Linear Operators

Recall that a linear transformation $T \in \text{L}(V)$ of a vector space into itself is called a (linear) operator. In this chapter we shall elaborate somewhat on the theory of operators. In so doing, we will define several important types of operators, and we will also prove some important diagonalization theorems. Much of this material is directly useful in physics and engineering as well as in mathematics. While some of this chapter overlaps with Chapter 8, we assume that the reader has studied at least Section 8.1.

10.1 LINEAR FUNCTIONALS AND ADJOINTS

Recall that in Theorem 9.3 we showed that for a finite-dimensional real inner product space $V$, the mapping $u \mapsto L_u = \langle u, \, \cdot \rangle$ was an isomorphism of $V$ onto $V^*$. This mapping had the property that $L_{au}v = \langle au, v \rangle = a\langle u, v \rangle = aL_u v$, and hence $L_{au} = aL_u$ for all $u \in V$ and $a \in \mathbb{R}$. However, if $V$ is a complex space with a Hermitian inner product, then $L_{au}v = \langle au, v \rangle = a^*\langle u, v \rangle = a^*L_u v$, and hence $L_{au} = a^*L_u$ which is not even linear (this was the definition of an anti-linear (or conjugate linear) transformation given in Section 9.2). Fortunately, there is a closely related result that holds even for complex vector spaces.

Let $V$ be finite-dimensional over $\mathbb{C}$, and assume that $V$ has an inner product $\langle \, , \rangle$ defined on it (this is just a positive definite Hermitian form on $V$). Thus for any $X, Y \in V$ we have $\langle X, Y \rangle \in \mathbb{C}$. For example, with respect to the
standard basis \{e_i\} for \(\mathbb{C}^n\) (which is the same as the standard basis for \(\mathbb{R}^n\)), we have \(X = \sum x^i e_i\) and hence (see Example 2.13)

\[
\langle X, Y \rangle = \langle \sum_i x^i e_i, \sum_j y^j e_j \rangle = \sum_{i,j} x^i y^j \delta_{ij} = \sum_{i,j} x^i y^j \delta_{ij} = \sum_{i,j} x^i y^j = X^* Y.
\]

Note that we are temporarily writing \(X^*T\) rather than \(X^\dagger\). We will shortly explain the reason for this (see Theorem 10.2 below). In particular, for any \(T \in L(V)\) and \(X \in V\) we have the vector \(TX \in V\), and hence it is meaningful to write expressions of the form \(\langle TX, Y \rangle\) and \(\langle X, TY \rangle\).

Since we are dealing with finite-dimensional vector spaces, the Gram-Schmidt process (Theorem 2.21) guarantees that we can always work with an orthonormal basis. Hence, let us consider a complex inner product space \(V\) with basis \{e_i\} such that \(\langle e_i, e_j \rangle = \delta_{ij}\). Then, just as we saw in the proof of Theorem 9.1, we now see that for any \(u = \sum u^j e_j \in V\) we have

\[
\langle e_i, u \rangle = \langle e_i, \sum_j u^j e_j \rangle = \sum_j u^j \langle e_i, e_j \rangle = \sum_j u^j \delta_{ij} = u^i
\]

and thus

\[
u = \sum_i \langle e_i, u \rangle e_i.
\]

Now consider the vector \(Te_j\). Applying the result of the previous paragraph we have

\[
Te_j = \sum_i \langle e_i, Te_j \rangle e_i.
\]

But this is precisely the definition of the matrix \(A = (a_{ij})\) that represents \(T\) relative to the basis \(\{e_i\}\). In other words, this extremely important result shows that the matrix elements \(a_{ij}\) of the operator \(T \in L(V)\) are given by

\[
a_{ij} = \langle e_i, Te_j \rangle.
\]

It is important to note however, that this definition depended on the use of an orthonormal basis for \(V\). To see the self-consistency of this definition, we go back to our original definition of \((a_{ij})\) as \(Te_j = \sum_k e_k a_{kj}\). Taking the scalar product of both sides of this equation with \(e_i\) yields (using the orthonormality of the \(e_i\))

\[
\langle e_i, Te_j \rangle = \langle e_i, \sum_k e_k a_{kj} \rangle = \sum_k a_{kj} \langle e_i, e_k \rangle = \sum_k a_{kj} \delta_{ik} = a_{ij}.
\]

We now prove the complex analogue of Theorem 9.3.
Theorem 10.1  Let $V$ be a finite-dimensional inner product space over $\mathbb{C}$. Then, given any linear functional $L$ on $V$, there exists a unique $u \in V$ such that $Lv = \langle u, v \rangle$ for all $v \in V$.

Proof  Let $\{e_i\}$ be an orthonormal basis for $V$ and define $u = \sum (Le_i)^* e_i$. Now define the linear functional $L_u$ on $V$ by $L_u v = \langle u, v \rangle$ for every $v \in V$. Then, in particular, we have

$$L_u e_i = \langle u, e_i \rangle = \langle \sum (Le_i)^* e_i, e_i \rangle = \sum_{j} (Le_j)^* e_i, e_i \rangle = \sum_{j} (Le_j) \delta_{ji} = Le_i.$$ 

Since $L$ and $L_u$ agree on a basis for $V$, they must agree on any $v \in V$, and hence $L = L_u = \langle u, \cdot \rangle$.

As to the uniqueness of the vector $u$, suppose $u' \in V$ has the property that $Lv = \langle u', v \rangle$ for every $v \in V$. Then $Lv = \langle u, v \rangle = \langle u', v \rangle$ so that $\langle u - u', v \rangle = 0$. Since $v$ was arbitrary we may choose $v = u - u'$. Then $\langle u - u', u - u' \rangle = 0$ which implies that (since the inner product is just a positive definite Hermitian form) $u - u' = 0$ or $u = u'$.

The importance of finite-dimensionality in this theorem is shown by the following example.

Example 10.1  Let $V = \mathbb{R}[x]$ be the (infinite-dimensional) space of all polynomials over $\mathbb{R}$, and define an inner product on $V$ by

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

for every $f, g \in V$. We will give an example of a linear functional $L$ on $V$ for which there does not exist a polynomial $h \in V$ with the property that $Lf = \langle h, f \rangle$ for all $f \in V$.

To show this, define the nonzero linear functional $L$ by

$$Lf = f(0) .$$

($L$ is nonzero since, e.g., $L(a + x) = a$.) Now suppose there exists a polynomial $h \in V$ such that $Lf = f(0) = \langle h, f \rangle$ for every $f \in V$. Then, in particular, we have

$$L(xf) = 0f(0) = 0 = \langle h, xf \rangle$$

for every $f \in V$. Choosing $f = xh$ we see that

$$0 = \langle h, x^2 h \rangle = \int_{0}^{1} x^2 h^2 \, dx .$$
Since the integrand is strictly positive, this forces $h$ to be the zero polynomial. Thus we are left with $L f = \langle h, f \rangle = \langle 0, f \rangle = 0$ for every $f \in V$, and hence $L = 0$. But this contradicts the fact that $L \neq 0$, and hence no such polynomial $h$ can exist.

Note the fact that $V$ is infinite-dimensional is required when we choose $f = x h$. The reason for this is that if $V$ consisted of all polynomials of degree $\leq$ some positive integer $N$, then $f = x h$ could have degree $> N$.

Now consider an operator $T \in L(V)$, and let $u$ be an arbitrary element of $V$. Then the mapping $L_{u v}: V \to \mathbb{C}$ defined by $L_{u v} = \langle u, v \rangle$ for every $v \in V$ is a linear functional on $V$. Applying Theorem 10.1, we see that there exists a unique $u' \in V$ such that $\langle u, v \rangle = L_{u v} = \langle u', v \rangle$ for every $v \in V$. We now define the mapping $T^*: V \to V$ by $T^* u = u'$. In other words, we define the adjoint $T^*$ of an operator $T \in L(V)$ by

$$\langle T^* u, v \rangle = \langle u, v \rangle$$

for all $u, v \in V$. The mapping $T^*$ is unique because $u'$ is unique for a given $u$. Thus, if $T^* u = u' = T^* u$, then $(T^* - T^*) u = 0$ for every $u \in V$, and hence $T^* - T^* = 0$ or $T^* = T^*$.

Note further that

$$\langle T u, v \rangle = \langle v, T u \rangle^* = \langle T^* v, u \rangle^* = \langle u, T^* v \rangle^* .$$

However, it follows from the definition that $\langle u, T^* v \rangle = \langle T^* u, v \rangle$. Therefore the uniqueness of the adjoint implies that $T^{**} = T$.

Let us show that the map $T^*$ is linear. For all $u_1, u_2, v \in V$ and $a, b \in \mathbb{C}$ we have

$$\langle T^* (a u_1 + b u_2), v \rangle = \langle a u_1 + b u_2, T v \rangle = a \ast \langle u_1, T v \rangle + b \ast \langle u_2, T v \rangle$$

$$= a \ast \langle T^* u_1, v \rangle + b \ast \langle T^* u_2, v \rangle$$

$$= \langle a T^* u_1, v \rangle + \langle b T^* u_2, v \rangle$$

$$= \langle a T^* u_1 + b T^* u_2, v \rangle .$$

Since this is true for every $v \in V$, we must have

$$T^* (a u_1 + b u_2) = a T^* u_1 + b T^* u_2 .$$

Thus $T^*$ is linear and $T^* \in L(V)$.  

If \( \{e_i\} \) is an orthonormal basis for \( V \), then the matrix elements of \( T \) are given by \( a_{ij} = \langle e_i, Te_j \rangle \). Similarly, the matrix elements \( b_{ij} \) of \( T^\dagger \) are related to those of \( T \) because

\[
b_{ij} = \langle e_i, T^\dagger e_j \rangle = \langle Te_i, e_j \rangle = \langle e_j, Te_i \rangle^* = a_{ji}^* .
\]

In other words, if \( A \) is the matrix representation of \( T \) relative to the orthonormal basis \( \{e_i\} \), then \( A^*T \) is the matrix representation of \( T^\dagger \). This explains the symbol and terminology for the Hermitian adjoint used in the last chapter. Note that if \( V \) is a real vector space, then the matrix representation of \( T^\dagger \) is simply \( A^T \), and we may denote the corresponding operator by \( T^T \).

We summarize this discussion in the following theorem, which is valid only in finite-dimensional vector spaces. (It is also worth pointing out that \( T^\dagger \) depends on the particular inner product defined on \( V \).)

**Theorem 10.2**  Let \( T \) be a linear operator on a finite-dimensional complex inner product space \( V \). Then there exists a unique linear operator \( T^\dagger \) on \( V \) defined by \( \langle T^\dagger u, v \rangle = \langle u, Tv \rangle \) for all \( u, v \in V \). Furthermore, if \( A \) is the matrix representation of \( T \) relative to an orthonormal basis \( \{e_i\} \), then \( A^*T \) is the matrix representation of \( T^\dagger \) relative to this same basis. If \( V \) is a real space, then the matrix representation of \( T^\dagger \) is simply \( A^T \).

**Example 10.2**  Let us give an example that shows the importance of finite-dimensionality in defining an adjoint operator. Consider the space \( V = \mathbb{R}[x] \) of all polynomials over \( \mathbb{R} \), and let the inner product be as in Example 10.1. Define the differentiation operator \( D \in L(V) \) by \( Df = df/dx \). We show that there exists no adjoint operator \( D^\dagger \) that satisfies \( \langle Df, g \rangle = \langle f, D^\dagger g \rangle \).

Using \( \langle Df, g \rangle = \langle f, D^\dagger g \rangle \), we integrate by parts to obtain

\[
\langle f, D^\dagger g \rangle = \langle Df, g \rangle = \int_0^1 (Df)g \, dx = \int_0^1 [D(fg) - fDg] \, dx
\]

\[
= (fg)(1) - (fg)(0) - \langle f, Dg \rangle .
\]

Rearranging, this general result may be written as

\[
\langle f, (D + D^\dagger)g \rangle = (fg)(1) - (fg)(0) .
\]

We now let \( f = x^2(1 - x)^2p \) for any \( p \in V \). Then \( f(1) = f(0) = 0 \) so that we are left with
Since this is true for every \( p \in V \), it follows that \( x^2(1-x)^2(D + D^\dagger)g = 0 \). But \( x^2(1-x)^2 > 0 \) except at the endpoints, and hence we must have \( (D + D^\dagger)g = 0 \) for all \( g \in V \), and thus \( D + D^\dagger = 0 \). However, the above general result then yields

\[
0 = \langle f, (D + D^\dagger)g \rangle = (fg)(1) - (fg)(0)
\]

which is certainly not true for every \( f, g \in V \). Hence \( D^\dagger \) must not exist.

We leave it to the reader to find where the infinite-dimensionality of \( V = \mathbb{R}[x] \) enters into this example. \( \Box \)

While this example shows that not every operator on an infinite-dimensional space has an adjoint, there are in fact some operators on some infinite-dimensional spaces that do indeed have an adjoint. A particular example of this is given in Exercise 10.1.3. In fact, the famous Riesz representation theorem asserts that any continuous linear functional on a Hilbert space does indeed have an adjoint. While this fact should be well known to anyone who has studied quantum mechanics, we defer further discussion until Chapter 12 (see Theorem 12.26).

As defined previously, an operator \( T \in \mathcal{L}(V) \) is **Hermitian** (or **self-adjoint**) if \( T^\dagger = T \). The elementary properties of the adjoint operator \( T^\dagger \) are given in the following theorem. Note that if \( V \) is a real vector space, then the properties of the matrix representing an adjoint operator simply reduce to those of the transpose. Hence, a real Hermitian operator is represented by a (real) symmetric matrix.

**Theorem 10.3** Suppose \( S, T \in \mathcal{L}(V) \) and \( c \in \mathbb{C} \). Then

(a) \((S + T)^\dagger = S^\dagger + T^\dagger\).

(b) \((cT)^\dagger = c^*T^\dagger\).

(c) \((ST)^\dagger = T^\dagger S^\dagger\).

(d) \(T^{\dagger\dagger} = (T^\dagger)^\dagger = T\).

(e) \(I^\dagger = I \) and \( 0^\dagger = 0 \).

(f) \((T^\dagger)^{-1} = (T^{-1})^\dagger\).

**Proof** Let \( u, v \in V \) be arbitrary. Then, from the definitions, we have

(a) \( \langle (S + T)^\dagger u, v \rangle = \langle u, (S + T)v \rangle = \langle u, Sv + Tv \rangle = \langle u, Sv \rangle + \langle u, Tv \rangle = \langle S^\dagger u, v \rangle + \langle T^\dagger u, v \rangle = \langle (S^\dagger + T^\dagger)u, v \rangle \).

(b) \( \langle (cT)^\dagger u, v \rangle = \langle cT v, u \rangle = c\langle u, Tv \rangle = c\langle T^\dagger u, v \rangle = \langle c^* T^\dagger u, v \rangle \).
(c) \((ST)^\dagger u, v) = \langle u, (ST)v \rangle = \langle u, S(Tv) \rangle = \langle S^\dagger u, T v \rangle = (T^\dagger(S^\dagger)u, v).\)

(d) This was shown in the discussion preceding Theorem 10.2.

(e) \(\langle lu, v \rangle = \langle u, v \rangle = \langle u, Iv \rangle = (I^\dagger u, v).\)

(f) \(I = I^\dagger = (T^\dagger T^\dagger)\dagger = (T^\dagger)^\dagger T^\dagger so that (T^\dagger)^\dagger = (T^\dagger)^{-1}.\)

The proof is completed by noting that the adjoint and inverse operators are unique. ■

**Corollary** If \(T \in L(V)\) is nonsingular, then so is \(T^\dagger.\)

**Proof** This follows from Theorems 10.3(f) and 5.10. ■

We now group together several other useful properties of operators for easy reference.

**Theorem 10.4** (a) Let \(V\) be an inner product space over either \(\mathbb{R}\) or \(\mathbb{C}\), let \(T \in L(V)\), and suppose that \(\langle u, Tv \rangle = 0\) for all \(u, v \in V\). Then \(T = 0.\)

(b) Let \(V\) be an inner product space over \(\mathbb{C}\), let \(T \in L(V)\), and suppose that \(\langle u, Tu \rangle = 0\) for all \(u \in V\). Then \(T = 0.\)

(c) Let \(V\) be a real inner product space, let \(T \in L(V)\) be Hermitian, and suppose that \(\langle u, Tu \rangle = 0\) for all \(u \in V\). Then \(T = 0.\)

**Proof** (a) Let \(u = Tv\). Then, by definition of the inner product, we see that \((Tv, Tv) = 0\) implies \(Tv = 0\) for all \(v \in V\) which implies that \(T = 0.\)

(b) For any \(u, v \in V\) we have (by hypothesis)

\[
0 = \langle u + v, T(u + v) \rangle
= \langle u, Tu \rangle + \langle u, Tv \rangle + \langle v, Tu \rangle + \langle v, Tv \rangle
= 0 + \langle v, Tu \rangle + \langle v, Tv \rangle + 0
= \langle u, Tv \rangle + \langle v, Tu \rangle
\]

Since \(v\) is arbitrary, we may replace it with \(iv\) to obtain

\[
0 = i\langle u, Tv \rangle - i\langle v, Tu \rangle.
\]

Dividing this by \(i\) and adding to \((*)\) results in \(0 = \langle u, Tv \rangle\) for any \(u, v \in V\). By (a), this implies that \(T = 0.\)

(c) For any \(u, v \in V\) we have \(\langle u + v, T(u + v) \rangle = 0\) which also yields \((*)\). Therefore, using \((*)\), the fact that \(T^\dagger = T\), and the fact that \(V\) is real, we obtain
0 = \langle T^\dagger u, v \rangle + \langle v, Tu \rangle = \langle Tu, v \rangle + \langle v, Tu \rangle = 2\langle v, Tu \rangle.

Since this holds for any \( u, v \in V \) we have \( T = 0 \) by (a). (Note that in this particular case, \( T^\dagger = T^T \).) ■

Exercises

1. Suppose \( S, T \in \mathbb{L}(V) \).
   (a) If \( S \) and \( T \) are Hermitian, show that \( ST \) and \( TS \) are Hermitian if and only if \([S, T] = ST - TS = 0\).
   (b) If \( T \) is Hermitian, show that \( S^\dagger TS \) is Hermitian for all \( S \).
   (c) If \( S \) is nonsingular and \( S^\dagger TS \) is Hermitian, show that \( T \) is Hermitian.

2. Consider \( V = \mathbb{M}_n(\mathbb{C}) \) with the inner product \( \langle A, B \rangle = \text{Tr}(B^\dagger A) \). For each \( M \in V \), define the operator \( T_M \in \mathbb{L}(V) \) by \( T_M(A) = MA \). Show that \( (T_M)^\dagger = T_M^\dagger \).

3. Consider the space \( V = \mathbb{C}[x] \). If \( f = \sum a_i x^i \in V \), we define the complex conjugate of \( f \) to be the polynomial \( f^* = \sum a_i^* x^i \in V \). In other words, if \( t \in \mathbb{R} \), then \( f^*(t) = (f(t))^* \). We define an inner product on \( V \) by
   \[
   \langle f, g \rangle = \int_0^1 f^*(t)g(t) \, dt .
   \]
   For each \( f \in V \), define the operator \( T_f \in \mathbb{L}(V) \) by \( T_f(g) = fg \). Show that \( (T_f)^\dagger = T_{f^*} \).

4. Let \( V \) be the space of all real polynomials of degree \( \leq 3 \), and define an inner product on \( V \) by
   \[
   \langle f, g \rangle = \int_0^1 f(x)g(x) \, dx .
   \]
   For any \( t \in \mathbb{R} \), find a polynomial \( h_t \in V \) such that \( \langle h_t, f \rangle = f(t) \) for all \( f \in V \).

5. If \( V \) is as in the previous exercise and \( D \) is the usual differentiation operator on \( V \), find \( D^\dagger \).

6. Let \( V = \mathbb{C}^2 \) with the standard inner product.
   (a) Define \( T \in \mathbb{L}(V) \) by \( T e_1 = (1, -2), T e_2 = (i, -1) \). If \( v = (z_1, z_2) \), find \( T^\dagger v \).
(b) Define \( T \in L(V) \) by \( T \mathbf{e}_1 = (1 + i, 2), \ T \mathbf{e}_2 = (i, i) \). Find the matrix representation of \( T^\dagger \) relative to the usual basis for \( V \). Is it true that \([T, T^\dagger] = 0\)?

7. Let \( V \) be a finite-dimensional inner product space and suppose \( T \in L(V) \). Show that \( \text{Im } T^\dagger = (\text{Ker } T)^\perp \).

8. Let \( V \) be a finite-dimensional inner product space, and suppose \( E \in L(V) \) is idempotent, i.e., \( E^2 = E \). Prove that \( E^\dagger = E \) if and only if \([E, E^\dagger] = 0\).

9. For each of the following inner product spaces \( V \) and \( L \in V^* \), find a vector \( u \in V \) such that \( Lv = \langle u, v \rangle \) for all \( v \in V \):

   (a) \( V = \mathbb{R}^3 \) and \( L(x, y, z) = x - 2y + 4z \).

   (b) \( V = \mathbb{C}^2 \) and \( L(z_1, z_2) = z_1 - z_2 \).

   (c) \( V \) is the space of all real polynomials of degree \( \leq 2 \) with inner product as in Exercise 4, and \( Lf = f(0) + Df(1) \). (Here \( D \) is the usual differentiation operator.)

10. (a) Let \( V = \mathbb{R}^2 \), and define \( T \in L(V) \) by \( T(x, y) = (2x + y, x - 3y) \). Find \( T^\dagger(3, 5) \).

   (b) Let \( V = \mathbb{C}^2 \), and define \( T \in L(V) \) by \( T(z_1, z_2) = (2z_1 + iz_2, (1 - i)z_1) \).

   Find \( T^\dagger(3 - i, 1 + i2) \).

   (c) Let \( V \) be as in Exercise 9(c), and define \( T \in L(V) \) by \( Tf = 3f + Df \).

   Find \( T^\dagger f \) where \( f = 3x^2 - x + 4 \).

10.2 ISOMETRIC AND UNITARY OPERATORS

Let \( V \) be a complex inner product space with the induced norm. Another important class of operators \( U \in L(V) \) is that for which \( \|Uv\| = |v| \) for all \( v \in V \). Such operators are called isometric because they preserve the length of the vector \( v \). Furthermore, for any \( v, w \in V \) we see that

\[ |Uv - Uw| = |U(v - w)| = |v - w| \]

so that \( U \) preserves distances as well. This is sometimes described by saying that \( U \) is an isometry.

If we write out the norm as an inner product and assume that the adjoint operator exists, we see that an isometric operator satisfies

\[ \langle v, v \rangle = \langle Uv, Uv \rangle = \langle v, (U^\dagger U)v \rangle \]
and hence $\langle v, (U^\dagger U - 1)v \rangle = 0$ for any $v \in V$. But then from Theorem 10.4(b) it follows that

$$U^\dagger U = 1 .$$

In fact, this is sometimes taken as the definition of an isometric operator. Note that this applies equally well to an infinite-dimensional space.

If $V$ is finite-dimensional, then (Theorems 3.21 and 5.13) it follows that $U^\dagger = U^{-1}$, and hence

$$U^\dagger U = UU^\dagger = 1 .$$

Any operator that satisfies either $U^\dagger U = UU^\dagger = 1$ or $U^\dagger = U^{-1}$ is said to be unitary. It is clear that a unitary operator is necessarily isometric. If $V$ is simply a real space, then unitary operators are called orthogonal.

Because of the importance of isometric and unitary operators in both mathematics and physics, it is worth arriving at both of these definitions from a slightly different viewpoint that also aids in our understanding of these operators. Let $V$ be a complex vector space with an inner product defined on it. We say that an operator $U$ is unitary if $|Uv| = |v|$ for all $v \in V$, and in addition, it has the property that it is a mapping of $V$ onto itself. Since $|Uv| = |v|$, we see that $Uv = 0$ if and only if $v = 0$, and hence $\text{Ker } U = \{0\}$. Therefore $U$ is one-to-one and $U^{-1}$ exists (Theorem 5.5). Since $U$ is surjective, the inverse is defined on all of $V$ also. Note that there has been no mention of finite-dimensionality. This was avoided by requiring that the mapping be surjective.

Starting from $|Uv| = |v|$, we may write $\langle v, (U^\dagger U)v \rangle = \langle v, v \rangle$. As we did in the proof of Theorem 10.4, if we first substitute $v = v_1 + v_2$ and then $v = v_1 + iv_2$, divide the second of these equations by $i$ and then add to the first, we find that $\langle v_1, (U^\dagger U)v_2 \rangle = \langle v_1, v_2 \rangle$. Since this holds for all $v_1, v_2 \in V$, it follows that $U^\dagger U = 1$. If we now multiply this equation from the left by $U$ we have $UU^\dagger U = U$, and hence $(UU^\dagger)(Uv) = Uv$ for all $v \in V$. But as $v$ varies over all of $V$, so does $Uv$ since $U$ is surjective. We then define $v' = Uv$ so that $(UU^\dagger)v' = v'$ for all $v' \in V$. This shows that $U^\dagger U = 1$ implies $UU^\dagger = 1$. What we have just done then, is show that a surjective norm-preserving operator $U$ has the property that $U^\dagger U = UU^\dagger = 1$. It is important to emphasize that this approach is equally valid in infinite-dimensional spaces.

We now define an isometric operator $\Omega$ to be an operator defined on all of $V$ with the property that $|\Omega v| = |v|$ for all $v \in V$. This differs from a unitary operator in that we do not require that $\Omega$ also be surjective. Again, the requirement that $\Omega$ preserve the norm tells us that $\Omega$ has an inverse (since it must be one-to-one), but this inverse is not necessarily defined on the whole of $V$. For example, let $\{e_i\}$ be an orthonormal basis for $V$, and define the “shift operator” $\Omega$ by

$$\Omega(e_i) = e_{i+1} .$$
This $\Omega$ is clearly defined on all of $V$, but the image of $\Omega$ is not all of $V$ since it does not include the vector $e_1$. Thus, $\Omega^{-1}$ is not defined on $e_1$.

Exactly as we did for unitary operators, we can show that $\Omega^+\Omega = 1$ for an isometric operator $\Omega$. If $V$ happens to be finite-dimensional, then obviously $\Omega\Omega^+ = 1$. Thus, on a finite-dimensional space, an isometric operator is also unitary.

Finally, let us show an interesting relationship between the inverse $\Omega^+$ of an isometric operator and its adjoint $\Omega^\dagger$. From $\Omega^+\Omega = 1$, we may write $\Omega^+(\Omega v) = v$ for every $v \in V$. If we define $\Omega v = v'$, then for every $v' \in \text{Im } \Omega$ we have $v = \Omega^+v'$, and hence

$$\Omega^+v' = \Omega^{-1}v' \text{ for } v' \in \text{Im } \Omega.$$  

On the other hand, if $w' \in (\text{Im } \Omega)^\perp$, then automatically $\langle w', \Omega v \rangle = 0$ for every $v \in V$. Therefore this may be written as $\langle \Omega^+w', v \rangle = 0$ for every $v \in V$, and hence (choose $v = \Omega^+w'$)

$$\Omega^+w' = 0 \text{ for } w' \in (\text{Im } \Omega)^\perp.$$ 

In other words, we have

$$\Omega^\dagger = \begin{cases} \Omega^{-1} & \text{on } \text{Im } \Omega \\ 0 & \text{on } (\text{Im } \Omega)^\perp \end{cases}. $$

For instance, using our earlier example of the shift operator, we see that $\langle e_1, e_i \rangle = 0$ for $i \neq 1$, and hence $e_1 \in (\text{Im } \Omega)^\perp$. Therefore $\Omega^+(e_1) = 0$, so that we clearly cannot have $\Omega\Omega^+ = 1$.

Our next theorem summarizes some of this discussion.

**Theorem 10.5** Let $V$ be a complex finite-dimensional inner product space. Then the following conditions on an operator $U \in \mathcal{L}(V)$ are equivalent:

(a) $U^\dagger = \overline{U}$.

(b) $\langle Uv, Uw \rangle = \langle v, w \rangle$ for all $v, w \in V$.

(c) $\|Uv\| = \|v\|$.

**Proof** (a) $\Rightarrow$ (b): $\langle Uv, Uw \rangle = \langle v, (U^\dagger U)w \rangle = \langle v, Uw \rangle = \langle v, w \rangle$.

(b) $\Rightarrow$ (c): $\|Uv\| = \|Uv\|^{1/2} = \|v\|^{1/2} = \|v\|$.

(c) $\Rightarrow$ (a): $\langle v, (U^\dagger U)v \rangle = \langle Uv, Uv \rangle = \langle v, v \rangle = \langle v, Iv \rangle$, and therefore $\langle v, (U^\dagger U - I)v \rangle = 0$. Hence (by Theorem 10.4(b)) we must have $U^\dagger U = I$, and thus $U^\dagger = U^{-1}$ (since $V$ is finite-dimensional).
From part (c) of this theorem we see that $U$ preserves the length of any vector. In particular, $U$ preserves the length of a unit vector, hence the designation “unitary.” Note also that if $v$ and $w$ are orthogonal, then $\langle v, w \rangle = 0$ and hence $\langle Uv, Uw \rangle = \langle v, w \rangle = 0$. Thus $U$ maintains orthogonality as well.

Condition (b) of this theorem is sometimes described by saying that a unitary transformation preserves inner products. In general, we say that a linear transformation (i.e., a vector space homomorphism) $T$ of an inner product space $V$ onto an inner product space $W$ (over the same field) is an inner product space isomorphism of $V$ onto $W$ if it also preserves inner products. Therefore, one may define a unitary operator as an inner product space isomorphism.

It is also worth commenting on the case of unitary operators defined on a real vector space. Since in this case the adjoint reduces to the transpose, we have $U^\dagger = U^T = U^{-1}$. If $V$ is a real vector space, then an operator $T = L(V)$ that satisfies $T^T = T^{-1}$ is said to be an orthogonal transformation. It should be clear that Theorem 10.5 also applies to real vector spaces if we replace the adjoint by the transpose. We will have more to say about orthogonal transformations below.

**Theorem 10.6** Let $V$ be finite-dimensional over $\mathbb{C}$ (resp. $\mathbb{R}$). A linear transformation $U \in L(V)$ is unitary (resp. orthogonal) if and only if it takes an orthonormal basis for $V$ into an orthonormal basis for $V$.

**Proof** We consider the case where $V$ is complex, leaving the real case to the reader. Let $\{e_i\}$ be an orthonormal basis for $V$, and assume that $U$ is unitary. Then from Theorem 10.5(b) we have

$$\langle Ue_i, Ue_j \rangle = \langle e_i, e_j \rangle = \delta_{ij}$$

so that $\{Ue_i\}$ is also an orthonormal set. But any orthonormal set is linearly independent (Theorem 2.19), and hence $\{Ue_i\}$ forms a basis for $V$ (since there are as many of the $Ue_i$ as there are $e_i$).

Conversely, suppose that both $\{e_i\}$ and $\{Ue_i\}$ are orthonormal bases for $V$ and let $v, w \in V$ be arbitrary. Then

$$\langle v, w \rangle = \langle \sum_i v^i e_i, \sum_j w^j e_j \rangle = \sum_{i,j} v^i \overline{w^j} \delta_{ij} = \sum_{i,j} v^i \overline{w^j} \delta_{ij} = \sum_{j} v^j \overline{w^j} \delta_{ij} \quad \text{and} \quad \langle v, w \rangle = \sum_{i,j} v^i \overline{w^j} \delta_{ij}$$

However, we also have
\[ \langle Uv, Uw \rangle = \langle U(\sum_i v^i e_i), U(\sum_j w^j e_j) \rangle = \sum_{i,j} v^i w^j \langle U e_i, U e_j \rangle = \sum_{i,j} v^i w^j \delta_{ij} = \sum_i v^i \overline{w}^i = \langle v, w \rangle. \]

This shows that \( U \) is unitary (Theorem 10.5).

**Corollary**  Let \( V \) and \( W \) be finite-dimensional inner product spaces over \( \mathbb{C} \). Then there exists an inner product space isomorphism of \( V \) onto \( W \) if and only if \( \dim V = \dim W \).

**Proof**  Clearly \( \dim V = \dim W \) if \( V \) and \( W \) are isomorphic. On the other hand, let \( \{ e_1, \ldots, e_n \} \) be an orthonormal basis for \( V \), and let \( \{ \overline{e}_1, \ldots, \overline{e}_n \} \) be an orthonormal basis for \( W \). (These bases exist by Theorem 2.21.) We define the (surjective) linear transformation \( U \) by the requirement \( U e_i = \overline{e}_i \). \( U \) is unique by Theorem 5.1. Since \( \langle U e_i, U e_j \rangle = \langle \overline{e}_i, \overline{e}_j \rangle = \delta_{ij} = \langle e_i, e_j \rangle \), the proof of Theorem 10.6 shows that \( U \) preserves inner products. In particular, we see that \( |Uv| = |v| \) for every \( v \in V \), and hence \( \text{Ker} \ U = \{0\} \) (by property (N1) of Theorem 2.17). Thus \( U \) is also one-to-one (Theorem 5.5).

From Theorem 10.2 we see that a complex matrix \( A \) represents a unitary operator relative to an orthonormal basis if and only if \( A^\dagger = A^{-1} \). We therefore say that a complex matrix \( A \) is a **unitary matrix** if \( A^\dagger = A^{-1} \). In the special case that \( A \) is a real matrix with the property that \( A^T = A^{-1} \), then we say that \( A \) is an **orthogonal matrix**. (These classes of matrices were also discussed in Section 8.1.) The reason for this designation is shown in the next example, which is really nothing more than another way of looking at what we have done so far.

**Example 10.3**  Suppose \( V = \mathbb{R}^n \) and \( X \in V \). In terms of an orthonormal basis \( \{ e_i \} \) for \( V \) we may write \( X = \sum_i x^i e_i \). Now suppose we are given another orthonormal basis \( \{ \overline{e}_i \} \) related to the first basis by \( \overline{e}_i = A(e_i) = \sum_j a_{ij} e_j \) for some real matrix \( (a_{ij}) \). Relative to this new basis we have \( A(X) = \overline{X} = \sum_i x^i \overline{e}_i \) where \( x^i = \sum_j a_{ij} \overline{x}^j \) (see Section 5.4). Then

\[
\|X\|^2 = (\sum_i x^i e_i, \sum_j x^j e_j) = \sum_{i,j} x^i \overline{x}^j \left( e_i, e_j \right) = \sum_{i,j} x^i x^j \delta_{ij} = \sum_{i,j} (x^i)^2 = \sum_{i,j,k} a_{ij} a_{ik} \overline{x}^j \overline{x}^k = \sum_{i,j,k} \overline{x}^j A^T A \overline{x}^k = \sum_{j,k} (A^T A)_{jk} \overline{x}^j \overline{x}^k.
\]

If \( A \) is orthogonal, then \( A^T = A^{-1} \) so that \( (A^T A)_{jk} = \delta_{jk} \) and we are left with
\[ |X|^2 = \sum_i (x^i)^2 = \sum_i (\bar{x}^i)^2 = |ar{X}|^2 \]

so that the length of \( X \) is unchanged under an orthogonal transformation. An equivalent way to see this is to assume that \( A \) simply represents a rotation so that the length of a vector remains unchanged by definition. This then forces \( A \) to be an orthogonal transformation (see Exercise 10.2.2).

Another way to think of orthogonal transformations is the following. We saw in Section 2.4 that the angle \( \theta \) between two vectors \( X, Y \in \mathbb{R}^n \) is defined by

\[
\cos \theta = \frac{\langle X, Y \rangle}{||X|| ||Y||}.
\]

Under the orthogonal transformation \( A \), we then have \( \bar{X} = A(X) \) and also

\[
\cos \bar{\theta} = \frac{\langle \bar{X}, \bar{Y} \rangle}{||\bar{X}|| ||\bar{Y}||}.
\]

But \( ||\bar{X}|| = ||X|| \) and \( ||\bar{Y}|| = ||Y|| \), and in addition,

\[
\langle X, Y \rangle = \langle \sum_i x^i e_i, \sum_j y^j e_j \rangle = \sum_i x^i y^j = \sum_{i,j,k} a_{ij} \bar{x}^i a_{jk} \bar{y}^k
\]

\[
= \sum_{j,k} \delta_{jk} \bar{x}^j \bar{y}^k = \sum_{j} \bar{x}^j \bar{y}^j = \langle \bar{X}, \bar{Y} \rangle
\]

so that \( \theta = \bar{\theta} \) (this also follows from the real vector space version of Theorem 10.5). Therefore an orthogonal transformation also preserves the angle between two vectors, and hence is nothing more than a rotation in \( \mathbb{R}^n \).

**Theorem 10.7** The following conditions on a matrix \( A \) are equivalent:

(a) \( A \) is unitary.

(b) The rows \( A_i \) of \( A \) form an orthonormal set.

(c) The columns \( A^j \) of \( A \) form an orthonormal set.

**Proof** We begin by noting that, using the usual inner product on \( \mathbb{C}^n \), we have

\[
(AA^\dagger)_{ij} = \sum_k a_{ik} a^\dagger_{kj} = \sum_k a_{ik} a_{jk}^* = \sum_k a_{jk}^* a_{ik} = \langle A_j, A_i \rangle
\]

and

\[
(A^\dagger A)_{ij} = \sum_k a_{ik}^* a_{kj} = \sum_k a_{ik}^* a_{kj} = \langle A_i^\dagger, A_j \rangle.
\]

Now, if \( A \) is unitary, then \( AA^\dagger = I \) implies \( (AA^\dagger)_{ij} = \delta_{ij} \) which then implies that \( \langle A_j, A_i \rangle = \delta_{ij} \) so that (a) is equivalent to (b). Similarly, we must have
\[(A^\dagger A)_{ij} = \delta_{ij} = (A^\dagger, A^\dagger) \] so that (a) is also equivalent to (c). Therefore (b) must also be equivalent to (c). □

Note that the equivalence of (b) and (c) in this theorem means that the rows of A form an orthonormal set if and only if the columns of A form an orthonormal set. But the rows of A are just the columns of \(A^T\), and hence A is unitary if and only if \(A^T\) is unitary.

It should be obvious that this theorem applies just as well to orthogonal matrices. Looking at this in the other direction, we see that in this case \(A^T = A^{-1}\) so that \(A^TA = AA^T = I\), and therefore

\[(A^T A)_{ij} = \sum_k a^T_{ik} a_{kj} = \sum_k a_{ki} a_{kj} = \delta_{ij} \]

\[(AA^T)_{ij} = \sum_k a_{ik} a^T_{kj} = \sum_k a_{ik} a_{jk} = \delta_{ij} \, .\]

Viewing the standard (orthonormal) basis \(\{e_i\}\) for \(\mathbb{R}^n\) as row vectors, we have

\[A_i = \sum_j a_{ij} e_j, \quad \text{and hence} \]

\[\langle A_i, A_j \rangle = (\sum_k a_{ik} e_k, \sum_r a_{jr} e_r) = \sum_{k,r} a_{ik} a_{jr} \langle e_k, e_r \rangle = \sum_{k,r} a_{ik} a_{jr} \delta_{kr} = \sum_k a_{ik} a_{jk} = \delta_{ij} \, .\]

Furthermore, it is easy to see that a similar result holds for the columns of A.

Our next theorem details several useful properties of orthogonal and unitary matrices.

**Theorem 10.8**  (a) If A is an orthogonal matrix, then \(\det A = \pm 1\).

(b) If U is a unitary matrix, then \(|\det U| = 1\). Alternatively, \(\det U = e^{i\phi}\) for some real number \(\phi\).

**Proof**  (a) We have \(AA^T = I\), and hence (from Theorems 4.8 and 4.1)

\[1 = \det I = \det(AA^T) = (\det A)(\det A^T) = (\det A)^2 \]

so that \(\det A = \pm 1\).

(b) If \(UU^\dagger = I\) then, as above, we have

\[1 = \det I = \det(UU^\dagger) = (\det U)(\det U^\dagger) = (\det U)(\det U^T)^* = (\det U)(\det U)^* = |\det U|^2 \, .\]
Since the absolute value is defined to be positive, this shows that $|\det U| = 1$ and hence $\det U = e^{i\phi}$ for some real $\phi$. ■

**Example 10.4**  Let us take a look at rotations in $\mathbb{R}^2$ as shown, for example, in the figure below. Recall from Example 10.3 that if we have two bases $\{e_i\}$ and $\{\bar{e}_i\}$, then they are related by a transition matrix $A = (a_{ij})$ defined by $\bar{e}_i = \sum a_{ij} e_j$. In addition, if $X = \sum x^i e_i = \sum \bar{x}^i \bar{e}_i$, then $\bar{x}^i = \sum a_{ij} x^j$. If both $\{e_i\}$ and $\{\bar{e}_i\}$ are orthonormal bases, then

$$\langle e_i, \bar{e}_j \rangle = \langle e_i, \sum_k e_k a_{kj} \rangle = \sum_k a_{kj} \langle e_i, e_k \rangle = \sum_k a_{kj} \delta_{ik} = a_{ij}.$$  

Using the usual dot product on $\mathbb{R}^2$ as our inner product (see Section 2.4, Lemma 2.3) and referring to the figure below, we see that the elements $a_{ij}$ are given by (also see Section 0.6 for the trigonometric identities)

$$a_{11} = e_1 \cdot \bar{e}_1 = |e_1| |\bar{e}_1| \cos \theta = \cos \theta$$
$$a_{12} = e_1 \cdot \bar{e}_2 = |e_1| |\bar{e}_2| \cos(\pi/2 + \theta) = -\sin \theta$$
$$a_{21} = e_2 \cdot \bar{e}_1 = |e_2| |\bar{e}_1| \cos(\pi/2 - \theta) = \sin \theta$$
$$a_{22} = e_2 \cdot \bar{e}_2 = |e_2| |\bar{e}_2| \cos \theta = \cos \theta$$

Thus the matrix $A$ is given by

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

We leave it to the reader to compute directly that $A^T A = AA^T = I$ and $\det A = +1$.  ■

**Example 10.5**  Referring to the previous example, we can show that any (real) 2 x 2 orthogonal matrix with $\det A = +1$ has the form
\[(a_{ij}) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \]

for some \( \theta \in \mathbb{R} \). To see this, suppose \( A \) has the form

\[
\begin{pmatrix} a & b \\ c & d \end{pmatrix}
\]

where \( a, b, c, d \in \mathbb{R} \). Since \( A \) is orthogonal, its rows form an orthonormal set, and hence we have

\[ a^2 + b^2 = 1, \quad c^2 + d^2 = 1, \quad ac + bd = 0, \quad ad - bc = 1 \]

where the last equation follows from \( \det A = 1 \).

If \( a = 0 \), then the first of these equations yields \( b = \pm 1 \), the third then yields \( d = 0 \), and the last yields \( -c = 1/b = \pm 1 \) which is equivalent to \( c = -b \). In other words, if \( a = 0 \), then \( A \) has either of the following forms:

\[
\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.
\]

The first of these is of the required form if we choose \( \theta = -90^\circ = -\pi/2 \), and the second is of the required form if we choose \( \theta = +90^\circ = +\pi/2 \).

Now suppose that \( a \neq 0 \). From the third equation we have \( c = -bd/a \), and substituting this into the second equation, we find \( (a^2 + b^2)d^2 = a^2 \). Using the first equation, this becomes \( a^2 = d^2 \) or \( a = \pm d \). If \( a = -d \), then the third equation yields \( b = c \), and hence the last equation yields \( -a^2 - b^2 = 1 \) which is impossible. Therefore \( a = d \), the third equation then yields \( c = -b \), and we are left with

\[
\begin{pmatrix} a & -c \\ c & a \end{pmatrix}.
\]

Since \( \det A = a^2 + c^2 = 1 \), there exists a real number \( \theta \) such that \( a = \cos \theta \) and \( c = \sin \theta \) which gives us the desired form for \( A \). //

**Exercises**

1. Let \( \text{GL}(n, \mathbb{C}) \) denote the subset of \( M_n(\mathbb{C}) \) consisting of all nonsingular matrices, \( \text{U}(n) \) the subset of all unitary matrices, and \( \text{L}(n) \) the set of all
nonsingular lower-triangular matrices.

(a) Show that each of these three sets forms a group.

(b) Show that any nonsingular $n \times n$ complex matrix can be written as a product of a nonsingular upper-triangular matrix and a unitary matrix. [Hint: Use Exercises 5.4.14 and 3.7.7.]

2. Let $V = \mathbb{R}^n$ with the standard inner product, and suppose the length of any $X \in V$ remains unchanged under $A \in \text{L}(V)$. Show that $A$ must be an orthogonal transformation.

3. Let $V$ be the space of all continuous complex-valued functions defined on $[0, 2\pi]$, and define an inner product on $V$ by
   \[ \langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f^*(x)g(x) \, dx. \]

   Suppose there exists $h \in V$ such that $|h(x)| = 1$ for all $x \in [0, 2\pi]$, and define $T_h \in \text{L}(V)$ by $T_h f = hf$. Prove that $T$ is unitary.

4. Let $W$ be a finite-dimensional subspace of an inner product space $V$, and recall that $V = W \oplus W^\perp$ (see Exercise 2.5.11). Define $U \in \text{L}(V)$ by
   \[ U(w_1 + w_2) = w_1 - w_2 \]

   where $w_1 \in W$ and $w_2 \in W^\perp$.

   (a) Prove that $U$ is a Hermitian operator.

   (b) Let $V = \mathbb{R}^2$ have the standard inner product, and let $W \subset V$ be spanned by the vector $(1, 0, 1)$. Find the matrix of $U$ relative to the standard basis for $V$.

5. Let $V$ be a finite-dimensional inner product space. An operator $\Omega \in \text{L}(V)$ is said to be a partial isometry if there exists a subspace $W$ of $V$ such that $|\Omega w| = |w|$ for all $w \in W$, and $|\Omega w| = 0$ for all $w \in W^\perp$. Let $\Omega$ be a partial isometry and suppose $\{w_1, \ldots, w_k\}$ is an orthonormal basis for $W$.

   (a) Show that $\langle \Omega u, \Omega v \rangle = \langle u, v \rangle$ for all $u, v \in W$. [Hint: Use Exercise 2.4.7.]

   (b) Show that $\{\Omega w_1, \ldots, \Omega w_k\}$ is an orthonormal basis for $\text{Im} \, \Omega$.

   (c) Show there exists an orthonormal basis $\{v_i\}$ for $V$ such that the first $k$ columns of $[\Omega]_V$ form an orthonormal set, and the remaining columns are zero.

   (d) Let $\{u_1, \ldots, u_r\}$ be an orthonormal basis for $(\text{Im} \, \Omega)^\perp$. Show that $\{\Omega w_1, \ldots, \Omega w_k, u_1, \ldots, u_r\}$ is an orthonormal basis for $V$.

   (e) Suppose $T \in \text{L}(V)$ satisfies $T(\Omega w_i) = w_i$ (for $1 \leq i \leq k$) and $Tu_i = 0$ (for $1 \leq i \leq r$). Show that $T$ is well-defined, and that $T = \Omega^\dagger$. 

6. Let $V$ be a complex inner product space, and suppose $H \in L(V)$ is Hermitian. Show that:
(a) $|v + iHv| = |v - iHv|$ for all $v \in V$.
(b) $u + iHu = v + iHv$ if and only if $u = v$.
(c) $1 + iH$ and $1 - iH$ are nonsingular.
(d) If $V$ is finite-dimensional, then $U = (1 - iH)(1 + iH)^{-1}$ is a unitary operator. (U is called the Cayley transform of $H$. This result is also true in an infinite-dimensional Hilbert space but the proof is considerably more difficult.)

10.3 NORMAL OPERATORS

We now turn our attention to characterizing the type of operator on $V$ for which there exists an orthonormal basis of eigenvectors for $V$. We begin by taking a look at some rather simple properties of the eigenvalues and eigenvectors of the operators we have been discussing.

To simplify our terminology, we remark that a complex inner product space is also called a unitary space, while a real inner product space is sometimes called a Euclidean space. If $H$ is an operator such that $H^\dagger = -H$, then $H$ is said to be anti-Hermitian (or skew-Hermitian). Furthermore, if $P$ is an operator such that $P = S^\dagger S$ for some operator $S$, then we say that $P$ is positive (or positive semidefinite or nonnegative). If $S$ also happens to be nonsingular (and hence $P$ is also nonsingular), then we say that $P$ is positive definite. Note that a positive operator is necessarily Hermitian since $(S^\dagger S)^\dagger = S^\dagger S$. The reason that $P$ is called positive is shown in part (d) of the following theorem.

**Theorem 10.9** (a) The eigenvalues of a Hermitian operator are real.
(b) The eigenvalues of an isometry (and hence also of a unitary transformation) have absolute value one.
(c) The eigenvalues of an anti-Hermitian operator are pure imaginary.
(d) A positive (positive definite) operator has eigenvalues that are real and nonnegative (positive).

**Proof** (a) If $H$ is Hermitian, $v \neq 0$, and $Hv = \lambda v$, we have

$$
\lambda \langle v, v \rangle = \langle v, \lambda v \rangle = \langle v, Hv \rangle = \langle H^\dagger v, v \rangle = \langle Hv, v \rangle \\
= \langle \lambda v, v \rangle = \lambda^* \langle v, v \rangle.
$$

But $\langle v, v \rangle \neq 0$, and hence $\lambda = \lambda^*$. 


(b) If $\Omega$ is an isometry, $v \neq 0$, and $\Omega v = \lambda v$, then we have (using Theorem 2.17)

$$|v| = |\Omega v| = |\lambda v| = |\lambda| |v| .$$

But $|v| \neq 0$, and hence $|\lambda| = 1$.

(c) If $H^\dagger = -H$, $v \neq 0$, and $Hv = \lambda v$, then

$$\lambda \langle v, v \rangle = \langle v, \lambda v \rangle = \langle v, Hv \rangle = \langle H^\dagger v, v \rangle = \langle -Hv, v \rangle = \langle -\lambda v, v \rangle = -\lambda^* \langle v, v \rangle .$$

But $\langle v, v \rangle \neq 0$, and hence $\lambda = -\lambda^*$. This shows that $\lambda$ is pure imaginary.

(d) Let $P = S^\dagger S$ be a positive definite operator. If $v \neq 0$, then the fact that $S$ is nonsingular means that $Sv \neq 0$, and hence $\langle Sv, Sv \rangle = |Sv|^2 > 0$. Then, for $Pv = (S^\dagger S)v = \lambda v$, we see that

$$\lambda \langle v, v \rangle = \langle v, \lambda v \rangle = \langle v, Pv \rangle = \langle v, (S^\dagger S)v \rangle = \langle Sv, Sv \rangle .$$

But $\langle v, v \rangle = |v|^2 > 0$ also, and therefore we must have $\lambda > 0$.

If $P$ is positive, then $S$ is singular and the only difference is that now for $v \neq 0$ we have $\langle Sv, Sv \rangle = |Sv|^2 \geq 0$ which implies that $\lambda \geq 0$. ■

We say that an operator $N$ is normal if $N^\dagger N = NN^\dagger$. Note this implies that for any $v \in V$ we have

$$\|Nv\|^2 = \langle Nv, Nv \rangle = \langle (N^\dagger N)v, v \rangle = \langle (NN^\dagger)v, v \rangle = \langle N^\dagger v, N^\dagger v \rangle = \|N^\dagger v\|^2 .$$

Now let $\lambda$ be a complex number. It is easy to see that if $N$ is normal then so is $N - \lambda 1$ since (from Theorem 10.3)

$$(N - \lambda 1)^\dagger (N - \lambda 1) = (N^\dagger - \lambda^* 1)(N - \lambda 1) = N^\dagger N - \lambda N^\dagger - \lambda^* N + \lambda^* \lambda 1$$

$$= (N - \lambda 1)(N^\dagger - \lambda^* 1) = (N - \lambda 1)(N - \lambda 1)^\dagger .$$

Using $N - \lambda 1$ instead of $N$ in the previous result we obtain

$$|Nv - \lambda v|^2 = |N^\dagger v - \lambda^* v|^2 .$$

Since the norm is positive definite, this equation proves the next theorem.
Theorem 10.10  Let N be a normal operator and let \( \lambda \) be an eigenvalue of N. Then \( Nv = \lambda v \) if and only if \( N^\dagger v = \lambda^* v \).

In words, if \( v \) is an eigenvector of a normal operator N with eigenvalue \( \lambda \), then \( v \) is also an eigenvector of \( N^\dagger \) with eigenvalue \( \lambda^* \). (Note it is always true that if \( \lambda \) is an eigenvalue of an operator T, then \( \lambda^* \) will be an eigenvalue of \( T^\dagger \). See Exercise 10.3.6.)

Corollary  If N is normal and \( Nv = 0 \) for some \( v \in V \), then \( N^\dagger v = 0 \).

Proof  This follows from Theorem 10.10 by taking \( \lambda = \lambda^* = 0 \). Alternatively, using \( N^\dagger N = NN^\dagger \) along with the fact that \( Nv = 0 \), we see that

\[
\langle N^\dagger v, N^\dagger v \rangle = \langle v, (NN^\dagger)v \rangle = \langle v, (N^\dagger N)v \rangle = 0 .
\]

Since the inner product is positive definite, this requires that \( N^\dagger v = 0 \).

Theorem 10.11  (a) Eigenvectors belonging to distinct eigenvalues of a Hermitian operator are orthogonal.

(b) Eigenvectors belonging to distinct eigenvalues of an isometric operator are orthogonal. Hence the eigenvectors of a unitary operator are orthogonal.

(c) Eigenvectors belonging to distinct eigenvalues of a normal operator are orthogonal.

Proof  As we note after the proof, Hermitian and unitary operators are special cases of normal operators, and hence parts (a) and (b) follow from part (c). However, it is instructive to give independent proofs of parts (a) and (b). Assume that T is an operator on a unitary space, and \( T v_i = \lambda_i v_i \) for \( i = 1, 2 \) with \( \lambda_1 \neq \lambda_2 \). We may then also assume without loss of generality that \( \lambda_1 \neq 0 \).

(a) If \( T = T^\dagger \), then (using Theorem 10.9(a))

\[
\lambda_2(v_1, v_2) = \langle v_1, \lambda_2 v_2 \rangle = \langle v_1, T v_2 \rangle = \langle T^\dagger v_1, v_2 \rangle = \langle T v_1, v_2 \rangle = \langle \lambda_1 v_1, v_2 \rangle = \lambda_1 \lambda_2 \langle v_1, v_2 \rangle .
\]

But \( \lambda_1 \neq \lambda_2 \), and hence \( \langle v_1, v_2 \rangle = 0 \).

(b) If T is isometric, then \( T^\dagger T = 1 \) and we have

\[
\langle v_1, v_2 \rangle = \langle v_1, (T^\dagger T)v_2 \rangle = \langle T v_1, v_2 \rangle = \lambda_1^* \lambda_2 \langle v_1, v_2 \rangle .
\]

But by Theorem 10.9(b) we have \( |\lambda_i|^2 = \lambda_i^* \lambda_i = 1 \), and thus \( \lambda_i^* = 1/\lambda_i \). Therefore, multiplying the above equation by \( \lambda_i \), we see that \( \lambda_i \langle v_1, v_2 \rangle = \)
\[ \lambda_2 \langle v_1, v_2 \rangle \] and hence, since \( \lambda_1 \neq \lambda_2 \), this shows that \( \langle v_1, v_2 \rangle = 0 \).

(c) If \( T \) is normal, then

\[ \langle v_1, Tv_2 \rangle = \lambda_2 \langle v_1, v_2 \rangle \]

while on the other hand, using Theorem 10.10 we have

\[ \langle v_1, Tv_2 \rangle = (T^* v_1, v_2) = (\lambda_1 v_1, v_2) = \lambda_1 \langle v_1, v_2 \rangle . \]

Thus \( \langle v_1, v_2 \rangle = 0 \) since \( \lambda_1 \neq \lambda_2 \).

We note that if \( H^* = H \), then \( H^* H = HH = HH^* \) so that any Hermitian operator is normal. Furthermore, if \( U \) is unitary, then \( U^* U = UU^* ( = 1) \) so that \( U \) is also normal.

A Hermitian operator \( T \) defined on a real inner product space is said to be symmetric. This is equivalent to requiring that with respect to an orthonormal basis, the matrix elements \( a_{ij} \) of \( T \) are given by

\[ a_{ij} = \langle e_i, Te_j \rangle = \langle Te_i, e_j \rangle = \langle e_i, Te_j \rangle = a_{ji} . \]

Therefore, a symmetric operator is represented by a real symmetric matrix. It is also true that antisymmetric operators (i.e., \( T^T = -T \)) and anti-Hermitian operators (\( H^* = -H \)) are normal. Therefore, part (a) and the unitary case in part (b) in the above theorem are really special cases of part (c).

**Theorem 10.12** (a) Let \( T \) be an operator on a unitary space \( V \), and let \( W \) be a \( T \)-invariant subspace of \( V \). Then \( W^\perp \) is invariant under \( T^\dagger \).

(b) Let \( U \) be a unitary operator on a unitary space \( V \), and let \( W \) be a \( U \)-invariant subspace of \( V \). Then \( W^\perp \) is also invariant under \( U \).

**Proof** (a) For any \( v \in W \) we have \( Tv \in W \) since \( W \) is \( T \)-invariant. Let \( w \in W^\perp \) be arbitrary. We must show that \( T^\dagger w \in W^\perp \). But this is easy because

\[ \langle T^\dagger w, v \rangle = \langle w, Tv \rangle = 0 \]

by definition of \( W^\perp \). Thus \( T^\dagger w \in W^\perp \) so that \( W^\perp \) is invariant under \( T^\dagger \).

(b) The fact that \( U \) is unitary means \( U^{-1} = U^\dagger \) exists, and hence \( U \) is non-singular. In other words, for any \( v' \in W \) there exists \( v \in W \) such that \( Uv = v' \). Now let \( w \in W^\perp \) be arbitrary. Then

\[ \langle Uw, v' \rangle = \langle Uw, Uv \rangle = \langle w, (U^\dagger U)v \rangle = \langle w, v \rangle = 0 \]

by definition of \( W^\perp \). Thus \( Uw \in W^\perp \) so that \( W^\perp \) is invariant under \( U \).
Recall from the discussion in Section 7.7 that the algebraic multiplicity of a given eigenvalue is the number of times the eigenvalue is repeated as a root of the characteristic polynomial. We also defined the geometric multiplicity as the number of linearly independent eigenvectors corresponding to this eigenvalue (i.e., the dimension of its eigenspace).

**Theorem 10.13** Let $H$ be a Hermitian operator on a finite-dimensional unitary space $V$. Then the algebraic multiplicity of any eigenvalue $\lambda$ of $H$ is equal to its geometric multiplicity.

**Proof** Let $V_\lambda = \{v \in V : Hv = \lambda v\}$ be the eigenspace corresponding to the eigenvalue $\lambda$. Furthermore, $V_\lambda$ is obviously invariant under $H$ since $Hv = \lambda v \in V_\lambda$ for every $v \in V_\lambda$. By Theorem 10.12(a), we then have that $V_\lambda^\perp$ is also invariant under $H^\dagger = H$. Furthermore, by Theorem 2.22 we see that $V = V_\lambda \oplus V_\lambda^\perp$. Applying Theorem 7.20, we may write $H = H_1 \oplus H_2$ where $H_1 = H|V_\lambda$ and $H_2 = H|V_\lambda^\perp$.

Let $A$ be the matrix representation of $H$, and let $A_i$ be the matrix representation of $H_i$ ($i = 1, 2$). By Theorem 7.20, we also have $A = A_1 \oplus A_2$. Using Theorem 4.14, it then follows that the characteristic polynomial of $A$ is given by

$$
\det(xI - A) = \det(xI - A_1) \det(xI - A_2).
$$

Now, $H_1$ is a Hermitian operator on the finite-dimensional space $V_\lambda$ with only the single eigenvalue $\lambda$. Therefore $\lambda$ is the only root of $\det(xI - A_1) = 0$, and hence it must occur with an algebraic multiplicity equal to the dimension of $V_\lambda$ (since this is just the size of the matrix $A_1$). In other words, if $\dim V_\lambda = m$, then $\det(xI - A_1) = (x - \lambda^m)$. On the other hand, $\lambda$ is not an eigenvalue of $A_2$ by definition, and hence $\det(xI - A_2) \neq 0$. This means that $\det(xI - A)$ contains $(x - \lambda)$ as a factor exactly $m$ times. \[\blacksquare\]

**Corollary** Any Hermitian operator $H$ on a finite-dimensional unitary space $V$ is diagonalizable.

**Proof** Since $V$ is a unitary space, the characteristic polynomial of $H$ will factor into (not necessarily distinct) linear terms. The conclusion then follows from Theorems 10.13 and 7.26. \[\blacksquare\]

In fact, from Theorem 8.2 we know that any normal matrix is unitarily similar to a diagonal matrix. This means that given any normal operator $T \in L(V)$, there is an orthonormal basis for $V$ that consists of eigenvectors of $T$. 
We develop this result from an entirely different point of view in the next section.

**Exercises**

1. Let \( V \) be a unitary space and suppose \( T \in L(V) \). Define \( T_+ = (1/2)(T + T^\dagger) \) and \( T_- = (1/2i)(T - T^\dagger) \).
   (a) Show that \( T_+ \) and \( T_- \) are Hermitian, and that \( T = T_+ + iT_- \).
   (b) If \( T'_+ \) and \( T'_- \) are Hermitian operators such that \( T = T'_+ + iT'_- \), show that \( T'_+ = T_+ \) and \( T'_- = T_- \).
   (c) Prove that \( T \) is normal if and only if \([T_+, T_-] = 0\).

2. Let \( N \) be a normal operator on a finite-dimensional inner product space \( V \).
   Prove \( \ker N = \ker N^\dagger \) and \( \operatorname{im} N = \operatorname{im} N^\dagger \). \([\text{Hint: Prove that } (\operatorname{im} N^\dagger)^\perp = \ker N, \text{ and hence that } \operatorname{im} N^\dagger = (\ker N)^\perp.\]

3. Let \( V \) be a finite-dimensional inner product space, and suppose \( T \in L(V) \) is both positive and unitary. Prove that \( T = 1 \).

4. Let \( H \in M_n(\mathbb{C}) \) be Hermitian. Then for any nonzero \( x \in \mathbb{C}^n \) we define the Rayleigh quotient to be the number

\[
R(x) = \frac{\langle x, Hx \rangle}{\|x\|^2}.
\]

Prove that \( \max\{R(x): x \neq 0\} \) is the largest eigenvalue of \( H \), and that \( \min\{R(x): x \neq 0\} \) is the smallest eigenvalue of \( H \).

5. Let \( V \) be a finite-dimensional unitary space, and suppose \( E \in L(V) \) is such that \( E^2 = E = E^\dagger \). Prove that \( V = \operatorname{im} E \oplus (\operatorname{im} E)^\perp \).

6. If \( V \) is finite-dimensional and \( T \in L(V) \) has eigenvalue \( \lambda \), show that \( T^\dagger \) has eigenvalue \( \lambda^* \).

**10.4 DIAGONALIZATION OF NORMAL OPERATORS**

We now turn to the problem of diagonalizing operators. We will discuss several of the many ways to approach this problem. Because most commonly used operators are normal, we first treat this general case in detail, leaving unitary and Hermitian operators as obvious special cases. Next, we go back
and consider the real and complex cases separately. In so doing, we will gain much insight into the structure of orthogonal and unitary transformations. While this problem was treated concisely in Chapter 8, we present an entirely different viewpoint in this section to acquaint the reader with other approaches found in the literature. If the reader has studied Chapter 8, he or she should keep in mind the rational and Jordan forms while reading this section, as many of our results (such as Theorem 10.16) follow almost trivially from our earlier work. We begin with some more elementary facts about normal transformations.

**Theorem 10.14** Let \( V \) be a unitary space.

(a) If \( T \in \text{L}(V) \) and \((T^\dagger T)v = 0\) for some \( v \in V \), then \( Tv = 0 \).

(b) If \( H \) is Hermitian and \( H^kv = 0 \) for \( k \geq 1 \), then \( Hv = 0 \).

(c) If \( N \) is normal and \( N^kv = 0 \) for \( k \geq 1 \), then \( Nv = 0 \).

(d) If \( N \) is normal, and if \((N - \lambda I)^kv = 0\) where \( k \geq 1 \) and \( \lambda \in \mathbb{C} \), then \( Nv = \lambda v \).

**Proof**

(a) Since \((T^\dagger T)v = 0\), we have \( 0 = \langle v, (T^\dagger T)v \rangle = \langle Tv, Tv \rangle \) which implies that \( Tv = 0 \) because the inner product is positive definite.

(b) We first show that if \( H^{2m}v = 0 \) for some positive integer \( m \), then \( Hv = 0 \). To see this, let \( T = H^{2m-1} \) and note that \( T^\dagger = T \) because \( H \) is Hermitian (by induction from Theorem 10.3(c)). Then \( T^\dagger T = TT = H^{2m} \), and hence

\[
0 = \langle H^{2m}v, v \rangle = \langle (T^\dagger T)v, v \rangle = \langle Tv, Tv \rangle
\]

which implies that \( 0 = Tv = H^{2m-1}v \). Repeating this process, we must eventually obtain \( Hv = 0 \).

Now, if \( H^kv = 0 \), then \( H^{2m}v = 0 \) for any \( 2^m \geq k \), and therefore applying the above argument, we see that \( Hv = 0 \).

(c) Define the Hermitian operator \( H = N^\dagger N \). Since \( N \) is normal, we see that

\[
(N^\dagger N)^2 = N^\dagger NN^\dagger N = N^\dagger N^2
\]

and by induction,

\[
(N^\dagger N)^k = N^\dagger N^k .
\]

By hypothesis, we then find that

\[
H^kv = (N^\dagger N)^kv = (N^\dagger N^k)v = 0
\]
and hence \((N^*N)v = Hv = 0\) by part (b). But then \(Nv = 0\) by part (a).

(d) Since \(N\) is normal, it follows that \(N - \lambda I\) is normal, and therefore by part (c) we have \((N - \lambda I)v = 0\). ■

Just as we did for operators, we say that a matrix \(N\) is normal if \(N^*N = NN^*\). We now wish to show that any normal matrix can be diagonalized by a unitary similarity transformation. Another way to phrase this is as follows. We say that two matrices \(A, B \in \mathbb{M}_n(\mathbb{C})\) are unitarily similar (or equivalent) if there exists a unitary matrix \(U \in \mathbb{M}_n(\mathbb{C})\) such that \(A = U^*BU = U^{-1}BU\). Thus, we wish to show that any normal matrix is unitarily similar to a diagonal matrix. This extremely important result is quite easy to prove with what has already been shown. Let us first prove this in the case of normal operators over the complex field. (See Theorem 8.2 for another approach.)

**Theorem 10.15** Let \(N\) be a normal operator on a finite-dimensional unitary space \(V\). Then there exists an orthonormal basis for \(V\) consisting of eigenvectors of \(N\) in which the matrix of \(N\) is diagonal.

**Proof** Let \(\lambda_1, \ldots, \lambda_r\) be the distinct eigenvalues of the normal operator \(N\). (These all exist in \(\mathbb{C}\) by Theorems 6.12 and 6.13.) Then (by Theorem 7.13) the minimal polynomial \(m(x)\) for \(N\) must be of the form

\[m(x) = (x - \lambda_1)^{n_1} \cdots (x - \lambda_r)^{n_r}\]

where each \(n_i \geq 1\). By the primary decomposition theorem (Theorem 7.23), we can write \(V = W_1 \oplus \cdots \oplus W_r\) where \(W_i = \text{Ker}(N - \lambda_i I)^{n_i}\). In other words,

\[(N - \lambda_i I)^{n_i} v_i = 0\]

for every \(v_i \in W_i\). By Theorem 10.14(d), we then have \(Nv_i = \lambda_i v_i\), so that every \(v_i \in W_i\) is an eigenvector of \(N\) with eigenvalue \(\lambda_i\).

Now, the inner product on \(V\) induces an inner product on each subspace \(W_i\), in the usual and obvious way, and thus by the Gram-Schmidt process (Theorem 2.21), each \(W_i\) has an orthonormal basis relative to this induced inner product. Note that by the last result of the previous paragraph, this basis must consist of eigenvectors of \(N\).

By Theorem 10.11(c), vectors in distinct \(W_i\) are orthogonal to each other. Therefore, according to Theorem 2.15, the union of the bases of the \(W_i\) forms a basis for \(V\), which thus consists entirely of eigenvectors of \(N\). By Theorem 7.14 then, the matrix of \(N\) is diagonal in this basis. (Alternatively, we see that the matrix elements \(n_{ij}\) of \(N\) relative to the eigenvector basis \(\{e_i\}\) are given by \(n_{ij} = \langle e_i, Ne_j \rangle = \langle e_i, \lambda_j e_j \rangle = \lambda_j \delta_{ij}\).) ■
Corollary 1  Let N be a normal matrix over \(\mathbb{C}\). Then there exists a unitary matrix \(U\) such that \(U^*NU = U^\dagger NU\) is diagonal. Moreover, the columns of \(U\) are just the eigenvectors of \(N\), and the diagonal elements of \(U^\dagger NU\) are the eigenvalues of \(N\).

Proof  The normal matrix \(N\) defines an operator on a finite-dimensional unitary space \(V\) with the standard orthonormal basis, and therefore by Theorem 10.15, \(V\) has an orthonormal basis of eigenvectors in which the matrix \(N\) is diagonal. By Theorem 10.6, any such change of basis in \(V\) is accomplished by a unitary transformation \(U\), and by Theorem 5.18, the matrix of the operator relative to this new basis is related to the matrix \(N\) in the old basis by the similarity transformation \(U^*NU = U^\dagger NU\).

Now note that the columns of \(U\) are precisely the eigenvectors of \(N\) (see the discussion preceding Example 7.4). We also recall that Theorem 7.14 tells us that the diagonal elements of the diagonal form of \(N\) are exactly the eigenvalues of \(N\).

Corollary 2  A real symmetric matrix can be diagonalized by an orthogonal matrix.

Proof  Note that a real symmetric matrix \(A\) may be considered as an operator on a finite-dimensional real inner product space \(V\). If we think of \(A\) as a complex matrix that happens to have all real elements, then \(A\) is Hermitian and hence has all real eigenvalues. This means that all the roots of the minimal polynomial for \(A\) lie in \(\mathbb{R}\). If \(\lambda_1, \ldots, \lambda_r\) are the distinct eigenvalues of \(A\), then we may proceed exactly as in the proof of Theorem 10.15 and Corollary 1 to conclude that there exists a unitary matrix \(U\) that diagonalizes \(A\). In this case, since \(W_i = \text{Ker}(A - \lambda_i I)^{n_i}\) and \(A - \lambda_i I\) is real, it follows that the eigenvectors of \(A\) are real and hence \(U\) is actually an orthogonal matrix.

Corollary 2 is also proved from an entirely different point of view in Exercise 10.4.9. This alternative approach has the advantage of presenting a very useful geometric picture of the diagonalization process.

Example 10.6  Let us diagonalize the real symmetric matrix

\[
A = \begin{pmatrix}
2 & -2 \\
-2 & 5
\end{pmatrix}.
\]

The characteristic polynomial of \(A\) is
\[ \Delta_A(x) = \det(xI - A) = (x - 2)(x - 5) - 4 = (x - 1)(x - 6) \]

and therefore the eigenvalues of \( A \) are 1 and 6. To find the eigenvectors of \( A \), we must solve the matrix equation \((\lambda_i I - A)v_i = 0\) for the vector \( v_i \). For \( \lambda_1 = 1 \) we have \( v_i = (x_i, y_i) \), and hence we find the homogeneous system of equations

\[
\begin{align*}
-x_1 + 2y_1 &= 0 \\
2x_1 - 4y_1 &= 0.
\end{align*}
\]

These imply that \( x_1 = 2y_1 \), and hence a nonzero solution is \( v_1 = (2, 1) \). For \( \lambda_2 = 6 \) we have the equations

\[
\begin{align*}
4x_2 + 2y_2 &= 0 \\
2x_2 - y_2 &= 0
\end{align*}
\]

which yields \( v_2 = (1, -2) \).

Note that \( \langle v_1, v_2 \rangle = 0 \) as it should according to Theorem 10.11, and that \( |v_1| = \sqrt{5} = |v_2| \). We then take the normalized basis vectors to be \( e_i = v_i / \sqrt{5} \) which are also eigenvectors of \( A \). Finally, \( A \) is diagonalized by the orthogonal matrix \( P \) whose columns are just the \( e_i \):

\[
P = \begin{pmatrix}
\frac{2}{\sqrt{5}} & 1/\sqrt{5} \\
1/\sqrt{5} & -2/\sqrt{5}
\end{pmatrix}.
\]

We leave it to the reader to show that

\[
P^T AP = \begin{pmatrix}
1 & 0 \\
0 & 6
\end{pmatrix}.
\]

Another important point to notice is that Theorem 10.15 tells us that even though an eigenvalue \( \lambda \) of a normal operator \( N \) may be degenerate (i.e., have algebraic multiplicity \( k > 1 \)), it is always possible to find \( k \) linearly independent eigenvectors belonging to \( \lambda \). The easiest way to see this is to note that from Theorem 10.8 we have \( |\det U| = 1 \neq 0 \) for any unitary matrix \( U \). This means that the columns of the diagonalizing matrix \( U \) (which are just the eigenvectors of \( N \)) must be linearly independent. This is in fact another proof that the algebraic and geometric multiplicities of a normal (and hence Hermitian) operator must be the same.

We now consider the case of real orthogonal transformations as independent operators, not as a special case of normal operators. First we need a gen-
eral definition. Let $V$ be an arbitrary finite-dimensional vector space over any field $\mathcal{F}$, and suppose $T \in L(V)$. A nonzero $T$-invariant subspace $W \subseteq V$ is said to be **irreducible** if the only $T$-invariant subspaces contained in $W$ are $\{0\}$ and $W$.

**Theorem 10.16**  
(a) Let $V$ be a finite-dimensional vector space over an algebraically closed field $\mathcal{F}$, and suppose $T \in L(V)$. Then every irreducible $T$-invariant subspace $W$ of $V$ is of dimension 1.
(b) Let $V$ be a finite-dimensional vector space over $\mathbb{R}$, and suppose $T \in L(V)$. Then every irreducible $T$-invariant subspace $W$ of $V$ is of dimension either 1 or 2.

**Proof**  
(a) Let $W$ be an irreducible $T$-invariant subspace of $V$. Then the restriction $T_W$ of $T$ to $W$ is just a linear transformation on $W$, where $T_W(w) = Tw \in W$ for every $w \in W$. Since $\mathcal{F}$ is algebraically closed, the characteristic polynomial of $T_W$ has at least one root (i.e., eigenvalue $\lambda$) in $\mathcal{F}$. Therefore $T$ has at least one (nonzero) eigenvector $v \in W$ such that $Tv = \lambda v \in W$. If we define $S(v)$ to be the linear span of $\{v\}$, then $S(v)$ is also a $T$-invariant subspace of $W$, and hence $S(v) = W$ because $W$ is irreducible. Therefore $W$ is spanned by the single vector $v$, and hence $\dim W = 1$.

(b) Let $W$ be an irreducible $T$-invariant subspace of $V$, and let $m(x)$ be the minimal polynomial for $T_W$. Therefore, the fact that $W$ is irreducible (so that $W$ is not a direct sum of $T$-invariant subspaces) along with the primary decomposition theorem (Theorem 7.23) tells us that we must have $m(x) = f(x)^n$ where $f(x) \in \mathbb{R}[x]$ is a prime polynomial. Furthermore, if $n$ were greater than 1, then we claim that $\ker f(T)^{n-1}$ would be a $T$-invariant subspace of $W$ (Theorem 7.18) that is different from $\{0\}$ and $W$.

To see this, first suppose that $\ker f(T)^{n-1} = \{0\}$. Then the linear transformation $f(T)^{n-1}$ is one-to-one, and hence $f(T)^{n-1}(W) = W$. But then

$$0 = f(T)^n(W) = f(T)f(T)^{n-1}(W) = f(T)(W).$$

However, $f(T)W \neq 0$ by definition of $m(x)$, and hence this contradiction shows that we cannot have $\ker f(T)^{n-1} = \{0\}$. Next, if we had $\ker f(T)^{n-1} = W$, this would imply that $f(T)^{n-1}(W) = 0$ which contradicts the definition of minimal polynomial. Therefore we must have $n = 1$ and $m(x) = f(x)$.

Since $m(x) = f(x)$ is prime, it follows from the corollary to Theorem 6.15 that we must have either $m(x) = x - a$ or $m(x) = x^2 + ax + b$ with $a^2 - 4ab < 0$. If $m(x) = x - a$, then there exists an eigenvector $v \in W$ with $Tv = av \in W$, and hence $S(v) = W$ as in part (a). If $m(x) = x^2 + ax + b$, then for any nonzero $w \in W$ we have
10.4 DIAGONALIZATION OF NORMAL OPERATORS

\[ 0 = m(T)w = T^2w + aTw + bw \]
and hence
\[ T^2w = T(Tw) = -aTw - bw \in W . \]

Thus \( S(w, Tw) \) is a \( T \)-invariant subspace of \( W \) with dimension either 1 or 2. However \( W \) is irreducible, and therefore we must have \( W = S(w, Tw) \). \( \blacksquare \)

**Theorem 10.17** Let \( V \) be a finite-dimensional Euclidean space, let \( T \in \mathbb{L}(V) \) be an orthogonal transformation, and let \( W \) be an irreducible \( T \)-invariant subspace of \( V \). Then one of the following two conditions holds:

(a) \( \dim W = 1 \), and for any nonzero \( w \in W \) we have \( Tw = \pm w \).

(b) \( \dim W = 2 \), and there exists an orthonormal basis \( \{e_1, e_2\} \) for \( W \) such that the matrix representation of \( T_W \) relative to this basis has the form

\[
\begin{pmatrix}
\cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{pmatrix}.
\]

**Proof** That \( \dim W \) equals 1 or 2 follows from Theorem 10.16(b). If \( \dim W = 1 \), then (since \( W \) is \( T \)-invariant) there exists \( \lambda \in \mathbb{R} \) such that \( Tw = \lambda w \) for any (fixed) \( w \in W \). But \( T \) is orthogonal so that

\[ \|w\| = \|Tw\| = \|\lambda w\| = |\lambda|\|w\| \]

and hence \( |\lambda| = 1 \). This shows that \( Tw = \lambda w = \pm w \).

If \( \dim W = 2 \), then the desired form of the matrix of \( T_W \) follows essentially from Example 10.5. Alternatively, we know that \( W \) has an orthonormal basis \( \{e_1, e_2\} \) by the Gram-Schmidt process. If we write \( Te_1 = ae_1 + be_2 \), then \( \|Te_1\| = \|e_1\| = 1 \) implies that \( a^2 + b^2 = 1 \). If we also write \( Te_2 = ce_1 + de_2 \), then similarly \( c^2 + d^2 = 1 \). Using \( \langle Te_1, Te_2 \rangle = \langle e_1, e_2 \rangle = 0 \) we find \( ac + bd = 0 \), and hence \( c = -bd/a \). But then \( 1 = d^2(1 + b^2/a^2) = d^2/a^2 \) so that \( a^2 = d^2 \) and \( c^2 = b^2 \). This means that \( Te_2 = \pm (be_1 + ae_2) \). If \( Te_2 = -be_1 + ae_2 \), then the matrix of \( T \) is of the form

\[
\begin{pmatrix}
a & -b \\
b & a
\end{pmatrix}
\]

and we may choose \( \theta \in \mathbb{R} \) such that \( a = \cos \theta \) and \( b = \sin \theta \) (since \( \det T = a^2 + b^2 = 1 \)). However, if \( Te_2 = be_1 - ae_2 \), then the matrix of \( T \) is

\[
\begin{pmatrix}
a & b \\
b & -a
\end{pmatrix}
\]
which satisfies the equation \( x^2 - 1 = (x - 1)(x + 1) = 0 \) (and has \( \det T = -1 \)). But if \( T \) satisfied this equation, then (by the primary decomposition theorem (Theorem 7.23)) \( W \) would be a direct sum of subspaces, in contradiction to the assumed irreducibility of \( W \). Therefore only the first case can occur.

This theorem becomes quite useful when combined with the next result.

**Theorem 10.18** Let \( T \) be an orthogonal operator on a finite-dimensional Euclidean space \( V \). Then \( V = W_1 \oplus \cdots \oplus W_r \) where each \( W_i \) is an irreducible \( T \)-invariant subspace of \( V \) such that vectors belonging to distinct subspaces \( W_i \) and \( W_j \) are orthogonal.

*Proof* If \( \dim V = 1 \) there is nothing to prove, so we assume \( \dim V > 1 \) and that the theorem is true for all spaces of dimension less than \( \dim V \). Let \( W_i \) be a nonzero \( T \)-invariant subspace of least dimension. Then \( W_i \) is necessarily irreducible. By Theorem 2.22 we know that \( V = W_i \oplus W_i^\perp \) where \( \dim W_i^\perp < \dim V \), and hence we need only show that \( W_i^\perp \) is also \( T \)-invariant. But this follows from Theorem 10.12(b) applied to real unitary transformations (i.e., orthogonal transformations). This also means that \( T(W_i^\perp) \subset W_i^\perp \), and hence \( T \) is an orthogonal transformation on \( W_i^\perp \) (since it takes vectors in \( W_i^\perp \) to vectors in \( W_i^\perp \)). By induction, \( W_i^\perp \) is a direct sum of pairwise orthogonal irreducible \( T \)-invariant subspaces, and therefore so is \( V = W_i \oplus W_i^\perp \).

From Theorem 10.18, we see that if we are given an orthogonal transformation \( T \) on a finite-dimensional Euclidean space \( V \), then \( V = W_1 \oplus \cdots \oplus W_r \) is the direct sum of pairwise orthogonal irreducible \( T \)-invariant subspaces \( W_i \). But from Theorem 10.17, we see that any such subspace \( W_i \) is of dimension either 1 or 2. Moreover, Theorem 10.17 also showed that if \( \dim W_i = 1 \), then the matrix of \( T|W_i \) is either (1) or \((-1)\), and if \( \dim W_i = 2 \), then the matrix of \( T|W_i \) is just the rotation matrix \( R_i \) given by

\[
R_i = \begin{pmatrix} \cos \theta_i & -\sin \theta_i \\ \sin \theta_i & \cos \theta_i \end{pmatrix}.
\]

Since each \( W_i \) has an orthonormal basis and the bases of distinct \( W_i \) are orthogonal, it follows that we can find an orthonormal basis for \( V \) in which the matrix of \( T \) takes the block diagonal form (see Theorem 7.20)

\[
(1) \oplus \cdots \oplus (1) \oplus (-1) \oplus \cdots \oplus (-1) \oplus R_1 \oplus \cdots \oplus R_m.
\]
These observations prove the next theorem.

**Theorem 10.19** Let $T$ be an orthogonal transformation on a finite-dimensional Euclidean space $V$. Then there exists an orthonormal basis for $V$ in which the matrix representation of $T$ takes the block diagonal form

$$M_1 \oplus \cdots \oplus M_r$$

where each $M_i$ is one of the following: $(+1)$, $(-1)$, or $R_i$.

**Exercises**

1. Prove that any nilpotent normal operator is necessarily the zero operator.

2. Let $A$ and $B$ be normal operators on a finite-dimensional unitary space $V$. For notational simplicity, let $v_a$ denote an eigenvector of $A$ corresponding to the eigenvalue $a$, let $v_b$ be an eigenvector of $B$ corresponding to the eigenvalue $b$, and let $v_{ab}$ denote a simultaneous eigenvector of $A$ and $B$, i.e., $Av_{ab} = av_{ab}$ and $Bv_{ab} = bv_{ab}$.

   (a) If there exists a basis for $V$ consisting of simultaneous eigenvectors of $A$ and $B$, show that the commutator $[A, B] = AB - BA = 0$.

   (b) If $[A, B] = 0$, show that there exists a basis for $V$ consisting entirely of simultaneous eigenvectors of $A$ and $B$. In other words, if $[A, B] = 0$, then $A$ and $B$ can be simultaneously diagonalized. *(Hint: There are several ways to approach this problem. One way follows easily from Exercise 8.1.3. Another intuitive method is as follows. First assume that at least one of the operators, say $A$, is nondegenerate. Show that $Bv_a$ is an eigenvector of $A$, and that $Bv_a = bv_a$ for some scalar $b$. Next assume that both $A$ and $B$ are degenerate. Then $Av_{a,i} = av_{a,i}$ where the $v_{a,i}$ ($i = 1, \ldots, m_a$) are linearly independent eigenvectors corresponding to the eigenvalue $a$ of multiplicity $m_a$. What does the matrix representation of $A$ look like in the $\{v_{a,i}\}$ basis? Again consider $Bv_{a,i}$. What does the matrix representation of $B$ look like? Now what happens if you diagonalize $B$?)*

3. If $N_1$ and $N_2$ are commuting normal operators, show that the product $N_1N_2$ is normal.

4. Let $V$ be a finite-dimensional complex (real) inner product space, and suppose $T \in L(V)$. Prove that $V$ has an orthonormal basis of eigenvectors of $T$ with corresponding eigenvalues of absolute value 1 if and only if $T$ is unitary (Hermitian and orthogonal).
5. For each of the following matrices A, find an orthogonal or unitary matrix 
P and a diagonal matrix D such that \( P^\dagger A P = D \):

(a) \[
\begin{pmatrix}
1 & 2 \\
2 & 1
\end{pmatrix}
\]
(b) \[
\begin{pmatrix}
0 & -1 \\
1 & 0
\end{pmatrix}
\]
(c) \[
\begin{pmatrix}
2 & 3 - i3 \\
3 + i3 & 5
\end{pmatrix}
\]
(d) \[
\begin{pmatrix}
0 & 2 & 2 \\
2 & 0 & 2 \\
2 & 2 & 0
\end{pmatrix}
\]
(e) \[
\begin{pmatrix}
2 & 1 & 1 \\
1 & 2 & 1 \\
1 & 1 & 2
\end{pmatrix}
\]

6. Let A, B and C be normal operators on a finite-dimensional unitary space, 
and assume that \([A, B] = 0\) but \([B, C] \neq 0\). If all of these operators are 
nondegenerate (i.e., all eigenvalues have multiplicity equal to 1), is it true 
that \([A, C] \neq 0\)? Explain. What if any of these are degenerate?

7. Let V be a finite-dimensional unitary space and suppose \( A \in L(V) \).
(a) Prove that \( \text{Tr}(AA^\dagger) = 0 \) if and only if \( A = 0 \).
(b) Suppose \( N \in L(V) \) is normal and \( AN = NA \). Prove that \( AN^\dagger = N^\dagger A \).

8. Let A be a positive definite real symmetric matrix on an \( n \)-dimensional 
Euclidean space V. Using the single variable formula (where \( a > 0 \))
\[
\int_{-\infty}^{\infty} \exp(-ax^2/2)dx = (2\pi/a)^{1/2}
\]
show that
\[
\int_{-\infty}^{\infty} \exp(-1/2)\langle \vec{x}, A\vec{x} \rangle d^n x = (2\pi)^{n/2}(\det A)^{-1/2}
\]
where \( d^n x = dx_1 \cdots dx_n \). [Hint: First consider the case where A is 
diagonal.]

9. (This is an independent proof of Corollary 2 of Theorem 10.15.) Let \( A = 
(a_{ij}) \in M_3(\mathbb{R}) \) be a real symmetric matrix. Thus \( A: \mathbb{R}^3 \to \mathbb{R}^3 \) is a Hermitian 
linear operator with respect to the inner product \( \langle , \rangle \). Prove there exists an 
ortogonal basis of eigenvectors of A using the following approach. (It 
should be clear after you have done this that the same proof will work in 
\( \mathbb{R}^n \) just as well.)
(a) Let \( S^2 \) be the unit sphere in \( \mathbb{R}^3 \), and define \( f: S^2 \to \mathbb{R} \) by
\[
f(x) = \langle Ax, x \rangle.
\]
Let \( M = \sup f(x) \) and \( m = \inf f(x) \) where the sup and inf are taken over \( S^2 \).
Show that there exist points \( x_1, x_1' \in S^2 \) such that \( f(x_1) = M \) and \( f(x_1') = m \). [Hint: Use Theorem A15.]

(b) Let \( C = x(t) \) be any curve on \( S^2 \) such that \( x(0) = x_1 \), and let a dot denote differentiation with respect to the parameter \( t \). Note that \( \dot{x}(t) \) is tangent to \( C \), and hence also to \( S^2 \). Show that \( \langle Ax_1, \dot{x}(0) \rangle = 0 \), and thus deduce that \( Ax_1 \) is normal to the tangent plane at \( x_1 \). [Hint: Consider \( df(x(t))/dt \) \( t=0 \) and note that \( C \) is arbitrary.]

(c) Show that \( \langle \dot{x}(t), x(t) \rangle = 0 \), and hence conclude that \( Ax_1 = \lambda_1 x_1 \). [Hint: Recall that \( S^2 \) is the unit sphere.]

(d) Argue that \( Ax_1' = \lambda_1' x_1' \), and in general, that any critical point of \( f(x) = \langle Ax, x \rangle \) on the unit sphere will be an eigenvector of \( A \) with critical value (i.e., eigenvalue) \( \lambda_i = \langle Ax_i, x_i \rangle \). (A critical point of \( f(x) \) is a point \( x_0 \) where \( df/dx = 0 \), and the critical value of \( f \) is just \( f(x_0) \).)

(e) Let \( [x_1] \) be the 1-dimensional subspace of \( \mathbb{R}^3 \) spanned by \( x_1 \). Show that \( [x_1] \) and \( [x_1]^\perp \) are both \( A \)-invariant subspaces of \( \mathbb{R}^3 \), and hence that \( A \) is Hermitian on \( [x_1]^\perp \subset \mathbb{R}^2 \). Note that \( [x_1]^\perp \) is a plane through the origin of \( S^2 \).

(f) Show that \( f \) now must achieve its maximum at a point \( x_2 \) on the unit circle \( S^1 \subset [x_1]^\perp \), and that \( Ax_2 = \lambda_2 x_2 \) with \( \lambda_2 \leq \lambda_1 \).

(g) Repeat this process again by considering the space \( [x_2]^\perp \subset [x_1]^\perp \), and show there exists a vector \( x_3 \in [x_2]^\perp \) with \( Ax_3 = \lambda_3 x_3 \) and \( \lambda_3 \leq \lambda_2 \leq \lambda_1 \).

### 10.5 THE SPECTRAL THEOREM

We now turn to another major topic of this chapter, the so-called spectral theorem. This important result is actually nothing more than another way of looking at Theorems 8.2 and 10.15. We begin with a simple version that is easy to understand and visualize if the reader will refer back to the discussion prior to Theorem 7.29.

**Theorem 10.20** Suppose \( A \in M_n(\mathbb{C}) \) is a diagonalizable matrix with distinct eigenvalues \( \lambda_1, \ldots, \lambda_r \). Then \( A \) can be written in the form

\[
A = \lambda_1 E_1 + \cdots + \lambda_r E_r
\]

where the \( E_i \) are \( n \times n \) matrices with the following properties:

(a) Each \( E_i \) is idempotent (i.e., \( E_i^2 = E_i \)).

(b) \( E_i E_j = 0 \) for \( i \neq j \).
(c) $E_1 + \cdots + E_r = I$.
(d) $AE_i = E_i A$ for every $E_i$.

Proof Since $A$ is diagonalizable by assumption, let $D = P^{-1} A P$ be the diagonal form of $A$ for some nonsingular matrix $P$ (whose columns are just the eigenvectors of $A$). Remember that the diagonal elements of $D$ are just the eigenvalues $\lambda_i$ of $A$. Let $P_i$ be the $n \times n$ diagonal matrix with diagonal element $1$ wherever a $\lambda_i$ occurs in $D$, and $0$'s everywhere else. It should be clear that the collection $\{P_i\}$ obeys properties (a) – (c), and that

$$P^{-1} A P = D = \lambda_1 P_1 + \cdots + \lambda_r P_r.$$ 

If we now define $E_i = PP_iP^{-1}$, then we have

$$A = P D P^{-1} = \lambda_1 E_1 + \cdots + \lambda_r E_r$$

where the $E_i$ also obey properties (a) – (c) by virtue of the fact that the $P_i$ do. Using (a) and (b) in this last equation we find

$$A E_i = (\lambda_1 E_1 + \cdots + \lambda_r E_r) E_i = \lambda_i E_i$$

and similarly it follows that $E_i A = \lambda_i E_i$ so that each $E_i$ commutes with $A$, i.e., $E_i A = A E_i$. □

By way of terminology, the collection of eigenvalues $\lambda_1, \ldots, \lambda_r$ is called the spectrum of $A$, the sum $E_1 + \cdots + E_r = I$ is called the resolution of the identity induced by $A$, and the expression $A = \lambda_1 E_1 + \cdots + \lambda_r E_r$ is called the spectral decomposition of $A$. These definitions also apply to arbitrary normal operators as in Theorem 10.22 below.

Corollary Let $A$ be diagonalizable with spectral decomposition as in Theorem 10.20. If $f(x) \in \mathbb{C}[x]$ is any polynomial, then

$$f(A) = f(\lambda_1) E_1 + \cdots + f(\lambda_r) E_r.$$ 

Proof Using properties (a) – (c) in Theorem 10.20, it is easy to see that for any $m > 0$ we have

$$A^m = \lambda_1^m E_1 + \cdots + \lambda_r^m E_r.$$ 

The result for arbitrary polynomials now follows easily from this result. □
Before turning to our proof of the spectral theorem, we first prove a simple but useful characterization of orthogonal projections.

**Theorem 10.21**  Let V be an inner product space and suppose \( E \in \mathbb{L}(V) \). Then \( E \) is an orthogonal projection if and only if \( E^2 = E = E^\dagger \).

**Proof**  We first assume that \( E \) is an orthogonal projection. By definition this means that \( E^2 = E \), and hence we must show that \( E^\dagger = E \). From Theorem 7.27 we know that \( V = \text{Im} \ E \oplus \text{Ker} \ E = \text{Im} \ E \oplus (\text{Im} \ E)^\perp \). Suppose \( v, w \in V \) are arbitrary. Then we may write \( v = v_1 + v_2 \) and \( w = w_1 + w_2 \) where \( v_1, w_1 \in \text{Im} \ E \) and \( v_2, w_2 \in (\text{Im} \ E)^\perp \). Therefore

\[
\langle v, Ew \rangle = \langle v_1 + v_2, w_1 \rangle = \langle v_1, w_1 \rangle + \langle v_2, w_1 \rangle = \langle v_1, w_1 \rangle
\]

and

\[
\langle v, E^\dagger w \rangle = \langle Ev, w \rangle = \langle v_1, w_1 + w_2 \rangle = \langle v_1, w_1 \rangle + \langle v_1, w_2 \rangle = \langle v_1, w_1 \rangle.
\]

In other words, \( \langle v, (E - E^\dagger)w \rangle = 0 \) for all \( v, w \in V \), and hence \( E = E^\dagger \) (by Theorem 10.4(a)).

On the other hand, if \( E^2 = E = E^\dagger \), then we know from Theorem 7.27 that \( E \) is a projection of \( V \) on \( \text{Im} \ E \) in the direction of \( \text{Ker} \ E \), i.e., \( V = \text{Im} \ E \oplus \text{Ker} \ E \). Therefore, we need only show that \( \text{Im} \ E \) and \( \text{Ker} \ E \) are orthogonal subspaces. To show this, let \( w \in \text{Im} \ E \) and \( w' \in \text{Ker} \ E \) be arbitrary. Then \( Ew = w \) and \( Ew' = 0 \) so that

\[
\langle w', w \rangle = \langle w', Ew \rangle = \langle E^\dagger w', w \rangle = \langle Ew', w \rangle = 0.
\]

(This was also proved independently in Exercise 10.3.5.)

We are now in a position to prove the spectral theorem for normal operators. In order to distinguish projection operators from their matrix representations in this theorem, we denote the operators by \( \pi_i \) and the corresponding matrices by \( E_i \).

**Theorem 10.22 (Spectral Theorem for Normal Operators)**  Let \( V \) be a finite-dimensional unitary space, and let \( N \) be a normal operator on \( V \) with distinct eigenvalues \( \lambda_1, \ldots, \lambda_r \). Then

(a) \( N = \lambda_1 \pi_1 + \cdots + \lambda_r \pi_r \) where each \( \pi_i \) is the orthogonal projection of \( V \) onto a subspace \( W_i = \text{Im} \pi_i \).

(b) \( \pi_i \pi_j = 0 \) for \( i \neq j \).
(c) $\pi_1 + \cdots + \pi_r = 1$.

(d) $V = W_1 \oplus \cdots \oplus W_r$ where the subspaces $W_i$ are mutually orthogonal.

(e) $W_j = \text{Im} \pi_j = \text{Ker}(N - \lambda_j I)$ is the eigenspace corresponding to $\lambda_j$.

**Proof** Choose any orthonormal basis $\{e_i\}$ for $V$, and let $A$ be the matrix representation of $N$ relative to this basis. As discussed following Theorem 7.6, the normal matrix $A$ has the same eigenvalues as the normal operator $N$. By Corollary 1 of Theorem 10.15 we know that $A$ is diagonalizable, and hence applying Theorem 10.20 we may write

$$A = \lambda_1 E_1 + \cdots + \lambda_r E_r$$

where $E_i^2 = E_i$, $E_i E_j = 0$ if $i \neq j$, and $E_1 + \cdots + E_r = I$. Furthermore, $A$ is diagonalized by a unitary matrix $P$, and as we saw in the proof of Theorem 10.20, $E_i = PP_i P^\dagger$ where each $P_i$ is a real diagonal matrix. Since each $P_i$ is clearly Hermitian, this implies that $E_i^\dagger = E_i$, and hence each $E_i$ is an orthogonal projection (Theorem 10.21).

Now define $\pi_i \in L(V)$ as that operator whose matrix representation relative to the basis $\{e_i\}$ is just $E_i$. From the isomorphism between linear transformations and their representations (Theorem 5.13), it should be clear that

$$N = \lambda_1 \pi_1 + \cdots + \lambda_r \pi_r$$

$$\pi_i^\dagger = \pi_i$$

$$\pi_i^2 = \pi_i$$

$$\pi_i \pi_j = 0 \text{ for } i \neq j$$

$$\pi_1 + \cdots + \pi_r = 1.$$  

Since $\pi_i^2 = \pi_i = \pi_i^\dagger$, Theorem 10.21 tells us that each $\pi_i$ is an orthogonal projection of $V$ on the subspace $W_i = \text{Im} \pi_i$. Since $\pi_1 + \cdots + \pi_r = 1$, we see that for any $v \in V$ we have $v = \pi_1 v + \cdots + \pi_r v$ so that $V = W_1 + \cdots + W_r$. To show that this sum is direct suppose, for example, that

$$w_1 \in W_1 \cap (W_2 + \cdots + W_r).$$

This means that $w_1 = w_2 + \cdots + w_r$ where $w_i \in W_i$ for each $i = 1, \ldots, r$. Since $w_i \in W_i = \text{Im} \pi_i$, it follows that there exists $v_i \in V$ such that $\pi_i v_i = w_i$ for each $i$. Then

$$w_i = \pi_i v_i = \pi_i^2 v_i = \pi_i w_i$$

and if $i \neq j$, then $\pi_i \pi_j = 0$ implies
\[ \pi_i \omega_j = (\pi_i \pi_i) \omega_j = 0 . \]

Applying \( \pi_i \) to \( \omega_1 = \omega_2 + \cdot \cdot \cdot + \omega_r \), we obtain \( \omega_1 = \pi_i \omega_i = 0 \). Hence we have shown that \( \omega_i \cap (\omega_2 + \cdot \cdot \cdot + \omega_r) = \{0\} \). Since this argument can clearly be applied to any of the \( \omega_i \), we have proved that \( V = \omega_1 \oplus \cdot \cdot \cdot \oplus \omega_r \).

Next we note that for each \( i, \pi_i \) is the orthogonal projection of \( V \) on \( \omega_i = \text{Im} \pi_i \) in the direction of \( \omega_i^\perp = \text{Ker} \pi_i \), so that \( V = \omega_i \oplus \omega_i^\perp \). Therefore, since \( V = \omega_1 \oplus \cdot \cdot \cdot \oplus \omega_r \), it follows that for each \( j \neq i \) we must have \( \omega_j \subset \omega_i^\perp \), and hence the subspaces \( \omega_i \) must be mutually orthogonal. Finally, the fact that \( \omega_i = \text{Ker}(N - \lambda_i 1) \) was proved in Theorem 7.29.

The observant reader will have noticed the striking similarity between the spectral theorem and Theorem 7.29. In fact, part of Theorem 10.22 is essentially a corollary of Theorem 7.29. This is because a normal operator is diagonalizable, and hence satisfies the hypotheses of Theorem 7.29. However, note that in the present case we have used the existence of an inner product in our proof, whereas in Chapter 7, no such structure was assumed to exist. We leave it to the reader to use Theorems 10.15 and 7.28 to construct a simple proof of the spectral theorem that makes no reference to any matrix representation of the normal operator (see Exercise 10.5.1).

**Theorem 10.23** Let \( \sum_{i=1}^{r} \lambda_i \omega_i \) be the spectral decomposition of a normal operator \( N \) on a finite-dimensional unitary space. Then for each \( i = 1, \ldots, r \) there exists a polynomial \( f_i(x) \in \mathbb{C}[x] \) such that \( f_i(\lambda_j) = \delta_{ij} \) and \( f_i(N) = \omega_i \).

**Proof** For each \( i = 1, \ldots, r \) we must find a polynomial \( f_i(x) \in \mathbb{C}[x] \) with the property that \( f_i(\lambda_j) = \delta_{ij} \). It should be obvious that the polynomials \( f_i(x) \) defined by

\[
 f_i(x) = \prod_{j \neq i} \frac{x - \lambda_j}{\lambda_j - \lambda_i}
\]

have this property. From the corollary to Theorem 10.20 we have \( p(N) = \sum_i p(\lambda_i) \omega_i \) for any \( p(x) \in \mathbb{C}[x] \), and hence

\[
 f_i(N) = \sum_i f_i(\lambda_j) \omega_j = \sum_i \delta_{ij} \omega_j = \omega_i
\]

as required. ■
Exercises

1. Use Theorems 10.15 and 7.28 to construct a proof of Theorem 10.22 that makes no reference to any matrix representations.

2. Let $N$ be an operator on a finite-dimensional unitary space. Prove that $N$ is normal if and only if $N^* = g(N)$ for some polynomial $g$. [Hint: If $N$ is normal with eigenvalues $\lambda_1, \ldots, \lambda_r$, use Exercise 6.4.2 to show the existence of a polynomial $g$ such that $g(\lambda_i) = \lambda_i^*$ for each $i$.]

3. Let $T$ be an operator on a finite-dimensional unitary space. Prove that $T$ is unitary if and only if $T$ is normal and $|\lambda| = 1$ for every eigenvalue $\lambda$ of $T$.

4. Let $H$ be a normal operator on a finite-dimensional unitary space. Prove that $H$ is Hermitian if and only if every eigenvalue of $H$ is real.

10.6 THE MATRIX EXPONENTIAL SERIES

We now use Theorem 10.20 to prove a very useful result, namely, that any unitary matrix $U$ can be written in the form $e^{iH}$ for some Hermitian matrix $H$. Before proving this however, we must first discuss some of the theory of sequences and series of matrices. In particular, we must define just what is meant by expressions of the form $e^{iH}$. If the reader already knows something about sequences and series of numbers, then the rest of this section should present no difficulty. However, for those readers who may need some review, we have provided all of the necessary material in Appendix B.

Let $\{S_r\}$ be a sequence of complex matrices where each $S_r \in M_n(\mathbb{C})$ has entries $s_{ij}^{(r)}$. We say that $\{S_r\}$ converges to the limit $S = (s_{ij}) \in M_n(\mathbb{C})$ if each of the $n^2$ sequences $\{s_{ij}^{(r)}\}$ converges to a limit $s_{ij}$. We then write $S_r \rightarrow S$ or $\lim_{r \rightarrow \infty} S_r = S$ (or even simply $\lim S_r = S$). In other words, a sequence $\{S_r\}$ of matrices converges if and only if every entry of $S_r$ forms a convergent sequence.

Similarly, an infinite series of matrices

$$\sum_{r=1}^{\infty} A_r$$

where $A_r = (a_{ij}^{(r)})$ is said to be convergent to the sum $S = (s_{ij})$ if the sequence of partial sums

$$S_m = \sum_{r=1}^{m} A_r$$
converges to S. Another way to say this is that the series $\sum A_r$ converges to S if and only if each of the $n^2$ series $\sum a^{(r)}_{ij}$ converges to $s_{ij}$ for each $i, j = 1, \ldots, n$. We adhere to the convention of leaving off the limits in a series if they are infinite.

Our next theorem proves several intuitively obvious properties of sequences and series of matrices.

**Theorem 10.24** (a) Let $\{S_r\}$ be a convergent sequence of $n \times n$ matrices with limit $S$, and let $P$ be any $n \times n$ matrix. Then $PS_r \rightarrow PS$ and $S_rP \rightarrow SP$.

(b) If $S_r \rightarrow S$ and $P$ is nonsingular, then $P^{-1}S_rP \rightarrow P^{-1}SP$.

(c) If $\sum A_r$ converges to $A$ and $P$ is nonsingular, then $\sum P^{-1}A_rP$ converges to $P^{-1}AP$.

**Proof** (a) Since $S_r \rightarrow S$, we have $\lim s^{(r)}_{ij} = s_{ij}$ for all $i, j = 1, \ldots, n$. Therefore

$$\lim(PS_r)_{ij} = \lim(\sum_{k} p_{ik}s^{(r)}_{kj}) = \sum_{k} p_{ik}\lim s^{(r)}_{kj} = \sum_{k} p_{ik}s_{kj} = (PS)_{ij} .$$

Since this holds for all $i, j = 1, \ldots, n$ we must have $PS_r \rightarrow PS$. It should be obvious that we also have $S_rP \rightarrow SP$.

(b) As in part (a), we have

$$\lim(P^{-1}S_rP)_{ij} = \lim(\sum_{k,m} p^{-1}_{ik}s^{(r)}_{km}p_{mj})$$

$$= \sum_{k,m} p^{-1}_{ik}p_{mj}\lim s^{(r)}_{km}$$

$$= \sum_{k,m} p^{-1}_{ik}p_{mj}s_{km}$$

$$= (P^{-1}SP)_{ij} .$$

Note that we may use part (a) to formally write this as

$$\lim(P^{-1}S_rP) = P^{-1}\lim(S_rP) = P^{-1}SP .$$

(c) If we write the $m$th partial sum as

$$S_m = \sum_{r=1}^{m} P^{-1}A_rP = P^{-1}\left(\sum_{r=1}^{m} A_r\right)P$$

then we have
\[ \lim_{m \to \infty} (S_m)_{ij} = \sum_{k,l} \lim \left\{ p^{-1}_{ik} \left( \sum_{r=1}^m a^{(r)}_{kl} \right) p_{lj} \right\} = \sum_{k,l} p^{-1}_{ik} p_{lj} \lim \sum_{r=1}^m a^{(r)}_{kl} = \sum_{k,l} p^{-1}_{ik} p_{lj} a_{kl} = P^{-1}AP. \]

**Theorem 10.25** For any \( A = (a_{ij}) \in M_n(\mathbb{C}) \) the following series converges:
\[
\sum_{r=0}^{\infty} \frac{A^r}{r!} = I + A + \frac{A^2}{2!} + \cdots + \frac{A^r}{r!} + \cdots.
\]

**Proof** Choose a positive real number \( M > \max\{n, |a_{ij}|\} \) where the max is taken over all \( i, j = 1, \ldots , n \). Then \( |a_{ij}| < M \) and \( n < M < M^2 \). Now consider the term \( A^2 = (b_{ij}) = (\sum_k a_{ik} a_{kj}) \). We have (by Theorem 2.17, property (N3))
\[
|b_{ij}| \leq \sum_{k=1}^n |a_{ik}||a_{kj}| < \sum_{k=1}^n M^2 = nM^2 < M^4.
\]
Proceeding by induction, suppose that for \( A^r = (c_{ij}) \), it has been shown that \( |c_{ij}| < M^{2r} \). Then \( A^{r+1} = (d_{ij}) \) where
\[
|d_{ij}| \leq \sum_{k=1}^n |a_{ik}||c_{kj}| < nMM^{2r} = nM^{2r+1} < M^{2(r+1)}.
\]
This proves that \( A^r = (a^{(r)}_{ij}) \) has the property that \( |a^{(r)}_{ij}| < M^{2r} \) for every \( r \geq 1 \).

Now, for each of the \( n^2 \) terms \( i, j = 1, \ldots, n \) we have
\[
\sum_{r=0}^{\infty} \frac{|a^{(r)}_{ij}|}{r!} < \sum_{r=0}^{\infty} \frac{M^{2r}}{r!} = \exp(M^2)
\]
so that each of these \( n^2 \) series (i.e., for each \( i, j = 1, \ldots, n \)) must converge (Theorem B26(a)). Hence the series \( I + A + A^2/2! + \cdots \) must converge (Theorem B20). \( \blacksquare \)

We call the series in Theorem 10.25 the **matrix exponential series**, and denote its sum by \( e^A = \exp A \). In general, the series for \( e^A \) is extremely difficult, if not impossible, to evaluate. However, there are important exceptions.
Example 10.7  Let $A$ be the diagonal matrix

$$A = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$ 

Then it is easy to see that

$$A^r = \begin{pmatrix} \lambda_1^r & 0 & \cdots & 0 \\ 0 & \lambda_2^r & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^r \end{pmatrix},$$

and hence

$$\exp A = I + A + \frac{A^2}{2!} + \cdots = \begin{pmatrix} e^{\lambda_1} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\lambda_n} \end{pmatrix}.$$ 

Example 10.8  Consider the $2 \times 2$ matrix

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

and let

$$A = \theta J = \begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix}$$

where $\theta \in \mathbb{R}$. Then noting that $J^2 = -I$, we see that $A^2 = -\theta^2 I$, $A^3 = -\theta^3 J$, $A^4 = \theta^4 I$, $A^5 = \theta^5 J$, $A^6 = -\theta^6 I$, and so forth. From elementary calculus we know that

$$\sin \theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots$$

and

$$\cos \theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots$$

and hence
\[ e^A = I + A + A^2 / 2! + \cdots = I + \theta J - \theta^2 I / 2! + \theta^3 J / 3! + \theta^4 I / 4! + \theta^5 J / 5! - \theta^6 I / 6! + \cdots = I(1 - \theta^2 / 2! + \theta^4 / 4! - \cdots) + J(\theta - \theta^3 / 3! + \theta^5 / 5! - \cdots) = (\cos \theta)I + (\sin \theta)J. \]

In other words, using the explicit forms of I and J we see that

\[ e^A = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \]

so that \( e^{\theta J} \) represents a rotation in \( \mathbb{R}^2 \) by an angle \( \theta \).

**Theorem 10.26** Let \( A \in M_n(\mathbb{C}) \) be diagonalizable, and let \( \lambda_1, \ldots, \lambda_r \) be the distinct eigenvalues of \( A \). Then the matrix power series

\[ \sum_{s=0}^{\infty} a_s A^s \]

converges if and only if the series

\[ \sum_{s=0}^{\infty} a_s \lambda_i^s \]

converges for each \( i = 1, \ldots, r \).

**Proof** Since \( A \) is diagonalizable, choose a nonsingular matrix \( P \) such that \( D = P^{-1}AP \) is diagonal. It is then easy to see that for every \( s \geq 1 \) we have

\[ a_s D^s = a_s P^{-1} A^s P = P^{-1} a_s A^s P \]

where the \( n \) diagonal entries of \( D^s \) are just the numbers \( \lambda_i^s \). By Theorem 10.24(c), we know that \( \sum a_s A^s \) converges if and only if \( \sum a_s D^s \) converges. But by definition of series convergence, \( \sum a_s D^s \) converges if and only if \( \sum a_s \lambda_i^s \) converges for every \( i = 1, \ldots, r \).

**Theorem 10.27** Let \( f(x) = a_0 + a_1 x + a_2 x^2 + \cdots \) be any power series with coefficients in \( \mathbb{C} \), and let \( A \in M_n(\mathbb{C}) \) be diagonalizable with spectral decomposition \( A = \lambda_1 E_1 + \cdots + \lambda_r E_r \). Then, if the series

\[ f(A) = a_0 I + a_1 A + a_2 A^2 + \cdots \]

converges, its sum is
\[ f(A) = f(\lambda_1)E_1 + \cdots + f(\lambda_r)E_r. \]

**Proof** As in the proof of Theorem 10.20, let the diagonal form of \( A \) be

\[ D = P^{-1}AP = \lambda_1 P_1 + \cdots + \lambda_r P_r \]

so that \( E_i = PP_iP^{-1} \). Now note that

\[
P^{-1}f(A)P = a_0 P^{-1}P + a_1 P^{-1}AP + a_2 P^{-1}APP^{-1}AP + \cdots \\
= f(P^{-1}AP) \\
= a_0 I + a_1 D + a_2 D^2 + \cdots \\
= f(D).
\]

Using properties (a) – (c) of Theorem 10.20 applied to the \( P_i \), it is easy to see that \( D^k = \lambda^{k}_1 P_1 + \cdots + \lambda^{k}_r P_r \) and hence

\[ f(D) = f(\lambda_1)P_1 + \cdots + f(\lambda_r)P_r. \]

Then if \( f(A) = \sum A_r \) converges, so does \( \sum P^{-1}A_r P = P^{-1}f(A)P = f(D) \) (Theorem 10.24(c)), and we have

\[ f(A) = f(PDP^{-1}) = Pf(D)P^{-1} = f(\lambda_1)E_1 + \cdots + f(\lambda_r)E_r. \]

**Example 10.9** Consider the exponential series \( e^A \) where \( A \) is diagonalizable. Then, if \( \lambda_1, \ldots, \lambda_k \) are the distinct eigenvalues of \( A \), we have the spectral decomposition \( A = \lambda_1 E_1 + \cdots + \lambda_k E_k \). Using \( f(A) = e^A \), Theorem 10.27 yields

\[ e^A = e^{\lambda_1}E_1 + \cdots + e^{\lambda_k}E_k \]

in agreement with Example 10.7. 

We can now prove our earlier assertion that a unitary matrix \( U \) can be written in the form \( e^{iH} \) for some Hermitian matrix \( H \).

**Theorem 10.28** Every unitary matrix \( U \) can be written in the form \( e^{iH} \) for some Hermitian matrix \( H \). Conversely, if \( H \) is Hermitian, then \( e^{iH} \) is unitary.

**Proof** By Theorem 10.9(b), the distinct eigenvalues of \( U \) may be written in the form \( e^{i\lambda_1}, \ldots, e^{i\lambda_k} \) where each \( \lambda_i \) is real. Since \( U \) is also normal, it fol-
follows from Corollary 1 of Theorem 10.15 that there exists a unitary matrix P such that $P^\dagger UP = P^{-1}UP$ is diagonal. In fact

$$P^{-1}UP = e^{i\lambda_1}P_1 + \cdots + e^{i\lambda_k}P_k$$

where the $P_i$ are the idempotent matrices used in the proof of Theorem 10.20. From Example 10.7 we see that the matrix $e^{i\lambda_1}P_1 + \cdots + e^{i\lambda_k}P_k$ is just $e^{iD}$ where

$$D = \lambda_1 P_1 + \cdots + \lambda_k P_k$$

is a diagonal matrix with the $\lambda_i$ as diagonal entries. Therefore, using Theorem 10.24(c) we see that

$$U = Pe^{iD}P^{-1} = e^{iPDP^{-1}} = e^{iH}$$

where $H = PDP^{-1}$. Since $D$ is a real diagonal matrix it is clearly Hermitian, and since $P$ is unitary (so that $P^{-1} = P^\dagger$), it follows that $H^\dagger = (PDP^\dagger)^\dagger = PDP^\dagger = H$ so that $H$ is Hermitian also.

Conversely, suppose $H$ is Hermitian with distinct real eigenvalues $\lambda_1, \ldots, \lambda_k$. Since $H$ is also normal, there exists a unitary matrix $P$ that diagonalizes $H$. Then as above, we may write this diagonal matrix as

$$P^{-1}HP = \lambda_1 P_1 + \cdots + \lambda_k P_k$$

so that (from Example 10.7 again)

$$P^{-1}e^{iHP} = e^{iP^{-1}HP} = e^{i\lambda_1}P_1 + \cdots + e^{i\lambda_k}P_k .$$

Using the properties of the $P_i$, it is easy to see that the right hand side of this equation is diagonal and unitary since using

$$(e^{i\lambda_1}P_1 + \cdots + e^{i\lambda_k}P_k)^\dagger = e^{-i\lambda_1}P_1 + \cdots + e^{-i\lambda_k}P_k$$

we have

$$(e^{i\lambda_1}P_1 + \cdots + e^{i\lambda_k}P_k)^\dagger(e^{i\lambda_1}P_1 + \cdots + e^{i\lambda_k}P_k) = I$$

and

$$(e^{i\lambda_1}P_1 + \cdots + e^{i\lambda_k}P_k)(e^{i\lambda_1}P_1 + \cdots + e^{i\lambda_k}P_k)^\dagger = I .$$

Therefore the left hand side must also be unitary, and hence (using $P^{-1} = P^\dagger$)
\[
I = (P^{-1} e^{iH} P)^\dagger (P^{-1} e^{iH} P)
= P^\dagger (e^{iH})^\dagger PP^{-1} e^{iH} P
= P^\dagger (e^{iH})^\dagger e^{iH} P
\]
so that \( PP^\dagger = I = (e^{iH})^\dagger e^{iH} \). Similarly we see that \( e^{iH}(e^{iH})^\dagger = I \), and thus \( e^{iH} \) is unitary.

While this theorem is also true in infinite dimensions (i.e., in a Hilbert space), its proof is considerably more difficult. The reader is referred to the books listed in the bibliography for this generalization.

Given a constant matrix \( A \), we now wish to show that

\[
\frac{d e^{tA}}{dt} = A e^{tA} .
\]  

To see this, we first define the derivative of a matrix \( M = M(t) \) to be that matrix whose elements are just the derivatives of the corresponding elements of \( M \). In other words, if \( M(t) = (m_{ij}(t)) \), then \( (dM/dt)_{ij} = dm_{ij}/dt \). Now note that (with \( M(t) = tA \))

\[
e^{tA} = I + tA + (tA)^2/2! + (tA)^3/3! + \cdots
\]
and hence (since the \( a_{ij} \) are constant) taking the derivative with respect to \( t \) yields the desired result:

\[
\frac{d e^{tA}}{dt} = 0 + A + tA^2 + (tA)^2 A/2! + \cdots
= A \{ I + tA + (tA)^2/2! + \cdots \}
= A e^{tA} .
\]

Next, given two matrices \( A \) and \( B \) (of compatible sizes), we recall that their commutator is the matrix \([A, B] = AB - BA = -[B, A] \). If \([A, B] = 0 \), then \( AB = BA \) and we say that \( A \) and \( B \) commute. Now consider the function \( f(x) = e^{xA}B e^{-xA} \). Leaving it to the reader to verify that the product rule for derivatives also holds for matrices, we obtain (note that \( A e^{xA} = e^{xA}A \))

\[
df/dx = A e^{xA}B e^{-xA} - e^{xA}B e^{-xA} A = Af - fA = [A, f]
\]
\[
d^2 f/dx^2 = [A, df/dx] = [A, [A, f]]
\]
\[\vdots\]
Expanding \( f(x) \) in a Taylor series about \( x = 0 \), we find (using \( f(0) = B \))

\[
f(x) = f(0) + (df/dx)_0 x + (d^2 f/dx^2)_0 x^2 / 2! + \cdots = B + [A, B] x + [A, [A, B]] x^2 / 2! + \cdots.
\]

Setting \( x = 1 \), we finally obtain

\[
e^A B e^{-A} = B + [A, B] + [A, [A, B]] / 2! + [A, [A, [A, B]]] / 3! + \cdots \tag{2}
\]

Note that setting \( B = I \) shows that \( e^A e^{-A} = I \) as we would hope.

In the particular case that both \( A \) and \( B \) commute with their commutator \([A, B]\), then we find from (2) that \( e^A B e^{-A} = B + [A, B] \) and hence \( e^A B = Be^A + [A, B] e^A \) or

\[
[A^A, B] = [A, B] e^A. \tag{3}
\]

**Example 10.10** We now show that if \( A \) and \( B \) are two matrices that both commute with their commutator \([A, B]\), then

\[
e^A e^B = \exp\{A + B + [A, B]/2\}. \tag{4}
\]

(This is sometimes referred to as *Weyl’s formula.*) To prove this, we start with the function \( f(x) = e^{x A} e^{x B} e^{-x (A+B)} \). Then

\[
df/dx = e^{x A} A e^{x B} e^{-x (A+B)} + e^{x A} e^{x B} B e^{-x (A+B)} - e^{x A} e^{x B} (A + B) e^{-x (A+B)}
\]

\[
= e^{x A} e^{x B} e^{-x (A+B)} - e^{x A} e^{x B} A e^{-x (A+B)}
\]

\[
= e^{x A} [A, e^{x B}] e^{-x (A+B)}. \tag{5}
\]

As a special case, note \([A, B] = 0\) implies \( df/dx = 0 \) so that \( f \) is independent of \( x \). Since \( f(0) = I \), it follows that we may choose \( x = 1 \) to obtain \( e^A e^B e^{- (A+B)} = I \) or \( e^A e^B = e^{A+B} \) (as long as \([A, B] = 0\)).

From (3) we have (replacing \( A \) by \( xB \) and \( B \) by \( A \)) \([A, e^{xB}] = x[A, B] e^{xB}\). Using this along with the fact that \( A \) commutes with the commutator \([A, B]\) (so that \( e^{xA}[A, B] = [A, B] e^{xA}\)), we have

\[
df/dx = xe^{xA}[A, B] e^{xB} e^{-x (A+B)} = x[A, B] f.
\]

Since \( A \) and \( B \) are independent of \( x \), we may formally integrate this from 0 to \( x \) to obtain

\[
\ln f(x)/f(0) = [A, B] x^2 / 2.
\]
Using \( f(0) = 1 \), this is \( f(x) = \exp\{[A, B]x^2/2\} \) so that setting \( x = 1 \) we find

\[
e^A e^B e^{-(A+B)} = \exp\{[A, B]/2\}.
\]

Finally, multiplying this equation from the right by \( e^{A+B} \) and using the fact that \([A, B]/2, A + B] = 0 \) yields (4).

**Exercises**

1. (a) Let \( N \) be a normal operator on a finite-dimensional unitary space. Prove that

\[
det e^N = e^{Tr N}.
\]

(b) Prove this holds for any \( N \in M_n(\mathbb{C}) \). [Hint: Use either Theorem 8.1 or the fact (essentially proved at the end of Section 8.6) that the diagonalizable matrices are dense in \( M_n(\mathbb{C}) \).]

2. If the limit of a sequence of unitary operators exists, is it also unitary? Why?

3. Let \( T \) be a unitary operator. Show that the sequence \( \{T^n: n = 0, 1, 2, \ldots \} \) contains a subsequence \( \{T^{n_k}: k = 0, 1, 2, \ldots \} \) that converges to a unitary operator. [Hint: You will need the fact that the unit disk in \( \mathbb{C}^2 \) is compact (see Appendix A).]

**10.7 POSITIVE OPERATORS**

Before proving the main result of this section (the polar decomposition theorem), let us briefly discuss functions of a linear transformation. We have already seen two examples of such a function. First, the exponential series \( e^A \) (which may be defined for operators exactly as for matrices) and second, if \( A \) is a normal operator with spectral decomposition \( A = \sum \lambda_i E_i \), then we saw that the linear transformation \( p(A) \) was given by \( p(A) = \sum p(\lambda_i) E_i \) where \( p(x) \) is any polynomial in \( \mathbb{C}[x] \) (Corollary to Theorem 10.20).

In order to generalize this notion, let \( N \) be a normal operator on a unitary space, and hence \( N \) has spectral decomposition \( \sum \lambda_i E_i \). If \( f \) is an arbitrary complex-valued function (defined at least at each of the \( \lambda_i \)), we define a linear transformation \( f(N) \) by

\[
f(N) = \sum f(\lambda_i) E_i.
\]
What we are particularly interested in is the function \( f(x) = \sqrt{x} \) defined for all real \( x \geq 0 \) as the positive square root of \( x \).

Recall (see Section 10.3) that we defined a positive operator \( P \) by the requirement that \( P = S^*S \) for some operator \( S \). It is then clear that \( P^* = P \), and hence \( P \) is normal. From Theorem 10.9(d), the eigenvalues of \( P = \sum \lambda_j E_j \) are real and non-negative, and we can define \( \sqrt{P} \) by

\[
\sqrt{P} = \sum \sqrt{\lambda_j} E_j
\]

where each \( \lambda_j \geq 0 \).

Using the properties of the \( E_j \), it is easy to see that \((\sqrt{P})^2 = P\). Furthermore, since \( E_j \) is an orthogonal projection, it follows that \( E_j^* = E_j \) (Theorem 10.21), and therefore \((\sqrt{P})^* = \sqrt{P} \) so that \( \sqrt{P} \) is Hermitian. Note that since \( P = S^*S \) we have

\[
\langle Pv, v \rangle = \langle (S^*S)v, v \rangle = \langle Sv, Sv \rangle = \|Sv\|^2 \geq 0.
\]

Just as we did in the proof of Theorem 10.23, let us write \( v = \sum \lambda_j E_j \) where the nonzero \( \lambda_j \) are mutually orthogonal. Then

\[
\sqrt{P}(v) = \sum \sqrt{\lambda_j} E_j v = \sum \sqrt{\lambda_j} \lambda_j
\]

and hence we also have (using \( \langle \lambda_j, \lambda_k \rangle = 0 \) for \( j \neq k \))

\[
\langle \sqrt{P}(v), v \rangle = \langle \sum_j \sqrt{\lambda_j} \lambda_j \rangle \sum_k \lambda_k \lambda_k = \sum_{j,k} \sqrt{\lambda_j} \lambda_k \langle \lambda_j, \lambda_k \rangle = \sum_j \sqrt{\lambda_j} \| \lambda_j \|^2 \geq 0.
\]

In summary, we have shown that \( \sqrt{P} \) satisfies

(a) \( (\sqrt{P})^2 = P \)
(b) \( (\sqrt{P})^* = \sqrt{P} \)
(c) \( \langle \sqrt{P}(v), v \rangle \geq 0 \)

and it is natural to ask about the uniqueness of any operator satisfying these three properties. For example, if we let \( T = \sum \pm \sqrt{\lambda_j} E_j \), then we still have \( T^2 = \sum \lambda_j E_j = P \) regardless of the sign chosen for each term. Let us denote the fact that \( \sqrt{P} \) satisfies properties (b) and (c) above by the expression \( \sqrt{P} = 0 \). In other words, by the statement \( A = 0 \) we mean that \( A^* = A \) and \( \langle Av, v \rangle \geq 0 \) for every \( v \in V \) (i.e., \( A \) is a positive Hermitian operator).

We now claim that if \( P = T^2 \) and \( T \geq 0 \), then \( T = \sqrt{P} \). To prove this, we first note that \( T \geq 0 \) implies \( T^* = T \) (property (b)), and hence \( T \) must also be normal. Now let \( \sum \mu_i F_i \) be the spectral decomposition of \( T \). Then
If \( v_i \neq 0 \) is an eigenvector of \( T \) corresponding to \( \mu_i \), then property (c) tells us that (using the fact that each \( \mu_i \) is real since \( T \) is Hermitian)

\[
0 \leq \langle Tv_i, v_i \rangle = \langle \mu_i v_i, v_i \rangle = \mu_i \|v_i\|^2 .
\]

But \( \|v_i\| > 0 \), and hence \( \mu_i \geq 0 \). In other words, any operator \( T \geq 0 \) has nonnegative eigenvalues. Since each \( \mu_i \) is distinct and nonnegative, so is each \( \mu_i^2 \), and hence each \( \mu_i^2 \) must be equal to some \( \lambda_i \). Therefore the corresponding \( F_i \) and \( E_i \) must be equal (by Theorem 10.22(e)). By suitably numbering the eigenvalues, we may write \( \mu_i^2 = \lambda_i \), and thus \( \mu_i = \sqrt{\lambda_i} \). This shows that

\[
T = \sum \mu_i F_i = \sum \sqrt{\lambda_i} E_i = \sqrt{P}
\]
as claimed.

We summarize this discussion in the next result which gives us three equivalent definitions of a \textbf{positive transformation}.

\textbf{Theorem 10.29}  Let \( P \) be an operator on a unitary space \( V \). Then the following conditions are equivalent:

(a) \( P = T^2 \) for some unique Hermitian operator \( T \geq 0 \).
(b) \( P = S^\dagger S \) for some operator \( S \).
(c) \( P^\dagger = P \) and \( \langle Pv, v \rangle \geq 0 \) for every \( v \in V \).

\textbf{Proof}  (a) \( \Rightarrow \) (b): If \( P = T^2 \) and \( T^\dagger = T \), then \( P = TT = T^\dagger T \).
(b) \( \Rightarrow \) (c): If \( P = S^\dagger S \), then \( P^\dagger = P \) and \( \langle Pv, v \rangle = \langle S^\dagger Sv, v \rangle = \langle Sv, Sv \rangle = \|Sv\|^2 \geq 0 \).
(c) \( \Rightarrow \) (a): Note that property (c) is just our statement that \( P \geq 0 \). Since \( P^\dagger = P \), we see that \( P \) is normal, and hence we may write \( P = \sum \lambda_i E_i \). Defining \( T = \sum \sqrt{\lambda_i} E_i \), we have \( T^\dagger = T \) (since every \( E_i \) is Hermitian), and the preceding discussion shows that \( T \geq 0 \) is the unique operator with the property that \( P = T^2 \).  

We remark that in the particular case that \( P \) is positive definite, then \( P = S^\dagger S \) where \( S \) is nonsingular. This means that \( P \) is also nonsingular.

Finally, we are in a position to prove the last result of this section, the so-called polar decomposition (or factorization) of an operator. While we state and prove this theorem in terms of matrices, it should be obvious by now that it applies just as well to operators.
**Theorem 10.30 (Polar Decomposition)**  If $A \in M_n(\mathbb{C})$, then there exist unique positive Hermitian matrices $H_1$, $H_2 \in M_n(\mathbb{C})$ and (not necessarily unique) unitary matrices $U_1$, $U_2 \in M_n(\mathbb{C})$ such that $A = U_1H_1 = H_2U_2$. Moreover, $H_1 = (A^\dagger A)^{1/2}$ and $H_2 = (AA^\dagger)^{1/2}$. In addition, the matrices $U_1$ and $U_2$ are uniquely determined if and only if $A$ is nonsingular.

**Proof**  Let $\lambda_1^2, \ldots, \lambda_n^2$ be the eigenvalues of the positive Hermitian matrix $A^\dagger A$, and assume the $\lambda_i$ are numbered so that $\lambda_i > 0$ for $i = 1, \ldots, k$ and $\lambda_i = 0$ for $i = k + 1, \ldots, n$ (see Theorem 10.9(d)). (Note that if $A$ is nonsingular, then $A^\dagger A$ is positive definite and hence $k = n$.) Applying Corollary 1 of Theorem 10.15, we let \{v_1, \ldots, v_n\} be the corresponding orthonormal eigenvectors of $A^\dagger A$. For each $i = 1, \ldots, k$ we define the vector $w_i = Av_i/\lambda_i$. Then

$$
\langle w_i, w_j \rangle = \langle Av_i/\lambda_i, Av_j/\lambda_j \rangle = \langle v_i, A^\dagger Av_j/\lambda_i \lambda_j \rangle = \langle v_i, v_j \rangle \lambda_j^2/\lambda_i \lambda_j = \delta_{ij} \lambda_j^2/\lambda_i \lambda_j
$$

so that $w_1, \ldots, w_k$ are also orthonormal. We now extend these to an orthonormal basis \{w_1, \ldots, w_n\} for $\mathbb{C}^n$. If we define the columns of the matrices $V$, $W \in M_n(\mathbb{C})$ by $V^\dagger = v_i$ and $W^\dagger = w_i$, then $V$ and $W$ will be unitary by Theorem 10.7.

Defining the Hermitian matrix $D \in M_n(\mathbb{C})$ by

$$
D = \text{diag}(\lambda_1, \ldots, \lambda_n)
$$

it is easy to see that the equations $Av_i = \lambda_i w_i$ may be written in matrix form as $AV = WD$. Using the fact that $V$ and $W$ are unitary, we define $U_1 = WV^\dagger$ and $H_1 = VD$ to obtain

$$
A = WDV^\dagger = (WV^\dagger)(VDV^\dagger) = U_1H_1.
$$

Since $\det(\lambda I - VDV^\dagger) = \det(\lambda I - D)$, we see that $H_1$ and $D$ have the same nonnegative eigenvalues, and hence $H_1$ is a positive Hermitian matrix. We can now apply this result to the matrix $A^\dagger$ to write $A^\dagger = \tilde{U}^\dagger \tilde{H}_1$ or $A = \tilde{H}_1 \tilde{U}_1^\dagger = \tilde{H}_1 \tilde{U}_1^\dagger$. If we define $H_2 = \tilde{H}_1$ and $U_2 = \tilde{U}_1^\dagger$, then we obtain $A = H_2U_2$ as desired.

We now observe that using $A = U_1H_1$ we may write

$$
A^\dagger A = H_1U_1^\dagger U_1H_1 = (H_1)^2
$$

and similarly

$$
AA^\dagger = H_2U_2^\dagger U_2H_2 = (H_2)^2
$$
so that \( H_1 \) and \( H_2 \) are unique even if \( A \) is singular. Since \( U_1 \) and \( U_2 \) are unitary, they are necessarily nonsingular, and hence \( H_1 \) and \( H_2 \) are nonsingular if \( A = U_1 H_1 = H_2 U_2 \) is nonsingular. In this case, \( U_1 = AH_1^{-1} \) and \( U_2 = H_2^{-1}A \) will also be unique. On the other hand, suppose \( A \) is singular. Then \( k \neq n \) and \( w_k, \ldots, w_n \) are not unique. This means that \( U_1 = WV \) (and similarly \( U_2 \)) is not unique. In other words, if \( U_1 \) and \( U_2 \) are unique, then \( A \) must be nonsingular.

**Exercises**

1. Let \( V \) be a unitary space and let \( E \in \mathcal{L}(V) \) be an orthogonal projection.
   (a) Show directly that \( E \) is a positive transformation.
   (b) Show that \( \|Ev\| \leq \|v\| \) for all \( v \in V \).

2. Prove that if \( A \) and \( B \) are commuting positive transformations, then \( AB \) is also positive.

3. This exercise is related to Exercise 7.5.5. Prove that any representation of a finite group is equivalent to a unitary representation as follows:
   (a) Consider the matrix \( X = \sum_{a \in G} D^\dagger(a)D(a) \). Show that \( X \) is Hermitian and positive definite, and hence that \( X = S^2 \) for some Hermitian \( S \).
   (b) Show that \( D(a)^\dagger XD(a) = X \).
   (c) Show that \( U(a) = SD(a)S^{-1} \) is a unitary representation.

**Supplementary Exercises for Chapter 10**

1. Let \( T \) be a linear transformation on a space \( V \) with basis \( \{e_1, \ldots, e_n\} \). If \( T(e_i) = \sum_j a_{ij} e_j \) for all \( i = 1, \ldots, n \) and \( T(e_i) \neq ce_i \) for any scalar \( c \), show that \( T \) is not normal.

2. Let \( A \) be a fixed \( n \times n \) matrix, and let \( B \) be any \( n \times n \) matrix such that \( A = B^2 \). Assume that \( B \) is similar to a diagonal matrix and has nonnegative eigenvalues \( \lambda_1, \ldots, \lambda_n \). Let \( p(x) \) be a polynomial such that \( p(\lambda_i^2) = \lambda_i \) for each \( i = 1, \ldots, n \). Show that \( p(A) = B \) and hence \( B \) is unique. How does this relate to our discussion of \( \sqrt{P} \) for a positive operator \( P \)?

3. Describe all operators that are both unitary and positive.

4. Is it true that for any \( A \in \mathcal{M}_n(\mathbb{C}) \), \( AA^\dagger \) and \( A^\dagger A \) are unitarily similar? Explain.
5. In each case, indicate whether or not the statement is true or false and give your reason.
   (a) For any $A \in M_n(\mathbb{C})$, $AA^\dagger$ has all real eigenvalues.
   (b) For any $A \in M_n(\mathbb{C})$, the eigenvalues of $AA^\dagger$ are of the form $|\lambda|^2$ where $\lambda$ is an eigenvalue of $A$.
   (c) For any $A \in M_n(\mathbb{C})$, the eigenvalues of $AA^\dagger$ are nonnegative real numbers.
   (d) For any $A \in M_n(\mathbb{C})$, $AA^\dagger$ has the same eigenvalues as $A^\dagger A$ if $A$ is nonsingular.
   (e) For any $A \in M_n(\mathbb{C})$, $\text{Tr}(AA^\dagger) = |\text{Tr } A|^2$.
   (f) For any $A \in M_n(\mathbb{C})$, $AA^\dagger$ is unitarily similar to a diagonal matrix.
   (g) For any $A \in M_n(\mathbb{C})$, $AA^\dagger$ has $n$ linearly independent eigenvectors.
   (h) For any $A \in M_n(\mathbb{C})$, the eigenvalues of $AA^\dagger$ are the same as the eigenvalues of $A^\dagger A$.
   (i) For any $A \in M_n(\mathbb{C})$, the Jordan form of $AA^\dagger$ is the same as the Jordan form of $A^\dagger A$.
   (j) For any $A \in M_n(\mathbb{C})$, the null space of $A^\dagger A$ is the same as the null space of $A$.

6. Let $S$ and $T$ be normal operators on $V$. Show that there are bases $\{u_i\}$ and $\{v_i\}$ for $V$ such that $[S]_u = [T]_v$ if and only if there are orthonormal bases $\{u'_i\}$ and $\{v'_i\}$ such that $[S]_{u'} = [T]_{v'}$.

7. Let $T$ be normal and let $k > 0$ be an integer. Show that there is a normal $S$ such that $S^k = T$.

8. Let $N$ be normal and let $p(x)$ be a polynomial over $\mathbb{C}$. Show that $p(N)$ is also normal.

9. Let $N$ be a normal operator on a unitary space $V$, let $W = \text{Ker } N$, and let $\tilde{N}$ be the transformation induced by $N$ on $V/W$. Show that $\tilde{N}$ is normal. Show that $\tilde{N}^\dagger$ is also normal.

10. Discuss the following assertion: For any linear transformation $T$ on a unitary space $V$, $TT^\dagger$ and $T^\dagger T$ have a common basis of eigenvectors.

11. Show that if $A$ and $B$ are real symmetric matrices and $A$ is positive definite, then $p(x) = \det(B - xA)$ has all real roots.