^2 - \frac{1}{4}\|x-y\|^2$ if $F = R$;

(b) $\langle x, y\rangle = \frac{1}{4}\sum_{k=1}^{4} i^k \|x + i^k y\|^2$ if $F = C$, where $i = \sqrt{-1}$.

21. Let A be an $n \times n$ matrix. Define

$$A_1 = \frac{1}{2}(A + A^*) \quad\text{and}\quad A_2 = \frac{1}{2i}(A - A^*).$$

(a) Prove that $A_1^* = A_1$, $A_2^* = A_2$, and $A = A_1 + iA_2$. Would it be reasonable to define A_1 and A_2 to be the real and imaginary parts, respectively, of the matrix A?

(b) Let A be an $n \times n$ matrix. Prove that if $A = B_1 + iB_2$, where $B_1^* = B_1$ and $B_2^* = B_2$, then $B_1 = A_1$ and $B_2 = A_2$.

22. Let V be a vector space over F, where F is either R or C. Whether or not V is an inner product space, we may still define a *norm* $\|\cdot\|$ as a real-valued function on V satisfying the following three conditions for all $x, y \in \mathsf{V}$ and $a \in F$.

(1) $\|x\| \ge 0$, and $\|x\| = 0$ if and only if $x = 0$.

(2) $\|ax\| = |a|\cdot\|x\|$.

(3) $\|x + y\| \le \|x\| + \|y\|$.

Prove that the following are norms on the given vector spaces V.

(a) $\mathsf{V} = \mathsf{M}_{m\times n}(F)$; $\|A\| = \max_{i,j} |A_{ij}|$ for all $A \in \mathsf{V}$

(b) $\mathsf{V} = \mathsf{C}([0,1])$; $\|f\| = \max_{t\in[0,1]} |f(t)|$ for all $f \in \mathsf{V}$

(c) $\mathsf{V} = \mathsf{C}([0,1])$; $\|f\| = \int_0^1 |f(t)|\,dt$ for all $f \in \mathsf{V}$

(d) $\mathsf{V} = \mathsf{R}^2$; $\|(a,b)\| = \max\{|a|, |b|\}$ for all $(a,b) \in \mathsf{V}$

Use Exercise 20 to show that there is no inner product $\langle\cdot,\cdot\rangle$ on R^2 such that $\|x\|^2 = \langle x, x\rangle$ for all $x \in \mathsf{R}^2$ if the norm is defined as in (d).

23. Let V be an inner product space, and define for each ordered pair of vectors the scalar $d(x,y) = \|x - y\|$, called the **distance** between x and y. Prove the following for all $x, y, z \in \mathsf{V}$.

(a) $d(x,y) \ge 0$.

(b) $d(x,y) = d(y,x)$.

(c) $d(x,y) \le d(x,z) + d(z,y)$.

(d) $d(x,x) = 0$.

(e) $d(x,y) \ne 0$ if $x \ne y$.

24. Let V be a real or complex vector space (possibly infinite-dimensional), and let β be a basis for V. For $x, y \in \mathsf{V}$ there exist $v_1, \ldots, v_n \in \beta$ such that

$$x = \sum_{i=1}^{n} a_i v_i \quad\text{and}\quad y = \sum_{i=1}^{n} b_i v_i.$$

Define

$$\langle x, y\rangle = \sum_{i=1}^{n} a_i \bar{b}_i.$$

(a) Prove that $\langle\cdot,\cdot\rangle$ is an inner product on V and that β is an orthonormal basis for V. Thus every real or complex vector space may be regarded as an inner product space.

(b) Prove that if $\mathsf{V} = \mathsf{R}^n$ or $\mathsf{V} = \mathsf{C}^n$ and β is the standard ordered basis, then the inner product defined above is the standard inner product.

25. Let $\|\cdot\|$ be a norm (as defined in Exercise 22) on a real vector space V satisfying the parallelogram law given in Exercise 11. Define

$$\langle x, y\rangle = \frac{1}{4}\left[\|x+y\|^2 - \|x-y\|^2\right].$$

Prove that $\langle\cdot,\cdot\rangle$ defines an inner product on V such that $\|x\|^2 = \langle x, x\rangle$ for all $x \in \mathsf{V}$.

26. Let $\|\cdot\|$ be a norm (as defined in Exercise 22) on a complex vector space V satisfying

$$\sum_{k=1}^{4} \|x + i^k y\|^2 = 4\left[\|x\|^2 + \|y\|^2\right],$$

where $i = \sqrt{-1}$. Define

$$\langle x, y\rangle = \frac{1}{4}\sum_{k=1}^{4} i^k \|x + i^k y\|^2.$$

Prove that $\langle \cdot, \cdot\rangle$ defines an inner product on V such that $\|x\|^2 = \langle x, x\rangle$ for all $x \in$ V.

27. Prove the converse of Exercise 26: Suppose that V is a complex inner product space. Then the corresponding norm satisfies the first equation of Exercise 26. (Note that in a similar manner Exercise 11 is the converse of Exercise 25.)

6.2 THE GRAM-SCHMIDT ORTHOGONALIZATION PROCESS AND ORTHOGONAL COMPLEMENTS

In previous chapters we have seen the special role of the standard ordered bases for C^n and R^n. The special properties of these bases stem from the fact that the basis vectors form an orthonormal set. Just as bases are the building blocks of vector spaces, bases that are also orthonormal sets are the building blocks of inner product spaces. We now name such bases.

Definition. *Let V be an inner product space. A subset of V is an* ***orthonormal basis*** *for V if it is an ordered basis that is orthonormal.*

Example 1

The standard ordered basis for F^n is an orthonormal basis for F^n. ∎

Example 2

The set

$$\left\{\left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right), \left(\frac{2}{\sqrt{5}}, \frac{-1}{\sqrt{5}}\right)\right\}$$

is an orthonormal basis for R^2. ∎

The next theorem and its corollaries illustrate why orthonormal sets and, in particular, orthonormal bases are so important.

Theorem 6.3. *Let V be an inner product space, and let $S = \{v_1, \ldots, v_k\}$ be an orthogonal set of nonzero vectors. If*

$$y = \sum_{i=1}^{k} a_i v_i,$$

then $a_j = \langle y, v_j\rangle / \|v_j\|^2$ for all j.

Proof. For $1 \le j \le k$, we have

$$\langle y, v_j\rangle = \left\langle \sum_{i=1}^{k} a_i v_i, v_j \right\rangle = \sum_{i=1}^{k} a_i \langle v_i, v_j\rangle = a_j \langle v_j, v_j\rangle$$
$$= a_j \|v_j\|^2. \quad ∎$$

The next corollary follows immediately from Theorem 6.3.

Corollary 1. *If, in addition to the hypotheses of Theorem 6.3, S is orthonormal, then*

$$y = \sum_{i=1}^{k} \langle y, v_i\rangle v_i.$$

If V possesses a finite orthonormal basis, then Corollary 1 allows us to compute the coefficients in a linear combination very easily (see Example 3).

Corollary 2. *Let V be an inner product space, and let S be an orthogonal set of nonzero vectors. Then S is linearly independent.*

Proof. Suppose that $v_1, \ldots, v_k \in S$ and

$$\sum_{i=1}^{k} a_i v_i = 0.$$

By Theorem 6.3, $a_j = \langle 0, v_j\rangle / \|v_j\|^2 = 0$ for all j. So S is linearly independent. ∎

Example 3

By Corollary 2, the orthonormal set

$$\left\{\frac{1}{\sqrt{2}}(1,1,0), \frac{1}{\sqrt{3}}(1,-1,1), \frac{1}{\sqrt{6}}(-1,1,2)\right\}$$

obtained in Example 8 of Section 6.1 is an orthonormal basis for R^3. Let $x = (2,1,3)$. The coefficients given by Corollary 1 of Theorem 6.3 that express x as a linear combination of the basis vectors are

$$a_1 = \frac{1}{\sqrt{2}}(2+1) = \frac{3}{\sqrt{2}}, \quad a_2 = \frac{1}{\sqrt{3}}(2-1+3) = \frac{4}{\sqrt{3}},$$

and

$$a_3 = \frac{1}{\sqrt{6}}(-2+1+6) = \frac{5}{\sqrt{6}}.$$

As a check we have

$$(2,1,3) = \frac{3}{2}(1,1,0) + \frac{4}{3}(1,-1,1) + \frac{5}{6}(-1,1,2). \quad ∎$$

Corollary 2 tells us that the vector space H in Section 6.1 contains an infinite linearly independent set and hence is not a finite-dimensional vector space.

Of course, we have not yet shown that every finite-dimensional inner product space possesses an orthonormal basis. The next theorem takes us most of the way in obtaining this result. It tells us how to construct an orthogonal set from a linearly independent set of vectors in such a way that both sets generate the same subspace.

Before stating this theorem, let us consider a simple case. Suppose that $\{w_1, w_2\}$ is a linearly independent subset of an inner product space (and hence a basis for some two-dimensional subspace). We want to construct an orthogonal set from $\{w_1, w_2\}$ that spans the same subspace. Figure 6.1 suggests that the set $\{v_1, v_2\}$, where $v_1 = w_1$ and $v_2 = w_2 - cw_1$ works if c is properly chosen.

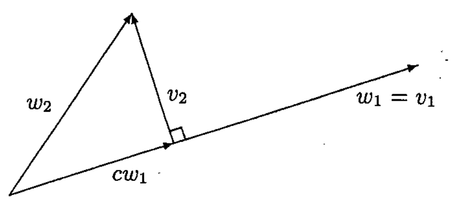

Figure 6.1

To find c, we need only solve the following equation.

$$0 = \langle v_2, w_1\rangle = \langle w_2 - cw_1, w_1\rangle = \langle w_2, w_1\rangle - c\langle w_1, w_1\rangle .$$

So

$$c = \frac{\langle w_2, w_1\rangle}{\|w_1\|^2}.$$

Thus

$$v_2 = w_2 - \frac{\langle w_2, w_1\rangle}{\|w_1\|^2} w_1.$$

This process can be extended to any finite linearly independent subset.

Theorem 6.4. *Let* V *be an inner product space, and let* $S = \{w_1, \ldots, w_n\}$ *be a linearly independent subset of* V*. Define* $S' = \{v_1, \ldots, v_n\}$*, where* $v_1 = w_1$*, and*

$$v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j\rangle}{\|v_j\|^2} v_j \quad \text{for } 2 \le k \le n. \tag{1}$$

Then S' *is an orthogonal set of nonzero vectors such that* $\text{span}(S') = \text{span}(S)$.

Proof. The proof is by induction on n. Let $S_n = \{w_1, \ldots, w_n\}$. If $n = 1$, then the theorem is proved by taking $S_1' = S_1$; i.e., $v_1 = w_1 \neq 0$. Assume then that the set $S_{k-1}' = \{v_1, \ldots, v_{k-1}\}$ with the desired properties has been constructed by the repeated use of (1). We show that the set $S_k' = \{v_1, \ldots, v_{k-1}, v_k\}$ also has the desired properties, where v_k is obtained from S_{k-1}' by (1). If $v_k = 0$, then (1) implies that $w_k \in \text{span}(S_{k-1}') = \text{span}(S_{k-1})$, which contradicts the assumption that S_k is linearly independent. For $1 \le i \le k-1$ we have from (1) that

$$\begin{aligned}\langle v_k, v_i\rangle &= \langle w_k, v_i\rangle - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j\rangle}{\|v_j\|^2}\langle v_j, v_i\rangle \\ &= \langle w_k, v_i\rangle - \frac{\langle w_k, v_i\rangle}{\|v_i\|^2}\|v_i\|^2 = 0,\end{aligned}$$

since $\langle v_j, v_i\rangle = 0$ if $i \neq j$ by the induction assumption that S_{k-1}' is orthogonal. Hence S_k' is an orthogonal set of nonzero vectors. Now by (1) we have that $\text{span}(S_k') \subseteq \text{span}(S_k)$. But by Corollary 2 of Theorem 6.3, S_k' is linearly independent; so $\dim(\text{span}(S_k')) = \dim(\text{span}(S_k)) = k$. Hence $\text{span}(S_k') = \text{span}(S_k)$. ∎

The construction of $\{v_1, \ldots, v_n\}$ by the use of (1) is called the **Gram-Schmidt orthogonalization process.**

Example 4

In R^3, let $w_1 = (1,1,0)$, $w_2 = (2,0,1)$, and $w_3 = (2,2,1)$. Then $\{w_1, w_2, w_3\}$ is linearly independent. We use (1) to compute the orthogonal vectors v_1, v_2, and v_3. Take $v_1 = w_1 = (1,1,0)$. Then $\|v_1\|^2 = 2$; so

$$\begin{aligned}v_2 &= w_2 - \frac{\langle w_2, v_1\rangle}{\|v_1\|^2} v_1 \\ &= (2,0,1) - \frac{2}{2}(1,1,0) \\ &= (1,-1,1).\end{aligned}$$

Finally,

$$\begin{aligned}v_3 &= w_3 - \frac{\langle w_3, v_1\rangle}{\|v_1\|^2} v_1 - \frac{\langle w_3, v_2\rangle}{\|v_2\|^2} v_2 \\ &= (2,2,1) - \frac{4}{2}(1,1,0) - \frac{1}{3}(1,-1,1) \\ &= \left(-\frac{1}{3}, \frac{1}{3}, \frac{2}{3}\right).\end{aligned}$$

∎

Theorem 6.5. *Let V be a nonzero finite-dimensional inner product space. Then V has an orthonormal basis β. Furthermore, if $\beta = \{v_1, v_2, \ldots, v_n\}$ and $x \in V$, then*

$$x = \sum_{i=1}^{n} \langle x, v_i \rangle v_i.$$

Proof. Let β_0 be an ordered basis for V. Apply Theorem 6.4 to obtain an orthogonal set β' of nonzero vectors with $\text{span}(\beta') = \text{span}(\beta_0) = V$. By dividing each vector in β' by its length, we obtain an orthonormal set β that generates V. By Corollary 2 of Theorem 6.3 β is linearly independent, and therefore β is an orthonormal basis for V. The remainder of the theorem follows from Corollary 1 to Theorem 6.3. ∎

We now have an alternate method for computing the matrix representation of a linear operator.

Corollary. *Let V be a finite-dimensional inner product space with an orthonormal basis $\beta = \{v_1, \ldots, v_n\}$. Let T be a linear operator on V, and let $A = [T]_\beta$. Then for any i and j, $A_{ij} = \langle T(v_j), v_i \rangle$.*

Proof. From Theorem 6.5 we have

$$T(v_j) = \sum_{i=1}^{n} \langle T(v_j), v_i \rangle v_i.$$

Hence $A_{ij} = \langle T(v_j), v_i \rangle$. ∎

The scalars $\langle x, v_i \rangle$ associated with x have been studied extensively for special inner product spaces. Although the vectors $v_1, \ldots, v_n$ were chosen from an orthonormal basis, we consider more general sets β for the definition of the scalars $\langle x, v_i \rangle$.

Definition. *Let β be an orthonormal subset (possibly infinite) of an inner product space V, and let $x \in V$. We define the* ***Fourier coefficients of x relative to β*** *to be the scalars $\langle x, y \rangle$, where $y \in \beta$.*

In the nineteenth century the French mathematician Jean Baptiste Fourier was associated with the study of the scalars

$$\int_0^{2\pi} f(t) \sin nt \, dt \quad \text{and} \quad \int_0^{2\pi} f(t) \cos nt \, dt,$$

or more generally,

$$c_n = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-int} \, dt,$$

of a function f. In the context of Example 9 of Section 6.1, we see that $c_n = \langle f, f_n \rangle$, where $f_n(t) = e^{int}$; that is, c_n is the nth Fourier coefficient of a continuous function $f \in V$ relative to S. These coefficients are the "classical" Fourier coefficients of a function, and the literature concerning the behavior of these coefficients is extensive. We will learn more about these Fourier coefficients in the remainder of this chapter.

Example 5

In H define $f(t) = t$. We compute the Fourier coefficients of f relative to the orthonormal set S in Example 9 of Section 6.1. Using integration by parts, we have for $n \neq 0$,

$$\langle f, f_n \rangle = \frac{1}{2\pi} \int_0^{2\pi} t\overline{e^{int}} \, dt = \frac{1}{2\pi} \int_0^{2\pi} te^{-int} \, dt = \frac{-1}{in}.$$

And, for $n = 0$,

$$\langle f, 1 \rangle = \frac{1}{2\pi} \int_0^{2\pi} t(1) \, dt = \pi.$$

As a result of these computations we obtain an upper bound for the sum of a special infinite series by using Exercise 14 as follows:

$$\begin{aligned} \|f\|^2 &\geq \sum_{n=-k}^{-1} |\langle f, f_n \rangle|^2 + |\langle f, 1 \rangle|^2 + \sum_{n=1}^{k} |\langle f, f_n \rangle|^2 \\ &= \sum_{n=-k}^{-1} \frac{1}{n^2} + \pi^2 + \sum_{n=1}^{k} \frac{1}{n^2} \\ &= 2\sum_{n=1}^{k} \frac{1}{n^2} + \pi^2 \end{aligned}$$

for every k. Now, using the fact that $\|f\|^2 = \frac{4}{3}\pi^2$, we obtain

$$\frac{4}{3}\pi^2 \geq 2\sum_{n=1}^{k} \frac{1}{n^2} + \pi^2$$

or

$$\frac{\pi^2}{6} \geq \sum_{n=1}^{k} \frac{1}{n^2}.$$

Because this inequality holds for all k, we may let $k \to \infty$ to obtain

$$\frac{\pi^2}{6} \geq \sum_{n=1}^{\infty} \frac{1}{n^2}.$$

Other results may be produced by replacing f by other functions. ∎

We are now ready to proceed with the concept of an *orthogonal complement.*

Definition. *Let S be a subset of an inner product space V. We define $S^\perp$ (read "S perp") to be the set of all vectors in V that are orthogonal to every vector in S; that is, $S^\perp = \{x \in \mathsf{V} : \langle x, y\rangle = 0 \text{ for all } y \in S\}$. $S^\perp$ is called the* ***orthogonal complement*** *of S.*

It is easily seen that $S^\perp$ is a subspace of V for any subset S of V.

Example 6

The reader should verify that $\{0\}^\perp = \mathsf{V}$ and $\mathsf{V}^\perp = \{0\}$ for any inner product space V. ∎

Example 7

If $\mathsf{V} = \mathsf{R}^3$ and $S = \{x\}$, then $S^\perp$ is simply the set of all vectors that are perpendicular to x (see Exercise 5). ∎

Exercise 16 provides an interesting example of an orthogonal complement in an infinite-dimensional inner product space.

Proposition 6.6. *Let W be a finite-dimensional subspace of an inner product space V, and let $y \in \mathsf{V}$. Then there exist unique vectors $u \in \mathsf{W}$ and $z \in \mathsf{W}^\perp$ such that $y = u + z$. Furthermore, if $\{v_1, \ldots, v_k\}$ is an orthonormal basis of W, then*

$$u = \sum_{i=1}^{k} \langle y, v_i\rangle\, v_i.$$

Proof. Let $\{v_1, \ldots, v_k\}$ be an orthonormal basis for W, let u be as defined in the equation above, and let $z = y - u$. Clearly $u \in \mathsf{W}$.

To show that $z \in \mathsf{W}^\perp$, it suffices to show that z is orthogonal to each v_j. For any j we have

$$\begin{aligned}\langle z, v_j\rangle &= \left\langle \left(y - \sum_{i=1}^{k} \langle y, v_i\rangle\, v_i\right), v_j \right\rangle = \langle y, v_j\rangle - \sum_{i=1}^{k} \langle y, v_i\rangle \langle v_i, v_j\rangle \\ &= \langle y, v_j\rangle - \langle y, v_j\rangle = 0.\end{aligned}$$

To show uniqueness, suppose that $y = u + z = u' + z'$, where $u' \in \mathsf{W}$ and $z' \in \mathsf{W}^\perp$. Then $u - u' = z' - z \in \mathsf{W} \cap \mathsf{W}^\perp = \{0\}$. Therefore, $u = u'$ and $z = z'$. ∎

Corollary. *In the notation of Proposition 6.6, the vector u is the unique vector in W that is "closest" to y; that is, for any $x \in \mathsf{W}$, $\|y - x\| \geq \|y - u\|$, and this inequality is an equality if and only if $x = u$.*

Proof. As in Proposition 6.6 we have that $y = u + z$, where $z \in \mathsf{W}^\perp$. Let $x \in \mathsf{W}$. By Exercise 10 of Section 6.1 we have

$$\begin{aligned}\|y - x\|^2 = \|u + z - x\|^2 &= \|(u - x) + z\|^2 = \|u - x\|^2 + \|z\|^2 \\ &\geq \|z\|^2 = \|y - u\|^2.\end{aligned}$$

Now suppose that $\|y - x\| = \|y - u\|$. Then the inequality above becomes an equality and therefore, $\|u - x\|^2 + \|z\|^2 = \|z\|^2$. It follows that $\|u - x\| = 0$, and hence $x = u$. ∎

The vector u in the corollary is called the **orthogonal projection of y on W**. We will see the importance of orthogonal projections of vectors in the application to least squares in Section 6.3.

Example 8

Let $\mathsf{V} = \mathsf{P}_3(R)$ with the inner product

$$\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx \quad \text{for all } f, g \in \mathsf{V}.$$

Let $\mathsf{W} = \mathsf{P}_1(R) = \operatorname{span}(\{1, x\})$ and $f(x) = x^2$. To compute the orthogonal projection f_1 of f on W, we first apply the Gram-Schmidt process to $\{1, x\}$ and obtain an orthonormal basis $\{g_1, g_2\}$ of W, where

$$g_1(x) = 1 \quad \text{and} \quad g_2(x) = 2\sqrt{3}(x - \tfrac{1}{2}).$$

It is easy to compute $\langle f, g_1\rangle = \frac{1}{3}$ and $\langle f, g_2\rangle = \sqrt{3}/6$. Therefore,

$$f_1(x) = \left(\frac{1}{3}\right)1 + \frac{\sqrt{3}}{6}\left(2\sqrt{3}\left(x - \frac{1}{2}\right)\right) = -\frac{1}{6} + x. \quad ∎$$

It was shown (Corollary 2 to Theorem 1.10) that any linearly independent set in a finite-dimensional vector space can be extended to a basis. The next theorem provides an interesting analog for an orthonormal subset of an inner product space.

Theorem 6.7. *Suppose that $S = \{v_1, \ldots, v_k\}$ is an orthonormal set in an n-dimensional inner product space V. Then*

(a) *S can be extended to an orthonormal basis $\{v_1, \ldots, v_k, v_{k+1}, \ldots, v_n\}$ for V.*

(b) *If $\mathsf{W} = \operatorname{span}(S)$, then (in the notation above) $S_1 = \{v_{k+1}, \ldots, v_n\}$ is an orthonormal basis for $\mathsf{W}^\perp$.*

(c) *If W is any subspace of V, then $\dim(\mathsf{V}) = \dim(\mathsf{W}) + \dim(\mathsf{W}^\perp)$.*

Proof. (a) By Corollary 2 to Theorem 1.10 S can be extended to a basis $S' = \{v_1, \ldots, v_k, w_{k+1}, \ldots, w_n\}$ for V. Now apply the Gram- Schmidt process to S'. By Exercise 7, the first k vectors resulting from this process are the vectors of S. Dividing the last $n-k$ of these vectors by their lengths results in an orthonormal set. The result now follows.

(b) Because S_1 is orthonormal, it is linearly independent by Corollary 2 to Theorem 6.3. Since S_1 is clearly a subset of $\mathsf{W}^\perp$, we need only show that it spans $\mathsf{W}^\perp$. Note that for any $x \in \mathsf{V}$, we have

$$x = \sum_{i=1}^{n} \langle x, v_i \rangle\, v_i.$$

Now if $x \in \mathsf{W}^\perp$, then $\langle x, v_i \rangle = 0$ for $1 \le i \le k$. Therefore,

$$x = \sum_{i=k+1}^{n} \langle x, v_i \rangle\, v_i \in \operatorname{span}(S_1).$$

(c) Let W be a subspace of V with an orthonormal basis $\{v_1, \ldots, v_k\}$. By (a) and (b), we have

$$\dim(\mathsf{V}) = n = k + (n-k) = \dim(\mathsf{W}) + \dim(\mathsf{W}^\perp). \quad ■$$

Example 9

Let $\mathsf{W} = \operatorname{span}(\{e_1, e_2\})$ in F^3. Then $x = (a, b, c) \in \mathsf{W}^\perp$ if and only if $0 = \langle x, e_1 \rangle = a$ and $0 = \langle x, e_2 \rangle = b$. So $x = (0, 0, c)$, and therefore $\mathsf{W}^\perp = \operatorname{span}(\{e_3\})$. One can deduce the same result by noting that $e_3 \in \mathsf{W}^\perp$ and from (c) above that $\dim(\mathsf{W}^\perp) = 3 - 2 = 1$. ■

EXERCISES

1. Label the following statements as being true or false.
   (a) The Gram-Schmidt orthogonalization process allows us to construct an orthonormal set from an arbitrary set of vectors.
   (b) Every finite-dimensional inner product space has an orthonormal basis.
   (c) The orthogonal complement of any set is a subspace.
   (d) If $\{v_1, \ldots, v_n\}$ is a basis for an inner product space V, then for any $x \in \mathsf{V}$ the scalars $\langle x, v_i \rangle$ are the Fourier coefficients of x.
   (e) An orthonormal basis must be an ordered basis.
   (f) Every orthogonal set is linearly independent.
   (g) Every orthonormal set is linearly independent.

2. In each of the following parts, apply the Gram-Schmidt process to the given subset S of the inner product space V. Then find an orthonormal basis β for V and compute the Fourier coefficients of the given vector relative to β. Finally, use Theorem 6.5 to verify your result.
   (a) $\mathsf{V} = \mathsf{R}^3$, $S = \{(1,0,1), (0,1,1), (1,3,3)\}$, and $x = (1,1,2)$.
   (b) $\mathsf{V} = \mathsf{R}^3$, $S = \{(1,1,1), (0,1,1), (0,0,1)\}$, and $x = (1,0,1)$.
   (c) $\mathsf{V} = \mathsf{P}_2(R)$ with the inner product $\langle f, g \rangle = \int_0^1 f(t)g(t)\,dt$, $S = \{1, x, x^2\}$, and $f(x) = 1 + x$.
   (d) $\mathsf{V} = \operatorname{span}(S)$, where $S = \{(1, i, 0), (1-i, 2, 4i)\}$, and $x = (3+i, 4i, -4)$.
3. In R^2 let
$$\beta = \left\{ \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right) \right\}.$$
   Find the Fourier coefficients of $(3, 4)$ relative to β.
4. Let $S = \{(1, 0, i), (1, 2, 1)\}$ in C^3. Compute $S^\perp$.
5. Let $S_0 = \{x_0\}$, where x_0 is a nonzero vector in R^3. Describe $S_0^\perp$ geometrically. Now suppose that $S = \{x_1, x_2\}$ is a linearly independent subset of R^3. Describe $S^\perp$ geometrically.
6. Let V be an inner product space, and let W be a finite-dimensional subspace of V. If $x \notin \mathsf{W}$, prove that there exists $y \in \mathsf{V}$ such that $y \in \mathsf{W}^\perp$ but $\langle x, y \rangle \neq 0$. *Hint:* Use Proposition 6.6.
7. Prove that if $\{w_1, \ldots, w_n\}$ is an orthogonal set of nonzero vectors, then the vectors $v_1, \ldots, v_n$ derived from the Gram-Schmidt process satisfy $v_i = w_i$ for $i = 1, \ldots, n$. *Hint:* Use induction.
8. Let $\mathsf{W} = \operatorname{span}(\{(i, 0, 1)\})$ in C^3 with the standard inner product. Find orthonormal bases for W and $\mathsf{W}^\perp$.
9. Let W be a finite-dimensional subspace of an inner product space V. Prove that there exists a projection T on W along $\mathsf{W}^\perp$ that satisfies $\mathsf{N}(\mathsf{T}) = \mathsf{W}^\perp$. In addition, prove that $\|\mathsf{T}(x)\| \le \|x\|$ for all $x \in \mathsf{V}$. *Hint:* Use Proposition 6.6 and Exercise 10 of Section 6.1. (Projections are defined in the exercises of Section 2.1.)
10. Let A be an $n \times n$ matrix with complex entries. Prove that $AA^* = I$ if and only if the rows of A form an orthonormal basis for C^n.
11. Let W_1 and W_2 be subspaces of a finite-dimensional inner product space. Prove that $(\mathsf{W}_1 + \mathsf{W}_2)^\perp = \mathsf{W}_1^\perp \cap \mathsf{W}_2^\perp$ and $(\mathsf{W}_1 \cap \mathsf{W}_2)^\perp = \mathsf{W}_1^\perp + \mathsf{W}_2^\perp$. (See the definition of the sum of subsets of a vector space on page 21.)
12. Let V be an inner product space, S and S_0 be subsets of V, and W be a finite-dimensional subspace of V. Prove the following.

(a) $S_0 \subseteq S$ implies that $S^\perp \subseteq S_0^\perp$.
(b) $S \subseteq (S^\perp)^\perp$; so $\text{span}(S) \subseteq (S^\perp)^\perp$.
(c) $\mathsf{W} = (\mathsf{W}^\perp)^\perp$. *Hint:* Use Exercise 6.
(d) $\mathsf{V} = \mathsf{W} \oplus \mathsf{W}^\perp$ (see the exercises of Section 1.3).

13. (a) *Parseval's Identity.* Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis for V. For any $x, y \in \mathsf{V}$ prove that

$$\langle x, y\rangle = \sum_{i=1}^{n} \langle x, v_i\rangle \overline{\langle y, v_i\rangle}.$$

(b) Use (a) to prove that if β is an orthonormal basis of a finite-dimensional inner product space V over F with inner product $\langle \cdot, \cdot\rangle$, then for any $x, y \in \mathsf{V}$

$$\langle \phi_\beta(x), \phi_\beta(y)\rangle' = \langle [x]_\beta, [y]_\beta\rangle' = \langle x, y\rangle,$$

where $\langle \cdot, \cdot\rangle'$ is the standard inner product on F^n.

14. (a) *Bessel's Inequality.* Let V be an inner product space, and let $S = \{v_1, \ldots, v_n\}$ be an orthonormal subset of V. Prove that for any $x \in \mathsf{V}$ we have

$$\|x\|^2 \geq \sum_{i=1}^{n} |\langle x, v_i\rangle|^2.$$

Hint: Apply Proposition 6.6 to $x \in \mathsf{V}$ and $\mathsf{W} = \text{span}(S)$. Then use Exercise 10 of Section 6.1.

(b) In the context of (a), prove that Bessel's inequality is an equality if and only if $x \in \text{span}(S)$.

15. Let T be a linear operator on a finite-dimensional inner product space V. If $\langle \mathsf{T}(x), y\rangle = 0$ for all $x, y \in \mathsf{V}$, prove that $\mathsf{T} = \mathsf{T}_0$. In fact, prove this result if the equality holds for all x and y in some basis for V.

16. Let $\mathsf{V} = \mathsf{C}([-1, 1])$. Suppose that W_e and W_o denote the subspaces of V consisting of the even and odd functions, respectively. (See Exercise 22 of Section 1.3.) Prove that $\mathsf{W}_e^\perp = \mathsf{W}_o$, where the inner product on V is defined by

$$\langle f, g\rangle = \int_{-1}^{1} f(t)g(t)\,dt.$$

17. In each of the following parts, find the orthogonal projection of the given vector on the given subspace W of the inner product space V.

(a) $\mathsf{V} = \mathsf{R}^2$, $u = (2, 6)$, $\mathsf{W} = \{(x, y) : y = 4x\}$.
(b) $\mathsf{V} = \mathsf{R}^3$, $u = (2, 1, 3)$, $\mathsf{W} = \{(x, y, z) : x + 3y - 2z = 0\}$.
(c) $\mathsf{V} = \mathsf{P}(R)$ with the inner product $\langle f, g\rangle = \int_0^1 f(t)g(t)\,dt$, $h(x) = 4 + 3x - 2x^2$, $\mathsf{W} = \mathsf{P}_1(R)$.

18. In Exercise 17 find the distance from the given vector to the subspace W.

19. Let $\mathsf{V} = \mathsf{C}([-1, 1])$ with the inner product $\langle f, g\rangle = \int_{-1}^{1} f(t)g(t)\,dt$. Let W be the subspace $\mathsf{P}_2(R)$ with the standard ordered basis β.

(a) Show that the Gram-Schmidt process applied to β yields the *Legendre polynomials* 1, t, and $t^2 - \frac{1}{3}$.
(b) Use (a) to produce an orthonormal basis γ of W.
(c) Let $h(t) = e^t$. Use (b) to compute the "best" (closest) second-degree polynomial approximation of h on the interval $[-1, 1]$.

20. Let V be the vector space defined in Example 5 of Section 1.2, the space of all sequences σ in F ($F = R$ or $F = C$) such that $\sigma(n) \neq 0$ for only finitely many positive integers n. For $\sigma, \mu \in \mathsf{V}$, we define $\langle \sigma, \mu\rangle = \sum_{n=1}^{\infty} \sigma(n)\overline{\mu(n)}$. Since all but a finite number of terms of the series are zero, the series converges.

(a) Prove that $\langle \cdot, \cdot\rangle$ is an inner product on V, and hence V is an inner product space.
(b) For each positive integer n, let e_n be the sequence defined by $e_n(k) = \delta_{n,k}$, where $\delta_{n,k}$ is the Kronecker delta. Prove that $\{e_1, e_2, \ldots\}$ is an orthonormal basis for V.

6.3 THE ADJOINT OF A LINEAR OPERATOR

In Section 6.1 we defined the conjugate transpose A^* of a matrix A. For a linear operator T on an inner product space V, we now define a related linear operator on V called the *adjoint* of T, whose matrix representation with respect to any orthonormal basis β of V is $[\mathsf{T}]_\beta^*$. The analogy between conjugation of complex numbers and adjoints of linear operators will become apparent. We first need a preliminary result.

Let V be an inner product space, and let $y \in \mathsf{V}$. The function $\mathsf{g}\colon \mathsf{V} \to F$ defined by $\mathsf{g}(x) = \langle x, y\rangle$ is clearly linear. More interesting is the fact that if V is finite-dimensional, every linear transformation from V into F is of this form.

Theorem 6.8. *Let V be a finite-dimensional inner product space over F, and let $\mathsf{g}\colon \mathsf{V} \to F$ be a linear transformation. Then there exists a unique vector $y \in \mathsf{V}$ such that $\mathsf{g}(x) = \langle x, y\rangle$ for all $x \in \mathsf{V}$.*

Proof. Let $\beta = \{v_1, \ldots, v_n\}$ be an orthonormal basis for V, and let

$$y = \sum_{i=1}^{n} \overline{g(v_i)} v_i.$$

Let $h\colon V \to F$ be defined by $h(x) = \langle x, y\rangle$, which is clearly linear. Furthermore, for $1 \le j \le n$ we have

$$h(v_j) = \langle v_j, y\rangle = \left\langle v_j, \sum_{i=1}^{n} \overline{g(v_i)} v_i \right\rangle = \sum_{i=1}^{n} g(v_i)\langle v_j, v_i\rangle$$
$$= \sum_{i=1}^{n} g(v_i)\delta_{ji} = g(v_j).$$

Since g and h both agree on β, we have that g = h by the corollary to Theorem 2.6.

To show that y is unique, suppose that $g(x) = \langle x, y'\rangle$ for all x. Then $\langle x, y\rangle = \langle x, y'\rangle$ for all x, so by Theorem 6.1(d) we have $y = y'$. ∎

Example 1

Define $g\colon R^2 \to R$ by $g(a_1, a_2) = 2a_1 + a_2$; clearly g is a linear transformation. Let $\beta = \{e_1, e_2\}$, and let $y = g(e_1)e_1 + g(e_2)e_2 = 2e_1 + e_2 = (2, 1)$ as in the proof of Theorem 6.8. Then $g(a_1, a_2) = \langle (a_1, a_2), (2, 1)\rangle = 2a_1 + a_2$. ∎

Theorem 6.9. *Let V be a finite-dimensional inner product space, and let T be a linear operator on V. Then there exists a unique function $T^*\colon V \to V$ such that $\langle T(x), y\rangle = \langle x, T^*(y)\rangle$ for all $x, y \in V$. Furthermore, T^* is linear.*

Proof. Let $y \in V$. Define $g\colon V \to F$ by $g(x) = \langle T(x), y\rangle$ for all $x \in V$. We first show that g is linear. Let $x_1, x_2 \in V$ and $c \in F$. Then

$$g(cx_1 + x_2) = \langle T(cx_1 + x_2), y\rangle = \langle cT(x_1) + T(x_2), y\rangle$$
$$= c\langle T(x_1), y\rangle + \langle T(x_2), y\rangle = cg(x_1) + g(x_2).$$

Hence g is linear.

We now apply Theorem 6.8 to obtain a unique vector $y' \in V$ such that $g(x) = \langle x, y'\rangle$; that is, $\langle T(x), y\rangle = \langle x, y'\rangle$ for all $x \in V$. Defining $T^*\colon V \to V$ by $T^*(y) = y'$, we have $\langle T(x), y\rangle = \langle x, T^*(y)\rangle$.

To show that T^* is linear, let $y_1, y_2 \in V$ and $c \in F$. Then for any $x \in V$, we have

$$\langle x, T^*(cy_1 + y_2)\rangle = \langle T(x), cy_1 + y_2\rangle$$
$$= \bar{c}\langle T(x), y_1\rangle + \langle T(x), y_2\rangle$$
$$= \bar{c}\langle x, T^*(y_1)\rangle + \langle x, T^*(y_2)\rangle$$
$$= \langle x, cT^*(y_1) + T^*(y_2)\rangle.$$

Since x is arbitrary, we have $T^*(cy_1 + y_2) = cT^*(y_1) + T^*(y_2)$ by Theorem 6.1(d).

Finally we need to show that T^* is unique. Suppose that $U\colon V \to V$ is linear and that it satisfies $\langle T(x), y\rangle = \langle x, U(y)\rangle$ for all $x, y \in V$. Then $\langle x, T^*(y)\rangle = \langle x, U(y)\rangle$ for all $x, y \in V$, so $T^* = U$. ∎

The linear operator T^* described in Theorem 6.9 is called the **adjoint** of the operator T. The symbol T^* is read "T star."

Thus T^* is the unique operator on V satisfying $\langle T(x), y\rangle = \langle x, T^*(y)\rangle$ for all $x, y \in V$. Note that we also have

$$\langle x, T(y)\rangle = \overline{\langle T(y), x\rangle} = \overline{\langle y, T^*(x)\rangle} = \langle T^*(x), y\rangle;$$

so $\langle x, T(y)\rangle = \langle T^*(x), y\rangle$ for all $x, y \in V$. We may view these equations symbolically as adding a * to T when shifting its position inside the inner product symbol.

In the infinite-dimensional case the adjoint of a linear operator T may be defined to be the function T^* such that $\langle T(x), y\rangle = \langle x, T^*(y)\rangle$ for all $x, y \in V$. Although the uniqueness and linearity of T^* follow as before, the existence of the adjoint is not guaranteed (see Exercise 22). The reader should observe the necessity of the hypothesis of finite-dimensionality in the proof of Theorem 6.8. Many of the theorems we prove about adjoints, nevertheless, do not depend on V being finite-dimensional. *Thus for the remainder of this chapter we adopt the convention for the exercises that a reference to the adjoint of a linear operator on an infinite-dimensional inner product space assumes its existence unless stated otherwise.*

A useful result for computing adjoints is Theorem 6.10 below.

Theorem 6.10. *Let V be a finite-dimensional inner product space, and let β be an orthonormal basis for V. If T is a linear operator on V, then*

$$[T^*]_\beta = [T]_\beta^*.$$

Proof. Let $A = [T]_\beta$, $B = [T^*]_\beta$, and $\beta = \{v_1, \ldots, v_n\}$. Then from the corollary to Theorem 6.5 we have

$$B_{ij} = \langle T^*(v_j), v_i\rangle = \overline{\langle v_i, T^*(v_j)\rangle}$$
$$= \overline{\langle T(v_i), v_j\rangle} = \overline{A_{ji}} = (A^*)_{ij}.$$

Hence $B = A^*$. ∎

Corollary. *Let A be an $n \times n$ matrix. Then $L_{A^*} = (L_A)^*$.*

Proof. If β is the standard ordered basis for F^n, then by Theorem 2.16 we have $[L_A]_\beta = A$. Hence $[(L_A)^*]_\beta = [L_A]_\beta^* = A^* = [L_{A^*}]_\beta$, and so $(L_A)^* = L_{A^*}$. ∎

As an application of Theorem 6.10, we compute the adjoint of a specific linear operator.

Example 2

Let T be the linear operator on C^2 defined by $T(a_1, a_2) = (2ia_1 + 3a_2, a_1 - a_2)$. If β is the standard ordered basis for C^2, then

$$[T]_\beta = \begin{pmatrix} 2i & 3 \\ 1 & -1 \end{pmatrix}.$$

So

$$[T^*]_\beta = [T]^*_\beta = \begin{pmatrix} -2i & 1 \\ 3 & -1 \end{pmatrix}.$$

Hence

$$T^*(a_1, a_2) = (-2ia_1 + a_2, 3a_1 - a_2). \quad \blacksquare$$

The following theorem demonstrates the analogy between the conjugates of complex numbers and the adjoints of linear operators.

Theorem 6.11. *Let V be an inner product space, and let T and U be linear operators on V. Then*
(a) $(T + U)^* = T^* + U^*$;
(b) $(cT)^* = \bar{c}T^*$ *for any* $c \in F$;
(c) $(TU)^* = U^*T^*$;
(d) $T^{**} = T$;
(e) $I^* = I$.

Proof. We prove (a) and (d); the rest are proved similarly. Let $x, y \in V$.
(a) Since

$$\begin{aligned}\langle x, (T + U)^*(y)\rangle &= \langle (T + U)(x), y\rangle = \langle T(x) + U(x), y\rangle \\ &= \langle T(x), y\rangle + \langle U(x), y\rangle = \langle x, T^*(y)\rangle + \langle x, U^*(y)\rangle \\ &= \langle x, T^*(y) + U^*(y)\rangle = \langle x, (T^* + U^*)(y)\rangle,\end{aligned}$$

(a) follows.
(d) Similarly, since

$$\langle x, T(y)\rangle = \langle T^*(x), y\rangle = \langle x, T^{**}(y)\rangle,$$

(d) follows. $\blacksquare$

The same proof works in the infinite-dimensional case provided that the existence of T^* and U^* is assumed.

Corollary. *Let A and B be* $n \times n$ *matrices. Then*
(a) $(A + B)^* = A^* + B^*$;

(b) $(cA)^* = \bar{c}A^*$ *for all* $c \in F$;
(c) $(AB)^* = B^*A^*$;
(d) $A^{**} = A$;
(e) $I^* = I$.

Proof. We prove only (c); the remaining parts can be proved similarly.
Since $L_{(AB)^*} = (L_{AB})^* = (L_A L_B)^* = (L_B)^*(L_A)^* = L_{B^*}L_{A^*} = L_{B^*A^*}$, we have $(AB)^* = B^*A^*$. $\blacksquare$

In the proof above we relied on the corollary to Theorem 6.10. An alternative proof that holds even for nonsquare matrices can be given by appealing directly to the definition of the conjugate transposes of the matrices A and B (see Exercise 5).

Least Squares Approximation

Consider the following problem: An experimenter collects data by taking measurements $y_1, y_2, \ldots, y_m$ at times $t_1, t_2, \ldots, t_m$, respectively. For example, he or she may be measuring unemployment at various times during some period. Suppose that the data $(t_1, y_1), \ldots, (t_m, y_m)$ are plotted as points in the plane (see Figure 6.2). From this distribution, the experimenter feels that there exists an essentially linear relationship between y and t, say $y = ct + d$, and would like to find the constants c and d so that the line $y = ct + d$ represents the best possible *fit* to the data collected. One such estimate of fit is to calculate the error E that represents the sum of the squares of the

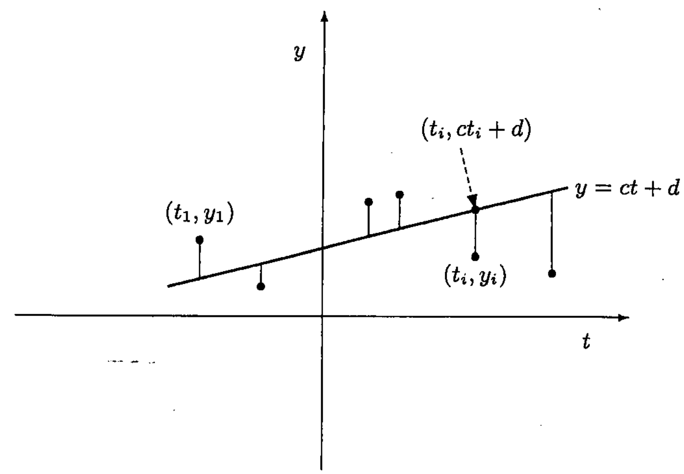

Figure 6.2

vertical distances from the points to the line; that is,

$$E = \sum_{i=1}^{m} (y_i - ct_i - d)^2.$$

Thus the problem is to find the constants c and d that minimize E. (For this reason the line $y = ct + d$ is called the **least squares line.**) If we let

$$A = \begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{pmatrix}, \quad x = \begin{pmatrix} c \\ d \end{pmatrix}, \quad \text{and} \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix},$$

then it follows that $E = \|y - Ax\|^2$.

We now develop a general method for finding an explicit vector $x_0 \in \mathsf{F}^n$ that minimizes E; that is, given an $m \times n$ matrix A, we find $x_0 \in \mathsf{F}^n$ such that $\|y - Ax_0\| \leq \|y - Ax\|$ for all vectors $x \in \mathsf{F}^n$. This method not only allows us to find the linear function that best fits the data but also the polynomial of any fixed degree that best fits the data.

First we need some notation and two simple lemmas. For $x, y \in \mathsf{F}^n$, let $\langle x, y \rangle_n$ denote the standard inner product of x and y in F^n. Notice that if x and y are regarded as column vectors, then $\langle x, y \rangle_n = y^* x$.

Lemma 1. *Let $A \in \mathsf{M}_{m\times n}(F)$, $x \in \mathsf{F}^n$, and $y \in \mathsf{F}^m$. Then*

$$\langle Ax, y \rangle_m = \langle x, A^* y \rangle_n .$$

Proof. By Exercise 5(b) we have

$$\langle Ax, y \rangle_m = y^*(Ax) = (y^* A)x = (A^* y)^* x = \langle x, A^* y \rangle_n . \quad \blacksquare$$

Lemma 2. *Let $A \in \mathsf{M}_{m\times n}(F)$. Then* $\text{rank}(A^*A) = \text{rank}(A)$.

Proof. By the dimension theorem we need only show that for $x \in \mathsf{F}^n$ we have $A^*Ax = 0$ if and only if $Ax = 0$. Clearly, $Ax = 0$ implies that $A^*Ax = 0$. So assume that $A^*Ax = 0$. Then $0 = \langle A^*Ax, x \rangle_n = \langle Ax, A^{**}x \rangle_m = \langle Ax, Ax \rangle_m$, so that $Ax = 0$. $\blacksquare$

Corollary. *If A is an $m \times n$ matrix such that* $\text{rank}(A) = n$, *then A^*A is invertible.*

Now consider the system $Ax = y$, where A is an $m \times n$ matrix and $y \in \mathsf{F}^m$. Define $\mathsf{W} = \{Ax : x \in \mathsf{F}^n\}$; that is, $\mathsf{W} = \mathsf{R}(\mathsf{L}_A)$. By the corollary to Proposition 6.6 there exists a unique vector in W, say Ax_0 where $x_0 \in \mathsf{F}^n$, that is closest to y. So $\|Ax_0 - y\| \leq \|Ax - y\|$ for all $x \in \mathsf{F}^n$.

To develop a practical method for finding such an x_0, we note from Proposition 6.6 and its corollary that $Ax_0 - y \in \mathsf{W}^\perp$; so $\langle Ax, Ax_0 - y \rangle_m = 0$ for all $x \in \mathsf{F}^n$. Thus by Lemma 1 we have that $\langle x, A^*(Ax_0 - y) \rangle_n = 0$ for all $x \in \mathsf{F}^n$; that is, $A^*(Ax_0 - y) = 0$. So we need only find a solution to $A^*Ax = A^*y$. If in addition we assume that $\text{rank}(A) = n$, then by Lemma 2 we have $x_0 = (A^*A)^{-1}A^*y$. We summarize this discussion in the following theorem.

Theorem 6.12. *Let $A \in \mathsf{M}_{m\times n}(F)$ and $y \in \mathsf{F}^m$. Then there exists $x_0 \in \mathsf{F}^n$ such that $(A^*A)x_0 = A^*y$ and $\|Ax_0 - y\| \leq \|Ax - y\|$ for all $x \in \mathsf{F}^n$. Furthermore, if* $\text{rank}(A) = n$, *then $x_0 = (A^*A)^{-1}A^*y$.*

To return to our experimenter, let us suppose that the data collected are $(1, 2), (2, 3), (3, 5)$, and $(4, 7)$. Then

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} \quad \text{and} \quad y = \begin{pmatrix} 2 \\ 3 \\ 5 \\ 7 \end{pmatrix};$$

hence

$$A^*A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} 30 & 10 \\ 10 & 4 \end{pmatrix}.$$

Thus

$$(A^*A)^{-1} = \frac{1}{20} \begin{pmatrix} 4 & -10 \\ -10 & 30 \end{pmatrix}.$$

Therefore

$$\begin{pmatrix} c \\ d \end{pmatrix} = x_0 = \frac{1}{20} \begin{pmatrix} 4 & -10 \\ -10 & 30 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 5 \\ 7 \end{pmatrix} = \begin{pmatrix} 1.7 \\ 0 \end{pmatrix}.$$

Thus the line $y = 1.7t$ is the least squares line. The error E may be computed directly as $\|Ax_0 - y\|^2 = 0.3$.

The method above may also be applied if the experimenter wants to fit a parabola $y = ct^2 + dt + e$ to the data. In this case, the appropriate matrix is

$$A = \begin{pmatrix} t_1^2 & t_1 & 1 \\ \vdots & \vdots & \vdots \\ t_m^2 & t_m & 1 \end{pmatrix}.$$

Finally, suppose in the linear case that the experimenter chose the times t_i $(1 \le i \le m)$ to satisfy

$$\sum_{i=1}^{m} t_i = 0.$$

Then the two columns of A would be orthogonal, so A^*A would be a diagonal matrix (see Exercise 17). In this case the computations are greatly simplified.

Minimal Solutions

In the context of the preceding discussion we showed that if $\text{rank}(A) = n$, then there exists a unique $x_0 \in \mathsf{F}^n$ such that Ax_0 is the point in W that is closest to y. Of course, if $\text{rank}(A) < n$, there will be infinitely many such vectors. It is often desirable to find such a vector of minimal norm. For what follows, we let $b = Ax_0$ as above. Then the system $Ax = b$ has at least one solution. A solution s is called a **minimal solution** if $\|s\| \le \|u\|$ for all other solutions u to $Ax = b$.

Theorem 6.13. *Let $A \in \mathsf{M}_{m\times n}(F)$ and $b \in \mathsf{F}^m$. Suppose that $Ax = b$ has at least one solution. Then the following are true.*

(a) *There exists exactly one minimal solution s of $Ax = b$, and $s \in \mathsf{R}(\mathsf{L}_{A^*})$.*
(b) *The vector s is the only solution to $Ax = b$ that lies in $\mathsf{R}(\mathsf{L}_{A^*})$; that is, if u satisfies $(AA^*)u = b$, then $s = A^*u$.*

Proof. (a) For simplicity of notation, we let $\mathsf{W} = \mathsf{R}(\mathsf{L}_{A^*})$ and $\mathsf{W}' = \mathsf{N}(\mathsf{L}_A)$. Let x be any solution to $Ax = b$. By Proposition 6.6 $x = s + y$ for some $s \in \mathsf{W}$ and $y \in \mathsf{W}^\perp$. But $\mathsf{W}^\perp = \mathsf{W}'$ by Exercise 12, and therefore, $b = Ax = As + Ay = As$. So s is a solution to $Ax = b$ that lies in W. To prove (a), we need only show that s is the unique minimal solution. Let v be any solution to $Ax = b$. By Theorem 3.9 we have that $v = s + u$, where $u \in \mathsf{W}'$. Since $s \in \mathsf{W}$, which equals $\mathsf{W}'^\perp$ by Exercise 12, we have by Exercise 10 of Section 6.1 that

$$\|v\|^2 = \|s+u\|^2 = \|s\|^2 + \|u\|^2 \ge \|s\|^2.$$

Thus s is a minimal solution. We can also see from the calculation above that if $\|u\| = \|s\|$, then $u = 0$ and $v = s$. Hence s is the unique minimal solution to $Ax = b$, proving (a).

(b) Assume that v is also a solution to $Ax = b$ that lies in W. Then

$$v - s \in \mathsf{W} \cap \mathsf{W}' = \mathsf{W} \cap \mathsf{W}^\perp = \{0\};$$

so $v = s$.

Finally, suppose that $(AA^*)u = b$, and let $v = A^*u$. Then $v \in \mathsf{W}$ and $Av = b$. Therefore, $s = v = A^*u$ by the discussion above. ∎

Example 3

Consider the system

$$\begin{aligned} x + 2y + z &= 4 \\ x - y + 2z &= -11 \\ x + 5y &= 19. \end{aligned}$$

Let

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & -1 & 2 \\ 1 & 5 & 0 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 4 \\ -11 \\ 19 \end{pmatrix}.$$

To find the minimal solution to this system, we must find a solution to $AA^*x = b$. Now

$$AA^* = \begin{pmatrix} 6 & 1 & 11 \\ 1 & 6 & -4 \\ 11 & -4 & 26 \end{pmatrix};$$

so we consider the system

$$\begin{aligned} 6x + y + 11z &= 4 \\ x + 6y - 4z &= -11 \\ 11x - 4y + 26z &= 19, \end{aligned}$$

for which a solution is

$$u = \begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}.$$

(Any solution will suffice.) Hence

$$s = A^*u = \begin{pmatrix} -1 \\ 4 \\ -3 \end{pmatrix}$$

is the minimal solution to the given system. ∎

EXERCISES

1. Label the following statements as being true or false. Assume that the underlying inner product spaces are finite-dimensional.
   (a) Every linear operator has an adjoint.
   (b) Every linear operator on V has the form $x \to \langle x, y \rangle$ for some $y \in \mathsf{V}$.
   (c) For every linear operator T on V and every basis β for V, we have $[\mathsf{T}^*]_\beta = ([\mathsf{T}]_\beta)^*$.
   (d) The adjoint of a linear operator is always unique.

(e) For any linear operators T and U and scalars a and b,

$$(a\mathsf{T}+b\mathsf{U})^* = a\mathsf{T}^* + b\mathsf{U}^*.$$

(f) For any $n \times n$ matrix A, $(\mathsf{L}_A)^* = \mathsf{L}_{A^*}$.
(g) For any linear operator T, $(\mathsf{T}^*)^* = \mathsf{T}$.

2. For each of the following inner product spaces V (over F) and linear transformations $\mathsf{g}\colon \mathsf{V} \to F$, find a vector y such that $\mathsf{g}(x) = \langle x, y\rangle$ for all $x \in \mathsf{V}$.
   (a) $\mathsf{V} = \mathsf{R}^3$, $\mathsf{g}(a_1, a_2, a_3) = a_1 - 2a_2 + 4a_3$
   (b) $\mathsf{V} = \mathsf{C}^2$, $\mathsf{g}(z_1, z_2) = z_1 - 2z_2$
   (c) $\mathsf{V} = \mathsf{P}_2(R)$ with $\langle f, h\rangle = \int_0^1 f(t)h(t)\,dt$, $\mathsf{g}(f) = f(0) + f'(1)$

3. For each of the following inner product spaces V and linear operators T on V, evaluate T^* at the given element of V.
   (a) $\mathsf{V} = \mathsf{R}^2$, $\mathsf{T}(a, b) = (2a + b, a - 3b)$, $x = (3, 5)$.
   (b) $\mathsf{V} = \mathsf{C}^2$, $\mathsf{T}(z_1, z_2) = (2z_1 + iz_2, (1 - i)z_1)$, $x = (3 - i, 1 + 2i)$.
   (c) $\mathsf{V} = \mathsf{P}_1(R)$ with $\langle f, g\rangle = \int_{-1}^1 f(t)g(t)\,dt$, $\mathsf{T}(f) = f' + 3f$,
   $f(t) = 4 - 2t$

4. Complete the proof of Theorem 6.11.

5. (a) Complete the proof of the corollary to Theorem 6.11 by using Theorem 6.11 as in the proof of (c).
   (b) State a result for nonsquare matrices that is analogous to the corollary to Theorem 6.11, and prove it using a matrix argument.

6. Let T be a linear operator on an inner product space V. Let $\mathsf{U}_1 = \mathsf{T}+\mathsf{T}^*$ and $\mathsf{U}_2 = \mathsf{T}\mathsf{T}^*$. Prove that $\mathsf{U}_1 = \mathsf{U}_1^*$ and $\mathsf{U}_2 = \mathsf{U}_2^*$.

7. Give an example of a linear operator T on an inner product space V such that $\mathsf{N}(\mathsf{T}) \neq \mathsf{N}(\mathsf{T}^*)$.

8. Let V be a finite-dimensional inner product space, and let T be a linear operator on V. Prove that if T is invertible, then T^* is invertible and $(\mathsf{T}^*)^{-1} = (\mathsf{T}^{-1})^*$.

9. Prove that if $\mathsf{V} = \mathsf{W} \oplus \mathsf{W}^\perp$ and T is the projection on W along $\mathsf{W}^\perp$, then $\mathsf{T} = \mathsf{T}^*$. *Hint:* Recall that $\mathsf{N}(\mathsf{T}) = \mathsf{W}^\perp$. (For definitions, see the exercises of Sections 1.3 and 2.1.)

10. Let T be a linear operator on an inner product space V. Prove that $\|\mathsf{T}(x)\| = \|x\|$ for all $x \in \mathsf{V}$ if and only if $\langle \mathsf{T}(x), \mathsf{T}(y)\rangle = \langle x, y\rangle$ for all $x, y \in \mathsf{V}$. *Hint:* Use Exercise 20 of Section 6.1.

11. For a linear operator T on an inner product space V, prove that $\mathsf{T}^*\mathsf{T} = \mathsf{T}_0$ implies $\mathsf{T} = \mathsf{T}_0$. Is the same result true if we assume that $\mathsf{T}\mathsf{T}^* = \mathsf{T}_0$?

12. Let V be a finite-dimensional inner product space, and let T be a linear operator on V. Prove that $\mathsf{R}(\mathsf{T}^*) = \mathsf{N}(\mathsf{T})^\perp$. *Hint:* Prove that $\mathsf{R}(\mathsf{T}^*)^\perp = \mathsf{N}(\mathsf{T})$, and then use Exercise 12(c) of Section 6.2.

13. Let T be a linear operator on a finite-dimensional vector space V. Prove the following.
    (a) $\mathsf{N}(\mathsf{T}^*\mathsf{T}) = \mathsf{N}(\mathsf{T})$. Deduce that $\text{rank}(\mathsf{T}^*\mathsf{T}) = \text{rank}(\mathsf{T})$.
    (b) $\text{rank}(\mathsf{T}) = \text{rank}(\mathsf{T}^*)$. Deduce from (a) that $\text{rank}(\mathsf{T}\mathsf{T}^*) = \text{rank}(\mathsf{T})$.
    (c) For any $n \times n$ matrix A, $\text{rank}(A^*A) = \text{rank}(AA^*) = \text{rank}(A)$.

14. Let V be an inner product space, and let $y, z \in \mathsf{V}$. Define $\mathsf{T}\colon \mathsf{V} \to \mathsf{V}$ by $\mathsf{T}(x) = \langle x, y\rangle z$ for all $x \in \mathsf{V}$. First prove that T is linear. Then show that T^* exists and find an explicit expression for it.

15. Let $\mathsf{T}\colon \mathsf{V} \to \mathsf{W}$ be a linear transformation, where V and W are finite-dimensional inner product spaces. Let $\langle\cdot,\cdot\rangle_1$ and $\langle\cdot,\cdot\rangle_2$ denote the inner products of V and W, respectively.
    (a) Prove that there exists a unique linear transformation $\mathsf{T}^*\colon \mathsf{W} \to \mathsf{V}$ such that $\langle \mathsf{T}(x), y\rangle_2 = \langle x, \mathsf{T}^*(y)\rangle_1$ for all $x \in \mathsf{V}$ and $y \in \mathsf{W}$.
    (b) Let β and γ be orthonormal bases for V and W, respectively. Prove that $[\mathsf{T}^*]_\gamma^\beta = ([\mathsf{T}]_\beta^\gamma)^*$.

16.† Let A be an $n \times n$ matrix. Prove that $\det(A^*) = \overline{\det(A)}$.

17. Suppose that A is an $m\times n$ matrix in which no two columns are identical. Prove that A^*A is a diagonal matrix if and only if every pair of columns of A is orthogonal.

18. For the data $(-3, 9), (-2, 6), (0, 2)$, and $(1, 1)$, find the line and the parabola that provide the least squares fits. Compute the error E in both cases.

19. In physics, *Hooke's law* states that (within certain limits) there is a linear relation between the length x of a spring and the force y applied to (or exerted by) the spring. That is, $y = cx + d$, where c is called the **spring constant**. Use the following data to estimate the spring constant. (The length is given in inches and the force is given in pounds.)

Length x	Force y
3.5	1.0
4.0	2.2
4.5	2.8
5.0	4.3

20. Find the minimal solution to

$$\begin{aligned} x + 2y - z &= 1 \\ 2x + 3y + z &= 2 \\ 4x + 7y - z &= 4. \end{aligned}$$

21. Show that for the problem of finding the least squares line $y = ct + d$ corresponding to the m observations $(t_1, y_1), \dots, (t_m, y_m)$, the equation $(A^*A)x_0 = A^*y$ of Theorem 6.12 takes the form of the *normal equations:*

$$\left(\sum_{i=1}^{m} t_i^2\right) c + \left(\sum_{i=1}^{m} t_i\right) d = \sum_{i=1}^{m} t_i y_i$$

and

$$\left(\sum_{i=1}^{m} t_i\right) c + md = \sum_{i=1}^{m} y_i.$$

These equations may also be obtained from the error E by setting the partial derivatives of E with respect to both c and d equal to zero.

22. Let V and $\{e_1, e_2, \dots\}$ be defined as in Exercise 20 of Section 6.2. Define $\mathsf{T}\colon \mathsf{V} \to \mathsf{V}$ by

$$\mathsf{T}(\sigma)(k) = \sum_{i=k}^{\infty} \sigma(i) \quad \text{for every positive integer } k.$$

Notice that the infinite series in the definition of T converges because $\sigma(i) \neq 0$ for only finitely many i.

(a) Prove that T is a linear operator on V.
(b) Prove that for any positive integer n, $\mathsf{T}(e_n) = \sum_{i=1}^{n} e_i$.
(c) Prove that T has no adjoint. *Hint:* By way of contradiction suppose that T^* exists. Prove that for any positive integer n, $\mathsf{T}^*(e_n)(k) \neq 0$ for infinitely many k.

6.4 NORMAL AND SELF-ADJOINT OPERATORS

We have seen the importance of diagonalizable operators in Chapter 5. For these operators it is necessary and sufficient for the vector space V to possess a basis of eigenvectors. As V is an inner product space in this chapter, it is reasonable to seek conditions that guarantee that V has an orthonormal basis of eigenvectors. A very important result that helps achieve our goal is Schur's theorem (Theorem 6.14). The formulation below is in terms of linear operators. The next section contains the more conventional matrix form. We begin with a lemma.

Lemma. *Let* T *be a linear operator on a finite-dimensional inner product space* V. *If* T *has an eigenvector, then so does* T^*.

Proof. Let $A = [\mathsf{T}]_\beta$, where β is an orthonormal basis for V. Let λ be an eigenvalue of T, and hence of A. Then $\det(A - \lambda I) = 0$. So by Exercise 16 of Section 6.3 and the corollary to Theorem 6.11, we also have that $\det(A^* - \overline{\lambda} I) = 0$. So $\overline{\lambda}$ is an eigenvalue of A^* and hence of T^*. In particular, T^* has an eigenvector. ∎

Recall (see the exercises of Section 2.1) that a subspace W of V is said to be **T-invariant** if T(W) is contained in W. If W is T-invariant, we may define the restriction $\mathsf{T_W}\colon \mathsf{W} \to \mathsf{W}$ by $\mathsf{T_W}(x) = \mathsf{T}(x)$ for all $x \in \mathsf{W}$. It is clear that $\mathsf{T_W}$ is a linear operator on W. Recall also from Section 5.2 that a polynomial is said to **split** if it factors into linear polynomials.

Theorem 6.14 (Schur). *Let* T *be a linear operator on a finite-dimensional inner product space* V. *Suppose that the characteristic polynomial of* T *splits. Then there exists an orthonormal basis* β *for* V *such that the matrix* $[\mathsf{T}]_\beta$ *is upper triangular.*

Proof. The proof is by induction on the dimension n of V. The result is immediate if $n = 1$. So suppose that the result is true for linear operators on $(n-1)$-dimensional inner product spaces whose characteristic polynomials split. By the lemma we can assume that T^* has a unit eigenvector z. Suppose that $\mathsf{T}^*(z) = \lambda z$ and that $\mathsf{W} = \mathrm{span}(\{z\})$. We show that $\mathsf{W}^\perp$ is T-invariant. If $y \in \mathsf{W}^\perp$ and $x = cz \in \mathsf{W}$, then

$$\begin{aligned} \langle \mathsf{T}(y), x\rangle = \langle \mathsf{T}(y), cz\rangle = \langle y, \mathsf{T}^*(cz)\rangle = \langle y, c\mathsf{T}^*(z)\rangle = \langle y, c\lambda z\rangle \\ = \overline{c\lambda}\,\langle y, z\rangle = \overline{c\lambda}(0) = 0. \end{aligned}$$

So $\mathsf{T}(y) \in \mathsf{W}^\perp$. It is easy to show (see Theorem 5.26) that the characteristic polynomial of $\mathsf{T}_{\mathsf{W}^\perp}$ divides the characteristic polynomial of T and hence splits. By Theorem 6.7(c) $\dim(\mathsf{W}^\perp) = n - 1$, so we may apply the induction hypothesis to $\mathsf{T}_{\mathsf{W}^\perp}$ and obtain an orthonormal basis γ of $\mathsf{W}^\perp$ such that $[\mathsf{T}_{\mathsf{W}^\perp}]_\gamma$ is upper triangular. Clearly, $\beta = \gamma \cup \{z\}$ is an orthonormal basis for V such that $[\mathsf{T}]_\beta$ is upper triangular. ∎

We now return to our original goal of finding an orthonormal basis of eigenvectors of a linear operator T on a finite-dimensional inner product space V. Note that if such an orthonormal basis β exists, then $[\mathsf{T}]_\beta$ is a diagonal matrix. Because diagonal matrices commute, we conclude that T and T^* commute. Thus *if* V *possesses an orthonormal basis of eigenvectors of* T, *then* $\mathsf{TT}^* = \mathsf{T}^*\mathsf{T}$.

Definitions. *Let* V *be an inner product space, and let* T *be a linear operator on* V. *We say that* T *is* **normal** *if* $\mathsf{TT}^* = \mathsf{T}^*\mathsf{T}$. *An* $n \times n$ *real or complex matrix* A *is* **normal** *if* $AA^* = A^*A$.

It follows immediately that T is normal if and only if $[\mathsf{T}]_\beta$ is normal, where β is an orthonormal basis.

Example 1

Let $\mathsf{T}\colon \mathsf{R}^2 \to \mathsf{R}^2$ be rotation by θ, where $0 < \theta < \pi$. The matrix representation of T in the standard ordered basis is given by

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

Note that $AA^* = I = A^*A$; so A and hence T is normal. ∎

Example 2

Suppose that A is a real *skew-symmetric* matrix; that is, $A^t = -A$. Then A is normal because both AA^t and A^tA are equal to $-A^2$. ∎

Clearly, the operator T in Example 1 does not even possess one eigenvector. So in the case of a real inner product space, we see that normality is not sufficient to guarantee an orthonormal basis of eigenvectors. All is not lost, however. We show that normality suffices if V is a *complex* inner product space.

Before we prove the promised result for normal operators, we need some general properties of normal operators.

Theorem 6.15. *Let* V *be an inner product space, and let* T *be a normal operator on* V. *Then the following are true.*

(a) $\|\mathsf{T}(x)\| = \|\mathsf{T}^*(x)\|$ *for all* $x \in \mathsf{V}$.
(b) $\mathsf{T} - c\mathsf{I}$ *is normal for every* $c \in F$.
(c) *If* x *is an eigenvector of* T, *then* x *is also an eigenvector of* T^*. *In fact, if* $\mathsf{T}(x) = \lambda x$, *then* $\mathsf{T}^*(x) = \overline{\lambda} x$.
(d) *If* λ_1 *and* λ_2 *are distinct eigenvalues of* T *with corresponding eigenvectors* x_1 *and* x_2, *then* x_1 *and* x_2 *are orthogonal.*

Proof. (a) For any $x \in \mathsf{V}$, we have

$$\|\mathsf{T}(x)\|^2 = \langle \mathsf{T}(x), \mathsf{T}(x)\rangle = \langle \mathsf{T}^*\mathsf{T}(x), x\rangle = \langle \mathsf{TT}^*(x), x\rangle = \langle \mathsf{T}^*(x), \mathsf{T}^*(x)\rangle = \|\mathsf{T}^*(x)\|^2.$$

The proof of (b) is left as an exercise.

(c) Suppose that $\mathsf{T}(x) = \lambda x$ for some $x \in \mathsf{V}$. Let $\mathsf{U} = \mathsf{T} - \lambda\mathsf{I}$. Then $\mathsf{U}(x) = 0$, and by (b) U is normal. Thus (a) implies that

$$0 = \|\mathsf{U}(x)\| = \|\mathsf{U}^*(x)\| = \|(\mathsf{T}^* - \overline{\lambda}\mathsf{I})(x)\| = \|\mathsf{T}^*(x) - \overline{\lambda}x\|.$$

Hence $\mathsf{T}^*(x) = \overline{\lambda}x$. So x is an eigenvector of T^*.

(d) Let λ_1 and λ_2 be distinct eigenvalues of T with corresponding eigenvectors x_1 and x_2. Then, using (c), we have

$$\lambda_1 \langle x_1, x_2\rangle = \langle \lambda_1 x_1, x_2\rangle = \langle \mathsf{T}(x_1), x_2\rangle = \langle x_1, \mathsf{T}^*(x_2)\rangle = \langle x_1, \overline{\lambda_2} x_2\rangle = \lambda_2 \langle x_1, x_2\rangle.$$

Since $\lambda_1 \neq \lambda_2$, we conclude that $\langle x_1, x_2\rangle = 0$. ∎

Theorem 6.16. *Let* T *be a linear operator on a finite-dimensional complex inner product space* V. *Then* T *is normal if and only if there exists an orthonormal basis for* V *consisting of eigenvectors of* T.

Proof. Suppose that T is normal. By the fundamental theorem of algebra (Theorem D.4) the characteristic polynomial of T splits. So we may apply Schur's theorem to obtain an orthonormal basis $\beta = \{v_1, \ldots, v_n\}$ for V such that $[\mathsf{T}]_\beta = A$ is upper triangular. We know that v_1 is an eigenvector of T because A is upper triangular. Assume that $v_1, \ldots, v_{k-1}$ are eigenvectors of T. We claim that v_k is also an eigenvector of T. It then follows by induction on k that all of the v_i's are eigenvectors of T, or equivalently, that A is a diagonal matrix.

We have

$$A = \begin{pmatrix} B & C \\ O & E \end{pmatrix} \quad \text{and} \quad A^* = \begin{pmatrix} B^* & O \\ C^* & E^* \end{pmatrix},$$

where B is a $(k-1) \times (k-1)$ diagonal matrix. Because A is upper triangular, $A_{jk} = 0$ for $j > k$. To show that v_k is an eigenvector of T, we need only show that $A_{jk} = 0$ for $j < k$. Note that by Theorem 6.15(c), $v_1, \ldots, v_{k-1}$ are also eigenvectors of T^*. But $A^* = [\mathsf{T}^*]_\beta$, so $C^* = O$. Thus $(A^*)_{kj} = 0$ for $j < k$, and so $A_{jk} = 0$ for $j < k$. Therefore, the vector v_k is an eigenvector of T; so by induction, all the vectors of β are eigenvectors of T.

The converse was already proved on page 350. ∎

Interestingly, as the next example shows, Theorem 6.16 does not extend to infinite-dimensional complex inner product spaces.

Example 3

Consider the inner product space H with the orthonormal set S from Example 9 in Section 6.1. Let $\mathsf{V} = \text{span}(S)$, and let T and U be linear operators on V defined by $\mathsf{T}(f) = f_1 f$ and $\mathsf{U}(f) = f_{-1}f$. So,

$$\mathsf{T}(f_k) = f_{k+1} \quad \text{and} \quad \mathsf{U}(f_k) = f_{k-1}$$

for all integers k. Then

$$\langle \mathsf{T}(f_i), f_j \rangle = \langle f_{i+1}, f_j \rangle = \delta_{(i+1),j} = \delta_{i,(j-1)} = \langle f_i, f_{j-1} \rangle = \langle f_i, \mathsf{U}(f_j) \rangle .$$

It follows that $\mathsf{U} = \mathsf{T}^*$. Furthermore, $\mathsf{TT}^* = \mathsf{I} = \mathsf{T}^*\mathsf{T}$, so T is normal.

We show that T has no eigenvectors. Suppose that f is an eigenvector of T, say, $\mathsf{T}(f) = \lambda f$ for some λ. Since V equals the span of S, we may write

$$f = \sum_{i=n}^{m} a_i f_i, \quad \text{where } a_m \neq 0.$$

Applying T to both sides of the preceding equation, we obtain

$$\sum_{i=n}^{m} a_i f_{i+1} = \sum_{i=n}^{m} \lambda a_i f_i.$$

Since $a_m \neq 0$, we can write f_{m+1} as a linear combination of $f_n, f_{n+1}, \ldots, f_m$. But this is a contradiction because S is linearly independent. ∎

Example 1 illustrates that normality is not sufficient to guarantee the existence of an orthonormal basis of eigenvectors for real inner product spaces. For real inner product spaces we must replace normality by the stronger condition that $\mathsf{T} = \mathsf{T}^*$.

Definitions. *Let* T *be a linear operator on an inner product space* V. *We say that* T *is* ***self-adjoint (Hermitian)*** *if* $\mathsf{T} = \mathsf{T}^*$. *An* $n \times n$ *real or complex matrix* A *is* ***self-adjoint (Hermitian)*** *if* $A = A^*$.

It follows immediately that T is self-adjoint if and only if $[\mathsf{T}]_\beta$ is self-adjoint, where β is an orthonormal basis. For real matrices, this condition reduces to the requirement that A is symmetric.

Before we state our main result for self-adjoint operators, we need some preliminary work.

By definition, a linear operator on a real inner product space has only real eigenvalues. The lemma that follows shows that the same can be said for self-adjoint operators on a complex inner product space. Similarly, the characteristic polynomial of every linear operator on a complex inner product space splits, and the same is true for self-adjoint operators on a real inner product space.

Lemma. *Let* T *be a self-adjoint operator on a finite-dimensional inner product space* V. *Then*

(a) *Every eigenvalue of* T *is real.*

(b) *Suppose that* V *is a real inner product space. Then the characteristic polynomial of* T *splits.*

Proof. (a) Suppose that $\mathsf{T}(x) = \lambda x$, for $x \neq 0$. Because a self-adjoint operator is also normal, we can apply Theorem 6.15(c) to obtain

$$\lambda x = \mathsf{T}(x) = \mathsf{T}^*(x) = \overline{\lambda} x.$$

So $\lambda = \overline{\lambda}$; that is, λ is real.

(b) Let $n = \dim(\mathsf{V})$, β be an orthonormal basis for V, and $A = [\mathsf{T}]_\beta$. Then A is self-adjoint. Define $\mathsf{T}_A \colon \mathsf{C}^n \to \mathsf{C}^n$ by $\mathsf{T}_A(x) = Ax$. Then T_A is a linear operator on C^n. Furthermore, T_A is self-adjoint because $[\mathsf{T}_A]_\gamma = A$, where γ is the standard ordered (orthonormal) basis for C^n. So by (a) the eigenvalues of T_A are real. By the fundamental theorem of algebra the characteristic polynomial of T_A splits into factors of the form $x - \lambda$. Since each λ is real, the characteristic polynomial splits over R. But T_A has the same characteristic polynomial as A, which has the same characteristic polynomial as T. Therefore, the characteristic polynomial of T splits. ∎

We are ready to establish one of the major results of this chapter.

Theorem 6.17. *Let* T *be a linear operator on a finite-dimensional real inner product space* V. *Then* T *is self-adjoint if and only if there exists an orthonormal basis* β *for* V *consisting of eigenvectors of* T.

Proof. Suppose that T is self-adjoint. By the lemma we may apply Schur's theorem to obtain an orthonormal basis β for V such that the matrix $A = [\mathsf{T}]_\beta$ is upper triangular. But

$$A^* = [\mathsf{T}]_\beta^* = [\mathsf{T}^*]_\beta = [\mathsf{T}]_\beta = A.$$

So A and A^* are both upper triangular, and therefore A is a diagonal matrix. Thus β must consist of eigenvectors of T.

The converse is left as an exercise. ∎

Theorem 6.17 is used extensively in many areas of mathematics and statistics. We restate this theorem in matrix form in the next section.

Example 4

As we noted earlier, real self-adjoint matrices are symmetric, and self-adjoint matrices are normal. The matrix A below is complex and symmetric.

$$A = \begin{pmatrix} i & i \\ i & 1 \end{pmatrix} \quad \text{and} \quad A^* = \begin{pmatrix} -i & -i \\ -i & 1 \end{pmatrix}$$

But A is not normal because $(AA^*)_{12} = 1 + i$, and $(A^*A)_{12} = 1 - i$. ∎

EXERCISES

1. Label the following statements as being true or false. Assume that the underlying inner product spaces are finite-dimensional.
   (a) Every self-adjoint operator is normal.
   (b) Operators and their adjoints have the same eigenvectors.
   (c) If T is an operator on an inner product space V, then T is normal if and only if $[\mathsf{T}]_\beta$ is normal, where β is any ordered basis for V.
   (d) A real or complex matrix A is normal if and only if L_A is normal.
   (e) The eigenvalues of a self-adjoint operator must all be real.
   (f) The identity and zero operators are self-adjoint.
   (g) Every normal operator is diagonalizable.
   (h) Every self-adjoint operator is diagonalizable.

2. For each of the linear operators below, determine whether it is normal, self-adjoint, or neither.
   (a) $\mathsf{T}\colon \mathsf{R}^2 \to \mathsf{R}^2$ defined by $\mathsf{T}(a,b) = (2a-2b, -2a+5b)$.
   (b) $\mathsf{T}\colon \mathsf{C}^2 \to \mathsf{C}^2$ defined by $\mathsf{T}(a,b) = (2a+ib, a+2b)$.
   (c) $\mathsf{T}\colon \mathsf{P}_2(R) \to \mathsf{P}_2(R)$ defined by $\mathsf{T}(f) = f'$, where $\langle f,g\rangle = \int_0^1 f(t)g(t)\,dt$.

   For (a), find an orthonormal basis for R^2 consisting of eigenvectors of T.

3. Let T and U be self-adjoint operators on an inner product space. Prove that TU is self-adjoint if and only if TU = UT.

4. Prove (b) of Theorem 6.15.

5. Let V be a complex inner product space, and let T be a linear operator on V. Define

$$\mathsf{T}_1 = \frac{1}{2}(\mathsf{T}+\mathsf{T}^*) \quad \text{and} \quad \mathsf{T}_2 = \frac{1}{2i}(\mathsf{T}-\mathsf{T}^*).$$

   (a) Prove that T_1 and T_2 are self-adjoint and that $\mathsf{T} = \mathsf{T}_1 + i\,\mathsf{T}_2$.
   (b) Suppose also that $\mathsf{T} = \mathsf{U}_1 + i\mathsf{U}_2$, where U_1 and U_2 are self-adjoint. Prove that $\mathsf{U}_1 = \mathsf{T}_1$ and $\mathsf{U}_2 = \mathsf{T}_2$.
   (c) Prove that T is normal if and only if $\mathsf{T}_1\mathsf{T}_2 = \mathsf{T}_2\mathsf{T}_1$.

6. Let T be a linear operator on an inner product space V, and let W be a T-invariant subspace of V. Prove the following.
   (a) If T is self-adjoint, then T_W is self-adjoint.
   (b) $\mathsf{W}^\perp$ is T^*-invariant.
   (c) If W is both T- and T^*-invariant, then $(\mathsf{T}_\mathsf{W})^* = (\mathsf{T}^*)_\mathsf{W}$.

   (d) If W is both T- and T^*-invariant and T is normal, then T_W is normal.

7. Let T be a normal operator on a finite-dimensional complex inner product space V, and let W be a subspace of V. Prove that if W is T-invariant, then W is also T^*-invariant. *Hint:* Use Exercise 24 of Section 5.4.

8. Let T be a normal operator on a finite-dimensional inner product space V. Prove that $\mathsf{N}(\mathsf{T}) = \mathsf{N}(\mathsf{T}^*)$ and $\mathsf{R}(\mathsf{T}) = \mathsf{R}(\mathsf{T}^*)$. *Hint:* Use Theorem 6.15 and Exercise 12 of Section 6.3.

9. Let T be a self-adjoint operator on a finite-dimensional inner product space V. Prove that for all $x \in \mathsf{V}$

$$\|\mathsf{T}(x) \pm ix\|^2 = \|\mathsf{T}(x)\|^2 + \|x\|^2.$$

   Deduce that $(\mathsf{T} - i\mathsf{I})$ is invertible and that $[(\mathsf{T}-i\mathsf{I})^{-1}]^* = (\mathsf{T}+i\mathsf{I})^{-1}$.

10. Assume that T is a linear operator on a complex (not necessarily finite-dimensional) inner product space V with an adjoint T^*. Prove the following.
   (a) If T is self-adjoint, then $\langle \mathsf{T}(x), x\rangle$ is real for all $x \in \mathsf{V}$.
   (b) If T satisfies $\langle \mathsf{T}(x), x\rangle = 0$ for all $x \in \mathsf{V}$, then $\mathsf{T} = \mathsf{T}_0$. *Hint:* Replace x by $x+y$ and then by $x+iy$ and expand the resulting inner products.
   (c) If $\langle \mathsf{T}(x), x\rangle$ is real for all $x \in \mathsf{V}$, then $\mathsf{T} = \mathsf{T}^*$.

11. Let T be a normal operator on a finite-dimensional real inner product space V whose characteristic polynomial splits. Prove that V has an orthonormal basis of eigenvectors of T. Hence prove that T is self-adjoint.

12. An $n \times n$ real matrix A is said to be a **Gramian** matrix if there exists a real (square) matrix B such that $A = B^tB$. Prove that A is a Gramian matrix if and only if A is symmetric and all of its eigenvalues are nonnegative. *Hint:* Apply Theorem 6.17 to L_A to obtain an orthonormal basis $\{v_1, \ldots, v_n\}$ of eigenvectors with the associated eigenvalues $\lambda_1, \ldots, \lambda_n$. Define the linear operator T by $\mathsf{U}(v_i) = \sqrt{\lambda_i}x_i$ and complete the proof.

The following definitions will be used in Exercises 13, 14, and 17 through 21.

Definitions. *A linear operator* T *on a finite-dimensional inner product space is called* ***positive definite*** [***positive semidefinite***] *if* T *is self-adjoint and* $\langle \mathsf{T}(x), x\rangle > 0$ [$\langle \mathsf{T}(x), x\rangle \geq 0$] *for all* $x \neq 0$.

13. Let T be a self-adjoint linear operator on an n-dimensional inner product space V, and let $A = [\mathsf{T}]_\beta$, where β is an orthonormal basis for V. Prove

(a) T is positive definite [semidefinite] if and only if all of its eigenvalues are positive [nonnegative].
(b) T is positive definite [semidefinite] if and only if L_A is also.
(c) T is positive definite if and only if

$$\sum_{i,j} A_{ij} a_j \bar{a}_i > 0 \text{ for all nonzero } n\text{-tuples } (a_1, \ldots, a_n).$$

This inequality is often used as the definition of a positive definite matrix. Changing the inequality to a nonstrict inequality gives the corresponding definition of a positive semidefinite matrix.

(d) T is positive semidefinite if and only if $A = B^*B$ for some square matrix B.
(e) If T and U are positive semidefinite operators such that $\mathsf{T}^2 = \mathsf{U}^2$, then $\mathsf{T} = \mathsf{U}$.
(f) Is the composite of two positive definite operators positive definite?

Results analogous to (a) through (e) hold for matrices as well as operators.

14. Let $\mathsf{T}\colon \mathsf{V} \to \mathsf{W}$ be a linear transformation, where V and W are finite-dimensional inner product spaces. Prove that $\mathsf{T}^*\mathsf{T}$ is positive semidefinite and $\text{rank}(\mathsf{T}^*\mathsf{T}) = \text{rank}(\mathsf{T})$. (See Exercise 15 of Section 6.3.)

15. *Simultaneous Diagonalization*

(a) Let V be a finite-dimensional real inner product space, and let U and T be self-adjoint linear operators on V such that $\mathsf{UT} = \mathsf{TU}$. Prove that there exists an orthonormal basis for V consisting of vectors that are eigenvectors of both U and T. (The complex version of this result appears as Exercise 10 of Section 6.6). *Hint:* For any eigenspace $\mathsf{W} = \mathsf{E}_\lambda$ of T we have that W is both T- and U-invariant. By Exercise 6 we have that $\mathsf{W}^\perp$ is both T- and U-invariant. Apply Theorem 6.17 and Proposition 6.6.
(b) State and prove the analogous result about commuting symmetric (real) matrices.

16. Prove the *Cayley-Hamilton theorem* for a complex $n\times n$ matrix A. That is, if $f(t)$ is the characteristic polynomial of A, prove that $f(A) = O$. *Hint:* By Schur's theorem show that you may assume that A is upper triangular, in which case

$$f(t) = \prod_{i=1}^{n} (A_{ii} - t).$$

Now if $\mathsf{T} = \mathsf{L}_A$, we have $(A_{jj}\mathsf{I} - \mathsf{T})(e_j) \in \text{span}(\{e_1, \ldots, e_{j-1}\})$ for $j \geq 2$, where $\{e_1, \ldots, e_n\}$ is the standard ordered basis for C^n. (The general case is proved in Section 5.4.)

Exercises 17 through 21 use the definition of *positive definite operator* that precedes Exercise 13.

17. Let T and U be positive definite operators on an inner product space V. Prove the following.

(a) $\mathsf{T} + \mathsf{U}$ is positive definite.
(b) If $c > 0$, then $c\mathsf{T}$ is positive definite.
(c) T^{-1} is positive definite.

18. Let V be an inner product space with inner product $\langle \cdot, \cdot \rangle$, and let T be a positive definite linear operator on V. Prove that $\langle x, y \rangle' = \langle \mathsf{T}(x), y \rangle$ defines another inner product on V.

19. Let V be a finite-dimensional inner product space, and let T and U be self-adjoint operators on V such that T is positive definite. Prove that both TU and UT are diagonalizable linear operators that have only real eigenvalues. *Hint:* Show that UT is self-adjoint with respect to the inner product $\langle x, y \rangle' = \langle \mathsf{T}(x), y \rangle$. To show that TU is self-adjoint, repeat the argument with T^{-1} in place of T.

20. The following result gives a converse to Exercise 18. Let V be a finite-dimensional inner product space with inner product $\langle \cdot, \cdot \rangle$, and let $\langle \cdot, \cdot \rangle'$ be any other inner product on V.

(a) Prove that there exists a unique linear operator T on V such that $\langle x, y \rangle' = \langle \mathsf{T}(x), y \rangle$ for all x and y in V. *Hint:* Let $\beta = \{v_1, \ldots, v_n\}$ be an orthonormal basis for V with respect to $\langle \cdot, \cdot \rangle$, and define a matrix A by $A_{ij} = \langle v_j, v_i \rangle'$ for all i and j. Let T be the unique linear operator on V such that $[\mathsf{T}]_\beta = A$.
(b) Prove that the operator T of (a) is positive definite with respect to both inner products.

21. Let U be a diagonalizable linear operator on a finite-dimensional inner product space V such that all of the eigenvalues of U are real. Prove that there exist positive definite linear operators T_1 and T_1' and self-adjoint linear operators T_2 and T_2' such that $\mathsf{U} = \mathsf{T}_2\mathsf{T}_1 = \mathsf{T}_1'\mathsf{T}_2'$. *Hint:* Let $\langle \cdot, \cdot \rangle$ be the inner product associated with V, β a basis of eigenvectors for U, $\langle \cdot, \cdot \rangle'$ the inner product on V with respect to which β is orthonormal (see Exercise 24(a) of Section 6.1), and T_1 the positive definite operator according to Exercise 20. Show that U is self-adjoint with respect to $\langle \cdot, \cdot \rangle'$ and $\mathsf{U} = \mathsf{T}_1^{-1}\mathsf{U}^*\mathsf{T}_1$ (the adjoint is with respect to $\langle \cdot, \cdot \rangle$). Let $\mathsf{T}_2 = \mathsf{T}_1{}^{-1}\mathsf{U}^*$.

22. The following argument gives another proof of Schur's theorem. Let T be a linear operator on a finite dimensional inner product space V.

(a) Suppose that β is an ordered basis for V such that $[\mathsf{T}]_\beta$ is an upper triangular matrix. Let γ be the orthonormal basis for V obtained by applying the Gram-Schmidt orthogonalization process to β and then dividing each resulting vector by its length. Prove that $[\mathsf{T}]_\gamma$ is an upper triangular matrix.

(b) Use Exercise 32 of Section 5.4 and (a) above to obtain an alternate proof of Schur's theorem.

6.5 UNITARY AND ORTHOGONAL OPERATORS AND THEIR MATRICES

In this section we continue our analogy between complex numbers and linear operators. Recall that the adjoint of a linear operator acts similarly to the conjugate of a complex number (see, for example, Theorem 6.11). A complex number z has length 1 if $z\overline{z} = 1$. In this section we study those linear operators T on an inner product space V such that $\mathsf{TT}^* = \mathsf{T}^*\mathsf{T} = \mathsf{I}$. We will see that these are precisely the linear operators that "preserve length" in the sense that $\|\mathsf{T}(x)\| = \|x\|$ for all $x \in \mathsf{V}$. As another characterization, we prove that on a finite-dimensional complex inner product space these are the normal operators whose eigenvalues all have absolute value 1.

In past chapters we were interested in studying those functions that preserve the structure of the underlying space. In particular, linear operators preserve the operations of vector addition and scalar multiplication, and isomorphisms preserve all the vector space structure. It is now natural to consider those linear operators T on an inner product space that preserve length. We will see that this condition guarantees, in fact, that T preserves the inner product.

Definitions. *Let* T *be a linear operator on an inner product space* V *(over F). If $\|\mathsf{T}(x)\| = \|x\|$ for all $x \in \mathsf{V}$, we call* T *a* ***unitary operator*** *if $F = C$ and an* ***orthogonal operator*** *if $F = R$.*

It should be noted that in the infinite-dimensional case, an operator that satisfies the norm requirement above is generally called an **isometry**. If, in addition, the operator is onto (the condition guarantees one-to-one), then the operator is called a **unitary** or **orthogonal operator**.

Clearly, any rotation or reflection in R^2 preserves length and hence is an orthogonal operator. We will study these operators in much more detail in Section 6.10.

Example 1

Let $h \in \mathsf{H}$, satisfy $|h(x)| = 1$ for all x. Define the linear operator T on H by $\mathsf{T}(f) = hf$. Then

$$\|\mathsf{T}(f)\|^2 = \|hf\|^2 = \frac{1}{2\pi}\int_0^{2\pi} h(t)f(t)\overline{h(t)f(t)}\,dt = \|f\|^2$$

since $|h(t)|^2 = 1$ for all t. So T is a unitary operator. ∎

Theorem 6.18. *Let* T *be a linear operator on a finite-dimensional inner product space* V*. Then the following are equivalent.*

(a) $\mathsf{TT}^* = \mathsf{T}^*\mathsf{T} = \mathsf{I}$.
(b) $\langle \mathsf{T}(x), \mathsf{T}(y)\rangle = \langle x, y\rangle$ *for all* $x, y \in \mathsf{V}$.
(c) *If β is an orthonormal basis for* V*, then $\mathsf{T}(\beta)$ is an orthonormal basis for* V.
(d) *There exists an orthonormal basis β for* V *such that $\mathsf{T}(\beta)$ is an orthonormal basis for* V.
(e) $\|\mathsf{T}(x)\| = \|x\|$ *for all* $x \in \mathsf{V}$.

Thus all the conditions above are equivalent to the definition of a unitary or orthogonal operator. From (a) it follows that unitary or orthogonal operators are normal.

Before proving the theorem, we first prove the following lemma. Compare this lemma to Exercise 10(b) of Section 6.4.

Lemma. *Let* U *be a self-adjoint operator on a finite-dimensional inner product space* V*. If $\langle x, \mathsf{U}(x)\rangle = 0$ for all $x \in \mathsf{V}$, then $\mathsf{U} = \mathsf{T}_0$.*

Proof. By either Theorem 6.16 or 6.17 we may choose an orthonormal basis β for V consisting of eigenvectors of U. If $x \in \beta$, then $\mathsf{U}(x) = \lambda x$ for some λ. Thus

$$0 = \langle x, \mathsf{U}(x)\rangle = \langle x, \lambda x\rangle = \overline{\lambda}\,\langle x, x\rangle\,;$$

so $\overline{\lambda} = 0$. Hence $\mathsf{U}(x) = 0$ for all $x \in \beta$, and thus $\mathsf{U} = \mathsf{T}_0$. ∎

Proof of Theorem 6.18. We prove first that (a) implies (b). Let $x, y \in \mathsf{V}$. Then $\langle x, y\rangle = \langle \mathsf{T}^*\mathsf{T}(x), y\rangle = \langle \mathsf{T}(x), \mathsf{T}(y)\rangle$.

Second, we prove that (b) implies (c). Let $\beta = \{v_1, \ldots, v_n\}$ be an orthonormal basis for V, so $\mathsf{T}(\beta) = \{\mathsf{T}(v_1), \ldots, \mathsf{T}(v_n)\}$. Now $\langle \mathsf{T}(v_i), \mathsf{T}(v_j)\rangle = \langle v_i, v_j\rangle = \delta_{ij}$. Therefore $\mathsf{T}(\beta)$ is an orthonormal basis for V.

That (c) implies (d) is obvious.

Next we prove that (d) implies (e). Let $x \in \mathsf{V}$, and let $\beta = \{v_1, \ldots, v_n\}$. Now

$$x = \sum_{i=1}^{n} a_i v_i$$

for some scalars a_i, and so

$$\|x\|^2 = \left\langle \sum_{i=1}^n a_i v_i, \sum_{j=1}^n a_j v_j \right\rangle = \sum_{i=1}^n \sum_{j=1}^n a_i \overline{a_j} \langle v_i, v_j \rangle$$
$$= \sum_{i=1}^n \sum_{j=1}^n a_i \overline{a_j} \delta_{ij} = \sum_{i=1}^n |a_i|^2$$

since β is orthonormal.

Applying the same manipulations to

$$\mathsf{T}(x) = \sum_{i=1}^n a_i \mathsf{T}(v_i)$$

and using the fact that $\mathsf{T}(\beta)$ is also orthonormal, we obtain

$$\|\mathsf{T}(x)\|^2 = \sum_{i=1}^n |a_i|^2.$$

Hence $\|\mathsf{T}(x)\| = \|x\|$.

Finally, we prove that (e) implies (a). For any $x \in \mathsf{V}$, we have

$$\langle x, x\rangle = \|x\|^2 = \|\mathsf{T}(x)\|^2 = \langle \mathsf{T}(x), \mathsf{T}(x)\rangle = \langle x, \mathsf{T}^*\mathsf{T}(x)\rangle.$$

So $\langle x, (\mathsf{I} - \mathsf{T}^*\mathsf{T})(x)\rangle = 0$ for all $x \in \mathsf{V}$. Let $\mathsf{U} = \mathsf{I} - \mathsf{T}^*\mathsf{T}$; then U is self-adjoint, and $\langle x, \mathsf{U}(x)\rangle = 0$ for all $x \in \mathsf{V}$. So by the lemma we have $\mathsf{T}_0 = \mathsf{U} = \mathsf{I} - \mathsf{T}^*\mathsf{T}$, and hence $\mathsf{T}^*\mathsf{T} = \mathsf{I}$. Since V is finite-dimensional, we may use Exercise 8 of Section 2.4 to conclude that $\mathsf{T}\mathsf{T}^* = \mathsf{I}$. ∎

It follows immediately from the definition that every eigenvalue of a unitary or orthogonal operator has absolute value 1. In fact, even more is true.

Corollary 1. *Let T be a linear operator on a finite-dimensional real inner product space V. Then V has an orthonormal basis of eigenvectors of T with corresponding eigenvalues of absolute value 1 if and only if T is both self-adjoint and orthogonal.*

Proof. Suppose that V has an orthonormal basis $\{v_1, \ldots, v_n\}$ such that $\mathsf{T}(v_i) = \lambda_i v_i$ and $|\lambda_i| = 1$ for all i. By Theorem 6.17 T is self-adjoint. Thus $(\mathsf{T}\mathsf{T}^*)(v_i) = \mathsf{T}(\lambda_i v_i) = \lambda_i \lambda_i v_i = \lambda_i^2 v_i = v_i$ for each i. So $\mathsf{T}\mathsf{T}^* = \mathsf{I}$, and again by Exercise 8 of Section 2.4, T is orthogonal by Theorem 6.18(a).

If T is self-adjoint, then by Theorem 6.17 we have that V possesses an orthonormal basis $\{v_1, \ldots, v_n\}$ such that $\mathsf{T}(v_i) = \lambda_i v_i$ for all i. If T is also orthogonal, we have

$$|\lambda_i| \cdot \|v_i\| = \|\lambda_i v_i\| = \|\mathsf{T}(v_i)\| = \|v_i\|;$$

so $|\lambda_i| = 1$ for every i. ∎

Corollary 2. *Let T be a linear operator on a finite-dimensional complex inner product space V. Then V has an orthonormal basis of eigenvectors of T with corresponding eigenvalues of absolute value 1 if and only if T is unitary.*

Proof. The proof is similar to the proof of Corollary 1. ∎

Example 2

Let $\mathsf{T}\colon \mathsf{R}^2 \to \mathsf{R}^2$ be a rotation by θ, where $0 < \theta < \pi$. It is clear geometrically that T "preserves length", that is, that $\|\mathsf{T}(x)\| = \|x\|$ for all $x \in \mathsf{R}^2$. The fact that rotations by a fixed angle preserve perpendicularity not only can be seen geometrically but now follows from (b) of Theoreom 6.18. Perhaps the fact that such a transformation preserves the inner product is not so obvious geometrically; however, we obtain this fact from (b) also. Finally, an inspection of the matrix

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

reveals that T is not self-adjoint for the given restriction on θ. As we mentioned earlier, this fact also follows from the geometric observation that T has no eigenvectors and from Theorem 6.15. It is seen easily from the matrix above that T^* is the rotation by $-\theta$. ∎

We now examine the matrices that represent unitary and orthogonal transformations.

Definitions. *Let A be a square matrix with entries in an arbitrary field F. A is called an* **orthogonal matrix** *if $A^tA = AA^t = I$. If $F = C$ or R and $A^*A = AA^* = I$, then A is called a* **unitary matrix.**

Since for a real matrix A we have $A^* = A^t$, a real unitary matrix is also orthogonal. In this case we call A **orthogonal** rather than unitary.

Note that the condition $AA^* = I$ is equivalent to the statement that the rows of A form an orthonormal basis for F^n because

$$\delta_{ij} = I_{ij} = (AA^*)_{ij} = \sum_{k=1}^n A_{ik}(A^*)_{kj} = \sum_{k=1}^n A_{ik}\overline{A_{jk}},$$

and the last term represents the inner product of the ith and jth rows of A.

A similar remark can be made about the columns of A and the condition $A^*A = I$.

It also follows from the definition above that a linear operator T on an inner product space V is unitary [orthogonal] if and only if $[\mathsf{T}]_\beta$ is unitary [orthogonal] for some orthonormal basis β for V.

Example 3

From Example 2 the matrix

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

is clearly orthogonal. One can easily see that the rows of the matrix form an orthonormal basis for R^2. Similarly, the columns of the matrix form an orthonormal basis for R^2. ∎

We know that for a complex normal [real symmetric] matrix A there is an orthonormal basis β for F^n consisting of eigenvectors of A. Hence A is similar to a diagonal matrix D. By Theorem 5.1 the matrix Q whose columns are the vectors in β is such that $D = Q^{-1}AQ$. But since the columns of Q are an orthonormal basis for F^n, it follows that Q is unitary [orthogonal]. In this case we say that A is **unitarily equivalent** [**orthogonally equivalent**] to D. It is easily seen (see Exercise 17) that this relation is an equivalence relation on $\mathsf{M}_{n\times n}(C)$ [$\mathsf{M}_{n\times n}(R)$]. More generally, *$A$ and B are unitarily equivalent [orthogonally equivalent] if and only if there exists a unitary [orthogonal] matrix P such that $A = P^*BP$.*

The preceding paragraph has proved half of each of the following two theorems.

Theorem 6.19. *Let A be a complex $n \times n$ matrix. Then A is normal if and only if A is unitarily equivalent to a diagonal matrix.*

Proof. By the remarks above we need only prove that if A is unitarily equivalent to a diagonal matrix, then A is normal.

Suppose that $A = P^*DP$, where P is a unitary matrix and D is a diagonal matrix. Then

$$AA^* = (P^*DP)(P^*DP)^* = (P^*DP)(P^*D^*P) = P^*DID^*P = P^*DD^*P.$$

Similarly, $A^*A = P^*D^*DP$. Since D is a diagonal matrix, however, we have $DD^* = D^*D$. Thus $AA^* = A^*A$. ∎

Theorem 6.20. *Let A be a real $n \times n$ matrix. Then A is symmetric if and only if A is orthogonally equivalent to a real diagonal matrix.*

Proof. The proof is similar to the proof of Theorem 6.19 and is left as an exercise. ∎

Example 4

Let

$$A = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{pmatrix}.$$

Since A is symmetric, Theorem 6.20 tells us that A is orthogonally equivalent to a diagonal matrix. We now find an orthogonal matrix P and a diagonal matrix D such that $P^tAP = D$.

To find P, we first obtain an orthonormal basis of eigenvectors. It is easy to show that the eigenvalues of A are 2 and 8. The set $\{(-1,1,0),(-1,0,1)\}$ is a basis for the eigenspace corresponding to 2. Because this set is not orthogonal, we apply the Gram-Schmidt process to obtain the orthogonal set $\{(-1,1,0), -\frac{1}{2}(1,1,-2)\}$. The set $\{(1,1,1)\}$ is a basis for the eigenspace corresponding to 8. Notice that $(1,1,1)$ is orthogonal to the preceding two vectors as predicted by Theorem 6.15(d). Taking the union of these two bases and normalizing the vectors, we obtain the following orthonormal basis for R^3 consisting of eigenvectors of A:

$$\left\{\frac{1}{\sqrt{2}}(-1,1,0), \frac{1}{\sqrt{6}}(1,1,-2), \frac{1}{\sqrt{3}}(1,1,1)\right\}.$$

Thus one possible choice is

$$P = \begin{pmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{pmatrix}. \quad ∎$$

Because of Schur's theorem (Theorem 6.14), the next result is immediate. As it is the matrix form of Schur's theorem, we also refer to it as Schur's theorem.

Theorem 6.21 (Schur). *Let $A \in \mathsf{M}_{n\times n}(F)$ be a matrix whose characteristic polynomial splits over F.*

(a) *If $F = C$, then A is unitarily equivalent to a complex upper triangular matrix.*
(b) *If $F = R$, then A is orthogonally equivalent to a real upper triangular matrix.*

Rigid Motions in the Plane

The purpose of this application is to characterize the so-called *rigid motions* of R^2. One may think intuitively of such a motion as a transformation that does not affect the shape of a figure under its action, hence the name *rigid*. For example, reflections, rotations, and translations ($x \to x + x_0$) are examples of rigid motions. In fact, we prove that every rigid motion is a composite of these three transformations. The general situation in R^n is handled in Section 6.10 and uses the results obtained here.

Definition. *Let* V *be a real inner product space. A function* $f: \mathsf{V} \to \mathsf{V}$ *is a* ***rigid motion*** *if*

$$\|f(x) - f(y)\| = \|x - y\|$$

for all $x, y \in \mathsf{V}$.

Although we prove a number of general results about rigid motions, our main result is in the setting of R^2.

Theorem 6.22. *Every rigid motion in* R^2 *is one of two types: a rotation (about the origin) followed by a translation, or a reflection (about the* x*-axis) followed by a rotation (about the origin) followed by a translation.*

Throughout we assume that f is a rigid motion on a real inner product space V and that $\mathsf{T}: \mathsf{V} \to \mathsf{V}$ is defined by

$$\mathsf{T}(x) = f(x) - f(0)$$

for all $x \in \mathsf{V}$.

Lemma 1. *For all* $x, y \in \mathsf{V}$
(a) $\|\mathsf{T}(x)\| = \|x\|$;
(b) $\|\mathsf{T}(x) - \mathsf{T}(y)\| = \|x - y\|$;
(c) $\langle \mathsf{T}(x), \mathsf{T}(y) \rangle = \langle x, y \rangle$;
(d) T *is linear.*
Hence T *is an orthogonal operator.*

Proof. (a) Because f is a rigid motion, we have

$$\|\mathsf{T}(x)\| = \|f(x) - f(0)\| = \|x - 0\| = \|x\|$$

for all $x \in \mathsf{V}$.

(b) For all $x, y \in \mathsf{V}$ we have

$$\begin{aligned}\|\mathsf{T}(x) - \mathsf{T}(y)\| &= \|(f(x) - f(0)) - (f(y) - f(0))\| \\ &= \|f(x) - f(y)\| = \|x - y\|.\end{aligned}$$

(c) For all $x, y \in \mathsf{V}$ we have

$$\|\mathsf{T}(x) - \mathsf{T}(y)\|^2 = \|\mathsf{T}(x)\|^2 - 2\langle \mathsf{T}(x), \mathsf{T}(y) \rangle + \|\mathsf{T}(y)\|^2$$

and

$$\|x - y\|^2 = \|x\|^2 - 2\langle x, y \rangle + \|y\|^2.$$

(c) now follows from (a) and (b) and the two equations above.

(d) For all $x, y \in \mathsf{V}$ and $a \in R$, we have by (b), (a), and (c):

$$\begin{aligned}\|\mathsf{T}(x + ay) - \mathsf{T}(x) - a\mathsf{T}(y)\|^2 &= \|[\mathsf{T}(x + ay) - \mathsf{T}(x)] - a\mathsf{T}(y)\|^2 \\ &= \|\mathsf{T}(x + ay) - \mathsf{T}(x)\|^2 + a^2\|\mathsf{T}(y)\|^2 - 2a\langle \mathsf{T}(x + ay) - \mathsf{T}(x), \mathsf{T}(y) \rangle \\ &= \|(x + ay) - x\|^2 + a^2\|y\|^2 - 2a[\langle \mathsf{T}(x + ay), \mathsf{T}(y) \rangle - \langle \mathsf{T}(x), \mathsf{T}(y) \rangle] \\ &= a^2\|y\|^2 + a^2\|y\|^2 - 2a[\langle x + ay, y \rangle - \langle x, y \rangle] \\ &= 2a^2\|y\|^2 - 2a[\langle x, y \rangle + a\|y\|^2 - \langle x, y \rangle] \\ &= 0. \quad \blacksquare\end{aligned}$$

Lemma 2. *The function* f *is an orthogonal operator followed by a translation.*

Proof. By Lemma 1, T is an orthogonal operator. If we define $\mathsf{U}: \mathsf{V} \to \mathsf{V}$ by $\mathsf{U}(x) = x + f(0)$ for all $x \in \mathsf{V}$, then U is a translation. So for any x,

$$\mathsf{UT}(x) = \mathsf{U}(\mathsf{T}(x)) = \mathsf{T}(x) + f(0) = f(x). \quad \blacksquare$$

Lemma 3. *If* V *is finite-dimensional, then* $\det(\mathsf{T}) = \pm 1$.

Proof. Let β be an orthonormal basis for V. Then by Theorem 6.10 and Exercise 16 of Section 6.3, we have

$$\det(\mathsf{T}^*) = \det([\mathsf{T}^*]_\beta) = \det([\mathsf{T}]_\beta^*) = \det([\mathsf{T}]_\beta) = \det(\mathsf{T}).$$

Because T is orthogonal by Lemma 1(a), we have that $\mathsf{I} = \mathsf{T}^*\mathsf{T}$ by Theorem 6.18(a). So

$$1 = \det(\mathsf{I}) = \det(\mathsf{T}^*\mathsf{T}) = \det(\mathsf{T}^*) \cdot \det(\mathsf{T}) = \det(\mathsf{T}) \cdot \det(\mathsf{T}) = \det(\mathsf{T})^2. \quad \blacksquare$$

Lemma 4. *Suppose that* $\mathsf{V} = \mathsf{R}^2$ *and that* β *is the standard ordered basis for* R^2. *Then there exists an angle* θ $(0 \le \theta < 2\pi)$ *such that*

$$[\mathsf{T}]_\beta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad \text{if } \det(\mathsf{T}) = 1$$

and

$$[\mathsf{T}]_\beta = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \quad \text{if } \det(\mathsf{T}) = -1.$$

Proof. Let $A = [\mathsf{T}]_\beta$. Because T is an orthogonal operator by Lemma 1, we conclude from Theorem 6.18(c) that $\mathsf{T}(\beta) = \{\mathsf{T}(e_1), \mathsf{T}(e_2)\}$ is an orthonormal basis for R^2. Because $\mathsf{T}(e_1)$ is a unit vector, there exists an angle θ $(0 \le \theta < 2\pi)$ such that $\mathsf{T}(e_1) = (\cos\theta, \sin\theta)$. Since $\mathsf{T}(e_2)$ is a unit vector and is orthogonal to $\mathsf{T}(e_1)$, there are only two possible choices for $\mathsf{T}(e_2)$. Either

$$\mathsf{T}(e_2) = (-\sin\theta, \cos\theta) \quad \text{or} \quad \mathsf{T}(e_2) = (\sin\theta, -\cos\theta).$$

If $\det(\mathsf{T}) = 1$, we must have the first case; if $\det(\mathsf{T}) = -1$, we must have the second case. $\blacksquare$

Proof of Theorem 6.22. By Lemma 2 we need only analyze the orthogonal operator T. By Lemma 3 det(T) = ±1. By Lemma 4, if det(T) = 1, we see that T is a rotation by θ. If det(T) = −1, then using

$$[\mathsf{T}]_\beta = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

we have that T is a reflection about the x-axis followed by a rotation. ∎

Conic Sections

As an application of Theorem 6.20, we consider the quadratic equation

$$ax^2 + 2bxy + cy^2 + dx + ey + f = 0. \tag{2}$$

For special choices of the coefficients in (2), we obtain the various conic sections. For example, if $a = c = 1$, $b = d = e = 0$, and $f = -1$, we obtain the circle $x^2 + y^2 = 1$ with center at the origin. The remaining conic sections, namely, the ellipse, parabola, and hyperbola, are obtained by other choices of the coefficients. Each of the preceding examples is easy to graph by the method of completing the square because the xy-term is absent. For example, the equation $x^2 + 2x + y^2 + 4y + 2 = 0$ may be rewritten as $(x+1)^2 + (y+2)^2 = 3$, which describes a circle with center at $(-1,-2)$ in the xy-coordinate system and radius $\sqrt{3}$. If we consider the transformation of coordinates $(x,y) \to (x',y')$, where $x' = x+1$ and $y' = y+2$, then our equation simplifies to $(x')^2 + (y')^2 = 3$. This change of variable allows us to eliminate the x- and y-terms.

We now concentrate solely on the elimination of the xy-term. To accomplish this, we consider the expression

$$ax^2 + 2bxy + cy^2 \tag{3}$$

which is called the **associated quadratic form** of (2). Quadratic forms are studied in more generality in Section 6.7.

If we let

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} x \\ y \end{pmatrix},$$

then (3) may be written as $X^tAX = \langle AX, X\rangle$. For example, the quadratic form $3x^2 + 4xy + 6y^2$ may be written as

$$X^t\begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix}X.$$

The fact that A is symmetric is crucial in our discussion. For, by Theorem 6.20, we may choose an orthogonal matrix P and a diagonal matrix D with real diagonal entries λ_1 and λ_2 such that $P^tAP = D$. Now define

$$X' = \begin{pmatrix} x' \\ y' \end{pmatrix}$$

by $X' = P^tX$ or, equivalently, by $PX' = PP^tX = X$. Then

$$X^tAX = (PX')^tA(PX') = X'^t(P^tAP)X' = X'^tDX' = \lambda_1(x')^2 + \lambda_2(y')^2.$$

Thus the transformation $(x,y) \to (x',y')$ allows us to eliminate the xy-term in (3) and hence in (2).

Furthermore, since P is orthogonal, we have by Lemma 3 to Theorem 6.22 (with $\mathsf{T} = \mathsf{L}_P$) that $\det(P) = \pm 1$. If $\det(P) = -1$, we may interchange the columns of P to obtain a matrix Q. Because the columns of P form an orthonormal basis of eigenvectors of A, the same is true of the columns of Q. Therefore,

$$Q^tAQ = \begin{pmatrix} \lambda_2 & 0 \\ 0 & \lambda_1 \end{pmatrix}.$$

Notice that $\det(Q) = -\det(P) = 1$. So, if $\det(P) = -1$, we can take Q for our new P; consequently, we assume that $\det(P) = 1$. By Lemma 4 to Theorem 6.22 (with $\mathsf{T} = \mathsf{L}_P$), it follows that matrix P represents a rotation.

In summary, the xy-term in (2) may be eliminated by a rotation of the x-axis and y-axis to new axes x' and y' given by $X = PX'$, where P is an orthogonal matrix and $\det(P) = 1$. Furthermore, the coefficients of $(x')^2$ and $(y')^2$ are the eigenvalues of

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

This result is a restatement of a result known as the *principal axis theorem* for R^2. The arguments above, of course, are easily extended to quadratic equations in n variables. For example, in the case $n = 3$, by special choices of the coefficients, we obtain the quadratic surfaces—the elliptic cone, the ellipsoid, the hyperbolic paraboloid, etc.

As an example, consider the quadratic equation

$$2x^2 - 4xy + 5y^2 - 36 = 0,$$

for which the associated quadratic form is $2x^2 - 4xy + 5y^2$. In the notation above

$$A = \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix},$$

so that the eigenvalues of A are 1 and 6 with associated eigenvectors

$$\begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -1 \\ 2 \end{pmatrix}.$$

As expected (from Theorem 6.15(d)), these vectors are orthogonal. The corresponding orthonormal basis of eigenvectors

$$\beta = \left\{ \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}, \begin{pmatrix} \frac{-1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} \right\}$$

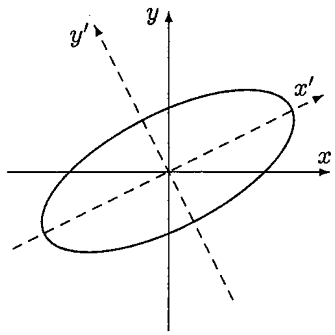

Figure 6.3

determines new axes x' and y' as in Figure 6.3. Hence if

$$P = \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix},$$

then

$$P^tAP = \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix}.$$

Under the transformation $X = PX'$ or

$$x = \frac{2}{\sqrt{5}}x' - \frac{1}{\sqrt{5}}y'$$
$$y = \frac{1}{\sqrt{5}}x' + \frac{2}{\sqrt{5}}y',$$

we have the new quadratic form $(x')^2 + 6(y')^2$. Thus the original equation $2x^2 - 4xy + 5y^2 = 36$ may be written in the form $(x')^2 + 6(y')^2 = 36$ relative to a new coordinate system with the x'- and y'-axes in the directions of the first and second elements of β, respectively. It is clear that this equation represents an ellipse (see Figure 6.3). Note that the matrix P above has the form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

where $\theta = \cos^{-1}\frac{2}{\sqrt{5}} \approx 26.6°$. So P is the matrix representation of a rotation of R^2 through the angle θ. Thus the change of variable $X = PX'$ can be accomplished by this rotation of the x- and y-axes. There is another possibility

for P, however. If the eigenvector of A corresponding to the eigenvalue 6 is taken to be $(1, -2)$ instead of $(-1, 2)$, then we obtain the matrix

$$\begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}$$

which is the matrix representation of a rotation through the angle $\theta = \sin^{-1}\left(-\frac{2}{\sqrt{5}}\right) \approx -63.4°$. This possibility produces the same ellipse as the one in Figure 6.3 but interchanges the names of the x'- and y'-axes.

EXERCISES

1. Label the following statements as being true or false. Assume that the underlying inner product spaces are finite-dimensional.
   (a) Every unitary operator is normal.
   (b) Every orthogonal operator is diagonalizable.
   (c) A matrix is unitary if and only if it is invertible.
   (d) If two matrices are unitarily equivalent, then they are also similar.
   (e) The sum of unitary matrices is unitary.
   (f) The adjoint of a unitary operator is unitary.
   (g) If T is an orthogonal operator on V, then $[\mathsf{T}]_\beta$ is an orthogonal matrix for any ordered basis β for V.
   (h) If all the eigenvalues of a linear operator are 1, then the operator must be unitary or orthogonal.
   (i) A linear operator may preserve the norm but not the inner product.

2. For each of the following matrices A, find an orthogonal or unitary matrix P and a diagonal matrix D such that $P^*AP = D$.

   (a) $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ (b) $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ (c) $\begin{pmatrix} 2 & 3-3i \\ 3+3i & 5 \end{pmatrix}$

   (d) $\begin{pmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{pmatrix}$ (e) $\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$

3. Prove that the composite of unitary [orthogonal] operators is unitary [orthogonal].

4. For $z \in C$ define $\mathsf{T}_z\colon C \to C$ by $\mathsf{T}_z(u) = zu$. Characterize those z for which T_z is normal, self-adjoint, or unitary.

5. Which of the following pairs of matrices are unitarily equivalent?

(a) $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ (b) $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}$

(c) $\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

(d) $\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -i \end{pmatrix}$

(e) $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$

6. Let V be the inner product space of complex-valued continuous functions on $[0, 1]$ with the inner product

$$\langle f, g\rangle = \int_0^1 f(t)\overline{g(t)}\,dt.$$

Let $h \in \mathsf{V}$, and define $\mathsf{T}\colon \mathsf{V} \to \mathsf{V}$ by $\mathsf{T}(f) = hf$. Prove that T is a unitary operator if and only if $|h(t)| = 1$ for $0 \le t \le 1$.

7. Prove that if T is a unitary operator on a finite-dimensional inner product space, then T has a unitary *square root*; that is, there exists a unitary operator U such that $\mathsf{T} = \mathsf{U}^2$.

8. Let T be a self-adjoint linear operator on a finite-dimensional inner product space V. Prove that $(\mathsf{T}+i\mathsf{I})(\mathsf{T}-i\mathsf{I})^{-1}$ is unitary using Exercise 9 of Section 6.4.

9. Let U be a linear operator on a finite-dimensional inner product space V. If $\|\mathsf{U}(x)\| = \|x\|$ for all x in some orthonormal basis for V, must U be unitary? Prove or give a counterexample.

10. Let A be an $n \times n$ real symmetric or complex normal matrix. Prove that

$$\operatorname{tr}(A) = \sum_{i=1}^{n} \lambda_i \quad \text{and} \quad \operatorname{tr}(A^*A) = \sum_{i=1}^{n} |\lambda_i|^2,$$

where the λ_i's are the (not necessarily distinct) eigenvalues of A.

11. Find an orthogonal matrix whose first row is $(\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$.

12. Let A be an $n \times n$ real symmetric or complex normal matrix. Prove that

$$\det(A) = \prod_{i=1}^{n} \lambda_i,$$

where the λ_i's are the (not necessarily distinct) eigenvalues of A.

13. Suppose that A and B are diagonalizable matrices. Prove or disprove that A is similar to B if and only if A and B are unitarily equivalent.

14. Let U be a unitary operator on an inner product space V, and let W be a finite-dimensional U-invariant subspace of V. Prove

(a) $\mathsf{U}(\mathsf{W}) = \mathsf{W}$;
(b) $\mathsf{W}^\perp$ is U-invariant.

Contrast (b) with Exercise 15.

15. Find an example of a unitary operator U on an inner product space and a U-invariant subspace W such that $\mathsf{W}^\perp$ is not U-invariant.

16. Prove that a matrix that is both unitary and upper triangular must be a diagonal matrix.

17. Show that "is unitarily equivalent to" is an equivalence relation on $\mathsf{M}_{n\times n}(C)$.

18. Let W be a finite-dimensional subspace of an inner product space V. By Theorem 6.7 and the exercises of Section 1.3, $\mathsf{V} = \mathsf{W} \oplus \mathsf{W}^\perp$. Define $\mathsf{U}\colon \mathsf{V} \to \mathsf{V}$ by $\mathsf{U}(v_1 + v_2) = v_1 - v_2$, where $v_1 \in \mathsf{W}$ and $v_2 \in \mathsf{W}^\perp$. Prove that U is a self-adjoint unitary operator.

19. Let V be a finite-dimensional inner product space. A linear operator U on V is called a **partial isometry** if there exists a subspace W of V such that $\|\mathsf{U}(x)\| = \|x\|$ for all $x \in \mathsf{W}$ and $\mathsf{U}(x) = 0$ for all $x \in \mathsf{W}^\perp$. Observe that W need *not* be U-invariant. Suppose that U is such an operator and $\{v_1, \ldots, v_k\}$ is an orthonormal basis for W. Prove the following.

(a) $\langle \mathsf{U}(x), \mathsf{U}(y)\rangle = \langle x, y\rangle$ for all $x, y \in \mathsf{W}$. *Hint:* Use Exercise 20 of Section 6.1.
(b) $\{\mathsf{U}(v_1), \ldots, \mathsf{U}(v_k)\}$ is an orthonormal basis for $\mathsf{R}(\mathsf{U})$.
(c) There exists an orthonormal basis γ for V such that the first k columns of $[\mathsf{U}]_\gamma$ form an orthonormal set and the remaining columns are zero.
(d) Let $\{w_1, \ldots, w_j\}$ be an orthonormal basis for $\mathsf{R}(\mathsf{U})^\perp$. Let $\beta = \{\mathsf{U}(v_1), \ldots, \mathsf{U}(v_k), w_1, \ldots, w_j\}$. Then β is an orthonormal basis for V.
(e) Let T be the linear operator on V that satisfies $\mathsf{T}(\mathsf{U}(v_i)) = v_i$ $(1 \le i \le k)$ and $\mathsf{T}(w_i) = 0$ $(1 \le i \le j)$. Prove that T is well-defined and that $\mathsf{T} = \mathsf{U}^*$. *Hint:* Show that $\langle \mathsf{U}(x), y\rangle = \langle x, \mathsf{T}(y)\rangle$ for all $x, y \in \beta$. There are four cases.
(f) Prove that U^* is a partial isometry.

This exercise is continued in Exercise 9 of Section 6.6.

20. Let A and B be $n \times n$ matrices that are unitarily equivalent.

(a) Prove that $\mathrm{tr}(A^*A) = \mathrm{tr}(B^*B)$.
(b) Use (a) to prove that

$$\sum_{i,j=1}^{n} |A_{ij}|^2 = \sum_{i,j=1}^{n} |B_{ij}|^2.$$

21. Find new coordinates x', y' so that the following quadratic forms can be written as $\lambda_1(x')^2 + \lambda_2(y')^2$.

(a) $x^2 + 4xy + y^2$
(b) $2x^2 + 2xy + 2y^2$
(c) $x^2 - 12xy - 4y^2$
(d) $3x^2 + 2xy + 3y^2$
(e) $x^2 - 2xy + y^2$

22. Consider the expression X^tAX, where $X^t = (x, y, z)$ and A is as defined in Exercise 2(e). Find a change of coordinates x', y', z' so that the expression above can be written in the form $\lambda_1(x')^2 + \lambda_2(y')^2 + \lambda_3(z')^2$.

23. Let $w_1, \ldots, w_n$ be linearly independent vectors in F^n, and let $u_1, \ldots, u_n$ be the orthogonal vectors obtained from $w_1, \ldots, w_n$ by the Gram-Schmidt orthogonalization process. Let $v_1, \ldots, v_n$ be the orthonormal basis obtained by setting

$$v_k = \frac{1}{\|u_k\|} u_k \qquad \text{for all } k.$$

(a) Solving (1) in Section 6.2 for w_k in terms of v_k, show that

$$w_k = \|u_k\| v_k + \sum_{j=1}^{k-1} \langle w_k, v_j \rangle\, v_j \quad (1 \le k \le n).$$

(b) Let A and Q denote the $n \times n$ matrices in which the kth columns are w_k and v_k, respectively. Define $R \in \mathsf{M}_{n\times n}(F)$ by

$$R_{jk} = \begin{cases} \|u_j\| & \text{if } j = k \\ \langle w_k, v_j \rangle & \text{if } j < k \\ 0 & \text{if } j > k. \end{cases}$$

Prove $A = QR$.

(c) Compute Q and R as in (b) for the 3×3 matrix whose columns are the vectors w_1, w_2, w_3, respectively, in Example 4 of Section 6.2.

(d) Since Q is unitary [orthogonal] and R is upper triangular in (b), we have shown that every invertible matrix is the product of a unitary [orthogonal] matrix and an upper triangular matrix. Suppose that $A \in \mathsf{M}_{n\times n}(F)$ is invertible and $A = Q_1R_1 = Q_2R_2$, where $Q_1, Q_2 \in \mathsf{M}_{n\times n}(F)$ are unitary and $R_1, R_2 \in \mathsf{M}_{n\times n}(F)$ are upper triangular. Prove that $D = R_2R_1^{-1}$ is a unitary diagonal matrix. *Hint:* Use Exercise 16.

(e) The QR factorization described in (b) provides an orthogonalization method for solving a linear system $Ax = b$ when A is invertible. Decompose A to QR, by the Gram-Schmidt process or other means, where Q is unitary and R is upper triangular. Then $QRx = b$, and hence $Rx = Q^*b$. This last system can be easily solved since R is upper triangular.

At one time, because of its great stability, this method for solving large systems of linear equations with a computer was being advocated as a better method than Gaussian elimination even though it requires about three times as much work. (Later, however, J. H. Wilkinson showed that if Gaussian elimination is done "properly," then it is nearly as stable as the orthogonalization method.)

Use the orthogonalization method and (c) to solve the system

$$\begin{aligned} x_1 + 2x_2 + 2x_3 &= 1 \\ x_1 \qquad\;\; + 2x_3 &= 11 \\ x_2 + x_3 &= -1. \end{aligned}$$

24. Suppose that β and γ are ordered bases for an n-dimensional real [complex] inner product space V. Prove that if Q is an orthogonal [unitary] $n \times n$ matrix that changes γ-coordinates into β-coordinates, then β is orthonormal if and only if γ is orthonormal.

25. Let V be a finite-dimensional complex [real] inner product space and let u be a unit vector in V. Define the **Householder** operator $\mathsf{H}_u \colon \mathsf{V} \to \mathsf{V}$ by $\mathsf{H}_u(x) = x - 2\langle x, u\rangle\, u$ for all $x \in \mathsf{V}$. Prove the following.

(a) H_u is linear.
(b) $\mathsf{H}_u(x) = x$ if and only if x is orthogonal to u.
(c) $\mathsf{H}_u(u) = -u$.
(d) $\mathsf{H}_u^* = \mathsf{H}_u$ and $\mathsf{H}_u^2 = \mathsf{I}$, and hence H_u is a unitary [orthogonal] operator on V.

(Note: If V is a real inner product space, then in the language of Section 6.10, H_u is a "reflection.")

26. Let V be a finite-dimensional inner product space over F. Let $x \neq y$ be nonzero vectors in V such that $\|x\| = \|y\|$.

(a) If $F = C$, prove that there exists a unit vector u in V and a complex number θ with $|\theta| = 1$ such that $\mathsf{H}_u(x) = \theta y$ (as defined in Exercise 25). *Hint:* Choose θ so that $\langle x, \theta y\rangle$ is real and set $u = \frac{\theta}{\|x-y\|}(x-y)$.

(b) If $F = R$, prove that there exists a unit vector u in V such that $\mathsf{H}_u(x) = y$.

6.6 ORTHOGONAL PROJECTIONS AND THE SPECTRAL THEOREM

In this section we rely heavily on Theorems 6.16 and 6.17 to develop an elegant representation of a normal operator (if $F = C$) or a self-adjoint operator (if $F = R$) T on a finite-dimensional inner product space. We prove that such an operator can be written in the form $\lambda_1\mathsf{T}_1 + \cdots + \lambda_k\mathsf{T}_k$, where $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of T and $\mathsf{T}_1, \ldots, \mathsf{T}_k$ are *orthogonal projections.* We must first develop some results about these special projections.

We assume that the reader is familiar with the results about direct sums developed at the end of Section 5.2. The special case where V is a direct sum of two subspaces is considered in the exercises of Section 1.3.

Recall from the exercises of Section 2.1 that if $\mathsf{V} = \mathsf{W}_1 \oplus \mathsf{W}_2$, then a linear operator T on V is the **projection on W_1 along W_2** if whenever $x = x_1 + x_2$, with $x_1 \in \mathsf{W}_1$ and $x_2 \in \mathsf{W}_2$, we have $\mathsf{T}(x) = x_1$. By Exercise 24 of Section 2.1, we have

$$\mathsf{R}(\mathsf{T}) = \mathsf{W}_1 = \{x \in \mathsf{V}\colon \mathsf{T}(x) = x\} \quad \text{and} \quad \mathsf{N}(\mathsf{T}) = \mathsf{W}_2.$$

So $\mathsf{V} = \mathsf{R}(\mathsf{T}) \oplus \mathsf{N}(\mathsf{T})$. Thus there is no ambiguity if we refer to T as a "projection on W_1" or simply as a "projection." In fact, it can be shown (see Exercise 16 of Section 2.3) that T is a projection if and only if $\mathsf{T} = \mathsf{T}^2$. Because $\mathsf{V} = \mathsf{W}_1 \oplus \mathsf{W}_2 = \mathsf{W}_1 \oplus \mathsf{W}_3$ does *not* imply that $\mathsf{W}_2 = \mathsf{W}_3$, we see that W_1 does not uniquely determine T. For an *orthogonal* projection T, however, T is uniquely determined by its range.

Definition. *Let* V *be an inner product space, and let* $\mathsf{T}\colon \mathsf{V} \to \mathsf{V}$ *be a projection. We say that* T *is an* ***orthogonal projection*** *if* $\mathsf{R}(\mathsf{T})^\perp = \mathsf{N}(\mathsf{T})$ *and* $\mathsf{N}(\mathsf{T})^\perp = \mathsf{R}(\mathsf{T})$.

Note that by Exercise 12(c) of Section 6.2 if V is finite-dimensional, we need only assume that one of the conditions above holds. For example, if $\mathsf{R}(\mathsf{T})^\perp = \mathsf{N}(\mathsf{T})$, then $\mathsf{R}(\mathsf{T}) = \mathsf{R}(\mathsf{T})^{\perp\perp} = \mathsf{N}(\mathsf{T})^\perp$.

Now assume that W is a finite-dimensional subspace of an inner product space V. In the notation of Proposition 6.6, we can define a function $\mathsf{T}\colon \mathsf{V} \to \mathsf{V}$ by $\mathsf{T}(y) = u$. It is easy to show that T is an orthogonal projection on W. We can say even more—there exists exactly one orthogonal projection on W. For if T and U are orthogonal projections on W, then $\mathsf{R}(\mathsf{T}) = \mathsf{W} = \mathsf{R}(\mathsf{U})$. Hence $\mathsf{N}(\mathsf{T}) = \mathsf{R}(\mathsf{T})^\perp = \mathsf{R}(\mathsf{U})^\perp = \mathsf{N}(\mathsf{U})$, and since all projections are uniquely determined by their range and null space, we have that $\mathsf{T} = \mathsf{U}$. We call T the **orthogonal projection** on W.

To understand the geometric difference between an arbitrary projection on W and the orthogonal projection on W, let $\mathsf{V} = \mathsf{R}^2$ and $\mathsf{W} = \text{span}\{(1,1)\}$. Define U and T as in Figure 6.4, where $\mathsf{T}(v)$ is the foot of a perpendicular from v on the line $y = x$ and $\mathsf{U}(a_1, a_2) = (a_1, a_1)$. Then T is the orthogonal projection on W, and U is a projection on W that is not orthogonal. Note that $v - \mathsf{T}(v) \in \mathsf{W}^\perp$, whereas $v - \mathsf{U}(v) \notin \mathsf{W}^\perp$.

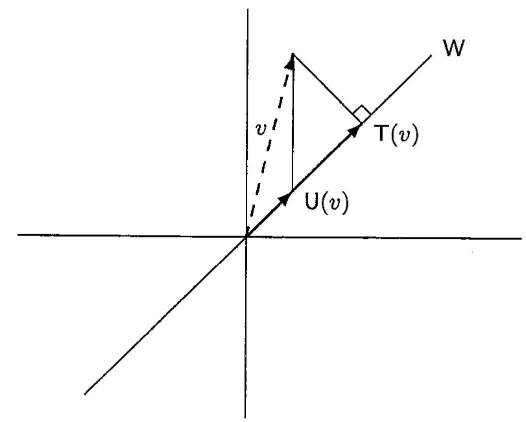

Figure 6.4

From Figure 6.4 we see that $\mathsf{T}(v)$ is the "best approximation in W to v"; that is, if $w \in \mathsf{W}$, then $\|w - v\| \geq \|\mathsf{T}(v) - v\|$. In fact, this approximation property characterizes T. These results follow immediately from the corollary to Proposition 6.6.

As an application to Fourier analysis, recall the inner product space H and the orthonormal set S in Example 9 of Section 6.1. Define a **trigonometric polynomial of degree** n to be a function $g \in \mathsf{H}$ of the form

$$g(t) = \sum_{j=-n}^{n} a_j f_j(t) = \sum_{j=-n}^{n} a_j e^{ijt},$$

where a_n or a_{-n} is nonzero.

Let $f \in \mathsf{H}$. We show that the best approximation to f by a trigonometric polynomial of degree less than or equal to n is the trigonometric polynomial whose coefficients are the Fourier coefficients of f relative to the orthonormal set S. For this result, let $\mathsf{W} = \text{span}(\{f_j \colon |j| \le n\})$, and let T be the orthogonal projection on W. The corollary to Proposition 6.6 tells us that the best approximation to f by an element of W is

$$\mathsf{T}(f) = \sum_{j=-n}^{n} \langle f, f_j \rangle f_j.$$

An algebraic characterization of orthogonal projections follows in the next theorem.

Theorem 6.23. *Let V be an inner product space, and let T be a linear operator on V. Then T is an orthogonal projection if and only if T has an adjoint T^* and $\mathsf{T}^2 = \mathsf{T} = \mathsf{T}^*$.*

Proof. Suppose that T is an orthogonal projection. Since $\mathsf{T}^2 = \mathsf{T}$ because T is a projection, we need only show that T^* exists and $\mathsf{T} = \mathsf{T}^*$. Now $\mathsf{V} = \mathsf{R}(\mathsf{T}) \oplus \mathsf{N}(\mathsf{T})$ and $\mathsf{R}(\mathsf{T})^\perp = \mathsf{N}(\mathsf{T})$. Let $x, y \in \mathsf{V}$. Then $x = x_1 + x_2$ and $y = y_1 + y_2$, where $x_1, y_1 \in \mathsf{R}(\mathsf{T})$ and $x_2, y_2 \in \mathsf{N}(\mathsf{T})$. Hence

$$\langle x, \mathsf{T}(y) \rangle = \langle x_1 + x_2, y_1 \rangle = \langle x_1, y_1 \rangle + \langle x_2, y_1 \rangle = \langle x_1, y_1 \rangle$$

and

$$\langle \mathsf{T}(x), y \rangle = \langle x_1, y_1 + y_2 \rangle = \langle x_1, y_1 \rangle + \langle x_1, y_2 \rangle = \langle x_1, y_1 \rangle.$$

So $\langle x, \mathsf{T}(y) \rangle = \langle \mathsf{T}(x), y \rangle$ for all $x, y \in \mathsf{V}$, and thus T^* exists and $\mathsf{T} = \mathsf{T}^*$.

Now suppose that $\mathsf{T}^2 = \mathsf{T} = \mathsf{T}^*$. Then T is a projection by Exercise 16 of Section 2.3, and hence we must show that $\mathsf{R}(\mathsf{T}) = \mathsf{N}(\mathsf{T})^\perp$ and $\mathsf{R}(\mathsf{T})^\perp = \mathsf{N}(\mathsf{T})$. Let $x \in \mathsf{R}(\mathsf{T})$ and $y \in \mathsf{N}(\mathsf{T})$. Then $x = \mathsf{T}(x) = \mathsf{T}^*(x)$, and so

$$\langle x, y \rangle = \langle \mathsf{T}^*(x), y \rangle = \langle x, \mathsf{T}(y) \rangle = \langle x, \mathit{0} \rangle = 0.$$

Therefore, $x \in \mathsf{N}(\mathsf{T})^\perp$, from which it follows that $\mathsf{R}(\mathsf{T}) \subseteq \mathsf{N}(\mathsf{T})^\perp$.

Let $y \in \mathsf{N}(\mathsf{T})^\perp$. We must show that $y \in \mathsf{R}(\mathsf{T})$, that is, $\mathsf{T}(y) = y$. Now

$$\begin{aligned} \|y - \mathsf{T}(y)\|^2 &= \langle y - \mathsf{T}(y), y - \mathsf{T}(y) \rangle \\ &= \langle y, y - \mathsf{T}(y) \rangle - \langle \mathsf{T}(y), y - \mathsf{T}(y) \rangle. \end{aligned}$$

Since $y - \mathsf{T}(y) \in \mathsf{N}(\mathsf{T})$, the first term must equal zero. But also

$$\langle \mathsf{T}(y), y - \mathsf{T}(y) \rangle = \langle y, \mathsf{T}^*(y - \mathsf{T}(y)) \rangle = \langle y, \mathsf{T}(y - \mathsf{T}(y)) \rangle = \langle y, \mathit{0} \rangle = 0.$$

Thus $y - \mathsf{T}(y) = \mathit{0}$; that is, $y = \mathsf{T}(y) \in \mathsf{R}(\mathsf{T})$. Hence $\mathsf{R}(\mathsf{T}) = \mathsf{N}(\mathsf{T})^\perp$.

Using the results above, we have that $\mathsf{R}(\mathsf{T})^\perp = \mathsf{N}(\mathsf{T})^{\perp\perp} \supseteq \mathsf{N}(\mathsf{T})$ (by Exercise 12(b) of Section 6.2). Now suppose that $x \in \mathsf{R}(\mathsf{T})^\perp$. For any $y \in \mathsf{V}$, we have $\langle \mathsf{T}(x), y \rangle = \langle x, \mathsf{T}^*(y) \rangle = \langle x, \mathsf{T}(y) \rangle = 0$. So $\mathsf{T}(x) = \mathit{0}$, and thus $x \in \mathsf{N}(\mathsf{T})$. Hence $\mathsf{R}(\mathsf{T})^\perp = \mathsf{N}(\mathsf{T})$. ∎

Let V be a finite-dimensional inner product space, W be a subspace of V, and T be the orthogonal projection on W. We may choose an orthonormal basis $\beta = \{v_1, \ldots, v_n\}$ for V such that $\{v_1, \ldots, v_k\}$ is a basis for W. Then $[\mathsf{T}]_\beta$ is a diagonal matrix with ones as the first k diagonal entries and zeros elsewhere. In fact, $[\mathsf{T}]_\beta$ has the form

$$\begin{pmatrix} I_k & O_1 \\ O_2 & O_3 \end{pmatrix}.$$

If U is any projection on W, we may choose a basis γ for V such that $[\mathsf{U}]_\gamma$ has the form above; however γ is not necessarily orthonormal.

We are now ready for the principal theorem of this section.

Theorem 6.24 (The Spectral Theorem). *Suppose that T is a linear operator on a finite-dimensional inner product space V over F with the distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. Assume that T is normal if $F = C$ and that T is self-adjoint if $F = R$. For each i ($1 \le i \le k$) let W_i be the eigenspace of T corresponding to the eigenvalue λ_i, and let T_i be the orthogonal projection on W_i. Then the following are true.*

(a) $\mathsf{V} = \mathsf{W}_1 \oplus \cdots \oplus \mathsf{W}_k$.
(b) *If W_i' denotes the direct sum of the subspaces W_j, $j \ne i$, then $\mathsf{W}_i^\perp = \mathsf{W}_i'$.*
(c) $\mathsf{T}_i\mathsf{T}_j = \delta_{ij}\mathsf{T}_i$ *for* $1 \le i, j \le k$.
(d) $\mathsf{I} = \mathsf{T}_1 + \cdots + \mathsf{T}_k$.
(e) $\mathsf{T} = \lambda_1\mathsf{T}_1 + \cdots + \lambda_k\mathsf{T}_k$.

Proof. (a) By Theorems 6.16 and 6.17, T is diagonalizable; so

$$\mathsf{V} = \mathsf{W}_1 \oplus \cdots \oplus \mathsf{W}_k$$

by Theorem 5.16.

(b) If $x \in \mathsf{W}_i$ and $y \in \mathsf{W}_j$ for some $i \ne j$, then $\langle x, y \rangle = 0$ by Theorem 6.15(d). It follows easily from this that $\mathsf{W}_i' \subseteq \mathsf{W}_i^\perp$. Now from (a) we have that

$$\dim(\mathsf{W}_i') = \sum_{j \ne i} \dim(\mathsf{W}_j) = \dim(\mathsf{V}) - \dim(\mathsf{W}_i).$$

On the other hand we have that $\dim(\mathsf{W}_i^\perp) = \dim(\mathsf{V}) - \dim(\mathsf{W}_i)$ by Theorem 6.7(c). Hence $\mathsf{W}_i' = \mathsf{W}_i^\perp$, proving (b).

(c) The proof of (c) is left as an exercise.

(d) Since T_i is the orthogonal projection on W_i, we have from (b) that $\mathsf{N}(\mathsf{T}_i) = \mathsf{R}(\mathsf{T}_i)^\perp = \mathsf{W}_i^\perp = \mathsf{W}_i'$. Hence for $x \in \mathsf{V}$ we have that $x = x_1 + \cdots + x_k$, where $x_i \in \mathsf{W}_i$ and $\mathsf{T}_i(x) = x_i$, proving (d).

(e) For $x \in \mathsf{V}$ write $x = x_1 + \cdots + x_k$, where $x_i \in \mathsf{W}_i$. Then

$$\begin{aligned} \mathsf{T}(x) &= \mathsf{T}(x_1) + \ldots + \mathsf{T}(x_k) = \lambda_1 x_1 + \cdots + \lambda_k x_k \\ &= \lambda_1\mathsf{T}_1(x) + \cdots + \lambda_k\mathsf{T}_k(x) = (\lambda_1\mathsf{T}_1 + \cdots + \lambda_k\mathsf{T}_k)(x). \end{aligned}$$ ∎

The set $\{\lambda_1, \ldots, \lambda_k\}$ of eigenvalues of T is called the **spectrum** of T, the sum $\mathsf{I} = \mathsf{T}_1 + \cdots + \mathsf{T}_k$ in (d) is called the **resolution of the identity operator induced by** T, and the sum $\mathsf{T} = \lambda_1\mathsf{T}_1 + \cdots + \lambda_k\mathsf{T}_k$ in (e) is called the **spectral decomposition** of T. Since the distinct eigenvalues of T are uniquely determined (up to order) by the subspaces W_i (and hence by the orthogonal projections T_i), the spectral decomposition of T is unique.

With the notation above, let β be the union of orthonormal bases of the W_i's and let $m_i = \dim(\mathsf{W}_i)$. (Thus m_i is the multiplicity of λ_i.) Then $[\mathsf{T}]_\beta$ has the form

$$\begin{pmatrix} \lambda_1 I_{m_1} & O & \cdots & O \\ O & \lambda_2 I_{m_2} & \cdots & O \\ \vdots & \vdots & & \vdots \\ O & O & \cdots & \lambda_k I_{m_k} \end{pmatrix};$$

that is, $[\mathsf{T}]_\beta$ is a diagonal matrix in which the diagonal entries are the eigenvalues λ_i of T, and each λ_i is repeated m_i times. If $\lambda_1\mathsf{T}_1 + \cdots + \lambda_k\mathsf{T}_k$ is the spectral decomposition of T, then it follows (from Exercise 7) that $g(\mathsf{T}) = g(\lambda_1)\mathsf{T}_1 + \cdots + g(\lambda_k)\mathsf{T}_k$ for any polynomial g. This fact is used below.

We now list several interesting corollaries of the spectral theorem; many more results are found in the exercises. For what follows we assume that T is a linear operator on a finite-dimensional inner product space V over F.

Corollary 1. *If $F = C$, then* T *is normal if and only if* $\mathsf{T}^* = g(\mathsf{T})$ *for some polynomial* g.

Proof. Suppose first that T is normal. Let $\mathsf{T} = \lambda_1\mathsf{T}_1 + \cdots + \lambda_k\mathsf{T}_k$ be the spectral decomposition of T. Taking the adjoint of both sides of the equation above, we have $\mathsf{T}^* = \overline{\lambda}_1\mathsf{T}_1 + \cdots + \overline{\lambda}_k\mathsf{T}_k$ since each T_i is self-adjoint. Using the Lagrange interpolation formula (see Section 1.6), we may choose a polynomial g such that $g(\lambda_i) = \overline{\lambda}_i$ for $1 \le i \le k$. Then

$$\begin{aligned} g(\mathsf{T}) &= g(\lambda_1)\mathsf{T}_1 + \cdots + g(\lambda_k)\mathsf{T}_k \\ &= \overline{\lambda}_1\mathsf{T}_1 + \cdots + \overline{\lambda}_k\mathsf{T}_k \\ &= \mathsf{T}^*. \end{aligned}$$

Conversely, if $\mathsf{T}^* = g(\mathsf{T})$ for some polynomial g, then T commutes with T^* since T commutes with every polynomial in T. So T is normal. ∎

Corollary 2. *If $F = C$, then* T *is unitary if and only if* T *is normal and* $|\lambda| = 1$ *for every eigenvalue* λ *of* T.

Proof. Suppose first that T is unitary and hence normal. Let λ be an eigenvalue of T with x as a corresponding eigenvector. Then $|\lambda| \cdot \|x\| = \|\lambda x\| = \|\mathsf{T}(x)\| = \|x\|$, and hence $|\lambda| = 1$ since $x \neq 0$.

Now suppose that $|\lambda| = 1$ for every eigenvalue λ of T, and let $\mathsf{T} = \lambda_1\mathsf{T}_1 + \cdots + \lambda_k\mathsf{T}_k$ be the spectral decomposition of T. Then by (c) of the spectral theorem

$$\begin{aligned} \mathsf{T}\mathsf{T}^* &= (\lambda_1\mathsf{T}_1 + \cdots + \lambda_k\mathsf{T}_k)(\overline{\lambda}_1\mathsf{T}_1 + \cdots + \overline{\lambda}_k\mathsf{T}_k) \\ &= |\lambda_1|^2\mathsf{T}_1 + \cdots + |\lambda_k|^2\mathsf{T}_k \\ &= \mathsf{T}_1 + \cdots + \mathsf{T}_k \\ &= \mathsf{I}. \end{aligned}$$

Hence T is unitary. ∎

Corollary 3. *If $F = C$ and* T *is normal, then* T *is self-adjoint if and only if every eigenvalue of* T *is real.*

Proof. Let $\mathsf{T} = \lambda_1\mathsf{T}_1 + \cdots + \lambda_k\mathsf{T}_k$ be the spectral decomposition of T. Suppose that every eigenvalue of T is real. Then $\mathsf{T}^* = \overline{\lambda}_1\mathsf{T}_1 + \cdots + \overline{\lambda}_k\mathsf{T}_k = \lambda_1\mathsf{T}_1 + \cdots + \lambda_k\mathsf{T}_k = \mathsf{T}$.

The converse has been proved in the lemma to Theorem 6.17. ∎

Corollary 4. *Let* T *be as in the spectral theorem with spectral decomposition* $\mathsf{T} = \lambda_1\mathsf{T}_1 + \cdots + \lambda_k\mathsf{T}_k$. *Then each* T_j *is a polynomial in* T.

Proof. Choose a polynomial g_j $(1 \le j \le k)$ such that $g_j(\lambda_i) = \delta_{ij}$. Then $g_j(\mathsf{T}) = g_j(\lambda_1)\mathsf{T}_1 + \cdots + g_j(\lambda_k)\mathsf{T}_k = \delta_{1j}\mathsf{T}_1 + \cdots + \delta_{kj}\mathsf{T}_k = \mathsf{T}_j$. ∎

EXERCISES

1. Label the following statements as being true or false. Assume that the underlying inner product spaces are finite-dimensional.
   (a) All projections are self-adjoint.
   (b) An orthogonal projection is uniquely determined by its range.
   (c) Every self-adjoint operator is a linear combination of orthogonal projections.
   (d) If T is a projection on W, then $\mathsf{T}(x)$ is the vector in W that is closest to x.
   (e) Every orthogonal projection is a unitary operator.
2. Let $\mathsf{V} = \mathsf{R}^2$, $\mathsf{W} = \mathrm{span}(\{(1,2)\})$, and β be the standard ordered basis for V. Compute $[\mathsf{T}]_\beta$, where T is the orthogonal projection on W. Do the same for $\mathsf{V} = \mathsf{R}^3$ and $\mathsf{W} = \mathrm{span}(\{(1,0,1)\})$.
3. For each of the matrices A in Exercise 2 of Section 6.5:

   (1) Verify that L_A possesses a spectral decomposition.

(2) For each eigenvalue of L_A, explicitly define the orthogonal projection on the corresponding eigenspace.

(3) Verify your results using the spectral theorem.

4. Let W be a finite-dimensional subspace of an inner product space V. Show that if T is the orthogonal projection on W, then I − T is the orthogonal projection on $\mathsf{W}^\perp$.

5. Let T be a linear operator on a finite-dimensional inner product space V.

(a) If T is an orthogonal projection, prove that $\|\mathsf{T}(x)\| \leq \|x\|$ for all $x \in \mathsf{V}$. Give an example of a projection for which this inequality does not hold. What can be concluded about a projection for which the inequality is actually an equality for all $x \in \mathsf{V}$?

(b) Suppose that T is a projection such that $\|\mathsf{T}(x)\| \leq \|x\|$ for $x \in \mathsf{V}$. Prove that T is an orthogonal projection.

6. Let T be a normal operator on a finite-dimensional inner product space. Prove that if T is a projection, then T is also an orthogonal projection.

7. Let T be a normal operator on a finite-dimensional complex inner product space V. Use the spectral decomposition $\lambda_1\mathsf{T}_1 + \cdots + \lambda_k\mathsf{T}_k$ of T to prove the following.

(a) If g is a polynomial, then

$$g(\mathsf{T}) = \sum_{i=1}^{k} g(\lambda_i)\mathsf{T}_i.$$

(b) If $\mathsf{T}^n = \mathsf{T}_0$ for some n, then $\mathsf{T} = \mathsf{T}_0$.

(c) Let U be a linear operator on V. Then U commutes with T if and only if U commutes with each T_i.

(d) There exists a normal operator U on V such that $\mathsf{U}^2 = \mathsf{T}$.

(e) T is invertible if and only if $\lambda_i \neq 0$ for $1 \leq i \leq k$.

(f) T is a projection if and only if every eigenvalue of T is 1 or 0.

(g) $\mathsf{T} = -\mathsf{T}^*$ if and only if every λ_i is an imaginary number.

8. Use Corollary 1 of the spectral theorem to show that if T is a normal operator on a complex finite-dimensional inner product space and U is a linear operator that commutes with T, then U commutes with T^*.

9. Referring to Exercise 19 of Section 6.5, prove the following facts about a partial isometry U.

(a) $\mathsf{U}^*\mathsf{U}$ is an orthogonal projection on W.

(b) $\mathsf{UU}^*\mathsf{U} = \mathsf{U}$.

10. *Simultaneous diagonalization.* Let U and T be normal operators on a finite-dimensional complex inner product space V such that TU = UT. Prove that there exists an orthonormal basis for V consisting of vectors that are eigenvectors of both T and U. *Hint:* Use the hint of Exercise 14 of Section 6.4 along with Exercise 8.

11. Prove (c) of the spectral theorem.

6.7* BILINEAR AND QUADRATIC FORMS

There is a certain class of scalar-valued functions of two variables defined on a vector space that is often considered in the study of such diverse subjects as geometry and multivariable calculus. This is the class of *bilinear forms.* We study the basic properties of this class with a special emphasis on symmetric bilinear forms, and we consider some of its applications to quadratic surfaces and multivariable calculus.

Throughout this section all bases should be regarded as ordered bases.

Bilinear Forms

Definition. *Let V be a vector space over a field F. A function H from the set V × V of ordered pairs of vectors to F is called a* **bilinear form** *on V if H is linear in each variable when the other variable is held fixed, that is, if*

(a) $H(ax_1 + x_2, y) = aH(x_1, y) + H(x_2, y)$ for all $x_1, x_2, y \in \mathsf{V}$ and $a \in F$

(b) $H(x, ay_1 + y_2) = aH(x, y_1) + H(x, y_2)$ for all $x, y_1, y_2 \in \mathsf{V}$ and $a \in F$.

We denote the set of all bilinear forms on V by $\mathcal{B}(\mathsf{V})$. Observe that an inner product on a vector space is a bilinear form if the underlying field is real but that it is not if the underlying field is complex.

Example 1

Define a function $H\colon \mathsf{R}^2 \times \mathsf{R}^2 \to R$ by

$$H\left(\begin{pmatrix}a_1\\a_2\end{pmatrix}, \begin{pmatrix}b_1\\b_2\end{pmatrix}\right) = 2a_1b_1 + 3a_1b_2 + 4a_2b_1 - a_2b_2 \quad \text{for } \begin{pmatrix}a_1\\a_2\end{pmatrix}, \begin{pmatrix}b_1\\b_2\end{pmatrix} \in \mathsf{R}^2.$$

We could verify directly that H is a bilinear form on R^2. However, it is more enlightening and less tedious to observe that if

$$A = \begin{pmatrix}2 & 3\\4 & -1\end{pmatrix}, \quad x = \begin{pmatrix}a_1\\a_2\end{pmatrix}, \quad \text{and} \quad y = \begin{pmatrix}b_1\\b_2\end{pmatrix},$$

then

$$H(x, y) = x^t A y.$$

The bilinearity of H now follows directly from the distributive property of matrix multiplication over matrix addition. ∎