chapter V

Orthogonal and unitary transformations, normal matrices

In this chapter we introduce an inner product based on an arbitrary positive definite symmetric bilinear form, or Hermitian form. On this basis the length of a vector and the concept of orthogonality can be defined. From this point on, we concentrate our attention on bases in which the vectors are mutually orthogonal and each is of length 1, the orthonormal bases. The Gram-Schmidt process for obtaining an orthonormal basis from an arbitrary basis is described.

Isometries are linear transformations which preserve length. They also preserve the inner product and therefore map orthonormal bases onto orthonormal bases. It is shown that a matrix representing an isometry has exactly the same properties as a matrix of transition representing a change of bases from one orthonormal basis to another. If the field of scalars is real, these matrices are said to be orthogonal; and if the field of scalars is complex, they are said to be unitary.

If A is an orthogonal matrix, we show that $A^T = A^{-1}$; and if A is unitary, we show that $A^* = A^{-1}$. Because of this fact a matrix representing a linear transformation and a matrix representing a bilinear form are transformed by exactly the same formula under a change of coordinates provided that the change is from one orthonormal basis to another. This observation unifies the discussions of Chapter III and IV.

The penalty for restricting our attention to orthonormal bases is that there is a corresponding restriction in the linear transformations and bilinear forms that can be represented by diagonal matrices. The necessary and sufficient condition that this be possible, expressed in terms of matrices, is that $A^*A = AA^*$. Matrices with this property are called *normal* matrices. Fortunately, the normal matrices constitute a large class of matrices and they happen to include as special cases most of the types that arise in physical problems.

Up to a certain point we can consider matrices with real coefficients to be special cases of matrices with complex coefficients. However, if we wish to restrict our attention to real vector spaces, then the matrices of transition must also be real. This restriction means that the situation for real vector spaces is not a special case of the situation for complex vector spaces. In particular, there are real normal matrices that are unitary similar to diagonal matrices but not orthogonal similar to diagonal matrices. The necessary and sufficient condition that a real matrix be orthogonal similar to a diagonal matrix is that it be symmetric.

The techniques for finding the diagonal normal form of a normal matrix and the unitary or orthogonal matrix of transition are, for the most part, not new. The eigenvalues and eigenvectors are found as in Chapter III. We show that eigenvectors corresponding to different eigenvalues are automatically orthogonal so all that needs to be done is to make sure that they are of length 1. However, something more must be done in the case of multiple eigenvalues. We are assured that there are enough eigenvectors, but we must make sure they are orthogonal. The Gram-Schmidt process provides the method for finding the necessary orthonormal eigenvectors.

1 | Inner Products and Orthogonal Bases

Even when speaking in abstract terms we have tried to draw an analogy between vector spaces and the geometric spaces we have encountered in 2- and 3-dimensional analytic geometry. For example, we have referred to lines and planes through the origin as subspaces; however, we have nowhere used the concept of distance. Some of the most interesting properties of vector spaces and matrices deal with the concept of distance. So in this chapter we introduce the concept of distance and explore the related properties.

For aesthetic reasons, and to show as clearly as possible that we need not have an a priori concept of distance, we use an approach which will emphasize the arbitrary nature of the concept of distance.

It is customary to restrict attention to the field of real numbers or the field of complex numbers when discussing vector space concepts related to distance. However, we need not be quite that restrictive. The scalar field F must be a subfield of the complex numbers with the property that, if $a \in F$, the conjugate complex $\bar{a}$ is also in F. Such a field is said to be *normal* over its real subfield. The real field and the complex field have this property, but so do many other fields. For most of the important applications of the material to follow the field of scalars is taken to be the real numbers or the field of complex numbers. Although most of the proofs given will be valid for any field normal over its real subfield, it will suffice to think in terms of the two most important cases.

In a vector space V of dimension n over the complex numbers (or a subfield of the complex numbers normal over its real subfield), let f be any fixed positive definite Hermitian form. For the purpose of the following development it does not matter which positive definite Hermitian form is chosen, but it will remain fixed for all the remaining discussion. Since this particular Hermitian form is now fixed, we write (α, β) instead of $f(\alpha, \beta)$. (α, β) is called the *inner product*, or *scalar product*, of α and β.

Since we have chosen a positive definite Hermitian form, $(\alpha, \alpha) \geq 0$ and $(\alpha, \alpha) > 0$ unless $\alpha = 0$. Thus $\sqrt{(\alpha, \alpha)} = \|\alpha\|$ is a well-defined non-negative real number which we call the *length* or *norm* of α. Observe that $\|a\alpha\| = \sqrt{(a\alpha, a\alpha)} = \sqrt{\bar{a}a(\alpha, \alpha)} = |a| \cdot \|\alpha\|$, so that multiplying a vector by a scalar a multiplies its length by $|a|$. We say that the *distance* between two vectors is the norm of their difference; that is, $d(\alpha, \beta) = \|\beta - \alpha\|$. We should like to show that this distance function has the properties we might reasonably expect a distance function to have. But first we have to prove a theorem that has interest of its own and many applications.

Theorem 1.1. *For any vectors* $\alpha, \beta \in V$, $|(\alpha, \beta)| \leq \|\alpha\| \cdot \|\beta\|$. *This inequality is known as Schwarz's inequality.*

PROOF. For t a real number consider the inequality

$$0 \leq \|(\alpha, \beta)t\alpha - \beta\|^2 = |(\alpha, \beta)|^2 \|\alpha\|^2 t^2 - 2t |(\alpha, \beta)|^2 + \|\beta\|^2. \tag{1.1}$$

If $\|\alpha\| = 0$, the fact that this inequality must hold for arbitrarily large t implies that $|(\alpha, \beta)| = 0$ so that Schwarz's inequality is satisfied. If $\|\alpha\| \neq 0$, take $t = 1/\|\alpha\|^2$. Then (1.1) is equivalent to Schwarz's inequality,

$$|(\alpha, \beta)| \leq \|\alpha\| \cdot \|\beta\|. \square \tag{1.2}$$

This proof of Schwarz's inequality does not make use of the assumption that the inner product is positive definite and would remain valid if the inner product were merely semi-definite. Using the assumption that the inner product is positive definite, however, an examination of this proof of Schwarz's inequality would reveal that equality can hold if and only if

$$\beta - \frac{(\alpha, \beta)}{(\alpha, \alpha)}\alpha = 0; \tag{1.3}$$

that is, if and only if β is a multiple of α.

If $\alpha \neq 0$ and $\beta \neq 0$, Schwarz's inequality can be written in the form

$$\frac{|(\alpha, \beta)|}{\|\alpha\| \cdot \|\beta\|} \leq 1. \tag{1.4}$$

In vector analysis the scalar product of two vectors is equal to the product of the lengths of the vectors times the cosine of the angle between them. The inequality (1.4) says, in effect, that in a vector space over the real numbers the ratio $\dfrac{(\alpha, \beta)}{\|\alpha\| \cdot \|\beta\|}$ can be considered to be a cosine. It would be a diversion for us to push this point much further. We do, however, wish to show that $d(\alpha, \beta)$ behaves like a distance function.

Theorem 1.2. *For* $d(\alpha, \beta) = \|\beta - \alpha\|$, *we have,*

(1) $d(\alpha, \beta) = d(\beta, \alpha)$,

(2) $d(\alpha, \beta) \geq 0$ *and* $d(\alpha, \beta) = 0$ *if and only if* $\alpha = \beta$,

(3) $d(\alpha, \beta) \leq d(\alpha, \gamma) + d(\gamma, \beta)$.

PROOF. (1) and (2) are obvious. (3) follows from Schwarz's inequality. To see this, observe that

$$\begin{aligned} \|\alpha + \beta\|^2 &= (\alpha + \beta, \alpha + \beta) \\ &= (\alpha, \alpha) + (\alpha, \beta) + (\beta, \alpha) + (\beta, \beta) \\ &= \|\alpha\|^2 + (\alpha, \beta) + \overline{(\alpha, \beta)} + \|\beta\|^2 \\ &\leq \|\alpha\|^2 + 2\,|(\alpha, \beta)| + \|\beta\|^2 \\ &\leq \|\alpha\|^2 + 2\,\|\alpha\| \cdot \|\beta\| + \|\beta\|^2 = (\|\alpha\| + \|\beta\|)^2. \end{aligned} \tag{1.5}$$

Replacing α by $\gamma - \alpha$ and β by $\beta - \gamma$, we have

$$\|\beta - \alpha\| \leq \|\gamma - \alpha\| + \|\beta - \gamma\|. \square \tag{1.6}$$

(3) is the familiar triangular inequality. It implies that the sum of two small vectors is also small. Schwarz's inequality tells us that the inner product of two small vectors is small. Both of these inequalities are very useful for these reasons.

According to Theorem 12.1 of Chapter IV and the definition of a positive definite Hermitian form, there exists a basis $A = \{\alpha_1, \ldots, \alpha_n\}$ with respect to which the representing matrix is the unit matrix. Thus,

$$(\alpha_i, \alpha_j) = \delta_{ij}. \tag{1.7}$$

Relative to this fixed positive definite Hermitian form, the inner product, every set of vectors that has this property is called an *orthonormal* set. The word "orthonormal" is a combination of the words "orthogonal" and "normal." Two vectors α and β are said to be *orthogonal* if $(\alpha, \beta) = (\beta, \alpha) = 0$. A vector α is *normalized* if it is of length 1; that is, if $(\alpha, \alpha) = 1$. Thus the vectors of an orthonormal set are mutually orthogonal and normalized. The basis A chosen above is an *orthonormal basis*. We shall see that orthonormal bases possess particular advantages for dealing with the properties of a vector space with an inner product. A vector space over the complex numbers with an inner product such as we have defined is called a *unitary space*. A vector space over the real numbers with an inner product is called a *Euclidean space*.

For $\alpha, \beta \in V$, let $\alpha = \sum_{i=1}^{n} x_i\alpha_i$ and $\beta = \sum_{i=1}^{n} y_i\alpha_i$. Then

$$\begin{aligned}(\alpha, \beta) &= \left(\sum_{i=1}^{n} x_i\alpha_i, \sum_{j=1}^{n} y_j\alpha_j\right) \\ &= \sum_{i=1}^{n} \bar{x}_i\left[\sum_{j=1}^{n} y_j(\alpha_i, \alpha_j)\right] \\ &= \sum_{i=1}^{n} \bar{x}_i y_i. \end{aligned} \tag{1.8}$$

If we represent α by the n-tuple $(x_1, \dots, x_n) = X$, and β by the n-tuple $(y_1, \dots, y_n) = Y$, the inner product can be written in the form

$$(\alpha, \beta) = \sum_{i=1}^{n} \bar{x}_i y_i = X^*Y. \tag{1.9}$$

This is a familiar formula in vector analysis where it is also known as the inner product, scalar, or dot product.

Theorem 1.3. *An orthonormal set is linearly independent.*

PROOF. Suppose that $\{\xi_1, \xi_2, \dots\}$ is an orthonormal set and that $\sum_i x_i\xi_i = 0$. Then $0 = (\xi_j, 0) = (\xi_j, \sum_i x_i\xi_i) = \sum_i x_i(\xi_j, \xi_i) = x_j$. Thus the set is linearly independent. □

It is an immediate consequence of Theorem 1.3 that an orthonormal set cannot contain more than n elements.

Since V has at least one orthonormal basis and orthonormal sets are linearly independent, some questions naturally arise. Are there other orthonormal bases? Can an orthonormal set be extended to an orthonormal basis? Can a linearly independent set be modified to form an orthonormal set? For infinite dimensional vector spaces the question of the existence of even one orthonormal basis is a non-trivial question. For finite dimensional vector spaces all these questions have nice answers, and the technique employed in giving these answers is of importance in infinite dimensional vector spaces as well.

Theorem 1.4. *If $A = \{\alpha_1, \dots, \alpha_s\}$ is any linearly independent set whatever in V, there exists an orthonormal set $X = \{\xi_1, \dots, \xi_s\}$ such that $\xi_k = \sum_{i=1}^{k} a_{ik}\alpha_i$.*

PROOF. (The Gram-Schmidt orthonormalization process). Since α_1 is an element of a linearly independent set $\alpha_1 \neq 0$, and therefore $\|\alpha_1\| > 0$.

$$\text{Let } \xi_1 = \frac{1}{\|\alpha_1\|}\alpha_1. \text{ Clearly, } \|\xi_1\| = 1.$$

Suppose, then, $\{\xi_1, \dots, \xi_r\}$ has been found so that it is an orthonormal set and such that each ξ_k is a linear combination of $\{\alpha_1, \dots, \alpha_k\}$. Let

$$\alpha'_{r=1} = \alpha_{r+1} - (\xi_1, \alpha_{r+1})\xi_1 - \cdots - (\xi_r, \alpha_{r+1})\xi_r. \tag{1.10}$$

Then for any ξ_i, $1 \leq i \leq r$, we have

$$(\xi_i, \alpha'_{r+1}) = (\xi_i, \alpha_{r+1}) - (\xi_i, \alpha_{r+1}) = 0. \tag{1.11}$$

Furthermore, since each ξ_k is a linear combination of the $\{\alpha_1, \ldots, \alpha_k\}$, α'_{r+1} is a linear combination of the $\{\alpha_1, \ldots, \alpha_{r+1}\}$. Also, α'_{r+1} is not zero since $\{\alpha_1, \ldots, \alpha_{r+1}\}$ is a linearly independent set and the coefficient of α_{r+1} in the representation of α'_{r+1} is 1. Thus we can define

$$\xi_{r+1} = \frac{1}{\|\alpha'_{r+1}\|} \alpha'_{r+1}. \tag{1.12}$$

Clearly, $\{\xi_1, \ldots, \xi_{r+1}\}$ is an orthonormal set with the desired properties. We can continue in this fashion until we exhaust the elements of A. The set $X = \{\xi_1, \ldots, \xi_s\}$ has the required properties. □

The Gram-Schmidt process is completely effective and the computations can be carried out exactly as they are given in the proof of Theorem 1.4. For example, let $A = \{\alpha_1 = (1, 1, 0, 1), \alpha_2 = (3, 1, 1, -1), \alpha_3 = (0, 1, -1, 1)\}$. Then

$$\xi_1 = \frac{1}{\sqrt{3}}(1, 1, 0, 1),$$

$$\alpha'_2 = (3, 1, 1, -1) - \frac{3}{\sqrt{3}}\frac{1}{\sqrt{3}}(1, 1, 0, 1) = (2, 0, 1, -2),$$

$$\xi_2 = \tfrac{1}{3}(2, 0, 1, -2),$$

$$\alpha'_3 = (0, 1, -1, 1) - \frac{2}{\sqrt{3}}\frac{1}{\sqrt{3}}(1, 1, 0, 1) - \frac{-3}{3}\frac{1}{3}(2, 0, 1, -2)$$

$$= \tfrac{1}{3}(0, 1, -2, -1),$$

$$\xi_3 = \frac{1}{\sqrt{6}}(0, 1, -2, -1).$$

It is easily verified that $\{\xi_1, \xi_2, \xi_3\}$ is an orthonormal set.

Corollary 1.5. *If $A = \{\alpha_1, \ldots, \alpha_n\}$ is a basis of V, the orthonormal set $X = \{\xi_1, \ldots, \xi_n\}$, obtained from A by the application of the Gram-Schmidt process, is an orthonormal basis of V.*

PROOF. Since X is orthonormal it is linearly independent. Since it contains n vectors it also spans V and is a basis. □

Theorem 1.4 and its corollary are used in much the same fashion in which we used Theorem 3.6 of Chapter I to obtain a basis (in this case an orthonormal basis) such that a subset spans a given subspace.

Theorem 1.6. *Given any vector α_1 of length* 1, *there is an orthonormal basis with α_1 as the first element.*

PROOF. Since the set $\{\alpha_1\}$ is linearly independent it can be extended to a basis with α_1 as the first element. Now, when the Gram-Schmidt process is applied, the first vector, being of length 1, is unchanged and becomes the first vector of an orthonormal basis. □

EXERCISES

In the following problems we assume that all n-tuples are representations of their vectors with respect to orthonormal bases.

1. Let $A = \{\alpha_1, \ldots, \alpha_4\}$ be an orthonormal basis of R^4 and let $\alpha, \beta \in V$ be represented by $(1, 2, 3, -1)$ and $(2, 4, -1, 1)$, respectively. Compute (α, β).

2. Let $\alpha = (1, i, 1 + i)$ and $\beta = (i, 1, i - 1)$ be vectors in C^3, where C is the field of complex numbers. Compute (α, β).

3. Show that the set $\{(1, i, 2), (1, i, -1), (1, -i, 0)\}$ is orthogonal in C^3.

4. Show that $(\alpha, 0) = (0, \alpha) = 0$ for all $\alpha \in V$.

5. Show that $\|\alpha + \beta\|^2 + \|\alpha - \beta\|^2 = 2\|\alpha\|^2 + 2\|\beta\|^2$.

6. Show that if the field of scalars is real and $\|\alpha\| = \|\beta\|$, then $\alpha - \beta$ and $\alpha + \beta$ are orthogonal, and conversely.

7. Show that if the field of scalars is real and $\|\alpha + \beta\|^2 = \|\alpha\|^2 + \|\beta\|^2$, then α and β are orthogonal, and conversely.

8. Verify Schwarz's inequality for the vectors α and β in Exercises 1 and 2.

9. The set $\{(1, -1, 1), (2, 0, 1), (0, 1, 1)\}$ is linearly independent, and hence a basis for F^3. Apply the Gram-Schmidt process to obtain an orthonormal basis.

10. Given the basis $\{(1, 0, 1, 0), (1, 1, 0, 0), (0, 1, 1, 1,), (0, 1, 1, 0)\}$ apply the Gram-Schmidt process to obtain an orthonormal basis.

11. Let W be a subspace of V spanned by $\{(0, 1, 1, 0), (0, 5, -3, -2), (-3, -3, 5, -7)\}$. Find an orthonormal basis for W.

12. In the space of real integrable functions let the inner product be defined by

$$\int_{-1}^{1} f(x)g(x)\,dx.$$

Find a polynomial of degree 2 orthogonal to 1 and x. Find a polynomial of degree 3 orthogonal to 1, x, and x^2. Are these two polynomials orthogonal?

13. Let $X = \{\xi_1, \ldots, \xi_m\}$ be a set of vectors in the n-dimensional space V. Consider the matrix $G = [g_{ij}]$ where

$$g_{ij} = (\xi_i, \xi_j).$$

Show that if X is linearly dependent, then the columns of G are also linearly dependent. Show that if X is linearly independent, then the columns of G are also linearly independent. Det G is known as the *Gramian* of the set X. Show that X is linearly dependent if and only if det $G = 0$. Choose an orthonormal basis in V and represent the vectors in X with respect to that basis. Show that G can be represented as the product of an $m \times n$ matrix and an $n \times m$ matrix. Show that det $G \geq 0$.

*2 | Complete Orthonormal Sets

We now develop some properties of orthonormal sets that hold in both finite and infinite dimensional vector spaces. These properties are deep and important in infinite dimensional vector spaces, but in finite dimensional vector spaces they could easily be developed in passing and without special terminology. It is of some interest, however, to borrow the terminology of infinite dimensional vector spaces and to give proofs, where possible, which are valid in infinite as well as finite dimensional vector spaces.

Let $X = \{\xi_1, \xi_2, \ldots\}$ be an orthonormal set and let α be any vector in V. The numbers $\{\alpha_i = (\xi_i, \alpha)\}$ are called the *Fourier coefficients* of α.

There is, first, the question of whether an expression like $\sum_i x_i\xi_i$ has any meaning in cases where infinitely many of the x_i are non-zero. This is a question of the convergence of an infinite series and the problem varies from case to case so that we cannot hope to deal with it in all generality. We have to assume for this discussion that all expressions like $\sum_i x_i\xi_i$ that we write down have meaning.

Theorem 2.1. *The minimum of* $\|\alpha - \sum_i x_i\xi_i\|$ *is attained if and only if all*

$$x_i = (\xi_i, x) = a_i.$$

PROOF.

$$\begin{aligned}
\|\alpha - \sum_i x_i\xi_i\|^2 &= (\alpha - \sum_i x_i\xi_i, \alpha - \sum_i x_i\xi_i) \\
&= (\alpha, \alpha) - \sum_i x_i\bar{a}_i - \sum_i \bar{x}_i a_i + \sum_i \bar{x}_i x_i \\
&= \sum_i \bar{a}_i a_i - \sum_i x_i\bar{a}_i - \sum_i \bar{x}_i a_i + \sum_i \bar{x}_i x_i + (\alpha, \alpha) - \sum_i \bar{a}_i a_i \\
&= \sum_i (\bar{a}_i - \bar{x}_i)(a_i - x_i) + \|\alpha\|^2 - \sum_i \bar{a}_i a_i \\
&= \sum_i |a_i - x_i|^2 + \|\alpha\|^2 - \sum_i |a_i|^2. \qquad (2.1)
\end{aligned}$$

Only the term $\sum_i |a_i - x_i|^2$ depends on the x_i and, being a sum of real squares, it takes on its minimum value of zero if and only if all $x_i = a_i$. □

Theorem 2.1 is valid for any orthonormal set X, whether it is a basis or not. If the norm is used as a criterion of smallness, then the theorem says that the best approximation of α in the form $\sum_i x_i\xi_i$ (using only the $\xi_i \in X$) is obtained if and only if all x_i are the Fourier coefficients.

Theorem 2.2 $\sum_i |a_i|^2 \leq \|\alpha\|^2$. *This inequality is known as Bessel's inequality.*

PROOF. Setting $x_i = a_i$ in equation (2.1) we have

$$\|\alpha\|^2 - \sum_i |a_i|^2 = \|\alpha - \sum_i a_i\xi_i\|^2 \geq 0. \square \tag{2.2}$$

It is desirable to know conditions under which the Fourier coefficients will represent the vector α. This means we would like to have $\alpha = \sum_i a_i\xi_i$. In a finite dimensional vector space the most convenient sufficient condition is that X be an orthonormal basis. In the theory of Fourier series and other orthogonal functions it is generally not possible to establish the validity of an equation like $\alpha = \sum_i a_i\xi_i$ without some modification of what is meant by convergence or a restriction on the set of functions under consideration. Instead, we usually establish a condition known as completeness. An orthonormal set is said to be *complete* if and only if it is not a subset of a larger orthonormal set.

Theorem 2.3. *Let $X = \{\xi_i\}$ be an orthonormal set. The following three conditions are equivalent:*

(1) *For each $\alpha, \beta \in V$,* $$(\alpha, \beta) = \sum_i \overline{(\xi_i, \alpha)}(\xi_i, \beta). \tag{2.3}$$

(2) *For each $\alpha \in V$,* $$\|\alpha\|^2 = \sum_i |(\xi_i, \alpha)|^2. \tag{2.4}$$

(3) *X is complete.*

Equations (2.3) *and* (2.4) *are both known as Parseval's identities.*

PROOF. Assume (1). Then $\|\alpha\|^2 = (\alpha, \alpha) = \sum_i \overline{(\xi_i, \alpha)}(\xi_i, \alpha)$

$$= \sum_i |(\xi_i, \alpha)|^2.$$

Assume (2). If X were not complete, it would be contained in a larger orthonormal set Y. But for any $\alpha_0 \in Y$, $\alpha_0 \notin X$, we would have

$$1 = \|\alpha_0\|^2 = \sum_i |(\xi_i, \alpha_0)|^2 = 0$$

because of (2) and the assumption that Y is orthonormal. Thus X is complete.

Now, assume (3). Let β be any vector in V and consider $\beta' = \beta - \sum_i (\xi_i, \beta)\xi_i$. Then

$$\begin{aligned}(\xi_i, \beta') &= \left(\xi_i, \beta - \sum_j (\xi_j, \beta)\xi_j\right) \\ &= (\xi_i, \beta) - \sum_j (\xi_j, \beta)(\xi_i, \xi_j) \\ &= (\xi_i, \beta) - (\xi_i, \beta) = 0;\end{aligned}$$

that is, β' is orthogonal to all $\xi_i \in X$. If $\|\beta'\| \neq 0$, then $X \cup \left\{\frac{1}{\|\beta'\|}\beta'\right\}$

would be a larger orthonormal set. Hence, $\|\beta'\| = 0$. Using the assumption that the inner product is positive definite we can now conclude that $\beta' = 0$. However, it is not necessary to use this assumption and we prefer to avoid using it. What we really need to conclude is that if α is any vector in V then $(\alpha, \beta') = 0$, and this follows from Schwarz's inequality. Thus we have

$$\begin{aligned} 0 = (\alpha, \beta') &= (\alpha, \beta - \sum_i (\xi_i, \beta)\xi_i) \\ &= (\alpha, \beta) - \sum_i (\xi_i, \beta)(\alpha, \xi_i) \\ &= (\alpha, \beta) - \sum_i \overline{(\xi_i, \alpha)}(\xi_i, \beta), \end{aligned}$$

or

$$(\alpha, \beta) = \sum_i \overline{(\xi_i, \alpha)}(\xi_i, \beta).$$

This completes the cycle of implications and proves that conditions (1), (2), and (3) are equivalent. □

Theorem 2.4. *The following two conditions are equivalent:*
(4) *The only vector orthogonal to all vectors in X is the zero vector.*
(5) *For each $\alpha \in V$, $\alpha = \sum_i (\xi_i, \alpha)\xi_i$.* (2.5)

PROOF. Assume (4). Let α be any vector in V and consider $\alpha' = \alpha - \sum_i (\xi_i, \alpha)\xi_i$. Then

$$\begin{aligned} (\xi_i, \alpha') &= (\xi_i, \alpha - \sum_j (\xi_j, \alpha)\xi_j) \\ &= (\xi_i, \alpha) - \sum_j (\xi_j, \alpha)(\xi_i, \xi_j) \\ &= (\xi_i, \alpha) - (\xi_i, \alpha) = 0; \end{aligned}$$

that is, α' is orthogonal to all $\xi_i \in X$. Thus $\alpha' = 0$ and $\alpha = \sum_i (\xi_i, \alpha)\xi_i$.
Now, assume (5) and let α be orthogonal to all $\xi_i \in X$. Then $\alpha = \sum_i (\xi_i, \alpha)\xi_i = 0$. □

Theorem 2.5. *The conditions* (4) *or* (5) *imply the conditions* (1), (2), *and* (3).

PROOF. Assume (5). Then

$$\begin{aligned} (\alpha, \beta) &= (\sum_i (\xi_i, \alpha)\xi_i, \sum_j (\xi_j, \beta)\xi_j) \\ &= \sum_i \overline{(\xi_i, \alpha)} \sum_j (\xi_j, \beta)(\xi_i, \xi_j) \\ &= \sum_i \overline{(\xi_i, \alpha)}(\xi_i, \beta). \ \square \end{aligned}$$

Theorem 2.6. *If the inner product is positive definite, the conditions* (1), (2), *or* (3) *imply the conditions* (4) *and* (5).

PROOF. In the proof that (3) implies (1) we showed that if $\alpha' = \alpha - \sum_i (\xi_i, \alpha)\xi_i$, then $\|\alpha'\| = 0$. If the inner product is positive definite, then $\alpha' = 0$ and, hence,

$$\alpha = \sum_i (\xi_i, \alpha)\xi_i. \square$$

The proofs of Theorems 2.3, 2.4, and 2.5 did not make use of the positive definiteness of the inner product and they remain valid if the inner product is merely non-negative semi-definite. Theorem 2.6 depends critically on the fact that the inner product is positive definite.

For finite dimensional vector spaces we always assume that the inner product is positive definite so that the three conditions of Theorem 2.3 and the two conditions of Theorem 2.4 are equivalent. The point of our making a distinction between these two sets of conditions is that there are a number of important inner products in infinite dimensional vector spaces that are not positive definite. For example, the inner product that occurs in the theory of Fourier series is of the form

$$(\alpha, \beta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{\alpha(x)}\beta(x)\, dx. \tag{2.6}$$

This inner product is non-negative semi-definite, but not positive definite if V is the set of integrable functions. Hence, we cannot pass from the completeness of the set of orthogonal functions to a theorem about the convergence of a Fourier series to the function from which the Fourier coefficients were obtained.

In using theorems of this type in infinite dimensional vector spaces in general and Fourier series in particular, we proceed in the following manner. We show that any $\alpha \in V$ can be approximated arbitrarily closely by finite sums of the form $\sum_i x_i\xi_i$. For the theory of Fourier series this theorem is known as the Weierstrass approximation theorem. A similar theorem must be proved for other sets of orthogonal functions. This implies that the minimum mentioned in Theorem 2.1 must be zero. This in turn implies that condition (2) of Theorem 2.3 holds. Thus Parseval's equation, which is equivalent to the completeness of an orthonormal set, is one of the principal theorems of any theory of orthogonal functions. Condition (5), which is the convergence of a Fourier series to the function which it represents, would follow if the inner product were positive definite. Unfortunately, this is usually not the case. To get the validity of condition (5) we must either add further conditions or introduce a different type of convergence.

EXERCISES

1. Show that if X is an orthonormal basis of a finite dimensional vector space, then condition (5) holds.

2. Let X be a finite set of mutually orthogonal vectors in V. Suppose that the only vector orthogonal to each vector in X is the zero vector. Show that X is a basis of V.

3 | The Representation of a Linear Functional by an Inner Product

For a fixed vector $\beta \in V$, (β, α) is a linear function of α. Thus there is a linear functional $\phi \in \hat{V}$ such that $\phi(\alpha) = (\beta, \alpha)$ for all α. We denote the linear functional defined in this way by ϕ_β. The following theorem is a converse of this observation.

Theorem 3.1. *Given a linear functional $\phi \in \hat{V}$, there exists a unique $\eta \in V$ such that $\phi(\alpha) = (\eta, \alpha)$ for all $\alpha \in V$.*

PROOF. Let $X = \{\xi_1, \ldots, \xi_n\}$ be an orthonormal basis of V, and let $\hat{X} = \{\phi_1, \ldots, \phi_n\}$ be the dual basis. Let $\phi \in \hat{V}$ have the representation $\phi = \sum_{i=1}^n y_i\phi_i$. Define $\eta = \sum_{i=1}^n \bar{y}_i\xi_i$. Then for each ξ_j, $(\eta, \xi_j) = (\sum_{i=1}^n \bar{y}_i\xi_i, \xi_j) = \sum_{i=1}^n y_i(\xi_i, \xi_j) = y_j = \sum_{i=1}^n y_i\phi_i(\xi_j) = \phi(\xi_j)$. But then $\phi(\alpha)$ and (η, α) are both linear functionals on V that coincide on the basis, and hence coincide on all of V.

If η_1 and η_2 are two choices such that $(\eta_1, \alpha) = (\eta_2, \alpha) = \phi(\alpha)$ for all $\alpha \in V$, then $(\eta_1 - \eta_2, \alpha) = 0$ for all $\alpha \in V$. For $\alpha = \eta_1 - \eta_2$ this means $(\eta_1 - \eta_2, \eta_1 - \eta_2) = 0$. Hence, $\eta_1 - \eta_2 = 0$ and the choice for η is unique. □

Call the mapping defined by this theorem η; that is, for each $\phi \in \hat{V}$, $\eta(\phi) \in V$ has the property that $\phi(\alpha) = (\eta(\phi), \alpha)$ for all $\alpha \in V$.

Theorem 3.2. *The correspondence between $\phi \in \hat{V}$ and $\eta(\phi) \in V$ is one-to-one and onto V.*

PROOF. In Theorem 3.1 we have already shown that $\eta(\phi)$ is well defined. Let β be any vector in V and let ϕ_β be the linear functional in $\hat{V}$ such that $\phi_\beta(\alpha) = (\beta, \alpha)$ for all α. Then $\beta = \eta(\phi_\beta)$ and the mapping is onto. Since (β, α), as a function of α, determines a unique linear functional ϕ_β the correspondence is one-to-one. □

Theorem 3.3. *If the inner product is symmetric, η is an isomorphism of $\hat{V}$ onto V.*

PROOF. We have already shown in Theorem 3.2 that η is one-to-one and onto. Let $\phi = \sum_i b_i\phi_i$ and consider $\beta = \sum_i b_i\eta(\phi_i)$. Then $(\beta, \alpha) = (\sum_i b_i\eta(\phi_i), \alpha) = (\alpha, \sum_i b_i\eta(\phi_i)) = \sum_i b_i(\alpha, \eta(\phi_i)) = \sum_i b_i(\eta(\phi_i), \alpha) = \sum_i b_i\phi_i(\alpha) = \phi(\alpha)$. Thus $\eta(\phi) = \beta = \sum_i b_i\eta(\phi_i)$ and η is linear. □

Notice that η is not linear if the scalar field is complex and the inner product is Hermitian. Then for $\phi = \sum_i b_i\phi_i$ we consider $\gamma = \sum_i \bar{b}_i\eta(\phi_i)$. We see that $(\gamma, \alpha) = (\sum_i \bar{b}_i\eta(\phi_i), \alpha) = \sum_i b_i(\eta(\phi_i), \alpha) = \sum_i b_i\phi_i(\alpha) = \phi(\alpha)$.

Thus $\eta(\phi) = \gamma = \sum_i \bar{b}_i \eta(\phi_i)$ and η is conjugate linear. It should be observed that even when η is conjugate linear it maps subspaces of $\hat{V}$ onto subspaces of V.

We describe this situation by saying that we can "represent a linear functional by an inner product." Notice that although we made use of a particular basis to specify the η corresponding to ϕ, the uniqueness shows that this choice is independent of the basis used. If V is a vector space over the real numbers, ϕ and η happen to have the same coordinates. This happy coincidence allows us to represent $\hat{V}$ in V and make V do double duty. This fact is exploited in courses in vector analysis. In fact, it is customary to start immediately with inner products in real vector spaces with orthonormal bases and not to mention $\hat{V}$ at all. All is well as long as things remain simple. As soon as things get a little more complicated, it is necessary to separate the structure of $\hat{V}$ superimposed on V. The vectors representing themselves in V are said to be *contravariant* and the vectors representing linear functionals in $\hat{V}$ are said to be *covariant.*

We can see from the proof of Theorem 3.1 that, if V is a vector space over the complex numbers, ϕ and the corresponding η will not necessarily have the same coordinates. In fact, there is no choice of a basis for which each ϕ and its corresponding η will have the same coordinates.

Let us examine the situation when the basis chosen in V is not orthonormal. Let $A = \{\alpha_1, \ldots, \alpha_n\}$ be any basis of V, and let $\hat{A} = \{\psi_1, \ldots, \psi_n\}$ be the corresponding dual basis of $\hat{V}$. Let $b_{ij} = (\alpha_i, \alpha_j)$. Since the inner product is Hermitian, $b_{ij} = \bar{b}_{ji}$, or $[b_{ij}] = B = B^*$. Since the inner product is positive definite, B has rank n. That is, B is non-singular. Let $\phi = \sum_{i=1}^n c_i \psi_i$ be an arbitrary linear functional in V. What are the coordinates of the corresponding η? Let $\eta = \sum_{i=1}^n y_i \alpha_i$. Then

$$\begin{aligned}
(\eta, \alpha_j) &= \left(\sum_{i=1}^n y_i \alpha_i, \alpha_j\right) \\
&= \sum_{i=1}^n \bar{y}_i (\alpha_i, \alpha_j) \\
&= \sum_{i=1}^n \bar{y}_i b_{ij} \\
&= \phi(\alpha_j) \\
&= \sum_{k=1}^n c_k \psi_k(\alpha_j) \\
&= c_j.
\end{aligned} \tag{3.1}$$

Thus, we have to solve the equations

$$\sum_{i=1}^n \bar{y}_i b_{ij} = \sum_{i=1}^n \bar{b}_{ji} \bar{y}_i = c_j,\ j = 1, \ldots, n. \tag{3.2}$$

In matrix form this becomes

$$\bar{B}\bar{Y} = C^T,$$

where

$$C = [c_1 \quad \cdots \quad c_n],$$

or

$$Y = B^{-1}C^* = (CB^{-1})^*. \tag{3.3}$$

Of course this means that it is rather complicated to obtain the coordinate representation of η from the coordinate representation of ϕ. But that is not the cause for all the fuss about covarient and contravariant vectors. After all, we have shown that η corresponds to ϕ independently of the basis used and the coordinates of η transform according to the same rules that apply to any other vector in V. The real difficulty stems from the insistence upon using (1.9) as the definition of the inner product, instead of using a definition not based upon coordinates.

If $\eta = \sum_{i=1}^n y_i\alpha_i$, and $\xi = \sum_{i=1}^n x_i\alpha_i$, we see that

$$\begin{aligned}(\eta, \xi) &= \left(\sum_{i=1}^n y_i\alpha_i, \sum_{j=1}^n x_j\alpha_j\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \bar{y}_i b_{ij} x_j \\ &= Y^*BX.\end{aligned} \tag{3.4}$$

Thus, if η represents the linear functional ϕ, we have

$$\begin{aligned}(\eta, \xi) &= Y^*BX \\ &= (CB^{-1})BX \\ &= CX \\ &= (C^*)^*X.\end{aligned} \tag{3.5}$$

Elementary treatments of vector analysis prefer to use C^* as the representation of η. This preference is based on the desire to use (1.9) as the definition of the inner product so that (3.5) is the representation of (η, ξ), rather than to use a coordinate-free definition which would lead to (η, ξ) being represented by (3.4). The elements of C^* are called the covariant components of η. We obtained C by representing ϕ in $\hat{V}$. Since the dual space is not available in such an elementary treatment, some kind of artifice must be used. It is then customary to introduce a *reciprocal basis* $A^* = \{\alpha_1^*, \ldots, \alpha_n^*\}$, where α_i^* has the property $(\alpha_i^*, \alpha_j) = \delta_{ij} = \phi_i(\alpha_j)$. A^* is the representation of the dual basis $\hat{A}$ in V. But C was the original representation of ϕ in terms of the dual basis. Thus, the insistence upon representing linear functionals by the inner product does not result in a single computational advantage. The confusion that it introduces is a severe price to pay to avoid introducing linear functionals and the dual space at the beginning.

4 | The Adjoint Transformation

Theorem 4.1. *For a given linear transformation σ on V, there is a unique linear transformation σ^* on V such that $(\sigma^*(\alpha), \beta) = (\alpha, \sigma(\beta))$ for all $\alpha, \beta \in V$.*

PROOF. Let σ be given. Then for a fixed α, $(\alpha, \sigma(\beta))$ is a linear function of β, that is, a linear functional on V. By Theorem 3.1 there is a unique $\eta \in V$ such that $(\alpha, \sigma(\beta)) = (\eta, \beta)$. Define $\sigma^*(\alpha)$ to be this η.

Now, $(a_1\alpha_1 + a_2\alpha_2, \sigma(\beta)) = \bar{a}_1(\alpha_1, \sigma(\beta)) + \bar{a}_2(\alpha_2, \sigma(\beta)) = \bar{a}_1(\sigma^*(\alpha_1), \beta) + \bar{a}_2(\sigma^*(\alpha_2), \beta) = (a_1\sigma^*(\alpha_1) + a_2\sigma^*(\alpha_2), \beta)$ so that $a_1\sigma^*(\alpha_1) + a_2\sigma^*(\alpha_2) = \sigma^*(a_1\alpha_1 + a_2\alpha_2)$ and σ^* is linear. □

Since for each α the choice for $\sigma^*(\alpha)$ is unique, σ^* is uniquely defined by σ. σ^* is called the *adjoint* of σ.

Theorem 4.2. *The relation between σ and σ^* is symmetric, that is, $(\sigma^*)^* = \sigma$.*

PROOF. Let σ be given. Then σ^* is defined uniquely by $(\sigma^*(\alpha), \beta) = (\alpha, \sigma(\beta))$ for all $\alpha, \beta \in V$. Then $(\sigma^*)^*$, which we denote by σ^{**}, is defined by $(\sigma^{**}(\alpha), \beta) = (\alpha, \sigma^*(\beta))$ for all $\alpha, \beta \in V$. Now the inner product is Hermitian so that $(\sigma^{**}(\alpha), \beta) = (\alpha, \sigma^*(\beta)) = \overline{(\sigma^*(\beta), \alpha)} = \overline{(\beta, \sigma(\alpha))} = (\sigma(\alpha), \beta)$. Thus $\sigma^{**}(\alpha) = \sigma(\alpha)$ for all $\alpha \in V$; that is, $\sigma^{**} = \sigma$. It then follows also that $(\sigma(\alpha), \beta) = (\alpha, \sigma^*(\beta))$. □

Let $A = [a_{ij}]$ be the matrix representing σ with respect to an orthonormal basis $X = \{\xi_i, \dots, \xi_n\}$ and let us find the matrix representing σ^*.

$$\begin{aligned}
(\sigma^*(\xi_j), \xi_k) &= (\xi_j, \sigma(\xi_k)) \\
&= \left(\xi_j, \sum_{i=1}^{n} a_{ik}\xi_i\right) \\
&= \sum_{i=1}^{n} a_{ik}(\xi_j, \xi_i) \\
&= a_{jk} \\
&= \sum_{i=1}^{n} a_{ji}(\xi_i, \xi_k) \\
&= \left(\sum_{i=1}^{n} \bar{a}_{ji}\xi_i, \xi_k\right). \qquad (4.1)
\end{aligned}$$

Since this equation holds for all ξ_k, $\sigma^*(\xi_j) = \sum_{i=1}^{n} \bar{a}_{ji}\xi_i$. Thus σ^* is represented by the conjugate transpose of A; that is, σ^* is represented by A^*.

The adjoint σ^* is closely related to the dual $\hat{\sigma}$ defined on page 142. σ is a linear transformation of V into itself, so the dual $\hat{\sigma}$ is a linear transformation of $\hat{V}$ into itself. Since η establishes a one-to-one correspondence between $\hat{V}$ and V, we can define a mapping of V into itself corresponding to $\hat{\sigma}$ on

$\hat{V}$. For any $\alpha \in V$ we can map α onto $\eta\{\hat{\sigma}[\eta^{-1}(\alpha)]\}$ and denote this mapping by $\eta(\hat{\sigma})$. Then for any $\alpha, \beta \in V$ we have

$$\begin{aligned}(\eta(\hat{\sigma})(\alpha), \beta) &= (\eta\{\hat{\sigma}[\eta^{-1}(\alpha)]\}, \beta) \\ &= \hat{\sigma}[\eta^{-1}(\alpha)](\beta) \\ &= \eta^{-1}(\alpha)[\sigma(\beta)] \\ &= (\alpha, \sigma(\beta)) \\ &= (\sigma^*(\alpha), \beta). \end{aligned} \tag{4.2}$$

Hence, $\eta(\hat{\sigma})(\alpha) = \sigma^*(\alpha)$ for all $\alpha \in V$, that is, $\eta(\hat{\sigma}) = \sigma^*$. The adjoint is a representation of the dual. Because the mapping η of $\hat{V}$ onto V is conjugate linear instead of linear, and because the vectors in $\hat{V}$ are represented by row matrices while those in V are represented by columns, the matrix representing σ^* is the transpose of the complex conjugate of the matrix representing $\hat{\sigma}$. Thus σ^* is represented by A^*.

We shall maintain the distinction between the dual $\hat{\sigma}$ defined on $\hat{V}$ and the adjoint σ^* defined on V. This distinction is not always made and quite often both terms are used for both purposes. Actually, this confusion seldom causes any trouble. However, it can cause trouble when discussing the matrix representation of $\hat{\sigma}$ or σ^*. If σ is represented by A, we have chosen also to represent $\hat{\sigma}$ by A with respect to the dual basis. If we had chosen to represent linear functionals by columns instead of rows, $\hat{\sigma}$ would have been represented by A^T. It would have been represented by A^T in either the real or the complex case. But the adjoint σ^* is represented by A^*. No convention will allow $\hat{\sigma}$ and σ^* to be represented by the same matrix in the complex case because the mapping η is conjugate linear. Because of this we have chosen to make clear the distinction between $\hat{\sigma}$ and σ^*, even to the extent of having the matrix representations look different. Furthermore, the use of rows to represent linear functionals has the advantage of making some of the formulas look simpler. However, this is purely a matter of choice and taste, and other conventions, used consistently, would serve as well.

Since we now have a model of $\hat{V}$ in V, we can carry over into V all the terminology and theorems on linear functionals in Chapter IV. In particular, we see that an orthonormal basis can also be considered to be its own dual basis since $(\xi_i, \xi_j) = \delta_{ij}$.

Recall that, when a basis is changed in V and P is the matrix of transition, $(P^T)^{-1}$ is the matrix of transition for the dual bases in $\hat{V}$. In mapping $\hat{V}$ onto V, $\overline{(P^T)^{-1}} = (P^*)^{-1}$ becomes the matrix of transition for the representation of dual basis in V. Since an orthonormal basis is dual to itself, if P is the matrix of transition from one orthonormal basis to another, then P must also be the matrix of transition for the dual basis; that is, $(P^*)^{-1} = P$. This important property of the matrices of transition from one orthonormal basis to another will be established independently in Section 6.

Let W be a subset of V. In Chapter IV-4, we defined $W^{\perp}$ to be the annihilator of W in $\hat{V}$. The mapping η of $\hat{V}$ onto V maps $W^{\perp}$ onto a subspace of V. It is easily seen that $\eta(W^{\perp})$ is precisely the set of all vectors orthogonal to every vector in W. Since we are in the process of dropping $\hat{V}$ as a separate space and identifying it with V, we denote the set of all vectors in orthogonal to all vectors in W by $W^{\perp}$ and call it the *annihilator* of W.

Theorem 4.3. *If W is a subspace of dimension p, $W^{\perp}$ is of dimension $n - \rho$. $W \cap W^{\perp} = \{0\}$. $W \oplus W^{\perp} = V$.*

PROOF. That $W^{\perp}$ is of dimension $n - \rho$ follows from Theorem 4.1 of Chapter IV. The other two assertions had no meaning in the context of Chapter IV. If $\alpha \in W \cap W^{\perp}$, then $\|\alpha\|^2 = (\alpha, \alpha) = 0$ so that $\alpha = 0$. Since $\dim \{W + W^{\perp}\} = \dim W + \dim W^{\perp} - \dim \{W \cap W^{\perp}\} = \rho + (n - \rho) = n$, $W \oplus W^{\perp} = V$. □

When W_1 and W_2 are subspaces of V such that their sum is direct and W_1 and W_2 are also orthogonal, we use the notation $W_1 \perp W_2$ to denote this sum. Actually, the fact that the sum is direct is a consequence of the fact that the subspaces are orthogonal. In this notation, the direct sum in the conclusion of Theorem 4.3 takes the form $V = W \perp W^{\perp}$.

Theorem 4.4. *Let W be a subspace invariant under σ. $W^{\perp}$ is then invariant under σ^*.*

PROOF. Let $\alpha \in W^{\perp}$. Then, for any $\beta \in W$, $(\sigma^*(\alpha), \beta) = (\alpha, \sigma(\beta)) = 0$ since $\sigma(\beta) \in W$. Thus $\sigma^*(\alpha) \in W^{\perp}$. □

Theorem 4.5. $K(\sigma^*) = \operatorname{Im}(\sigma)^{\perp}$.

PROOF. By definition $(\alpha, \sigma(\beta)) = (\sigma^*(\alpha), \beta)$. $(\alpha, \alpha(\beta)) = 0$ for all $\beta \in V$ if and only if $\alpha \in \operatorname{Im}(\sigma)^{\perp}$ and $(\sigma^*(\alpha), \beta) = 0$ for all $\beta \in V$ if and only if $\alpha \in K(\sigma^*)$. Thus $K(\sigma^*) = \operatorname{Im}(\sigma)^{\perp}$. □

Theorem 4.5 here is equivalent to Theorem 5.3 of Chapter IV.

Theorem 4.6. *If S and T are subspaces of V, then $(S + T)^{\perp} = S^{\perp} \cap T^{\perp}$ and $(S \cap T)^{\perp} = S^{\perp} + T^{\perp}$.*

PROOF. This theorem is equivalent to Theorem 4.4 of Chapter IV. □

Theorem 4.7. *For each conjugate bilinear form f, there is a linear transformation σ such that $f(\alpha, \beta) = (\alpha, \sigma(\beta))$ for all $\alpha, \beta \in V$.*

PROOF. For a fixed $\alpha \in V$, $f(\alpha, \beta)$ is linear in β. Thus by Theorem 3.1 there is a unique $\eta \in V$ such that $f(\alpha, \beta) = (\eta, \beta)$ for all $\beta \in V$. Define $\sigma^*(\alpha) = \eta$. σ^* is linear since $(\sigma^*(a_1\alpha_1 + a_2\alpha_2), \beta) = f(a_1\alpha_1 + a_2\alpha_2, \beta) = \bar{a}_1 f(\alpha_1, \beta) + \bar{a}_2 f(\alpha_2, \beta) = \bar{a}_1(\sigma^*(\alpha_1), \beta) + \bar{a}_2(\sigma^*(\alpha_2), \beta) = (a_1\sigma^*(\alpha_1) + a_2\sigma^*(\alpha_2), \beta)$. Let $\sigma^{**} = \sigma$ be the linear transformation of which σ^* is the adjoint. Then $f(\alpha, \beta) = (\sigma^*(\alpha), \beta) = (\alpha, \sigma(\beta))$. □

We shall call σ the *linear transformation associated* with the conjugate bilinear form f. The *eigenvalues* and *eigenvectors* of a conjugate bilinear form are defined to be the eigenvalues and eigenvectors of the associated linear transformation. Conversely, to each linear transformation σ there is associated a conjugate bilinear form $(\alpha, \sigma(\beta))$, and we shall also freely transfer terminology in the other direction. Thus a linear transformation will be called *symmetric*, or *skew-symmetric*, etc., if it is associated with a symmetric, or skew-symmetric bilinear form.

Theorem 4.8. *The conjugate bilinear form f and the linear transformation σ for which $f(\alpha, \beta) = (\alpha, \sigma(\beta))$ are represented by the same matrix with respect to an orthonormal basis.*

PROOF. Let $X = \{\xi_1, \ldots, \xi_n\}$ be an orthonormal basis and let $A = [a_{ij}]$ be the matrix representing σ with respect to this basis. Then $f(\xi_i, \xi_j) = (\xi_i, \sigma(\xi_j)) = (\xi_i, \sum_{k=1}^n a_{kj}\xi_k) = \sum_{k=k}^n a_{kj}(\xi_i, \xi_k) = a_{ij}$. □

A linear transformation is called *self-adjoint* if $\sigma^* = \sigma$. Clearly, a linear transformation is self-adjoint if and only if the matrix representing it (with respect to an orthonormal basis) is Hermitian. However, by means of Theorem 4.7 self-adjointness of a linear transformation can be related to the Hermitian character of a conjugate bilinear form without the intervention of matrices. Namely, if f is a Hermitian form then $(\sigma^*(\alpha), \beta) = (\alpha, \sigma(\beta)) = f(\alpha, \beta) = \overline{f(\beta, \alpha)} = \overline{(\beta, \sigma(\alpha))} = (\sigma(\alpha), \beta)$.

Theorem 4.9. *If σ and τ are linear transformations on V such that $(\sigma(\alpha), \beta) = (\tau(\alpha), \beta)$ for all $\alpha, \beta \in V$, then $\sigma = \tau$.*

PROOF. If $(\sigma(\alpha), \beta) - (\tau(\alpha), \beta) = ((\sigma - \tau)(\alpha), \beta) = 0$ for all α, β, then for each α and $\beta = (\sigma - \tau)(\alpha)$ we have $\|(\sigma - \tau)(\alpha)\|^2 = 0$. Hence, $(\sigma - \tau)(\alpha) = 0$ for all α and $\sigma = \tau$. □

Corollary 4.10. *If σ and τ are linear transformations on V such that $(\alpha, \sigma(\beta)) = (\alpha, \tau(\beta))$ for all $\alpha, \beta \in V$, then $\sigma = \tau$.* □

Theorem 4.9 provides an independent proof that the adjoint operator σ^* is unique. Corollary 4.10 shows that the linear transformation σ corresponding to the bilinear form f such that $f(\alpha, \beta) = (\alpha, \sigma(\beta))$ is also unique. Since, in turn, each linear transformation σ defines a bilinear form f by the formula $f(\alpha, \beta) = (\alpha, \sigma(\beta))$, this establishes a one-to-one correspondence between conjugate bilinear forms and linear transformations.

Theorem 4.11. *Let V be a unitary vector space. If σ and τ are linear transformations on V such that $(\sigma(\alpha), \alpha) = (\tau(\alpha), \alpha)$ for all $\alpha \in V$, then $\sigma = \tau$.*

PROOF. It can be checked that

$$\begin{aligned}(\sigma(\alpha), \beta) = \tfrac{1}{4}\{&(\sigma(\alpha + \beta), \alpha + \beta) - (\sigma(\alpha - \beta), \alpha - \beta) \\ &- i(\sigma(\alpha + i\beta), \alpha + i\beta) + i(\sigma(\alpha - i\beta), \alpha - i\beta)\}. \qquad (4.3)\end{aligned}$$

It follows from the hypothesis that $(\sigma(\alpha), \beta) = (\tau(\alpha), \beta)$ for all $\alpha, \beta \in V$. Hence, by Theorem 4.9, $\sigma = \tau$. □

It is curious to note that this theorem can be proved because of the relation (4.3), which is analogous to formula (12.4) of Chapter IV. But the analogue of formula (10.1) in the real case does not yield the same conclusion. In fact, if V is a vector space over the real numbers and σ is skew-symmetric, then $(\sigma(\alpha), \alpha) = (\alpha, \sigma^*(\alpha)) = (\alpha, -\sigma(\alpha)) = -(\alpha, \sigma(\alpha)) = -(\sigma(\alpha), \alpha)$ for all α. Thus $(\sigma(\alpha), \alpha) = 0$ for all α. In the real case the best analogue of this theorem is that if $(\sigma(\alpha), \alpha) = (\tau(\alpha), \alpha)$ for all $\alpha \in V$, then $\sigma + \sigma^* = \tau + \tau^*$, or σ and τ have the same symmetric part.

EXERCISES

1. Show that $(\sigma\tau)^* = \tau^*\sigma^*$.

2. Show that if $\sigma^*\sigma = 0$, then $\sigma = 0$.

3. Let σ be a skew-symmetric linear transformation on a vector space over the real numbers. Show that $\sigma^* = -\sigma$.

4. Let f be a skew-Hermitian form—that is, $f(\alpha, \beta) = \overline{-f(\beta, \alpha)}$—and let σ be the associated skew-Hermitian linear transformation. Show that $\sigma^* = -\sigma$.

5. Show that eigenvalues of a real skew-symmetric linear transformation are either 0 or pure imaginary. Show that the same is true for a skew-Hermitian linear transformation.

6. For what kind of linear transformation σ is it true that $(\xi, \sigma(\xi)) = 0$ for all $\xi \in V$?

7. For what kind of linear transformation σ is it true that $\sigma(\xi) \in \xi^\perp$ for all $\xi \in V$?

8. Show that if W is an invariant subspace under σ, then $W^\perp$ is an invariant subspace under σ^*.

9. Show that if σ is self-adjoint and W is invariant under σ, then $W^\perp$ is also invariant under σ.

10. Let π be the projection of V onto S along T. Let π^* be the adjoint of π. Show that π^* is the projection of V onto $T^\perp$ along $S^\perp$.

11. Let $W = \sigma(V)$. Show that $W^\perp$ is the kernel of σ^*.

12. Show that σ and σ^* have the same rank.

13. Let $W = \sigma(V)$. Show that $\sigma^*(V) = \sigma^*(W)$.

14. Show that $\sigma^*(V) = \sigma^*\sigma(V)$. Show that $\sigma(V) = \sigma\sigma^*(V)$.

15. Show that if $\sigma^*\sigma = \sigma^*\sigma$, then $\sigma^*(V) = \sigma(V)$.

16. Show that if $\sigma^*\sigma = \sigma\sigma^*$, then σ and σ^* have the same kernel.

17. Show that $\sigma + \sigma^*$ is self-adjoint.

18. Show that if $\sigma + \sigma^* = 0$, then σ is skew-symmetric, or skew-Hermitian.

19. Show that $\sigma - \sigma^*$ is skew-symmetric, or skew-Hermitian.

20. Show that every linear transformation is the sum of a self-adjoint transformation and a skew-Hermitian transformation.

21. Show that if $\sigma\sigma^* = \sigma^*\sigma$, then $\mathrm{Im}(\sigma)$ is an invariant subspace under σ. In fact, show that $\sigma^n(V) = \sigma(V)$ for all $n \geq 1$.

22. Show that if σ is a scalar transformation, that is $\sigma(\alpha) = a\alpha$, then $\sigma^*(\alpha) = \bar{a}\alpha$.

5 | Orthogonal and Unitary Transformations

Definition. A linear transformation of V into itself is called an *isometry* if it preserves length; that is, σ is an isometry if and only if $\|\sigma(\alpha)\| = \|\alpha\|$ for all $\alpha \in V$. An isometry in a vector space over the real numbers is called an *orthogonal transformation.* An isometry in a vector space over the complex numbers is called a *unitary transformation.* We try to save duplication and repetition by treating the real and complex cases together whenever possible.

Theorem 5.1. *A linear transformation σ of V into itself is an isometry if and only if it preserves the inner product; that is, if and only if $(\alpha, \beta) = (\sigma(\alpha), \sigma(\beta))$ for all $\alpha, \beta \in V$.*

PROOF. Certainly, if σ preserves the inner product then it preserves length since $\|\sigma(\alpha)\|^2 = (\sigma(\alpha), \sigma(\alpha)) = (\alpha, \alpha) = \|\alpha\|^2$.

The converse requires the separation of the real and complex cases. For an inner product over the real numbers we have

$$\begin{aligned}(\alpha, \beta) &= \tfrac{1}{2}\{(\alpha + \beta, \alpha + \beta) - (\alpha, \alpha) - (\beta, \beta)\} \\ &= \tfrac{1}{2}\{\|\alpha + \beta\|^2 - \|\alpha\|^2 - \|\beta\|^2\}. \end{aligned} \tag{5.1}$$

For an inner product over the complex numbers we have

$$\begin{aligned}(\alpha, \beta) &= \tfrac{1}{4}\{(\alpha + \beta, \alpha + \beta) - (\alpha - \beta, \alpha - \beta) \\ &\quad - i(\alpha + i\beta, \alpha + i\beta) + i(\alpha - i\beta, \alpha - i\beta)\} \\ &= \tfrac{1}{4}\{\|\alpha + \beta\|^2 - \|\alpha - \beta\|^2 - i\,\|\alpha + i\beta\|^2 + i\,\|\alpha - i\beta\|^2\}. \end{aligned} \tag{5.2}$$

In either case, any linear transformation which preserves length will preserve the inner product. □

Theorem 5.2. *A linear transformation σ of V into itself is an isometry if and only if it maps an orthonormal basis onto an orthonormal basis.*

PROOF. It follows immediately from Theorem 5.1 that if σ is an isometry, then σ maps every orthonormal set onto an orthonormal set and, therefore, an orthonormal basis onto an orthonormal basis.

On the other hand, let $X = \{\xi_1, \ldots, \xi_n\}$ be any orthonormal basis which is mapped by σ onto an orthonormal basis $\{\sigma(\xi_1), \ldots, \sigma(\xi_n)\}$. For an arbitrary vector $\alpha \in V$, $\alpha = \sum_{i=1}^{n} x_i\xi_i$, we have

$$\begin{aligned}\|\sigma(\alpha)\|^2 &= (\sigma(\alpha), \sigma(\alpha)) \\ &= \left(\sum_{i=1}^{n} x_i\sigma(\xi_i), \sum_{j=1}^{n} x_j\sigma(\xi_j)\right) \\ &= \sum_{i=1}^{n} \bar{x}_i \sum_{j=1}^{n} x_j(\sigma(\xi_i), \sigma(\xi_j)) \\ &= \sum_{i=1}^{n} \bar{x}_i \sum_{j=1}^{n} x_j\delta_{ij} \\ &= \sum_{i=1}^{n} \bar{x}_i x_i = \|\alpha\|^2. \end{aligned} \tag{5.3}$$

Thus σ preserves length and it is an isometry. □

Theorem 5.3. *σ is an isometry if and only if $\sigma^* = \sigma^{-1}$.*

PROOF. If σ is an isometry, then $(\sigma(\alpha), \sigma(\beta)) = (\alpha, \beta)$ for all $\alpha, \beta \in V$. By the definition of σ^*, $(\alpha, \beta) = (\sigma^*[\sigma(\alpha)], \beta) = (\sigma^*\sigma(\alpha), \beta)$. Since this equation holds for all $\beta \in V$, $\sigma^*\sigma(\alpha)$ is uniquely defined and $\sigma^*\sigma(\alpha) = \alpha$. Thus $\sigma^*\sigma$ is the identity transformation, that is, $\sigma^* = \sigma^{-1}$.

Conversely, suppose that $\sigma^* = \sigma^{-1}$. Then $(\sigma(\alpha), \sigma(\beta)) = (\sigma^*[\sigma(\alpha)], \beta) = (\sigma^*\sigma(\alpha), \beta) = (\alpha, \beta)$ for all $\alpha, \beta \in V$, and σ is an isometry. □

EXERCISES

1. Let σ be an isometry and let λ be an eigenvalue of σ. Show that $|\lambda| = 1$.
2. Show that the real eigenvalues of an isometry are ± 1.
3. Let $X = \{\xi_1, \xi_2\}$ be an orthonormal basis of V. Find an isometry that maps ξ_1 onto $\dfrac{1}{\sqrt{2}}(\xi_1 + \xi_2)$.
4. Let $X = \{\xi_1, \xi_2, \xi_3\}$ be an orthonormal basis of V. Find an isometry that maps ξ_1 onto $\frac{1}{3}(\xi_1 + 2\xi_2 + 2\xi_3)$.

6 | Orthogonal and Unitary Matrices

Let σ be an isometry and let $U = [u_{ij}]$ be a matrix representing σ with respect to an orthonormal basis $X = \{\xi_i, \dots, \xi_n\}$. Since σ is an isometry, the set $X' = \{\sigma(\xi_1), \dots, \sigma(\xi_n)\}$ must also be orthonormal. Thus

$$\begin{aligned}\delta_{ij} &= (\sigma(\xi_i), \sigma(\xi_j)) \\ &= \left(\sum_{k=1}^{n} u_{k_i}\xi_k, \sum_{l=1}^{n} u_{lj}\xi_l\right) \\ &= \sum_{k=1}^{n} \overline{u_{k_i}}\left(\sum_{l=1}^{n} u_{lj}(\xi_k, \xi_l)\right) \\ &= \sum_{k=1}^{n} \overline{u_{k_i}} u_{kj}. \end{aligned} \tag{6.1}$$

This is equivalent to the matrix equation $U^*U = I$, which also follows from Theorem 5.3.

It is also easily seen that if $U^*U = I$, then σ must map an orthonormal basis onto an orthonormal basis. By Theorem 5.2 σ is then an isometry. Thus,

Theorem 6.1. *A matrix U whose elements are complex numbers represents a unitary transformation (with respect to an orthonormal basis) if and only if $U^* = U^{-1}$. A matrix with this property is called a unitary matrix.* □

If the underlying field of scalars is the real numbers instead of the complex numbers, then U is real and $U^* = U^T$. Nothing else is really changed and we have the corresponding theorem for vector spaces over the real numbers.

Theorem 6.2. *A matrix U whose elements are real numbers represents an orthogonal transformation (with respect to an orthonormal basis) if and only if $U^T = U^{-1}$. A real matrix with this property is called an orthogonal matrix.* □

As is the case in Theorems 6.1 and 6.2, quite a bit of the discussion of unitary and orthogonal transformations and matrices is entirely parallel. To avoid unnecessary duplication we discuss unitary transformations and matrices and leave the parallel discussion for orthogonal transformations and matrices implicit. Up to a certain point, an orthogonal matrix can be considered to be a unitary matrix that happens to have real entries. This viewpoint is not quite valid because a unitary matrix with real coefficients represents a unitary transformation, an isometry on a vector space over the complex numbers. This viewpoint, however, leads to no trouble until we make use of the algebraic closure of the complex numbers, the property of complex numbers that every polynomial equation with complex coefficients possesses at least one complex solution.

It is customary to read equations (6.1) as saying that the columns of U are orthonormal. Conversely, if the columns of U are orthonormal, then $U^* = U^{-1}$ and U is unitary. Also, U^* as a left inverse is also a right inverse; that is, $UU^* = I$. Thus,

$$\sum_{k=1}^{n} u_{ik}\overline{u_{jk}} = \delta_{ij} = \sum_{k=1}^{n} \overline{u_{jk}}u_{ik}. \tag{6.2}$$

Thus U is unitary if and only if the rows of U are orthonormal. Hence,

Theorem 6.3. *Unitary and orthogonal matrixes are characterized by the property that their columns are orthonormal. They are equally characterized by the property that their rows are orthonormal.* □

Theorem 6.4. *The product of unitary matrices is unitary. The product of orthogonal matrices is orthogonal.*

PROOF. This follows immediately from the observation that unitary and orthogonal matrices represent isometries, and one isometry followed by another results in an isometry. □

A proof of Theorem 6.4 based on the characterizing property $U^* = U^{-1}$ (or $U^T = U^{-1}$ for orthogonal matrices) is just as brief. Namely, $(U_1U_2)^* = U_2^*U_1^* = U_2^{-1}U_1^{-1} = (U_1U_2)^{-1}$.

Now suppose that $X = \{\xi_1, \ldots, \xi_n\}$ and $X' = \{\xi_1', \ldots, \xi_n'\}$ are two orthonormal bases, and that $P = [p_{ij}]$ is the matrix of transition from the basis X to the basis X'. By definition,

$$\xi_j' = \sum_{i=1}^{n} p_{ij}\xi_i. \tag{6.3}$$

Thus,

$$\begin{aligned}(\xi_j', \xi_k') &= \left(\sum_{i=1}^{n} p_{ij}\xi_i, \sum_{s=1}^{n} p_{sk}\xi_s\right) \\ &= \sum_{i=1}^{n} \bar{p}_{ij} \sum_{s=1}^{n} p_{sk}(\xi_i, \xi_s) \\ &= \sum_{i=1}^{n} \bar{p}_{ij}p_{ik} = \delta_{jk}.\end{aligned} \tag{6.4}$$

This means the columns of P are orthonormal and P is unitary (or orthogonal). Thus we have

Theorem 6.5. *The matrix of transition from one orthonormal basis to another is unitary* (*or orthogonal if the underlying field is real*). □

We have seen that two matrices representing the same linear transformation with respect to different bases are similar. If the two bases are both orthonormal, then the matrix of transition is unitary (or orthogonal). In this case we say that the two matrices are *unitary similar* (or *orthogonal similar*). The matrices A and A' are unitary (orthogonal) similar if and only if there exists a unitary (orthogonal) matrix P such that $A' = P^{-1}AP = P^*AP$ ($A' = P^{-1}AP = P^TAP$).

If H and H' are matrices representing the same conjugate bilinear form with respect to different bases, they are Hermitian congruent and there exists a non-singular matrix P such that $H' = P^*HP$. P is the matrix of transition and, if the two bases are orthonormal, P is unitary. Then $H' = P^*HP = P^{-1}HP$. Hence, if we restrict our attention to orthonormal bases in vector spaces over the complex numbers, we see that matrices representing linear transformations and matrices representing conjugate bilinear forms transform according to the same rules; they are unitary similar.

If B and B' are matrices representing the same real bilinear form with respect to different bases, they are congruent and there exists a non-singular matrix P such that $B' = P^TBP$. P is the matrix of transition and, if the two bases are orthonormal, P is orthogonal. Then $B' = P^TBP = P^{-1}BP$. Hence, if we restrict our attention to orthonormal bases in vector spaces over the real numbers, we see that matrices representing linear transformations and matrices representing bilinear forms transform according to the same rules; they are orthogonal similar.

In our earlier discussions of similarity we sought bases with respect to which the representing matrix had a simple form, usually a diagonal form. We were not always successful in obtaining a diagonal form. Now we restrict the set of possible bases even further by demanding that they be orthonormal. But we can also restrict our attention to the set of matrices which are unitary (or orthogonal) similar to diagonal matrices. It is fortunate that this restricted class of matrices includes a rather wide range of cases occurring in some of the most important applications of matrices. The main goal of this chapter is to define and characterize the class of matrices unitary similar to diagonal matrices and to organize computational procedures by means of which these diagonal matrices and the necessary matrices of transition can be obtained. We also discuss the special cases so important in the applications of the theory of matrices.

EXERCISES

1. Test the following matrices for orthogonality. If a matrix is orthogonal, find its inverse.

$$(a)\ \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{-\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} \qquad (b)\ \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} \qquad (c)\ \begin{bmatrix} 0.6 & 0.8 \\ 0.8 & -0.6 \end{bmatrix}$$

2. Which of the following matrices are unitary?

$$(a)\ \begin{bmatrix} \frac{1+i}{2} & \frac{1-i}{2} \\ \frac{1-i}{2} & \frac{1+i}{2} \end{bmatrix} \qquad (b)\ \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \qquad (c)\ \begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix}$$

3. Find an orthogonal matrix with $(1/\sqrt{2}, 1/\sqrt{2})$ in the first column. Find an orthogonal matrix with $(\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$ in the first column.

4. Find a symmetric orthogonal matrix with $(\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$ in the first column. Compute its square.

5. The following matrices are all orthogonal. Describe the geometric effects in real Euclidean 3-space of the linear transformations they represent.

$$(a)\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad (b)\ \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (c)\ \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

$$(d)\ \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (e)\ \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Show that these five matrices, together with the identity matrix, each have different eigenvalues (provided θ is not 0° or 180°), and that the eigenvalues of any third-order orthogonal matrix must be one of these six cases.

6. If a matrix represents a rotation of R^2 around the origin through an angle of θ, then it has the form

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

Show that $A(\theta)$ is orthogonal. Knowing that $A(\theta) \cdot A(\psi) = A(\theta + \psi)$, prove that $\sin(\theta + \psi) = \sin\theta\cos\psi + \cos\theta\sin\psi$. Show that if U is an orthogonal 2×2 matrix, then $U^{-1}A(\theta)U = A(\pm\theta)$.

7. Find the matrix B representing the real quadratic form $q(x, y) = ax^2 + 2bxy + cy^2$. Show that the discriminant $D = ac - b^2$ is the determinant of B. Show that the discriminant is invariant under orthogonal coordinate changes, that is, changes of coordinates for which the matrix of transition is orthogonal.

7 | Superdiagonal Form

In this section we restrict our attention to vector spaces (and to matrices) over the field of complex numbers. We have already observed that not every matrix is similar to a diagonal matrix. Thus, it is also true that not every matrix is unitary similar to a diagonal matrix. We later restrict our attention to a class of matrices which are unitary similar to diagonal matrices. As an intermediate step we obtain a relatively simple form to which every matrix can be reduced by unitary similar transformations.

Theorem 2.1. *Let σ be any linear transformation of V, a finite dimensional vector space over the complex numbers, into itself. There exists an orthonormal basis of V with respect to which the matrix representing σ is in superdiagonal form; that is, every element below the main diagonal is zero.*

PROOF. The proof is by induction on n, the dimension of V. The theorem says there is an orthonormal basis $Y = \{\eta_1, \ldots, \eta_n\}$ such that $\sigma(\eta_k) = \sum_{i=1}^{k} a_{ik}\eta_i$, the important property being that the summation ends with the kth term. The theorem is certainly true for $n = 1$.

Assume the theorem is true for vector spaces of dimensions $<n$. Since V is a vector space over the complex numbers, σ has at least one eigenvalue. Let λ_1 be an eigenvalue for σ and let $\xi_1' \neq 0$, $\|\xi_1'\| = 1$, be a corresponding eigenvector. There exists a basis, and hence an orthonormal basis, with ξ_1' as the first element. Let the basis be $X' = \{\xi_1', \ldots, \xi_n'\}$ and let W be the subspace spanned by $\{\xi_2', \ldots, \xi_n'\}$. W is the subspace consisting of all vectors orthogonal to ξ_1'. For each $\alpha = \sum_{i=1}^n a_i\xi_i'$ define $\tau(\alpha) = \sum_{i=2}^n a_i\xi_i' \in W$. Then $\tau\sigma$ restricted to W is a linear transformation of W into itself. According to the induction assumption, there is an orthonormal basis $\{\eta_2, \ldots, \eta_n\}$ of W such that for each η_k, $\tau\sigma(\eta_k)$ is expressible in terms of $\{\eta_2, \ldots, \eta_k\}$ alone. We see from the way τ is defined that $\sigma(\eta_k)$ is expressible in terms of $\{\xi_1', \eta_2, \ldots, \eta_k\}$ alone. Let $\eta_1 = \xi_1'$. Then $Y = \{\eta_1, \eta_2, \ldots, \eta_n\}$ is the required basis.

ALTERNATE PROOF. The proof just given was designed to avoid use of the concept of adjoint introduced in Section 4. Using that concept, a very much simpler proof can be given. This proof also proceeds by induction on n. The assertion for $n = 1$ is established in the same way as in the first proof given. Assume the theorem is true for vector spaces of dimension $<n$. Since V is a vector space over the complex numbers, σ^* has at least one eigenvalue. Let λ_n be an eigenvalue for σ^* and let η_n, $\|\eta_n\| = 1$, be a corresponding eigenvector. Then by Theorem 4.4, $W = \langle\eta_n\rangle^\perp$ is an invariant subspace under σ. Since $\eta_n \neq 0$, W is of dimension $n - 1$. According to the induction assumption, there is an orthonormal basis $\{\eta_1, \ldots, \eta_{n-1}\}$ of W such that $\sigma(\eta_k) = \sum_{i=1}^k a_{ik}\eta_i$, for $k = 1, 2, \ldots, n - 1$. However, $\{\eta_1, \ldots, \eta_n\}$ is also an orthonormal basis of U and $\sigma(\eta_n) = \sum_{i=1}^k a_{ik}\eta_i$, for $k = 1, \ldots, n$. □

***Corollary* 7.2.** *Over the field of complex numbers, every matrix is unitary similar to a superdiagonal matrix.* □

Theorem 7.1 and Corollary 7.2 depend critically on the assumption that the field of scalars is the field of complex numbers. The essential feature of this condition is that it guarantees the existence of eigenvalues and eigenvectors. If the field of scalars is not algebraically closed, the theorem is simply not true.

***Corollary* 7.3.** *The diagonal terms of the superdiagonal matrix representing σ are the eigenvalues of σ.*

PROOF. If $A = [a_{ij}]$ is in superdiagonal form, then the characteristic polynomial is $(a_{11} - x)(a_{22} - x) \cdots (a_{nn} - x)$. □

EXERCISES

1. Let σ be a linear transformation mapping U into V. Let A be any basis of U whatever. Show that there is an orthonormal basis B of V such that the matrix

representing σ with respect to A and B is in superdiagonal form. (In this case where U and V need not be of the same dimension so that the matrix representing σ need not be square, by superdiagonal form we mean that all elements below the main diagonal are zeros.)

2. Let σ be a linear transformation on V and let $Y = \{\eta_1, \ldots, \eta_n\}$ be an orthonormal basis such that the matrix representing σ with respect to Y is in superdiagonal form. Show that the matrix representing σ^* with respect to Y is in subdiagonal form; that is, all elements above the main diagonal are zeros.

3. Let σ be a linear transformation on V. Show that there is an orthonormal basis Y of V such that the matrix representing σ with respect to Y is in subdiagonal form.

8 | Normal Matrices

It is possible to give a necessary and sufficient condition that a matrix be unitary similar to a diagonal matrix. The real value in establishing this condition is that several important types of matrices do satisfy this condition.

Theorem 8.1. *A matrix A in superdiagonal form is a diagonal matrix if and only if $A^*A = AA^*$.*

PROOF. Let $A = [a_{ij}]$ where $a_{ij} = 0$ if $i > j$. Suppose that $A^*A = AA^*$. This means, in particular, that

$$\sum_{j=1}^{n} \bar{a}_{ji}a_{ji} = \sum_{k=1}^{n} a_{ik}\bar{a}_{ik}. \tag{8.1}$$

But since $a_{ij} = 0$ for $i > j$, this reduces to

$$\sum_{j=1}^{i} |a_{ji}|^2 = \sum_{k=i}^{n} |a_{ik}|^2. \tag{8.2}$$

Now, if A were not a diagonal matrix, there would be a first index i for which there exists an index $k > i$ such that $a_{ik} \neq 0$. For this choice of the index i the sum on the left in (8.2) reduces to one term while the sum on the right contains at least two non-zero terms. Thus,

$$\sum_{j=1}^{i} |a_{ji}|^2 = |a_{ji}|^2 = \sum_{k=i}^{n} |a_{ik}|^2, \tag{8.3}$$

which is a contradiction. Thus A must be a diagonal matrix.

Conversely, if A is a diagonal matrix, then clearly $A^*A = AA^*$. □

A matrix A for which $A^*A = AA^*$ is called a *normal matrix.*

Theorem 8.2. *A matrix is unitary similar to a diagonal matrix if and only if it is normal.*

PROOF. If A is a normal matrix, then any matrix unitary similar to A is also normal. Namely, if U is unitary, then

$$\begin{aligned}(U^*AU)^*(U^*AU) &= U^*A^*UU^*AU \\ &= U^*A^*AU \\ &= U^*AA^*U \\ &= U^*AUU^*A^*U. \\ &= (U^*AU)(U^*AU)^*. \end{aligned} \tag{8.4}$$

Thus, if A is normal, the superdiagonal form to which it is unitary similar is also normal and, hence, diagonal. Conversely, if A is unitary similar to a diagonal matrix, it is unitary similar to a normal matrix and it is therefore normal itself. □

Theorem 8.3. *Unitary matrices and Hermitian matrices are normal.*

PROOF. If U is unitary then $U^*U = U^{-1}U = UU^{-1} = UU^*$. If H is Hermitian then $H^*H = HH = HH^*$. □

EXERCISES

1. Determine which of the following matrices are orthogonal, unitary, symmetric Hermitian, skew-symmetric, skew-Hermitian, or normal.

(a) $\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & i \\ i & 2 \end{bmatrix}$

(d) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ (e) $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ (f) $\begin{bmatrix} 1 & 1-i \\ 1+i & 3 \end{bmatrix}$

(g) $\frac{1}{3}\begin{bmatrix} 1 & -2 & 2 \\ 2 & 2 & 1 \\ 2 & -1 & -2 \end{bmatrix}$ (h) $\frac{1}{3}\begin{bmatrix} 1 & 2 & 2 \\ 2 & -2 & 1 \\ 2 & 1 & -2 \end{bmatrix}$ (i) $\begin{bmatrix} 0 & -1 & -2 \\ 1 & 0 & -3 \\ 2 & 3 & 0 \end{bmatrix}$

(j) $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ (k) $\begin{bmatrix} 0 & -1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{bmatrix}$

2. Which of the matrices of Exercise 1 are unitary similar to diagonal matrices?

3. Show that a real skew-symmetric matrix is normal.

4. Show that a skew-Hermitian matrix is normal.

5. Show by example that there is a skew-symmetric complex matrix which is not normal.

6. Show by example that there is a symmetric complex matrix which is not normal.

7. Find an example of a normal matrix which is not Hermitian or unitary.

8. Show that if $M = A + Bi$ where A and B are real and symmetric, then M is normal if and only if A and B commute.

9 | Normal Linear Transformations

Theorem 9.1. *If there exists an orthonormal basis consisting of eigenvectors of a linear transformation σ, then $\sigma^*\sigma = \sigma\sigma^*$.*

PROOF. Let $X = \{\xi_1, \ldots, \xi_n\}$ be an orthonormal basis consisting of eigenvectors of σ. Let λ_i be the eigenvalue corresponding to ξ_i. Then. $(\sigma^*(\xi_i), \xi_j) = (\xi_i, \sigma(\xi_j)) = (\xi_i, \lambda_j\xi_j) = \lambda_j\delta_{ij} = \lambda_i\delta_{ij} = \lambda_i(\xi_i, \xi_j) = (\bar{\lambda}_i\xi_i, \xi_j)$. For a fixed ξ_i this equation holds for all ξ_j and, hence, $(\sigma^*(\xi_i), \alpha) = (\bar{\lambda}_i\xi_i, \alpha)$ for all $\alpha \in V$. This means $\sigma^*(\xi_i) = \bar{\lambda}_i\xi_i$ and ξ_i is an eigenvector of σ^* with eigenvalue $\bar{\lambda}_i$. Then $\sigma\sigma^*(\xi_i) = \sigma(\bar{\lambda}_i\xi_i) = \bar{\lambda}_i\lambda_i\xi_i = \sigma^*\sigma(\xi_i)$. Since $\sigma\sigma^* = \sigma^*\sigma$ on a basis of V, $\sigma\sigma^* = \sigma^*\sigma$ on all of V. □

A linear transformation σ for which $\sigma^*\sigma = \sigma\sigma^*$ is called a *normal linear transformation.* Clearly, a linear transformation is normal if and only if the matrix representing it (with respect to an orthonormal basis) is normal.

In the proof of Theorem 9.1 the critical step is showing that an eigenvector of σ is also an eigenvector of σ^*. The converse is also true.

Theorem 9.2. *If ξ is an eigenvector of a normal linear transformation σ corresponding to the eigenvalue λ, then ξ is an eigenvector of σ^* corresponding to $\bar{\lambda}$.*

PROOF. Since σ is normal $(\sigma(\xi), \sigma(\xi)) = (\sigma^*\sigma(\xi), \xi) = (\sigma\sigma^*(\xi), \xi) = (\sigma^*(\xi), \sigma^*(\xi))$. Since ξ is an eigenvector of σ corresponding to λ, $\sigma(\xi) = \lambda\xi$ so that

$$\begin{aligned}
0 = \|\sigma(\xi) - \lambda\xi\|^2 &= (\sigma(\xi) - \lambda\xi, \sigma(\xi) - \lambda\xi) \\
&= (\sigma(\xi), \sigma(\xi)) - \bar{\lambda}(\xi, \sigma(\xi)) - \lambda(\sigma(\xi), \xi) + \bar{\lambda}\lambda(\xi, \xi) \\
&= (\sigma^*(\xi), \sigma^*(\xi)) - \bar{\lambda}(\sigma^*(\xi), \xi) - \lambda(\xi, \sigma^*(\xi)) + \bar{\lambda}\lambda(\xi, \xi) \\
&= (\sigma^*(\xi) - \bar{\lambda}\xi, \sigma^*(\xi) - \bar{\lambda}\xi) \\
&= \|\sigma^*(\xi) - \bar{\lambda}\xi\|^2. \qquad (9.1)
\end{aligned}$$

Thus $\sigma^*(\xi) - \bar{\lambda}\xi = 0$, or $\sigma^*(\xi) = \bar{\lambda}\xi$. □

Theorem 9.3. *For a normal linear transformation, eigenvectors corresponding to different eigenvalues are orthogonal.*

PROOF. Suppose $\sigma(\xi_1) = \lambda_1\xi_1$ and $\sigma(\xi_2) = \lambda_2\xi_2$ where $\lambda_1 \neq \lambda_2$. Then $\lambda_2(\xi_1, \xi_2) = (\xi_1, \lambda_2\xi_2) = (\xi_1, \sigma(\xi_2)) = (\sigma^*(\xi_1), \xi_2) = (\bar{\lambda}_1\xi_1, \xi_2) = \lambda_1(\xi_1, \xi_2)$. Thus $(\lambda_1 - \lambda_2)(\xi_1, \xi_2) = 0$. Since $\lambda_1 - \lambda_2 \neq 0$ we see that $(\xi_1, \xi_2) = 0$; that is, ξ_1 and ξ_2 are orthogonal. □

Theorem 9.4. *If σ is normal, then $(\sigma(\alpha), \sigma(\beta)) = (\sigma^*(\alpha), \sigma^*(\beta))$ for all $\alpha, \beta \in V$.*

PROOF. $(\sigma(\alpha), \sigma(\beta)) = (\sigma^*\sigma(\alpha), \beta) = (\sigma\sigma^*(\alpha), \beta) = (\sigma^*(\alpha), \sigma^*(\beta))$. □

Corollary 9.5. *If σ is normal, $\|\sigma(\alpha)\| = \|\sigma^*(\alpha))\|$ for all $\alpha \in V$.* □

Theorem 9.6. *If* $(\sigma(\alpha), \sigma(\beta)) = (\sigma^*(\alpha), \sigma^*(\beta))$ *for all* $\alpha, \beta \in V$, *then* σ *is normal.*

PROOF. $(\alpha, \sigma\sigma^*(\beta)) = (\sigma^*(\alpha), \sigma^*(\beta)) = (\sigma(\alpha), \sigma(\beta)) = (\alpha, \sigma^*\sigma(\beta))$ for all $\alpha, \beta \in V$. By Corollary 4.10, $\sigma\sigma^* = \sigma^*\sigma$ and σ is normal. □

Theorem 9.7. *If* $\|\sigma(\alpha)\| = \|\sigma^*(\alpha)\|$ *for all* $\alpha \in V$, *then* σ *is normal.*

PROOF. We must divide this proof into two cases:

1. V is a vector space over F, a subfield of the real numbers. Then

$$(\sigma(\alpha), \sigma(\beta)) = \tfrac{1}{4}\{\|\sigma(\alpha + \beta)\|^2 - \|\sigma(\alpha - \beta)\|^2\}.$$

It then follows from the hypothesis that $(\sigma(\alpha), \sigma(\beta)) = (\sigma^*(\alpha), \sigma^*(\beta))$ for all $\alpha, \beta \in V$, and σ is normal.

2. V is a vector space over F, a non-real normal subfield of the complex numbers. Let $a \in F$ be chosen so that $a \neq \bar{a}$. Then

$$(\sigma(\alpha), \sigma(\beta)) = \frac{1}{2(\bar{a} - a)}$$
$$\times \{\bar{a}\,\|\sigma(\alpha + \beta)\|^2 - \bar{a}\,\|\sigma(\alpha - \beta)\|^2 - \|\sigma(\alpha + a\beta)\|^2 + \|\sigma(\alpha - a\beta)\|^2\}.$$

Again, it follows that σ is normal. □

Theorem 9.8. *If* σ *is normal then* $K(\sigma) = K(\sigma^*)$.

PROOF. Since $\|\sigma(\alpha)\| = \|\sigma^*(\alpha)\|$, $\sigma(\alpha) = 0$ if and only if $\sigma^*(\alpha) = 0$. □

Theorem 9.9. *If* σ *is normal,* $K(\sigma) = \mathrm{Im}(\sigma)^{\perp}$.

PROOF. By Theorem 4.5 $K(\sigma^*) = \mathrm{Im}(\sigma)^{\perp}$, and by Theorem 9.8 $K(\sigma) = K(\sigma^*)$. □

Theorem 9.10. *If* σ *is normal,* $\mathrm{Im}\,\sigma = \mathrm{Im}\,\sigma^*$.

PROOF. $\mathrm{Im}\,\sigma = K(\sigma)^{\perp} = \mathrm{Im}\,\sigma^*$. □

Theorem 9.11. *If* σ *is a normal linear transformation and* W *is a set of eigenvectors of* σ, *then* $W^{\perp}$ *is an invariant subspace under* σ.

PROOF. $\alpha \in W^{\perp}$ if and only if $(\xi, \alpha) = 0$ for all $\xi \in W$. But then $(\xi, \sigma(\alpha)) = (\sigma^*(\xi), \alpha) = (\bar{\lambda}\xi, \alpha) = \lambda(\xi, \alpha) = 0$. Hence, $\sigma(\alpha) \in W^{\perp}$ and $W^{\perp}$ is invariant under σ. □

Notice it is not necessary that W be a subspace, it is not necessary that W contain all the eigenvectors corresponding to any particular eigenvalue, and it is not necessary that the eigenvectors in W correspond to the same eigenvalue. In particular, if ξ is an eigenvector of σ, then $\{\xi\}^{\perp}$ is an invariant subspace under σ.

Theorem 9.12. *Let* V *be a vector space with an inner product, and let* σ *be a normal linear transformation of* V *into itself. If* W *is a subspace which is invariant under both* σ *and* σ^*, *then* σ *is normal on* W.

PROOF. Let $\underline{\sigma}$ denote the linear transformation of W into itself induced by σ. Let $\underline{\sigma}^*$ denote the adjoint of $\underline{\sigma}$ on W. Then for all, $\alpha, \beta \in W$ we have

$$(\underline{\sigma}^*(\alpha), \beta) = (\alpha, \underline{\sigma}(\beta)) = (\alpha, \sigma(\beta)) = (\sigma^*(\alpha), \beta).$$

Since $((\underline{\sigma}^* - \sigma^*)(\alpha), \beta) = 0$ for all $\alpha, \beta \in W$, $\underline{\sigma}^*$ and σ^* coincide on W. Thus $\underline{\sigma}^*\underline{\sigma} = \sigma^*\sigma = \sigma\sigma^* = \underline{\sigma}\underline{\sigma}^*$ on W, and $\underline{\sigma}$ is normal. □

Theorem 9.13. *Let V be a finite dimensional vector space over the complex numbers, and let σ be a normal linear transformation on V. If W is invariant under σ, then W is invariant under σ^* and σ is normal on W.*

PROOF. By Theorem 4.4, $W^\perp$ is invariant under σ^*. Let $\overline{\sigma^*}$ be the restriction of σ^* with $W^\perp$ as domain and codomain. Since $W^\perp$ is also a finite dimensional vector space over the complex numbers, $\overline{\sigma^*}$ has at least one eigenvalue λ and corresponding to it a non-zero eigenvector ξ. Thus $\overline{\sigma^*}(\xi) = \lambda\xi = \sigma^*(\xi)$. Thus, we have found an eigenvector for σ^* in $W^\perp$.

Now proceed by induction. The theorem is certainly true for spaces of dimension 1. Assume the theorem holds for vector spaces of dimension $<n$. By Theorem 9.2, ξ is an eigenvector of σ. By Theorem 9.11, $\langle\xi\rangle^\perp$ is invariant under both σ and σ^*. By Theorem 9.12, σ is normal on $\langle\xi\rangle^\perp$. Since $\langle\xi\rangle \subset W^\perp$, $W \subset \langle\xi\rangle^\perp$. Since $\dim \langle\xi\rangle^\perp = n - 1$, the induction assumption applies. Hence, σ is normal on W and W is invariant under σ^*. □

Theorem 9.13 is also true for a vector space over any subfield of the complex numbers, but the proof is not particularly instructive and this more general form of Theorem 9.13 will not be needed later.

We should like to obtain a converse of Theorem 9.1 and show that a normal linear transformation has enough eigenvectors to make up an orthonormal basis. Such a theorem requires some condition to guarantee the existence of eigenvalues or eigenvectors. One of the most important general conditions is to assume we are dealing with vector spaces over the complex numbers.

Theorem 9.14. *If V is a finite dimensional vector space over the complex numbers and σ is a normal linear transformation, then V has an orthonormal basis consisting of eigenvectors of σ.*

PROOF. Let n be the dimension of V. The theorem is certainly true for $n = 1$, for if $\{\xi_1\}$ is a basis $\sigma(\xi_1) = a_1\xi_1$.

Assume the theorem holds for vector spaces of dimension $<n$. Since V is a finite dimensional vector space over the complex numbers, σ has at least one eigenvalue λ_1, and corresponding to it a non-zero eigenvector ξ_1 which we can take to be normalized. By Theorem 9.11, $\{\xi_1\}^\perp$ is an invariant subspace under σ. This means that σ acts like a linear transformation on $\{\xi_1\}^\perp$ when we confine our attention to $\{\xi_1\}^\perp$. But then σ is also normal on

$\{\xi_1\}^\perp$. Since $\{\xi_1\}^\perp$ is of dimension $n - 1$, our induction assumption applies and $\{\xi_1\}^\perp$ has an orthonormal basis $\{\xi_2, \ldots, \xi_n\}$ consisting of eigenvectors of σ. $\{\xi_1, \xi_2, \ldots, \xi_n\}$ is the required orthonormal basis of V consisting of eigenvectors of σ. □

We can observe from examining the proof of Theorem 9.1 that the conclusion that σ and σ^* commute followed immediately after we showed that the eigenvectors of σ were also eigenvectors of σ^*. Thus the following theorem follows immediately.

Theorem 9.15. *If there exists a basis (orthonormal or not) consisting of vectors which are eigenvectors for both σ and τ, then $\sigma\tau = \tau\sigma$.* □

Any possible converse to Theorem 9.15 requires some condition to ensure the existence of the necessary eigenvectors. In the following theorem we accomplish this by assuming that the field of scalars is the field of complex numbers, any set of conditions that would imply the existence of the eigenvectors could be substituted.

Theorem 9.16. *Let V be a finite dimensional vector space over the complex numbers and let σ and τ be normal linear transformations on V. If $\sigma\tau = \tau\sigma$, then there exists an orthonormal basis consisting of vectors which are eigenvectors for both σ and τ.*

PROOF. Suppose $\sigma\tau = \tau\sigma$. Let λ be an eigenvalue of σ and let $S(\lambda)$ be the eigenspace of σ consisting of all eigenvectors of σ corresponding to λ. Then for each $\xi \in S(\lambda)$ we have $\sigma\tau(\xi) = \tau\sigma(\xi) = \tau(\lambda\xi) = \lambda\tau(\xi)$. Hence, $\tau(\xi) \in S(\lambda)$. This shows that $S(\lambda)$ is an invariant subspace under τ; that is, τ confined to $S(\lambda)$ can be considered to be a normal linear transformation of $S(\lambda)$ into itself. By Theorem 9.14 there is an orthonormal basis of $S(\lambda)$ consisting of eigenvectors of τ. Being in $S(\lambda)$ they are also eigenvectors of σ. By Theorem 9.3 the basis vectors obtained in this way in eigenspaces corresponding to different eigenvalues of σ are orthogonal. Again, by Theorem 9.14 there is a basis of V consisting of eigenvectors of σ. This implies that the eigenspaces of σ span V and, hence, the entire orthonormal set obtained in this fashion is an orthonormal basis of V. □

As we have seen, self-adjoint linear transformations and isometries are particular cases of normal linear transformations. They can also be characterized by the nature of their eigenvalues.

Theorem 9.17. *Let V be a finite dimensional vector space over the complex numbers. A normal linear transformation σ on V is self-adjoint if and only if all its eigenvalues are real.*

PROOF. Suppose σ is self-adjoint. Let λ be an eigenvalue for σ and let ξ be an eigenvector corresponding to λ. Then $\|\sigma(\xi)\|^2 = (\sigma(\xi), \sigma(\xi)) = (\sigma^*(\xi), \sigma(\xi)) = \lambda^2 \|\xi\|^2$. Thus λ^2 is real non-negative and λ is real.

On the other hand, suppose σ is a normal linear transformation and that all its eigenvalues are real. Since σ is normal there exists a basis $X = \{\xi_1, \ldots, \xi_n\}$ of eigenvectors of σ. Let λ_i be the eigenvalue corresponding to ξ_i. Then $\sigma^*(\xi_i) = \bar{\lambda}_i \xi_i = \lambda_i \xi_i = \sigma(\xi_i)$. Since σ^* coincides with σ on a basis of V, $\sigma = \sigma^*$ on all of V. □

Theorem 9.18. *Let V be a finite dimensional vector space over the complex numbers. A normal linear transformation σ on V is an isometry if and only if all its eigenvalues are of absolute value* 1.

PROOF. Suppose σ is an isometry. Let λ be an eigenvalue of σ and let ξ be an eigenvector corresponding to λ. Then $\|\xi\|^2 = \|\sigma(\xi)\|^2 = (\sigma(\xi), \sigma(\xi)) = (\lambda\xi, \lambda\xi) = |\lambda|^2 (\xi, \xi)$. Hence $|\lambda|^2 = 1$.

On the other hand suppose σ is a normal linear transformation and that all its eigenvalues are of absolute value 1. Since σ is normal there exists a basis $X = \{\xi_1, \ldots, \xi_n\}$ of eigenvectors of σ. Let λ_i be the eigenvalue corresponding to ξ_i. Then $(\sigma(\xi_i), \sigma(\xi_j)) = (\lambda_i\xi_i, \lambda_j\xi_j) = \bar{\lambda}_i\lambda_j(\xi_i, \xi_j) = \delta_{ij}$. Hence, σ maps an orthonormal basis onto an orthonormal basis and it is an isometry. □

EXERCISES

1. Prove Theorem 9.2 directly from Corollary 9.5.

2. Show that if there exists an orthonormal basis such that σ and τ are both represented by diagonal matrices, then $\sigma\tau = \tau\sigma$.

3. Show that if σ and τ are normal linear transformations such that $\sigma\tau = \tau\sigma$, then there is an orthonormal basis of V such that the matrices representing σ and τ are both diagonal; that is, σ and τ can be diagonalized simultaneously.

4. Show that the linear transformation associated with a Hermitian form is self-adjoint.

5. Let f be a Hermitian form and let σ be the associated linear transformation. Let $X = \{\xi_1, \ldots, \xi_n\}$ be a basis of eigenvectors of σ (show that such a basis exists) and let $\{\lambda_1, \ldots, \lambda_n\}$ be the corresponding eigenvalues. Let $\alpha = \sum_{i=1}^n a_i\xi_i$ and $\beta = \sum_{i=1}^n b_i\xi_i$ be arbitrary vectors in V. Show that $f(\alpha, \beta) = \sum_{i=1}^n \bar{a}_i b_i \lambda_i$.

6. (Continuation) Let q be the Hermitian quadratic form associated with the Hermitian form f. Let S be the set of all unit vectors in V; that is, $\alpha \in S$ if and only if $\|\alpha\| = 1$. Show that the maximum value of $q(\alpha)$ for $\alpha \in S$ is the maximum eigenvalue, and the minimum value of $q(\alpha)$ for $\alpha \in S$ is the minimum eigenvalue. Show that $q(\alpha) \neq 0$ for all non-zero $\alpha \in V$ if all the eigenvalues of f are non-zero and of the same sign.

7. Let σ be a normal linear transformation and let $\{\lambda_1, \ldots, \lambda_k\}$ be the distinct eigenvalues of σ. Let M_i be the subspace of eigenvectors of σ corresponding to λ_i. Show that $V = M_1 \perp \cdots \perp M_k$.

8. (Continuation) Let π_i be the projection of V onto M_i along $M_i^{\perp}$. Show that $1 = \pi_1 + \cdots + \pi_k$. Show that $\sigma = \lambda_1\pi_1 + \cdots + \lambda_k\pi_k$. Show that $\sigma^r = \lambda_1{}^r\pi_1 + \cdots + \lambda_k{}^r\pi_k$. Show that if $p(x)$ is a polynomial, then $p(\sigma) = \sum_{i=1}^k p(\lambda_i)\pi_i$.

10 | Hermitian and Unitary Matrices

Although all the results we state in this section have already been obtained, they are sufficiently useful to deserve being summarized separately. In this section we are considering matrices whose entries are complex numbers.

Theorem 10.1. *If H is Hermitian, then*
(1) *H is unitary similar to a diagonal matrix D.*
(2) *The elements along the main diagonal of D are the eigenvalues of H.*
(3) *The eigenvalues of H are real.*
Conversely, if H is normal and all its eigenvalues are real, then H is Hermitian.

PROOF. We have already observed that a Hermitian matrix is normal so that (1) and (2) follow immediately. Since D is diagonal and Hermitian, $\bar{D} = D^* = D$ and the eigenvalues are real.

Conversely, if H is a normal matrix with real eigenvalues, then the diagonal form to which it is unitary similar must be real and hence Hermitian. Thus H itself must be Hermitian. □

Theorem 10.2. *If A is unitary, then*
(1) *A is unitary similar to a diagonal matrix D.*
(2) *The elements along the main diagonal of D are the eigenvalues of A.*
(3) *The eigenvalues of A are of absolute value 1.*
Conversely, if A is normal and all its eigenvalues are of absolute value 1, then A is unitary.

PROOF. We have already observed that a unitary matrix is normal so that (1) and (2) follow immediately. Since D is also unitary, $\bar{D}D = D^*D = I$ so that $|\lambda_i|^2 = \bar{\lambda}_i\lambda_i = 1$ for each eigenvalue λ_i.

Conversely, if A is a normal matrix with eigenvalues of absolute value 1, then from the diagonal form D we have $D^*D = \bar{D}D = I$ so that D and A are unitary. □

Corollary 10.3. *If A is orthogonal, then*
(1) *A is unitary similar to a diagonal matrix D.*
(2) *The elements along the main diagonal of D are the eigenvalues of A.*
(3) *The eigenvalues of A are of absolute value 1.* □

This is a conventional statement of this corollary and in this form it is somewhat misleading. If A is a unitary matrix that happens to be real, then this corollary says nothing that is not contained in Theorem 10.2. A little more information about A and its eigenvalues is readily available. For example, the characteristic equation is real so that the eigenvalues occur in conjugate pairs. An orthogonal matrix of odd order has at least one real eigenvalue, etc. If A is really an orthogonal matrix, representing an isometry in a vector space over the real numbers, then the unitary matrix mentioned in the corollary does not necessarily represent a permissible change of basis. An orthogonal matrix is not always orthogonal similar to a diagonal matrix. As an example, consider the matrix representing a 90° rotation in the Euclidean plane. However, properly interpreted, the corollary is useful.

EXERCISES

1. Find the diagonal matrices to which the following matrices are unitary similar. Classify each as to whether it is Hermitian, unitary, or orthogonal.

(*a*) $\begin{bmatrix} \frac{1+i}{2} & \frac{1-i}{2} \\ \frac{1-i}{2} & \frac{1+i}{2} \end{bmatrix}$ (*b*) $\begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix}$ (*c*) $\begin{bmatrix} 0.6 & -0.8 \\ 0.8 & 0.6 \end{bmatrix}$

(*d*) $\begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix}$ (*e*) $\begin{bmatrix} 1 & i & 0 \\ -i & 1 & i \\ 0 & -i & 1 \end{bmatrix}.$

2. Let A be an arbitrary square complex matrix. Since A^*A is Hermitian, there is a unitary matrix P such that P^*A^*AP is a diagonal matrix D. Let $F = P^*AP$. Show that $F^*F = D$. Show that D is real and the elements of D are non-negative.

3. Show that every complex matrix can be written as the sum of a real matrix and an imaginary matrix; that is, if M is complex, then $M = A + Bi$ where A and B are real. Show that M is Hermitian if and only if A is symmetric and B is skew-symmetric. Show that M is skew-Hermitian if and only if A is skew-symmetric and B is symmetric.

11 | Real Vector Spaces

We now wish to consider linear transformation and matrices in vector spaces over the real numbers. Much of what has been done for complex vector spaces can be carried over to real vectors spaces without any difficulty. We must be careful, however, when it comes to theorems depending on the existence of eigenvalues and eigenvectors. In particular, Theorems 7.1 and 7.2 do not carry over as stated. Those parts of Section 8 and 9 which depend on these theorems must be reexamined carefully before their implications for real vector spaces can be established.

An examination of the proof of Theorem 7.1 will reveal that the only use made of any special properties of the complex numbers not shared by the real numbers was at the point where it was asserted that each linear transformation has at least one eigenvalue. In stating a corresponding theorem for real vector spaces we have to add an assumption concerning the existence of eigenvalues. Thus we have the following modification of Theorem 7.1 for real vector spaces.

Theorem 11.1 *Let V be a finite dimensional vector space over the real numbers, and let σ be a linear transformation on V whose characteristic polynomial factors into real linear factors. Then there exists an orthonormal basis of V with respect to which the matrix representing σ is in superdiagonal form.*

PROOF. Let n be the dimension of V. The theorem is certainly true for $n = 1$.

Assume the theorem is true for real vector spaces of dimensions $<n$. Let λ_1 be an eigenvalue for σ and let $\xi_1' \neq 0$, $\|\xi_1'\| = 1$, be a corresponding eigenvector. There exists an orthonormal basis with ξ_1' as the first element. Let the basis be $X' = \{\xi_1', \dots, \xi_n'\}$ and let W be the subspace spanned by $\{\xi_2', \dots, \xi_n'\}$. For each $\alpha = \sum_{i=1}^n a_i\xi_i'$ define $\tau(\alpha) = \sum_{i=2}^n a_i\xi_i' \in W$. Then $\tau\sigma$ restricted to W is a linear transformation of W into itself.

In the proof of Theorem 7.1 we could apply the induction hypothesis to $\tau\sigma$ without any difficulty since the assumptions of Theorem 7.1 applied to all linear transformations on V. Now we are dealing with a set of linear transformations, however, whose characteristic polynomials factor into real linear factors. Thus we must show that the characteristic polynomial for $\tau\sigma$ factors into real linear factors.

First, consider $\tau\sigma$ as defined on all of V. Since $\tau\sigma(\xi_1') = \tau(\lambda_1\xi_1') = 0$, $\tau\sigma(\alpha) = \tau\sigma[\tau(\alpha)] = \tau\sigma\tau(\alpha)$ for all $\alpha \in V$. This implies that $(\tau\sigma)^k(\alpha) = \tau\sigma^k(\alpha)$ since any τ to the right of a σ can be omitted if there is a τ to the left of that σ.

Let $f(x)$ be the characteristic polynomial for σ. It follows from the observations of the previous paragraph that $\tau f(\tau\sigma) = \tau f(\sigma) = 0$ on V. But on W, τ acts like the identity transformation, so that $f(\tau\sigma) = 0$ when restricted to W. Hence, the minimum polynomial for $\tau\sigma$ on W divides $f(x)$. By assumption, $f(x)$ factors into real linear factors so that the minimum polynomial for $\tau\sigma$ on W must also factor into real linear factors. This means that the hypotheses of the theorem are satisfied for $\tau\sigma$ on W. By induction, there is an orthogonal basis $\{\eta_2, \dots, \eta_n\}$ of W such that for each η_k, $\tau\sigma(\eta_k)$ is expressible in terms of $\{\eta_2, \dots, \eta_k\}$ alone. We see from the way τ is defined that $\sigma(\eta_k)$ is expressible in terms of $\{\xi_1', \eta_2, \ldots, \eta_k\}$ alone. Let $\eta_1 = \xi_1'$. Then $Y = \{\eta_1, \eta_2, \ldots, \eta_n\}$ is the required basis. □

Since any $n \times n$ matrix with real entries represents some linear transformation with respect to any orthonormal basis, we have

Theorem 11.2. *Let A be a real matrix with real characteristic values. Then A is orthogonal similar to a superdiagonal matrix.* □

Now let us examine the extent to which Sections 8 and 9 apply to real vector spaces. Theorem 8.1 applies to matrices with coefficients in any subfield of the complex numbers and we can use it for real matrices without reservation. Theorem 8.2 does not hold for real matrices, however. To obtain the corresponding theorem over the real numbers we must add the assumption that the characteristic values are real. A normal matrix with real characteristic values is Hermitian and, being real, it must then be symmetric. On the other hand a real symmetric matrix has all real characteristic values. Hence, we have

Theorem 11.3. *A real matrix is orthogonal similar to a diagonal matrix if and only if it is symmetric.* □

Because of the importance of real quadratic forms, in many applications this is a very useful and important theorem, one of the most important of this chapter. We describe some of the applications in Chapter VI and show how this theorem is used.

Of the theorems in Section 9 only Theorems 9.14 and 9.16 fail to hold as stated for real vector spaces. As before, adding the assumption that all the characteristic values of the linear transformation σ are real to the condition that σ is normal amounts to assuming that σ is self-adjoint. Hence, the theorems corresponding to Theorems 9.14 and 9.16 are

Theorem 11.4. *If V is a finite dimensional vector space over the real numbers and σ is a self-adjoint linear transformation on V, then V has an orthonormal basis consisting of eigenvectors of σ.* □

Theorem 11.5. *Let V be a finite dimensional vector space over the real numbers and let σ and τ be self-adjoint linear transformations on V. If $\sigma\tau = \tau\sigma$, then there exists an orthonormal basis of V consisting of vectors which are eigenvectors for both σ and τ.* □

Theorem 9.18 must be modified by substituting the words "characteristic values" for "eigenvalues." Thus,

Theorem 11.6. *A normal linear transformation σ defined on a real vector space V is an isometry if and only if all its characteristic values are of absolute value* 1. □

EXERCISES

1. For those of the following matrices which are orthogonal similar to diagonal matrices, find the diagonal form.

(a) $\begin{bmatrix} 13 & 6 \\ 6 & -3 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 \\ 5 & 4 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix}$

(d) $\begin{bmatrix} 7 & 4 & -4 \\ 4 & -8 & -1 \\ -4 & -1 & -8 \end{bmatrix}$ (e) $\begin{bmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{bmatrix}$

(f) $\begin{bmatrix} 3 & -4 & 2 \\ -4 & -1 & 6 \\ 2 & 6 & -2 \end{bmatrix}$ (g) $\begin{bmatrix} -4 & 4 & -2 \\ 4 & -4 & 2 \\ -2 & 2 & -1 \end{bmatrix}$

(h) $\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (i) $\begin{bmatrix} 0 & 1 & 2 \\ 1 & -\frac{4}{3} & -1 \\ 2 & -1 & -\frac{15}{4} \end{bmatrix}$

(j) $\begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{bmatrix}$ (k) $\begin{bmatrix} 5 & 2 & 2 \\ 2 & 2 & -4 \\ 2 & -4 & 2 \end{bmatrix}$.

2. Which of the matrices of Exercise 1, Section 8, are orthogonal similar to diagonal matrices?

3. Let A and B be real symmetric matrices with A positive definite. There is a non-singular matrix P such that $P^TAP = I$. Show that P^TBP is symmetric. Show that there exists a non-singular matrix Q such that $Q^TAQ = I$ and Q^TBQ is a diagonal matrix.

4. Show that every real skew-symmetric matrix A has the form $A = P^TBP$ where P is orthogonal and B^2 is diagonal.

5. Show that if A and B are real symmetric matrices, and A is positive definite, then the roots of $\det(B - xA) = 0$ are all real.

6. Show that a real skew-symmetric matrix of positive rank is not orthogonal similar to a diagonal matrix.

7. Show that if A is a real 2×2 normal matrix with at least one element equal to zero, then it is symmetric or skew-symmetric.

8. Show that if A is a real 2×2 normal matrix with no zero element, then A is symmetric or a scalar multiple of an orthogonal matrix.

9. Let σ be a skew-symmetric linear transformation on the vector space V over the real numbers. The matrix A representing σ with respect to an orthonormal

basis is skew-symmetric. Show that the real characteristic values of A are zeros. The characteristic equation may have complex solutions. Show that all complex solutions are pure imaginary. Why are these solutions not eigenvalues of σ?

10. (Continuation) Show that σ^2 is symmetric. Show that the characteristic values of A^2 are real. Show that the non-zero eigenvalues of A^2 are negative. Let $-\mu^2$ be a non-zero eigenvalue of σ^2 and let ξ be a corresponding eigenvector. Define η to be $\frac{1}{\mu}\sigma(\xi)$. Show that $\sigma(\eta) = -\mu\xi$. Show that ξ and η are orthogonal. Show that η is also an eigenvector of σ^2 corresponding to $-\mu^2$.

11. (Continuation) Let σ be the skew-symmetric linear transformation considered in Exercises 9 and 10. Show that there exists an orthonormal basis of V such that the matrix representing σ has all zero elements except for a sequence of 2×2 matrices down the main diagonal of the form

$$\begin{bmatrix} 0 & -\mu_k \\ \mu_k & 0 \end{bmatrix},$$

where the numbers μ_k are defined as in Exercise 10.

12. Let σ be an orthogonal linear transformation on a vector space V over the real numbers. Show that the real characteristic values of σ are ± 1. Show that any eigenvector of σ corresponding to a real eigenvalue is also an eigenvector of σ^* corresponding to the same eigenvalue. Show that these eigenvectors are also eigenvectors of $\sigma + \sigma^*$ corresponding to the eigenvalues ± 2.

13. (Continuation) Show that $\sigma + \sigma^*$ is self-adjoint. Show that there exists a basis of eigenvectors of $\sigma + \sigma^*$. Show that if an eigenvector of $\sigma + \sigma^*$ is also an eigenvector of σ, then the corresponding eigenvalue is ± 2. Let 2μ be an eigenvalue of $\sigma + \sigma^*$ for which the corresponding eigenvector ξ is not an eigenvector of σ. Show that μ is real and that $|\mu| < 1$. Show that $(\xi, \sigma(\xi)) = \mu(\xi, \xi)$.

14. (Continuation) Define η to be $\dfrac{\sigma(\xi) - \mu\xi}{\sqrt{1 - \mu^2}}$. Show that ξ and η are orthogonal. Show that $\sigma(\xi) = \mu\xi + \sqrt{1 - \mu^2}\eta$, and $\sigma(\eta) = -\sqrt{1 - \mu^2}\xi + \mu\eta$.

15. (Continuation) Let σ be the orthogonal linear transformation considered in Exercises 12, 13, 14. Show that there exists an orthonormal basis of V such that the matrix representing σ has all zero elements except for a sequence of ± 1's and/or 2×2 matrices down the main diagonal of the form

$$\begin{bmatrix} \cos\theta_k & -\sin\theta_k \\ \sin\theta_k & \cos\theta_k \end{bmatrix},$$

where $\mu_k = \cos\theta_k$ are defined as in Exercise 13.

12 | The Computational Processes

We now summarize a complete set of computational steps which will effectively determine a unitary (or orthogonal) matrix of transition for diagonalizing a given normal matrix. Let A be a given normal matrix.

1. Determine the characteristic matrix $C(x) = A - xI$.
2. Compute the characteristic polynomial $f(x) = \det(A - xI)$.
3. Determine all eigenvalues of A by finding all the solutions of the characteristic equation $f(x) = 0$. In any but very special or contrived examples this step is tedious and lengthy. In an arbitrarily given example we can find at best only approximate solutions. In that case all the following steps are also approximate. In some applications special information derivable from the peculiarities of the application will give information about the eigenvalues or the eigenvectors without our having to solve the characteristic equation.
4. For each eigenvalue λ_i find the corresponding eigenvectors by solving the homogeneous linear equations

$$C(\lambda_i)X = 0. \tag{12.1}$$

Each such system of linear equations is of rank less than n. Thus the technique of Chapter II-7 is the recommended method.

5. Find an orthonormal basis consisting of eigenvectors of A. If the eigenvalues are distinct, Theorem 9.3 assures us that they are mutually orthogonal. Thus all that must be done is to normalize each vector and the required orthonormal basis is obtained immediately.

Even where a multiple eigenvalue λ_i occurs, Theorem 8.2 or Theorem 9.14 assures us that an orthonormal basis of eigenvectors exists. Thus, the nullity of $C(\lambda_i)$ must be equal to the algebraic multiplicity of λ_i. Hence, there is no difficulty in obtaining a basis of eigenvectors. The problem is that the different eigenvectors corresponding to the multiple eigenvalue λ_i are not automatically orthogonal; however, that is easily remedied. All we need to do is to take a basis of eigenvectors and use the Gram-Schmidt orthonormalization process in each eigenspace. The vectors obtained in this way will still be eigenvectors since they are linear combinations of eigenvectors corresponding to the same eigenvalue. Vectors from different eigenspaces will be orthogonal because of Theorem 9.3. Since eigenspaces are seldom of very high dimensions, the amount of work involved in applying the Gram-Schmidt process is usually quite nominal.

We now give several examples to illustrate the computational procedures and the various diagonalization theorems. Remember that these examples are contrived so that the characteristic equation can easily be solved. Randomly given examples of high order are very likely to result in vexingly difficult characteristic equations.

Example 1. A real symmetric matrix with distinct eigenvalues. Let

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

We first determine the characteristic matrix,

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

and then the characteristic polynomial,

$$f(x) = \det C(x) = -x^3 + 6x^2 - 3x - 10 = -(x+1)(x-2)(x-5).$$

The eigenvalues are $\lambda_1 = -1$, $\lambda_2 = 2$, $\lambda_3 = 5$.

Solving the equations $C(\lambda_i)X = 0$ we obtain the eigenvectors $\alpha_1 = (2, 2, 1)$, $\alpha_2 = (-2, 1, 2)$, $\alpha_3 = (1, -2, 2)$. Theorem 9.3 assures us that these eigenvectors are orthogonal, and upon checking we see that they are. Normalizing them, we obtain the orthonormal basis

$$X = \{\xi_1 = \tfrac{1}{3}(2, 2, 1),\ \xi_2 = \tfrac{1}{3}(-2, 1, 2),\ \xi_3 = \tfrac{1}{3}(1, -2, 2)\}.$$

The orthogonal matrix of transition is

$$P = \frac{1}{3}\begin{bmatrix} 2 & -2 & 1 \\ 2 & 1 & -2 \\ 1 & 2 & 2 \end{bmatrix}.$$

Example 2. A real symmetric matrix with repeated eigenvalues. Let

$$A = \begin{bmatrix} 5 & 2 & 2 \\ 2 & 2 & -4 \\ 2 & -4 & 2 \end{bmatrix}.$$

The corresponding characteristic matrix is

$$C(x) = \begin{bmatrix} 5-x & 2 & 2 \\ 2 & 2-x & -4 \\ 2 & -4 & 2-x \end{bmatrix},$$

and the characteristic polynomial is

$$f(x) = -x^3 + 9x^2 - 108 = -(x + 3)(x - 6)^2.$$

The eigenvalues are $\lambda_1 = -3$, $\lambda_2 = \lambda_3 = 6$.

Corresponding to $\lambda_1 = -3$, we obtain the eigenvector $\alpha_1 = (1, -2, -2)$. For $\lambda_2 = \lambda_3 = 6$ we find that the eigenspace $S(6)$ is of dimension 2 and is the set of all solutions of the equation

$$x_1 - 2x_2 - 2x_3 = 0.$$

Thus $S(6)$ has the basis $\{(2, 1, 0), (2, 0, 1)\}$. We can now apply the Gram-Schmidt process to obtain the orthonormal basis

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

Again, by Theorem 9.3 we are assured that α_1 is orthogonal to all vectors in $S(6)$, and to these vectors in particular. Thus,

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

is an orthonormal basis of eigenvectors. The orthogonal matrix of transition is

$$P = \begin{bmatrix} \frac{1}{3} & \frac{2}{\sqrt{5}} & \frac{2}{3\sqrt{5}} \\ \frac{-2}{3} & \frac{1}{\sqrt{5}} & \frac{-4}{3\sqrt{5}} \\ \frac{-2}{3} & 0 & \frac{5}{3\sqrt{5}} \end{bmatrix}.$$

It is worth noting that, whereas the eigenvector corresponding to an eigenvalue of multiplicity 1 is unique up to a factor of absolute value 1, the orthonormal basis of the eigenspace corresponding to a multiple eigenvalue is not unique. In this example, any vector orthogonal to $(1, -2, -2)$ must be in $S(6)$. Thus $\{\frac{1}{3}(2, 2, -1), \frac{1}{3}(2, -1, 2)\}$ would be another choice for an orthonormal basis for $S(6)$. It happens to result in a slightly simpler orthogonal matrix of transition (in this case a matrix over the rational numbers.)

Example 3. *A Hermitian matrix.* Let

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

Then

$$C(x) = \begin{bmatrix} 2 - x & 1 - i \\ 1 + i & 3 - x \end{bmatrix},$$

and $f(x) = x^2 - 5x + 4 = (x - 1)(x - 4) = 0$ is the characteristic equation. The eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = 4$. (The example is contrived so that the eigenvalues are rational, but the fact that they are real is assured by Theorem 10.1.) Corresponding to $\lambda_1 = 1$ we obtain the normalized eigenvector $\xi_1 = \dfrac{1}{\sqrt{3}}(-1 + i, 1)$, and corresponding to $\lambda_2 = 4$ we obtain the normalized eigenvector $\xi_2 = \dfrac{1}{\sqrt{3}}(1, 1 + i)$. The unitary matrix of transition is

$$U = \frac{1}{\sqrt{3}}\begin{bmatrix} -1 + i & 1 \\ 1 & 1 + i \end{bmatrix}.$$

Example 4. *An orthogonal matrix.* Let

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

This orthogonal matrix is real but not symmetric. Therefore, it is unitary similar to a diagonal matrix but it is not orthogonal similar to a diagonal matrix. We have

$$C(x) = \begin{bmatrix} \frac{1}{3} - x & -\frac{2}{3} & \frac{2}{3} \\ -\frac{2}{3} & \frac{1}{3} - x & \frac{2}{3} \\ -\frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} - x \end{bmatrix},$$

and, hence, $-x^3 + \frac{1}{3}x^2 - \frac{1}{3}x + 1 = -(x - 1)(x^2 + \frac{2}{3}x + 1) = 0$ is the characteristic equation. Notice that the real eigenvalues of an orthogonal matrix are particularly easy to find since they must be of absolute value 1. The eigenvalues are $\lambda_1 = 1$, $\lambda_2 = \dfrac{-1 + 2\sqrt{2}i}{3}$, and $\lambda_3 = \dfrac{-1 - 2\sqrt{2}i}{3}$. The corresponding normalized eigenvectors are $\xi_1 = \dfrac{1}{\sqrt{2}}(1, -1, 0)$, $\xi_2 = \frac{1}{2}(1, 1, \sqrt{2}i)$, and $\xi_3 = \frac{1}{2}(1, 1, -\sqrt{2}i)$. Thus, the unitary matrix of transition is

$$U = \begin{bmatrix} 1/\sqrt{2} & \frac{1}{2} & \frac{1}{2} \\ -1/\sqrt{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & i/\sqrt{2} & -i/\sqrt{2} \end{bmatrix}.$$

EXERCISES

1. Apply the computational methods outlined in this section to obtain the orthogonal or unitary matrices of transition to diagonalize each of the normal matrices given in Exercises 1 of Sections 8, 10, and 11.

2. Carry out the program outlined in Exercises 12 through 15 of Section 11. Consider the orthogonal linear transformation σ represented by the orthogonal matrix

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

Find an orthonormal basis of eigenvectors of $\sigma + \sigma^*$. Find the representation of σ with respect to this basis. Since $\sigma + \sigma^*$ has one eigenvalue of multiplicity 2, the pairing described in Exercise 14 of Section 11 is not necessary. If $\sigma + \sigma^*$ had an eigenvalue of multiplicity 4 or more, such a pairing would be required to obtain the desired form.