Chapter 1

Background Linear Algebra

This course requires some knowledge of linear algebra beyond what is normally taught in a beginning undergraduate course. In this section, we recall some general facts from basic linear algebra and introduce some additional facts needed in the course, including generalized eigenspaces and the Jordan canonical form of a linear map on a complex vector space. It is important that you familiarize yourself with these basic facts, which are also important in and of themselves.

1.1 Subspaces and Quotient Spaces

In what follows, all vector spaces will be defined over the field $\mathbb{F}$ of either real or complex numbers. In other words, $\mathbb{F}$ will denote either $\mathbb{R}$ or $\mathbb{C}$, so when we talk about a vector space V over $\mathbb{F}$, we mean that V is a vector space either over $\mathbb{R}$ or $\mathbb{C}$.

A vector space can of course be defined over any algebraic field $\mathbb{F}$, and not just $\mathbb{R}$ or $\mathbb{C}$, but we will limit ourselves to these two fields in order to simplify the exposition. Many of the results presented here carry over to vector spaces over arbitrary fields (the ones for $\mathbb{C}$ mostly carry over to algebraically closed fields), although the proofs may not necessarily be the same, especially for fields with prime characteristic.

Unless otherwise stated, all vector spaces will be finite-dimensional.

Let V be a vector space over $\mathbb{F}$, and let A and B be any nonempty subsets of V and $\lambda \in \mathbb{F}$. We will use the following notation:

$$A + B = \{a + b \,|\, a \in A \text{ and } b \in B\} \tag{1.1}$$

and

$$\lambda A = \{\lambda a \,|\, a \in A\}\,. \tag{1.2}$$

For simplicity, if v is a vector in V, we write

$$v + B := \{v\} + B.$$

This set is called the *translate* of B by the vector v. From (1.1) above, it is easy to see that

$$A + B = \bigcup_{a \in A} (a + B) = \bigcup_{b \in B} (b + A).$$

Note that a nonempty subset W of V is a vector subspace of V if and only if $W + W \subset W$ and $\lambda W \subset W$, for all $\lambda \in \mathbb{F}$.

Suppose now that W is a subspace of the vector space V. If v is any vector in V, the translate $v + W$ is called the *affine subspace with direction W through v*.

For example, suppose that V is the ordinary 3-dimensional Euclidean space $\mathbb{R}^3$ and W is the (x, y)-plane:

$$W = \left\{ \left(\begin{array}{c} x \\ y \\ 0 \end{array} \right) \,\middle|\, x, y \in \mathbb{R} \right\}$$

and if $v = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}$, then $v + W$ is the affine plane with equation $z = 3$.

As another example, if W is the one-dimensional subspace of $\mathbb{R}^3$ spanned by the vector

$$w = \begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}$$

(this is a straight line through the origin), and if

$$v = \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix},$$

then $v + W$ is the straight line in $\mathbb{R}^3$ through v and parallel to w; that is, it is the affine straight line specified by the parametric equations

$$x = 4 - t \quad , \quad y = 2t \quad , \quad z = 1 + 2t.$$

Now let W be a subspace of a vector space V. Two translates $v_1 + W$ and $v_2 + W$ of W coincide if and only of $v_1 - v_2 \in W$. To see this, first assume that $v_1 + W = v_2 + W$. Then $v_1 = v_1 + 0 \in v_1 + W = v_2 + W$, so, $v_1 = v_2 + w$ for some $w \in W$, whence $v_1 - v_2 = w \in W$. Conversely, suppose that $v_1 - v_2 = w \in W$. Then $v_1 + W = v_2 + w + W = v_2 + W$.

The set of all translates of W by vectors in V is denoted by V/W, and is called the *quotient space* of V by W. (We pronounce V/W as "V *mod* W.") Thus,

$$V/W = \{v + W \mid v \in V\}.$$

The set V/W carries a natural vector space structure with vector addition given by (1.1):

$$(v_1 + W) + (v_2 + W) = (v_1 + v_2) + (W + W) = v_1 + v_2 + W,$$

and scalar multiplication on V/W given by

$$\lambda(v + W) = \lambda v + W.$$

Note that this definition of scalar multiplication is slightly different from the definition of scalar multiplication of sets in (1.2) above. (The reason being that $0B = \{0\}$ for any nonempty subset B of V.) We will leave to the student the routine verification that the operations above give rise to a vector space structure on V/W. Note that in V/W the zero vector is $0 + W = W$. (Later, when we study quotient spaces in greater detail, we will abuse notation and simply denote the zero vector in V/W by 0.)

Proposition 1.1.1. *Let W be a subspace of V. Then* $\dim(V/W) = \dim V - \dim W$.

Proof. Let $B' = (w_1, \ldots, w_m)$ be any basis of W, and extend this to a basis $B = (w_1, \ldots, w_m, v_{m+1}, \ldots, v_n)$ of V. We claim that $(v_{m+1} + W, \ldots, v_n + W)$ is a basis of V/W. First we show that they span V/W. Let $v + W$ be an arbitrary element of V/W. Then $v = a_1 w_1 + \cdots + a_m w_m + a_{m+1} v_{m+1} + \cdots + a_n v_n$, for suitable scalars $a_1, \ldots, a_n$. Then

$$\begin{aligned} v + W &= a_1 w_1 + \cdots + a_m w_m + a_{m+1} v_{m+1} + \cdots + a_n v_n + W \\ &= a_{m+1} v_{m+1} + \cdots + a_n v_n + W \\ &= a_{m+1}(v_{m+1} + W) + \cdots + a_n(v_n + W), \end{aligned}$$

so $v + W$ is a linear combination of $(v_{m+1} + W, \ldots, v_n + W)$.

Next we show that $(v_{m+1} + W, \ldots, v_n + W)$ is a linearly independent set in V/W. Suppose that $a_{m+1}(v_{m+1} + W) + \cdots + a_n(v_n + W) = 0$. This is equivalent to

$a_{m+1}v_{m+1} + \cdots + a_nv_n + W = W$, so that $a_{m+1}v_{m+1} + \cdots + a_nv_n \in W$. Since $w_1, \ldots, w_m$ is a basis of W, we must have $a_{m+1}v_{m+1} + \cdots + a_nv_n = b_1w_1 + \cdots + b_mw_m$, for suitable scalars $b_1, \ldots, b_m$, and thus

$$-b_1w_1 - \cdots - b_mw_m + a_{m+1}v_{m+1} + \cdots + a_nv_n = 0;$$

Since $(w_1, \ldots, w_m, v_{m+1}, \ldots, v_n)$ is a basis of V, we see that, in particular, $a_{m+1} = \cdots = a_n = 0$. □

It is easy to check that the sum of subspaces of V is also a subspace of V. Explicitly, if $W_1, \ldots, W_k$ are subspaces of V, then $W_1 + \cdots + W_k$ is also a subspace of V. This sum is called a *direct sum* if, for any vectors $w_1 \in W_1, \ldots, w_k \in W_k$, the condition

$$w_1 + \cdots + w_k = 0$$

implies that $w_1 = 0, \ \ldots, w_k = 0$. In this case, we will use the notation $W_1 \oplus \cdots \oplus W_k$ to denote the direct sum. Note that $(w_1, \ldots, w_m)$ is a linearly independent set if and only if the subspace sum $\mathbb{F}w_1 + \cdots + \mathbb{F}w_m$ is direct.

Exercise 1.1.2. Prove that if U and W are subspaces of V, then the sum $U + W$ is direct if and only if $U \cap W = \{0\}$.

Example 1.1.3. Let $\langle\ ,\ \rangle$ be an inner product on a real vector space V. If W is a subspace of V, put $W^\perp = \{v \in V \mid \langle v, w\rangle = 0 \text{ for all } w \in W\}$. The subspace $W^\perp$ is called the *orthogonal complement* of W in V. We have $V = W \oplus W^\perp$. (See [Axl97], Theorem 6.29.)

Exercise 1.1.4. Let U and W be subspaces of V. Show that $\dim(U + W) = \dim U + \dim W - \dim(U \cap W)$. From this, show that $\dim(U \oplus W) = \dim U + \dim W$.

Given any subspace W of a vector space V, we can always find a subspace U of V such that $V = W \oplus U$. (U is called a *complementary subspace* to W.) The cases $W = \{0\}$ and $W = V$ being trivial, we can assume that $\{0\} \neq W \subsetneqq V$. Take any basis $(w_1, \ldots, w_m)$ of W, extend this to a basis $(w_1, \ldots, w_m, v_{m+1}, \ldots, v_n)$ of V, and put $U = \mathbb{F}v_{m+1} + \cdots + \mathbb{F}v_n$. Then it is clear that $V = U \oplus W$. Since there are infinitely many ways to complete a basis of W to a basis of V, it is also clear that, unless $W = \{0\}$ or $W = V$, the choice of a complementary subspace to W is not unique.

1.2 Linear Maps

Let V and W be vector spaces over $\mathbb{F}$. The set of all linear maps from V to W will be denoted by $\mathcal{L}(V, W)$. $\mathcal{L}(V, W)$ has a natural vector space structure given

by addition and scalar multiplication of linear maps: if S and T are in $\mathcal{L}(V,W)$ and $\lambda \in \mathbb{F}$, then the linear maps $S+T$ and λS in $\mathcal{L}(V,W)$ are given by

$$\begin{aligned}(S+T)(v) &= S(v)+T(v)\\ (\lambda S)(v) &= \lambda\, S(v) \qquad\qquad \text{for all } v \in V.\end{aligned}$$

It is not hard to prove that with these operations, $\mathcal{L}(V,W)$ is a vector space over $\mathbb{F}$.

Fix a basis $(v_1,\ldots,v_n)$ of V. Then any $T \in \mathcal{L}(V,W)$ is completely determined by its effect on the basis vectors v_j. For if $v \in V$, then we can write $v = c_1v_1 + \cdots + c_nv_n$ for scalars $c_1,\ldots,c_n$, whence

$$T(v) = c_1T(v_1) + \cdots + c_nT(v_n). \tag{1.3}$$

Conversely, given any vectors $w_1,\ldots,w_n$ in W, there is a unique linear map $T \in \mathcal{L}(V,W)$ such that $T(v_1) = w_1,\ldots,T(v_n) = w_n$. This is because any vector $v \in V$ can be written uniquely as $v = c_1v_1 + \cdots c_nv_n$; if we define the map $T : V \to W$ by (1.3), then it is easy to see that $T \in \mathcal{L}(V,W)$.

Abstract linear algebra is inextricably bound to matrix theory since any linear map may be represented by an appropriate matrix, and since the algebra of linear maps corresponds to the algebra of matrices.

More precisely, let us fix bases $B = (v_1,\ldots,v_n)$ and $B' = (w_1,\ldots,w_m)$ of vector spaces V and W, respectively. Recall that any $T \in \mathcal{L}(V,W)$ is uniquely determined by the basis images $T(v_1),\ldots,T(v_n)$ in W. Each of these vectors $T(v_j)$ is a unique linear combination of $w_1,\ldots,w_m$:

$$T(v_j) = \sum_{i=1}^{m} a_{ij}w_i \qquad\qquad (j = 1,\ldots,n). \tag{1.4}$$

We define the *matrix* $M_{B,B'}(T)$ *of* T *with respect to these bases* to be the $m \times n$ matrix whose (i,j)-entry is a_{ij};

$$M_{B',B}(T) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}. \tag{1.5}$$

We will frequently denote this matrix by $M(T)$ if the bases B and B' are clear from the context of the discussion.

Let $T \in \mathcal{L}(V,W)$. The *kernel*, or *nullspace*, of T is the subspace of V given by

$$\ker T = \{v \in V \mid T(v) = 0\}.$$

From linear algebra, we know that the linear map T is *injective*, or one-to-one, if and only if $\ker T = \{0\}$; in this case we say that T is a *linear isomorphism of* V *into* W.

The *range* of T is the subspace of W given by

$$T(V) = \{T(v) \,|\, v \in V\}.$$

We recall the definition that T is *surjective*, or *onto*, if $T(V) = W$.

The following is an easy to prove, yet important fact in linear algebra.

Theorem 1.2.1. *(The rank-Nullity Theorem) For any $T \in \mathcal{L}(V, W)$, we have*

$$\dim T(V) = \dim V - \dim(\ker T).$$

In particular, if T is a linear isomorphism from V onto W, then $\dim V = \dim W$.

Proof. Let $(v_1, \ldots, v_k)$ be a basis of $\ker T$, and extend this to a basis $(v_1, \ldots, v_k, v_{k+1}, \ldots, v_n)$ of V. Then $T(v_j) = 0$ for $1 \le j \le k$, and $(T(v_{k+1}), \ldots, T(v_n))$ is linearly independent: in fact, any relation $\sum_{j=k+1}^{n} a_j\, T(v_j) = 0$ implies that $\sum_{j=k+1}^{n} a_j\, v_j \in \ker T = \operatorname{span}\{v_1, \ldots, v_k\}$. Hence $a_{k+1} = \cdots = a_n = 0$. Moreover, since $T(V) = \operatorname{span}\{T(v_1), \ldots, T(v_n)\} = \operatorname{span}\{T(v_{k+1}), \ldots, T(v_n)\}$, we see that $(T(v_{k+1}), \ldots, T(v_n))$ is a basis of $T(V)$.

Since $\dim(\ker T) = k$ and $\dim T(V) = n - k$, the theorem follows. □

Making an abrupt and unforgivable change of notation for the moment, suppose that W is a subspace of a vector space V. The *quotient map* π from V onto the quotient space V/W is given by $\pi(v) = v + W$. It is obvious that π is linear and surjective, with kernel W. Using Theorem 1.2.1, this provides a completely trivial proof of Proposition 1.1.1.

1.3 The Matrix of a Linear Map

Again let us fix bases $B = \{v_1, \ldots, v_n\}$ and $B' = \{w_1, \ldots, w_m\}$ of V and W, respectively. From (1.4) and (1.5) we see that each $T \in \mathcal{L}(V, W)$ corresponds to a unique $m \times n$ matrix $M_{B',B}(T)$. The map $T \mapsto M_{B',B}(T)$ is from $\mathcal{L}(V, W)$ to the vector space $M_{m,n}(\mathbb{F})$ of $m \times n$ matrices with entries in $\mathbb{F}$ is easily checked to be linear (and onto), and hence is a linear isomorphism. Since $\dim M_{m,n} = mn$, we see that $\dim \mathcal{L}(V, W) = mn = nm = \dim V \cdot \dim W$.

Another useful property of the map $T \mapsto M_{B',B}(T)$ is that it is multiplicative. More precisely, suppose that V, W, and U are vector spaces with fixed bases B, B', and B'', respectively, and suppose that $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, U)$. Then the composite map $ST := S \circ T$ belongs to $\mathcal{L}(V, U)$, and we have

$$M_{B'',B}(ST) = M_{B'',B'}(S)\, M_{B',B}(T), \tag{1.6}$$

where the right hand side is a matrix product.

Exercise 1.3.1. Prove equation (1.6).

For simplicity, we'll denote the space of linear maps $\mathcal{L}(V, V)$ simply by $\mathcal{L}(V)$. An element of $\mathcal{L}(V)$ is called a *linear operator* on V.

Theorem 1.3.2. *Fix a basis $B = (v_1, \ldots, v_n)$ of V. Suppose that $T \in \mathcal{L}(V)$. Then the following are equivalent:*

1. *T is one-to-one.*
2. *T is onto.*
3. *The matrix $M(T) := M_{B,B}(T)$ with respect to the basis B is nonsingular.*

Proof. (1) $\iff$ (2): $\dim T(V) = \dim V - \dim(\ker T)$ so $\dim T(V) = \dim V$ if and only if $\dim(\ker T) = 0$.
(1) and (2) $\iff$ (3): If T is one-to-one and onto, it is invertible; that is, there is a unique linear map $S \in \mathcal{L}(V)$ such that $ST = TS = 1_V$, the identity map on V. If I_n is the identity $n \times n$ matrix, we see from (1.6) that

$$\begin{aligned} I_n &= M(ST) = M(S)M(T) \\ &= M(TS) = M(T)M(S), \end{aligned}$$

which shows that $M(T)$ is invertible. Conversely, assume $M(T)$ is an invertible $n \times n$ matrix; let S be the linear operator on V whose matrix is $M(T)^{-1}$. Then, again by (1.6),

$$\begin{aligned} I_n &= M(T)^{-1}M(T) = M(S)M(T) = M(ST) \\ &= M(T)M(T)^{-1} = M(T)M(S) = M(TS), \end{aligned}$$

which shows that $ST = TS = 1_V$. □

If P and Q are $n \times n$ matrices with P nonsingular, the *conjugate* of Q by P is the $n \times n$ matrix PQP^{-1}. Suppose now that $B = (v_1, \ldots, v_n)$ and $B' = (v'_1, \ldots, v'_n)$ are two bases of the same vector space V. The *change of basis matrix* from B to B' is $M_{B',B}(1_V)$: it gives us the coefficients in the linear combination expressing each v_j as a linear combination of the v'_i's. From (1.6), this change of basis matrix is nonsingular, with inverse $M_{B,B'}(1_V)$. For simplicity, let us denote this change of basis matrix by S. Let $T \in \mathcal{L}(V)$. If $M(T) := M_{B,B}(T)$ is the matrix of T with respect to the basis B, then its matrix with respect to B' is given by conjugating $M(T)$ by S. Explicitly, by (1.6)

$$\begin{aligned} M_{B',B'}(T) &= M_{B',B}(1_V)M_{B,B}(T)M_{B,B'}(1_V) \\ &= SM(T)S^{-1}. \end{aligned} \tag{1.7}$$

Example 1.3.3. Suppose that T is the linear operator on $\mathbb{R}^3$ whose matrix with respect to the standard basis $B_0 = (e_1, e_2, e_3)$ of $\mathbb{R}^3$ is

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

Consider the vectors

$$v_1 = \begin{pmatrix} 4 \\ 8 \\ 7 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} -2 \\ -3 \\ -2 \end{pmatrix}, \qquad v_3 = \begin{pmatrix} 3 \\ 5 \\ 4 \end{pmatrix}$$

The 3×3 matrix S whose columns are v_1, v_2, v_3 is invertible; its inverse can be calculated using the Gauss-Jordan method and is found to be

$$\begin{pmatrix} 4 & -2 & 3 \\ 8 & -3 & 5 \\ 7 & -2 & 4 \end{pmatrix}^{-1} = \begin{pmatrix} -2 & 2 & -1 \\ 3 & -5 & 4 \\ 5 & -6 & 4 \end{pmatrix}.$$

We therefore see that $B = (v_1, v_2, v_3)$ is a linearly independent set which thus forms a basis of $\mathbb{R}^3$, and the change of basis matrix from the standard basis B_0 to B is given by S. Hence the matrix of T with respect to the basis B is

$$\begin{aligned} SAS^{-1} &= \begin{pmatrix} 4 & -2 & 3 \\ 8 & -3 & 5 \\ 7 & -2 & 4 \end{pmatrix} \begin{pmatrix} 2 & -1 & 0 \\ -1 & 3 & 1 \\ 0 & 2 & 0 \end{pmatrix} \begin{pmatrix} -2 & 2 & -1 \\ 3 & -5 & 4 \\ 5 & -6 & 4 \end{pmatrix} \\ &= \begin{pmatrix} -22 & 28 & -18 \\ -74 & 91 & -59 \\ -57 & 69 & -44 \end{pmatrix}. \end{aligned}$$

The *transpose* of an $m \times n$ matrix $A = a_{ij}$ is the $n \times m$ matrix ${}^tA$ whose (i, j) entry is a_{ji}. Thus rows of A transform into the columns of ${}^tA$, and the columns of A transform into the rows of ${}^tA$. It's not hard to show that if A and B are $m \times n$ matrices and α is a scalar, then ${}^t(A + B) = {}^tA + {}^tB$, ${}^t(\alpha A) = \alpha\, {}^tA$. A somewhat longer but completely straightforward calculation shows that if A is an $m \times n$ matrix and B is an $n \times k$ matrix, then ${}^t(AB) = {}^tB\, {}^tA$.

The *dual space* of a vector space V is $\mathcal{L}(V, \mathbb{F})$ (where $\mathbb{F}$ is viewed as a one-dimensional vector space), and is denoted V^*. Its elements are called *linear functionals* on V. Any basis $(v_1, \ldots, v_n)$ of V gives rise to a *dual basis* $(f_1, \ldots, f_n)$ of V^* where each f_i is given by

$$f_i(v_j) = \delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Let V and W be vector spaces over $\mathbb{F}$, and let : $V \to W$ be a linear map. The *transpose* of T is the map ${}^tT : W^* \to V^*$ given by ${}^tT(\lambda) = \lambda \circ T$, for all $\lambda \in W^*$.

It is not hard to show that ${}^tT$ is a linear map from W^* to V^*. Suppose that $B = (v_1, \ldots, v_n)$ and $B' = (w_1, \ldots, w_m)$ are bases of V and W, respectively. Let $B^* = (f_1, \ldots, f_n)$ and $(B')^* = (h_1, \ldots, h_n)$ be the corresponding dual bases of V^* and W^*, respectively. We have the easily verified relation

$$M_{B^*,(B')^*}({}^tT) = {}^t(M_{B',B}(T))$$

where the right hand side denotes the transpose of the matrix $M_{B',B}(T)$.

Exercise 1.3.4. (a). Prove that T is injective iff ${}^tT$ is surjective.

(b). Using Part (a), prove that the ranges of T and ${}^tT$ have the same dimension.

(c). Prove that the row space and the column space of an $m \times n$ matrix A over $\mathbb{F}$ have the same dimension. (This dimension is called the *rank* of A. The dimension of the range of a linear mapping T is called the *rank* of T.)

1.4 Determinant and Trace

The *determinant* of an $n \times n$ matrix A is defined in various ways. (Most definitions of the determinant in standard linear algebra texts are non-intuitive. Axler's book [Axl97] develops all of linear algebra without resorting to the determinant until the very end, where it "comes naturally." For our purposes, since we're after different game, it's sufficient to provide two of the equivalent expressions of the determinant and state its most salient features.)

If $A = (a_{ij})$, let us recall that its determinant $\det A$ is the homogeneous degree n polynomial in the entries of A given by

$$\det A = \sum_{\sigma} \epsilon(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)},$$

where the sum runs through all the permutations σ of $\{1, \ldots, n\}$, and $\epsilon(\sigma)$ denotes the sign of the permutation σ.

Let's also recall that the determinant $\det A$ can also be expanded using minors along a given row or column, as follows. For each pair of indices i, j in $\{1, \ldots, n\}$, let A_{ij} denote the $(n-1) \times (n-1)$ submatrix obtained from A by deleting the ith row and jth column. Then the *minor expansion* of $\det A$ along the ith row is

$$\det A = \sum_{k=1}^{n} (-1)^{i+k} a_{ik} \det A_{ik}$$

and that along the jth column is

$$\det A = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det A_{kj}.$$

If A has real entries, $\det A$ has a geometrical significance. Let $v_1, \ldots, v_n$ denote the columns of A. These are vectors in $\mathbb{R}^n$, and $|\det A|$ turns out to be the n-dimensional volume of the parallelepiped whose sides are $v_1, \ldots, v_n$. This can be proved by induction on n, and can be seen at least for 3×3 matrices A, since $\det A$ is the triple scalar product $(v_1 \times v_2) \cdot v_3$. (The proof for 2×2 determinants is even easier and just uses cross products.)

Exercise 1.4.1 (Graduate Exercise). Prove this geometrical fact about the n-dimensional determinant.

The determinant is multiplicative in that, if A and B are square matrices of the same size, then $\det AB = \det A \cdot \det B$. We also recall that a square matrix A is nonsingular if and only if $\det A \neq 0$. It then follows by multiplicativity that $\det A^{-1} = (\det A)^{-1}$.

Let V be a vector space over $\mathbb{F}$ and suppose that $T \in \mathcal{L}(V)$. We define the *determinant of* T to be $\det M(T)$, where $M(T)$ is the matrix of T with respect to any basis B of V. The value of the determinant $\det T$ is independent of the choice of basis: if B' is any other basis of V and S is the change of basis matrix from B to B', then by (1.7), the matrix of T with respect to B' is $S\,M(T)\,S^{-1}$, and hence

$$\begin{aligned} \det(S\,M(T)\,S^{-1}) &= \det S \, \det M(T) \, \det S^{-1} \\ &= \det S \, \det M(T) \, (\det S)^{-1} \\ &= \det M(T). \end{aligned}$$

Theorem 1.4.2. *Let $T \in \mathcal{L}(V)$. Then T is invertible if and only if $\det T \neq 0$.*

Proof. By Theorem 1.3.2, T is invertible $\iff$ $M(T)$ is nonsingular $\iff$ $\det M(T) \neq 0 \iff \det T \neq 0$. □

Another useful quantity associated to a linear operator $T \in \mathcal{L}(V)$ is its *trace*. If $A = (a_{ij})$ is an square $n \times n$ matrix, the *trace* of A is defined to be the sum of its diagonal entries: $\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}$. The trace satisfies the following easily verified property:

Proposition 1.4.3. *Let A and B be $n \times n$ matrices. Then $\operatorname{tr} AB = \operatorname{tr} BA$.*

Proof. Let $A = (a_{ij}$ and $B = (b_{ij})$. Then from the definition of matrix product, $AB = (c_{ij})$, where

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Likewise, $BA = (d_{ij})$, with

$$d_{ij} = \sum_{k=1}^{n} b_{ik} a_{kj}.$$

Hence

$$\operatorname{tr} AB = \sum_{i=1}^{n} c_{ii} = \sum_{i=1}^{n}\sum_{k=1}^{n} a_{ik}b_{ki} \tag{1.8}$$

whereas

$$\operatorname{tr}(BA) = \sum_{i=1}^{n} d_{ii} = \sum_{i=1}^{n}\sum_{k=1}^{n} b_{ik}a_{ki}.$$

If we interchange the indices i and k in the above sum, we see that it equals the sum in (1.8). □

The *trace* of a linear operator $T \in \mathcal{L}(V)$ is, by definition, the trace of the matrix $M(T)$, where $M(T)$ is the matrix of T with respect to any basis of V. Now the matrix of T with respect to any other basis of V is given by $S\,M(T)\,S^{-1}$ for some matrix S, and it follows from Proposition 1.4.3 above that $\operatorname{tr}(S\,M(T)\,S^{-1}) = \operatorname{tr}(M(T)\,S^{-1}\,S) = \operatorname{tr} M(T)$. Thus the trace of T is a well-defined scalar, depending only on T and not on the choice of basis of V.

1.5 Eigenvalues and Invariant Subspaces

A scalar $\lambda \in \mathbb{F}$ is called an *eigenvalue* of a linear operator $T \in \mathcal{L}(V)$ if there is a *nonzero* vector $v \in V$ such that $T(v) = \lambda v$. The vector v is called an *eigenvector* of T corresponding to λ. If $\lambda \in \mathbb{C}$, the subspace $\ker(T - \lambda I_V) = \{v \in V \mid T(v) = \lambda v\}$ is called the *eigenspace of T corresponding to λ*.

Proposition 1.5.1. *Suppose that $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. Then the following are equivalent:*

1. *λ is an eigenvalue of T*
2. $\ker(T - \lambda I_V) \neq \{0\}$
3. $\det(T - \lambda I_V) = 0$.

Proof. An easy exercise. □

The polynomial $\det(\lambda I_V - T)$ in the indeterminate λ, with coefficients in $\mathbb{F}$, is called the *characteristic polynomial* of T; its roots in $\mathbb{F}$ are precisely the eigenvalues of T.

Linear operators on real vector spaces do not necessarily have real eigenvalues. For instance, consider the operator $T \in \mathcal{L}(\mathbb{R}^2)$ given by $Tx = Ax$ $(x \in \mathbb{R}^2)$, where

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

Then the characteristic polynomial of T is $\det(\lambda I_2 - T) = \lambda^2 + 1$, which has no real roots. Thus T has no real eigenvalues.

On the other hand, a linear operator on a complex vector space has at least one eigenvalue.

Theorem 1.5.2. *Let V be a nonzero vector space over $\mathbb{C}$ and let $T \in \mathcal{L}(V)$. Then T has at least one eigenvalue.*

Proof. The characteristic polynomial $\det(\lambda I_V - T)$ is a polynomial in λ of degree $\dim V > 0$, and so by Gauss's Fundamental Theorem of Algebra, has at least one complex root, which is, by Proposition 1.5.1, an eigenvalue of T. □

An easy consequence of the Fundamental Theorem of algebra is that any polynomial $p(z)$ of degree n has n complex roots, counting multiplicities, and has a unique linear factorization $p(z) = c \cdot \prod_{j=1}^{k} (z - \lambda_j)^{m_j}$, where $\lambda_1, \ldots, \lambda_k$ are the distinct roots of $p(z)$ and $m_1, \ldots, m_k$ are their respective multiplicities, with $m_1 + \cdots + m_k = n$. Applying this to the characteristic polynomial of T, we get $\det(\lambda I_V - T) = \prod_{j=1}^{k} (\lambda - \lambda_j)^{m_j}$. Here $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of T and $m_1, \ldots, m_k$ are called their respective *multiplicities*.

It is often useful to study an operator $T \in \mathcal{L}(V)$ by examining its invariant subspaces. A subspace W is said to be *invariant under* T, or *T-invariant* if $T(w) \in W$ for all $w \in W$. Thus the restriction $T|_W$ of the map T to W belongs to $\mathcal{L}(W)$. If v is an eigenvector of T, then the one-dimensional subspace $\mathbb{F}v$ is obviously a T-invariant subspace.

Exercise 1.5.3. If two operators S and T in $\mathcal{L}(V)$ commute, then show that both the kernel $\ker S$ and the range $S(V)$ are T-invariant.

Since T commutes with $T - \lambda I_V$ for all $\lambda \in \mathbb{F}$, Exercise 1.5.3 implies, in particular, that each eigenspace $\ker(T - \lambda I_V)$ is T-invariant.

Suppose that W is a T-invariant subspace of V. Then T induces a well-defined map T' on the quotient space V/W given by $T'(v + W) = T(v) + W$. It is easy to check that $T' \in \mathcal{L}(V/W)$. We have the commutative diagram

$$\begin{array}{ccc} V & \xrightarrow{T} & V \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ V/W & \xrightarrow[T']{} & V/W \end{array} \tag{1.9}$$

which says that $T'\pi = \pi T$ (as maps from V to V/W).

1.6 Upper Triangular Matrices

We would like to find a basis of V with respect to which the matrix of a given operator T is "nice," in some sense. Ideally, we want the matrix of T to be diagonal, if this is possible, or at least to have as many 0's as possible, arranged in an orderly fashion.

A square matrix is called *upper triangular* if all the entries below the main diagonal are 0. Such a matrix can then be represented as follows:

$$\begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Note that the determinant of any upper triangular matrix is the product of the diagonal entries. Thus, if A is the matrix above, then $\det A = \lambda_1 \cdots \lambda_n$.

Proposition 1.6.1. *Suppose that $T \in \mathcal{L}(V)$. Then the following are equivalent:*

1. *There is a basis of V for which the matrix of T is upper triangular*
2. *There is a nested sequence of subspaces $0 \subsetneqq V_1 \subsetneqq \cdots \subsetneqq V_n = V$ each of which is invariant under T.*

The proof is obvious. Note that the condition (2) implies that the subspace V_j must have dimension j.

It is a very useful fact that if V is a complex vector space, any linear map $T \in \mathcal{L}(V)$ has an upper triangular representation:

Theorem 1.6.2. *Let V be a vector space over $\mathbb{C}$ and let $T \in \mathcal{L}(V)$. Then there is a basis of V for which the matrix of T is upper triangular.*

Proof. By induction on $\dim V$. If $\dim V = 1$, there is nothing to prove. So let's assume that $\dim V = n > 1$. Then by Theorem 1.5.2, T has an eigenvalue λ. Let v_1 be an eigenvector corresponding to λ, and let W be the one-dimensional subspace $\mathbb{C}v_1$. Since W is T-invariant, we can consider the induced linear map T' on the complex vector space V/W given by $T'(v + W) = T(v) + W$. Now $\dim(V/W) = n - 1$, so by the induction hypothesis, there is a basis $(v_2 + W, \ldots, v_n + W)$ of V/W for which the matrix of T' is upper triangular. Note that for each j, $j = 2, \ldots, n$, $T'(v_j + W)$ is a linear combination of $v_2 + W, \ldots, v_j + W$; hence $T(v_j)$ is a linear combination of $v_1, v_2, \ldots, v_j$.

It remains to prove that $(v_1, v_2, \ldots, v_n)$ is a linearly independent set (and so is a basis of V). Once we prove this, it is clear that the matrix of T with respect to this basis is upper triangular.

Suppose that $\sum_{i=1}^n c_j v_j = 0$. Then $(\sum_{i=1}^n c_i v_i) + W = W \implies \sum_{i=2}^n (c_i v_i + W) = W \implies c_2 = \cdots = c_n = 0$, by the linear independence of $(v_2+W, \ldots, v_n+W)$, so we end up with $c_1 v_1 = 0$, which clearly implies that $c_1 = 0$. □

Note: Suppose that $n \geq 2$. Since the choice of the v_j's in the proof is not necessarily unique, the upper triangular matrix in the theorem above is not necessarily unique. What is unique, from the characteristic polynomial of T, are the diagonal entries and their multiplicities.

Let $p(z) = a_m z^m + a_{m-1} z^{m-1} + \cdots + a_1 z + a_0$ be any polynomial in the variable z. If $T \in \mathcal{L}(V)$, we put $p(T) = a_m T^m + a_{m-1} T^{m-1} + \cdots + a_1 T + a_0 I_V$. Then $p(T) \in \mathcal{L}(V)$, and if M is the matrix of T with respect to some basis of V, then $p(M) = a_m M^m + a_{m-1} M^{m-1} + \cdots + a_1 M + a_0 I_n$ is the matrix of $p(T)$ with respect to this basis.

Theorem 1.6.3. *(The Cayley-Hamilton Theorem) Suppose that V is a vector space over $\mathbb{C}$ and that $T \in \mathcal{L}(V)$. Let $p(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$ be the characteristic polynomial of T. Then the linear map $p(T)$ is identically zero on V.*

Proof. By induction on $\dim V$. If $\dim V = 1$, then the conclusion is obvious: $T = \lambda_1 I_V$ for some $\lambda_1 \in \mathbb{C}$, the characteristic polynomial of T is $p(\lambda) = \lambda - \lambda_1$, and $p(T) = T - \lambda_1 I_V \equiv 0$.

So assume that $n > 1$ and that the theorem holds for all linear maps on all vector spaces of dimension $< n$. Suppose that $\dim V = n$ and that $T \in \mathcal{L}(V)$. Choose a basis $(v_1, \ldots, v_n)$ of V for which T has upper triangular matrix

$$\begin{pmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}. \tag{1.10}$$

Then the characteristic polynomial $\det(\lambda I_V - T) = \prod_{j=1}^n (\lambda - \lambda_j)$, and the diagonal entries $\lambda_1, \ldots, \lambda_n$ above are the eigenvalues of T. (Of course, these λ_j are not necessarily distinct.) Let $V_1 = \mathbb{C}v_1$. Now V_1 is a T-invariant subspace, the quotient space $V' = V/V_1$ is easily seen to have basis $v_2 + V_1, \ldots, v_n + V_1$, and the matrix of the induced map $T' : V/V_1 \to V/V_1,\ v + V_1 \mapsto T(v) + V_1$ with respect to this basis is

$$\begin{pmatrix} \lambda_2 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}. \tag{1.11}$$

The characteristic polynomial of T' is thus $(\lambda - \lambda_2) \cdots (\lambda - \lambda_n)$. Since V' has dimension $n - 1$, the induction hypothesis implies that $(T' - \lambda_2 I_{V'}) \cdots (T' - \lambda_n I_{V'}) \equiv 0$ on V'.

Thus for any $v \in V$, we have

$$\begin{aligned}(T' - \lambda_2 I_{V'}) \cdots (T' - \lambda_n I_{V'})(v + V_1) &= (T - \lambda_2 I_V) \cdots (T - \lambda_n I_V)(v) + V_1 \\ &= V_1\end{aligned}$$

and so

$$(T - \lambda_2 I_V) \cdots (T - \lambda_n I_V)(v) \in V_1$$

Therefore $(T - \lambda_2 I_V) \cdots (T - \lambda_n I_V)(v) = cv_1$ for some $c \in \mathbb{C}$. Hence

$$\begin{aligned}(T - \lambda_1 I_V)(T - \lambda_2 I_V) \cdots (T - \lambda_n I_V)(v) &= (T - \lambda_1 I_V)(cv_1) \\ &= 0.\end{aligned}$$

Since $v \in V$ is arbitrary, we have shown that $(T - \lambda_1 I_V) \cdots (T - \lambda_n I_V) \equiv 0$ on V, proving the conclusion for V and completing the induction step. □

Remark 1.6.4. The Cayley-Hamilton Theorem also holds for linear operators on real vector spaces. In order to see this, we note that it suffices to prove the Cayley-Hamilton Theorem for real square matrices, due to the correspondence between linear maps on V and square matrices of size $\dim V$. But then any real square matrix can be considered to be a complex matrix, which by the Cayley-Hamilton Theorem, satisfies its characteristic polynomial.

We can summarize the statement of the Cayley-Hamilton Theorem by saying that a linear operator on a vector space *satisfies* its characteristic polynomial.

Exercise 1.6.5. Suppose that $T \in \mathcal{L}(V)$ is invertible. Show that there exists a polynomial $p(z)$ such that $T^{-1} = p(T)$.

Exercise 1.6.6. Suppose that V is a n-dimensional complex vector space, and that $T \in \mathcal{L}(V)$ has eigenvalues 4, 5. Prove that

$$(T - 4I_V)^{n-1}(T - 5I_V)^{n-1} = 0.$$

1.7 Generalized Eigenspaces

The *spectrum* of a linear operator T on a vector space V over $\mathbb{F}$ is the collection of all eigenvalues of T in $\mathbb{F}$. We saw, from Theorem 1.5.2, that if V is complex, then any $T \in \mathcal{L}(V)$ has a nonempty spectrum.

In order to derive further nice properties about linear operators, it will be useful to have at least one eigenvalue, so in this section, we'll assume that V is a vector space over $\mathbb{C}$ and that $T \in \mathcal{L}(V)$.

Choose a basis of V for which T has upper triangular matrix

$$\begin{pmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}. \tag{1.12}$$

As we had already observed, the characteristic polynomial $\det(\lambda I_V - T)$ equals $\prod_{j=1}^{n}(\lambda - \lambda_j)$, and the diagonal entries $\lambda_1, \ldots, \lambda_n$ above are the eigenvalues of T, which are not necessarily distinct. Note that the number of times each distinct eigenvalue λ_j appears in the diagonal of the matrix above equals its multiplicity m_j as a root of the characteristic polynomial of T.

As mentioned previously, the upper triangular representation (1.12) of T is not unique, except for the appearance of the eigenvalues (with the correct multiplicities) along the diagonal. Our goal is to obtain a particular upper triangular representation of T which is unique and useful in the sense that much of the behavior of T is apparent upon cursory examination of the matrix.

With this goal in mind, we define the generalized eigenspaces of T as follows. For any complex scalar λ, the *generalized eigenspace* of T corresponding to λ is the set

$$\{v \in V \mid (T - \lambda I_V)^k v = 0 \text{ for some } k \in \mathbb{Z}^+\}. \tag{1.13}$$

Note that the eigenspace $\ker(T - \lambda I_V)$ is a subset of the generalized eigenspace (1.13) above. The following result shows that the generalized eigenspace is a subspace of V.

Theorem 1.7.1. *Fix $\lambda \in \mathbb{C}$. Then there is an integer $m \geq 0$ such that*

$$\{0\} \subsetneqq \ker(T - \lambda I_V) \subsetneqq \ker(T - \lambda I_V)^2 \subsetneqq \cdots \subsetneqq \ker(T - \lambda I_V)^m = \\ = \ker(T - \lambda I_V)^{m+1} = \ker(T - \lambda I_V)^{m+2} = \cdots$$

Proof. For simplicity, let $S = T - \lambda I_V$. If $v \in \ker S^k$, then $S^{k+1}(v) = S(S^k(v)) = 0$, so $v \in \ker S^{k+1}$, so it follows that $\ker S^k \subset \ker S^{k+1}$, and we get a nested chain of subspaces $\{0\} \subset \ker S \subset \ker S^2 \subset \cdots$

Since V is finite-dimensional, the chain stops increasing after some point. Let m be the smallest nonnegative integer such that $\ker S^m = \ker S^{m+1}$. Then $\ker S^{m+1} = \ker S^{m+2}$: if $v \in \ker S^{m+2}$, then $S(v) \in \ker S^{m+1} = \ker S^m$, so $S^m(Sv) = 0$, and so $v \in \ker S^{m+1}$.

Arguing in the same manner, we see that $\ker S^{m+2} = \ker S^{m+3}$, etc. □

Note that the m in Theorem 1.7.1 must be $\leq \dim V$.

It follows that the generalized eigenspace of T corresponding to λ equals the kernel $\ker(T - \lambda I_V)^{\dim V}$, which is a subspace of V.

Corollary 1.7.2. *Let $S \in \mathcal{L}(V)$. Then there is an integer m such that*

$$V \supsetneqq S(V) \supsetneqq S^2(V) \supsetneqq \cdots \supsetneqq S^m(V) = S^{m+1}(V) = S^{m+2}(V) = \cdots \tag{1.14}$$

Proof. This follows immediately from the proof of the preceding theorem, once we observe that $V \supset S(V) \supset S^2(V) \supset \cdots$, and that $\dim S^k(V) = \dim V - \dim \ker(S^k)$. □

Exercise 1.7.3. Show that for any $T \in \mathcal{L}(V)$, we have

$$V = (\ker T^n) \oplus T^n(V),$$

where $n = \dim V$.

As an additional application of the upper triangular representation (1.12) of T, we can determine the dimension of each generalized eigenspace.

Proposition 1.7.4. *Suppose that $T \in \mathcal{L}(V)$ has characteristic polynomial given by $\det(\lambda I_V - T) = \prod_{j=1}^{n}(\lambda - \lambda_j)$ ($\lambda_1, \ldots, \lambda_n$ not necessarily distinct). Then the generalized eigenspace of T corresponding to λ_j has dimension equal to the multiplicity of λ_j.*

Proof. Note that m_j is the number of times λ_j appears in the diagonal in the upper triangular representation of T.

We prove this by induction on $\dim V$. If $\dim V = 1$, then the conclusion is trivial. Let $n \geq 2$ and assume that the conclusion holds for all linear operators on all complex vector spaces of dimension $< n$. Suppose that $\dim V = n$ and that $T \in \mathcal{L}(V)$.

Choose a basis $(v_1, \ldots, v_n)$ of V for which the matrix of T is upper triangular, of the form (1.10). For each j, we let V_j denote the generalized eigenspace of T corresponding to λ_j. (We're not assuming that the λ_j are distinct. If $\lambda_j = \lambda_k$, then obviously V_j will be the same as V_k.)

Let $W = \mathbb{C}v_1$, and let T' be the induced linear map on $V' = V/W$: $T'(v+W) = T(v) + W$. Then, with respect to the basis $(v_2 + W, \ldots, v_n + W)$ of V/W, the matrix of T' is upper triangular and is given by (1.11).

Now the induction hypothesis says that the generalized eigenspace V_j' of T' corresponding to λ_j has dimension m_j if $\lambda_j \neq \lambda_1$ and has dimension $m_1 - 1$ if $\lambda_j = \lambda_1$.

We therefore consider the two cases $\lambda_j \neq \lambda_1$ and $\lambda_j = \lambda_1$ separately.

Let $\pi : V \to V/W$ be the quotient map. We first note that, for any j, $\pi(V_j) \subset V_j'$. Indeed, for each $v \in V_j$, we have $(T - \lambda_j I_V)^N v = 0$, for sufficiently large N, so $0 = \pi\left((T - \lambda_j I_V)^N v\right) = (T' - \lambda_j I_{V'})^N(\pi(v))$, whence $\pi(v) \in V_j'$. Thus π maps V_j into V_j'.

Now let us first consider the case $\lambda_j \neq \lambda_1$. We claim that π maps V_j isomorphically onto V_j'.

Suppose that $v \in V_j$ belongs to $\ker \pi$. Then $v = cv_1$ for some $c \in \mathbb{C}$. The condition $(T - \lambda_j I_V)^N v = 0$ (for some N) then implies that $c(\lambda_1 - \lambda_j)^N v_1 = 0$, so that $c = 0$, and so $v = 0$. Hence π maps V_j injectively into V_j'.

Next let us show that π maps V_j onto V_j'. Let $v' \in V_j'$, and write $v' = \pi(v)$ for some $v \in V$. By assumption $(T' - \lambda_j I_{V'})^N(\pi(v)) = 0$ for some N. This yields $\pi\left((T - \lambda_j I_V)^N v\right) = 0$, so $(T - \lambda_j I_V)^N v = cv_1$ for some $c \in \mathbb{C}$. Now let $u = v - c(\lambda_1 - \lambda_j)^{-N} v_1$. Then $(T - \lambda_j I_V)^N u = 0$, so $u \in V_j$, and clearly, $\pi(u) = \pi(v) = v'$. Thus $\pi(V_j) = V_j'$, and so we can conclude that $\dim V_j = m_j$.

Finally we consider the case $\lambda_j = \lambda_1$. We claim that π maps V_1 onto V_1' with kernel $\mathbb{C}v_1$. Since $\mathbb{C}v_1 = \ker \pi = V_1 \cap \ker \pi$, it suffices to prove that $\pi(V_1) = V_1'$. Let $v' \in V_1'$. We have $v' = \pi(v)$ for some $v \in V$. Now the condition $(T' - \lambda_1 I_{V'})^N v' = 0$ for some N implies that $(T - \lambda_1 I_V)^N v = av_1$ for some $a \in \mathbb{C}$. Hence $(T - \lambda_1 I_V)^{N+1} v = 0$, and thus $v \in V_1$. This shows that $v' \in \pi(V_1)$, and so $\pi(V_1) = V_1'$.

We conclude in this case that $\dim V_1 = \dim V_1' + 1 = m_1$. This completes the proof of Proposition 1.7.4. □

Exercise 1.7.5. Suppose that V is a complex vector space of dimension n and $T \in \mathcal{L}(V)$ such that

$$\ker T^{n-2} \subsetneqq \ker T^{n-1}.$$

Prove that T has at most two distinct eigenvalues.

We now reorder the eigenvalues of T, if necessary, so that we now assume that the *distinct* eigenvalues of T are $\lambda_1, \ldots, \lambda_k$. Let us again consider the characteristic polynomial

$$p(\lambda) = \det(\lambda I_V - T) \tag{1.15}$$

and its factorization

$$\prod_{j=1}^{k} (\lambda - \lambda_j)^{m_j} \tag{1.16}$$

Our objective now is to show that V is the direct sum of the generalized eigenspaces corresponding to each λ_j.

Lemma 1.7.6. *Let $p_1(z), \ldots, p_k(z)$ be nonzero polynomials with coefficients in $\mathbb{C}$ sharing no common factor of degree ≥ 1. Then there are polynomials $q_1(z), \ldots, q_k(z)$ such that $p_1(z)q_1(z) + \cdots + p_k(z)q_k(z) = 1$.*

Proof. (Optional: requires some ring theory.) Let $\mathbb{C}[z]$ be the ring of polynomials in z with complex coefficients. Since $\mathbb{C}[z]$ is a Euclidean ring, it is a principal ideal domain, and so the ideal $\mathbb{C}[z]p_1(z) + \cdots + \mathbb{C}[z]p_k(z)$ is principal: $\mathbb{C}[z]p_1(z) + \cdots + \mathbb{C}[z]p_k(z) = \mathbb{C}[z]r(z)$, for some nonzero polynomial $r(z)$. Clearly, $r(z)$ divides all the $p_j(z)$, so $r(z)$ must be a degree 0 polynomial; i.e., a nonzero constant. Thus $\mathbb{C}[z]p_1(z) + \cdots + \mathbb{C}[z]p_k(z) = \mathbb{C}[z]$; in particular $1 \in \mathbb{C}[z]p_1(z) + \cdots + \mathbb{C}[z]p_k(z)$. □

Theorem 1.7.7. *V is a direct sum of the generalized eigenspaces of the eigenvalues of T. More precisely,*

$$V = \bigoplus_{j=1}^{k} \ker(T - \lambda_j I_V)^{m_j}.$$

For each j, the generalized eigenspace corresponding to λ_j equals $\ker((T - \lambda_j I_V)^{m_j}$. *Thus* $\dim(\ker(T - \lambda_j I_V)^{m_j}) = m_j$.

Proof. We let $p(\lambda)$ be the characteristic polynomial (1.15) of T, factored as in (1.16).

Suppose first that T has just one eigenvalue λ_1. Then $p(\lambda) = (\lambda - \lambda_1)^n$, and by the Cayley-Hamilton Theorem, $(T - \lambda_1 I_V)^n = 0$, so $V = \ker(T - \lambda_1 I_V)^n$, proving the theorem.

Thus we can assume that T has more than one eigenvalue.

For each j, let

$$p_j(\lambda) = \frac{p(\lambda)}{(\lambda - \lambda_j)^{m_j}} = \prod_{l \neq j} (\lambda - \lambda_l)^{m_l}.$$

Then by Lemma 1.7.6, there exist complex polynomials $q_1(\lambda), \ldots, q_k(\lambda)$ such that $p_1(\lambda)q_1(\lambda) + \cdots + p_k(\lambda)q_k(\lambda) = 1$. Replacing λ by T, we have $p_1(T)q_1(T) + \cdots + p_k(T)q_k(T) = I_V$. (Strictly speaking, we're applying the well-defined algebra homomorphism $p(\lambda) \mapsto p(T)$ from $\mathbb{C}(\lambda)$ to $\mathcal{L}(V)$.)

For each j, let V_j be the image $V_j = p_j(T)q_j(T)(V)$. Then V_j is a subspace of V, and for each $v \in V$, we have

$$v = I_V v = p_1(T)q_1(T)(v) + \cdots + p_k(T)q_k(T)(v) \in V_1 + \cdots + V_k.$$

Thus $V = V_1 + \cdots + V_k$.

(*Note:* The subspaces V_j here have not yet been proven to be generalized eigenspaces: that comes next!)

Now by the Cayley-Hamilton Theorem,

$$\begin{aligned}(T - \lambda_j I_V)^{m_j} V_j &= (T - \lambda_j I_V)^{m_j} p_j(T)q_j(T)(V) \\ &= q_j(T)p(T)(V) \\ &= \{0\}.\end{aligned}$$

This shows that $V_j \subset \ker(T - \lambda_j I_V)^{m_j}$.

Note that each of the subspaces V_j is T-invariant, since

$$T(V_j) = Tp_j(T)q_j(T)(V) = p_j(T)q_j(T)T(V) \subset p_j(T)q_j(T)(V) = V_j.$$

Moreover, for $i \neq j$, the restriction of $T - \lambda_i I_V$ to V_j is invertible. For, if $w \in V_j$ such that $(T - \lambda_i I_V)w = 0$, then $T(w) = \lambda_i w$, so $0 = (T - \lambda_j I_V)^{m_j}(w) = (\lambda_i - \lambda_j)^{m_j} w$, which implies that $w = 0$.

Next we prove that the sum $V = V_1 + \cdots V_k$ is direct. If this were not the case, there would exist vectors $v_1 \in V_1, \ldots, v_k \in V_k$, not all zero, such that $v_1 + \cdots + v_k = 0$. Assume that $v_j \neq 0$. Then since $T - \lambda_i I_V$ is invertible on V_j for $i \neq j$, we see that $p_j(T)$ is invertible on V_j and is identically zero on V_i for all other i. Thus

$$\begin{aligned} 0 &= p_j(T)(v_1 + \cdots + v_k) \\ &= p_j(T)(v_1) + \cdots + p_j(T)(v_k) \\ &= p_j(T)v_j. \end{aligned}$$

The last expression above is nonzero because $p_j(T)$ is invertible on V_j and $v_j \neq 0$. This contradiction shows that $V = V_1 \oplus \cdots \oplus V_k$.

Note that $\dim V_j \leq m_j$, for all j by Proposition 1.7.4, since V_j is contained in the generalized eigenspace of T corresponding to λ_j. Since we have a direct sum $V = V_1 \oplus \cdots \oplus V_k$ and $\dim V = m_1 + \cdots + m_n$, we must in fact have $\dim V_j = m_j$, for all j.

It remains to prove that $V_j = \ker(T - \lambda_j I_V)^{m_j}$ for all j. We already know that V_j is a subspace of $\ker(T - \lambda I_V)^{m_j}$, which is itself a subspace of the generalized eigenspace of T corresponding to λ_j. This generalized eigenspace has dimension m_j, by Proposition 1.7.4. Since $\dim V_j = m_j$, these subspaces all coincide, and in particular $V_j = \ker(T - \lambda I_V)^{m_j}$.

The last assertion follows from what we have shown above. □

1.8 The Jordan Canonical Form

An operator $N \in \mathcal{L}(V)$ is said to be *nilpotent* if $N^m = 0$ for some positive integer m.

Exercise 1.8.1. Let V be a complex vector space. Prove that N is nilpotent if and only if the only eigenvalue of N is 0.

Suppose that V is a vector space over $\mathbb{F}$, and that $N \in \mathcal{L}(V)$ is nilpotent. Then there is a basis of V with respect to which the matrix of T has the form

$$\begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix}, \tag{1.17}$$

where all the entries on or below the main diagonal are 0. (We call such a matrix *strictly upper triangular*.) If $\mathbb{F} = \mathbb{C}$, then this follows immediately from Theorem

1.6.2 and Exercise 1.8.1. For $\mathbb{F}$ aribitrary, we just consider a basis of $\ker N$, then extend this to a basis of $\ker N^2$, then extend that to a basis of $\ker N^3$, etc. If we continue this procedure, we end up with a basis of $V = \ker N^m$, for sufficiently large m. It is clear that the matrix of N with respect to this basis is strictly upper triangular.

Our goal in this section is to represent a linear operator $T \in \mathcal{L}(V)$ by a matrix which has as many 0's as possible. As a first step, we see that an immediate application of Theorem 1.7.7 is the following result.

Proposition 1.8.2. *Let T be a linear operator on a complex vector space V, with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. Then there is a basis of V with respect to which the matrix of T has the following block diagonal form*

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_k \end{pmatrix}, \tag{1.18}$$

where each A_j is an upper triangular matrix of the form

$$\begin{pmatrix} \lambda_j & & * \\ & \ddots & \\ 0 & & \lambda_j \end{pmatrix}. \tag{1.19}$$

Proof. By Theorem 1.7.7, we have $V = \bigoplus_{j=1}^{k} V_j$, where $V_j = \ker(T - \lambda_j I_V)^{m_j}$. Thus the restriction $(T - \lambda_j I_V)|_{V_j}$ is nilpotent and there is a basis of V_j for which the matrix of this restriction is strictly upper triangular, of the form (1.17). Hence the matrix of $T|V_j$ with respect to this basis is of the form (1.19). If we combine the bases of the V_j so obtained, we get a basis of V with respect to which the matrix of T has the form (1.18). □

We now try to modify the bases of the V_j so as to simplify the block diagonal matrix (1.18) further. As already noted each V_j is T-invariant, and in addition, the restriction $(T - \lambda_j I_V)|_{V_j}$ is nilpotent. Thus we need to find a suitable basis of V_j with respect to which the matrix of $(T - \lambda_j I_V)$ is suitably nice.

Proposition 1.8.3. *Let N be a nilpotent operator on a nonzero vector space V over $\mathbb{F}$. Then there are vectors $v_1, \ldots, v_k$ in V and nonnegative integers $r_1, \ldots, r_k$ such that*

1. *the list $(v_1, Nv_1, \ldots, N^{r_1}v_1, v_2, Nv_2, \ldots, N^{r_2}v_2, \ldots, v_k, Nv_k, \ldots, N^{r_k}v_k)$ is a basis of V*
2. *$(N^{r_1}v_1, N^{r_2}v_2, \ldots, N^{r_k}v_k)$ is a basis of $\ker N$.*

Proof. The proof is by induction on $\dim V$. If $\dim V = 1$, then N must be the zero operator 0, so the proposition is trivially true.

Let $\dim V > 1$, suppose that the proposition holds for all nilpotent linear operators on all vector spaces of dimension $< \dim V$, and let N be a nilpotent linear operator on V. Since the range $N(V)$ is an N-invariant subspace of dimension $< \dim V$, we apply the induction hypothesis to obtain vectors $w_1, \ldots, w_m$ and nonnegative integers $s_1, \ldots, s_m$ such that

(a) $(w_1, Nw_1, \ldots, N^{s_1}w_1, \ldots, w_m, Nw_m, \ldots, N^{s_m}v_m)$ is a basis of $N(V)$; and

(b) $(N^{s_1}w_1, \ldots, N^{s_m}w_m)$ is a basis of $(\ker N) \cap N(V)$.

Pick vectors $v_1, \ldots, v_m$ such that $Nv_j = w_j$ for $1 \leq j \leq m$. Also, if necessary, pick additional vectors $v_{m+1}, \ldots, v_k$ so as to complete the list in (b) above to a basis of $\ker N$.

Put $r_1 = s_1 + 1, \ldots, r_m = s_m + 1$ and let $r_{m+1} = \cdots = r_k = 0$. We first claim that the list

$$(v_1, Nv_1, \ldots, N^{r_1}v_1, v_2, Nv_2, \ldots, N^{r_2}v_2, \ldots, v_m, Nv_m, \ldots, N^{r_m}v_m, v_{m+1}, \ldots, v_k) \tag{1.20}$$

is linearly independent. Indeed, given the relation

$$\sum_{j=1}^{m}\sum_{l=0}^{r_j} a_{lj}N^l v_j \;+\; \sum_{t=m+1}^{k} b_t v_t \;=\; 0, \tag{1.21}$$

we apply the operator N to both sides to obtain

$$\begin{aligned} 0 &= \sum_{j=1}^{m}\sum_{l=0}^{r_j} a_{lj}N^{l+1}v_j \\ &= \sum_{j=1}^{m}\sum_{l=0}^{r_j} a_{lj}N^{l}w_j \\ &= \sum_{j=1}^{m}\sum_{l=0}^{s_j} a_{lj}N^{l}w_j. \end{aligned}$$

It follows by condition (a) above that $a_{lj} = 0$ for all $0 \leq l \leq s_j$; $1 \leq j \leq m$; that is, all the coefficients a_{lj} in the last sum above vanish. From (1.21), this leaves us with the relation

$$\begin{aligned} 0 &= a_{r_1 1}N^{r_1}v_1 + \cdots + a_{r_m m}N^{r_m}v_m + b_{m+1}v_{m+1} + \cdots + b_k v_k \\ &= a_{r_1 1}N^{s_1}w_1 + \cdots + a_{r_m m}N^{s_m}w_m + b_{m+1}v_{m+1} + \cdots + b_k v_k. \end{aligned}$$

But by condition b and the choice of $v_{m+1}, \ldots, v + k$, we see that all the coefficients above also vanish. It follows that the list (1.20) - which coincides with the list in conclusion (1) of the proposition - is linearly independent.

The list (1.20) is also a basis, since by the induction hypothesis (a), $\dim N(V) = s_1 + \cdots + s_m + m$, and $\dim(\ker N) = k$, so $\dim V = (\sum_{i=1}^m s_i) + m + k = (\sum_{i=1}^m r_i) + k$, which equals the number of vectors in (1.20). Condition (2) in the statement of the proposition is satisfied by construction, since $\ker N$ has basis $(N^{s_1} w_1, \ldots, N^{s_m} w_m, v_{m+1}, \ldots, v_k) = (N^{r_1} v_1, \ldots, N^{r_m} v_m, v_{m+1}, \ldots, v_k)$. □

Remark 1.8.4. The numbers k and $r_1, \ldots, r_k$ in Proposition 1.8.3 are unique in the following sense. Let us, without loss of generality, arrange the basis in Proposition 1.8.3 such that $r_1 \geq r_2 \geq \cdots \geq r_k$. Now suppose that

$$(u_1, \ldots, N^{l_1} u_1, u_2, \ldots, N^{l_2} u_2, \cdots, u_s, \cdots, N^{l_s} u_s)$$

is another basis of V satisfying the conclusions of the proposition, with $l_1 \geq \cdots \geq l_s$. Then $s = k$ and $l_1 = r_1, \ldots, l_k = r_k$. This can be proved by a simple induction argument, going from the range $N(V)$ to V. We leave the details to the reader.

Suppose $T \in \mathcal{L}(V)$. We call a basis of V a *Jordan basis for* T if the matrix of T with respect to this basis has block diagonal form

$$\begin{pmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_m \end{pmatrix} \tag{1.22}$$

where each diagonal block A_j has the form

$$\begin{pmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{pmatrix}. \tag{1.23}$$

The matrix (1.22) is then called the *Jordan canonical matrix* of T. The blocks A_j are called *Jordan blocks*.

Theorem 1.8.5. *Let V be a nonzero complex vector space and let $T \in \mathcal{L}(V)$. Then V has a Jordan basis for T.*

Proof. Assume that T has characteristic polynomial $p(\lambda) = \prod_{j=1}^k (\lambda - \lambda_j)^{m_j}$, where $\lambda_1, \cdots, \lambda_k$ are the distinct eigenvalues of T. By Theorem 1.7.7, V is the direct sum

$$V = \bigoplus_{j=1}^k V_j,$$

where V_j is the generalized eigenspace $\ker(T - \lambda_j I_V)^{m_j}$. V_j is T-invariant and of course the restriction $N_j := (T - \lambda_j I_V)|_{V_j} = T|_{V_j} - \lambda_j I_{V_j}$ is nilpotent. But

then we can apply Proposition 1.8.3 to N_j to obtain a basis of V_j for which the matrix of N_j has block form

$$\begin{pmatrix} R_1 & & & 0 \\ & R_2 & & \\ & & \ddots & \\ 0 & & & R_l \end{pmatrix} \tag{1.24}$$

where each diagonal block R_i is of the form

$$\begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix}. \tag{1.25}$$

Each block R_i above corresponds to the list $(N_j^{r_i} v_i, \ldots, N_j v_i, v_i)$ corresponding to the vector v_i in the basis given in Part (1) of Proposition 1.8.3. The linear span of $(N_j^{r_i} v_i, \ldots, N_j v_i, v_i)$ is invariant under $T|_{V_j} = N_j + \lambda_j I_{V_j}$, and on this linear span, the matrix of $T|_{V_j}$ is of the form 1.23.

Putting these bases together, we obtain a Jordan basis for T. □

Corollary 1.8.6. *Let A be an $n \times n$ matrix with complex entries. Then there is an $n \times n$ matrix S such that SAS^{-1} is of the form (1.22).*

Remark 1.8.7. Since the generalized eigenspace corresponding to λ_j has dimension m_j, the multiplicity of λ_j, the size of the collection of blocks corresponding to λ_j in the Jordan canonical form (1.22) is unique. Then, by Remark 1.8.4, for each λ_j, the number of Jordan blocks and their respective sizes is also unique.

1.9 The Jordan-Chevalley Decomposition

Suppose that V is a vector space of dimension n over $\mathbb{F}$ and that $T \in \mathcal{L}(V)$. Since $\dim(\mathcal{L}(V)) = n^2$, the operators $I_V, T, T^2, \ldots T^{n^2}$ are linearly dependent in $\mathcal{L}(V)$. We thus have a relation

$$Q(T) := a_0 I_V + a_1 T + a_2 T^2 + \cdots + a_{n^2} T^{n^2} = 0 \tag{1.26}$$

such that not all coefficients a_i are 0. One such relation, of course, is $p(T) = 0$, where $p(\lambda)$ is the characteristic polynomial of T.

A *monic* polynomial $p(z)$ is a polynomial in z whose highest degree coefficient is 1. Thus we may write $p(z) = z^m + a_{m-1} z^{m-1} + \cdots + a_1 z + a_0$.

Proposition 1.9.1. *Let $T \in \mathcal{L}(V)$, and let $p(z)$ be a monic polynomial of smallest positive degree such that $p(T) = 0$. If $s(z)$ is any polynomial such that $s(T) = 0$, then $p(z)$ divides $s(z)$.*

Proof. By the Euclidean algorithm (*i.e.* long division), we have

$$s(z) = q(z)p(z) + r(z),$$

where $q(z)$ and $r(z)$ are polynomials with $\deg r(z) < \deg p(z)$. Replacing z by T in the above we obtain

$$\begin{aligned} & s(T) = q(T)p(T) + r(T) \\ \implies \qquad & 0 = r(T), \end{aligned}$$

which by the minimality of p implies that $r(z) = 0$. □

It follows that there is only one such polynomial $p(z)$. We call this polynomial the *minimal polynomial* of $T \in \mathcal{L}(V)$, and denote it by $P_{\min}(z)$. From (1.26), any minimal polynomial of T has degree $\leq n^2$, and better yet, by the Cayley-Hamilton Theorem, it must have degree $\leq n$.

Corollary 1.9.2. *The minimal polynomial of T divides the characteristic polynomial of T.*

Proposition 1.9.3. *Let V be a complex vector space and $T \in \mathcal{L}(V)$. Then the roots of the minimal polynomial of T are precisely its eigenvalues.*

Proof. Let λ be an eigenvalue of T, and let v eigenvector corresponding to λ. Then since $v \neq 0$,

$$0 = P_{\min}(T)(v) = P_{\min}(\lambda)v \implies P_{\min}(\lambda) = 0.$$

Conversely, suppose that λ is a root of $P_{\min}$. Then by Proposition 1.9.2, λ is a root of the characteristic polynomial of T, whence λ is an eigenvalue of T. □

If the eigenvalues of T all have multiplicity 1; that is, if the characteristic polynomial of T is of the form $\chi(z) = \prod_{i=1}^{n}(z - \lambda_i)$, with the λ_i distinct, then, by Corollary 1.9.2 and Proposition 1.9.3, the characteristic polynomial coincides with the minimal polynomial. On the other hand, if T is scalar multiplication, $T = \lambda I_V$, then the minimal polynomial of T is $z - \lambda$, whereas its characteristic polynomial is $(z - \lambda)^n$.

Exercise 1.9.4. Let T be the linear operator on a 6-dimensional complex vector space whose Jordan matrix is

$$\begin{pmatrix} \lambda_1 & 1 & & & & 0 \\ & \lambda_1 & 1 & & \ddots & \\ & & \lambda_1 & & & \\ & \ddots & & \lambda_2 & 1 & \\ & & & & \lambda_2 & \\ 0 & & & & & \lambda_2 \end{pmatrix}. \tag{1.27}$$

Find the minimal polynomial of T. For any $T \in \mathcal{L}(V)$, formulate a theorem stating what the minimal polynomial is in terms of its Jordan matrix. Then prove your theorem.

Exercise 1.9.5. Suppose that V is a vector space over $\mathbb{C}$, and that $T \in \mathcal{L}(V)$ has characteristic polynomial $\chi(z) = \prod_{i=1}^{k}(z-\lambda_i)^{m_i}$ and minimal polynomial $P_{\min}(z) = \prod_{i=1}^{k}(z-\lambda_i)^{r_i}$. Suppose that V_i is the generalized eigenspace corresponding to λ_i. Prove that

$$r_i = \min\{r \,|\, (T-\lambda_i I_V)^r|_{V_i} = 0\}.$$

Exercise 1.9.6. Suppose that $T \in \mathcal{L}(V)$ and $v \in V$. Prove that there is a unique monic polynomial $s(z)$ of lowest degree such that $s(T)\,v = 0$. Then prove that $s(z)$ divides the minimal polynomial of T.

Exercise 1.9.7. Give an example of a linear operator on $\mathbb{C}^4$ whose characteristic polynomial is $z(z-1)^2(z-2)$ and whose minimal polynomial is $z(z-1)(z-2)$.

Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T. The operator T is said to be *semisimple* if the minimal polynomial of T is $(z-\lambda_1)\cdots(z-\lambda_k)$.

Proposition 1.9.8. *T is semisimple if and only if it is diagonalizable; that is, there is a basis of V consisting of eigenvectors of T.*

Proof. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T.

Suppose first that T is diagonalizable. Then there is a basis of V consisting of eigenvectors of T. Let $V_1 (= \ker(T-\lambda_1 I_V))$ be the eigenspace of T corresponding to λ_1, V_2 the eigenspace corresponding to λ_2, etc. Then we must have $V = V_1 \oplus \cdots \oplus V_k$. For each j, the restriction $(T-\lambda_j I_V)|_{V_j}$ is obviously identically 0. Hence $(T-\lambda_1 I_V)\cdots(T-\lambda_k I_V) \equiv 0$. so the minimal polynomial of T is $(z-\lambda_1)\cdots(z-\lambda_k)$, and T is semisimple.

Next, we assume that T is semisimple, and try to prove that T is diagonalizable. If V_j is the generalized eigenspace corresponding to λ_j, we have the direct decomposition

$$V = \bigoplus_{j=1}^{k} V_j$$

V_j is invariant under $T - \lambda_i I_V$, and if $i \neq j$, the restriction $(T - \lambda_i I_V)|V_j$ is invertible, for if $v \in V_j$ satisfies $(T - \lambda_i I_V)(v) = 0$, then $v \in V_j \cap V_i = \{0\}$. Thus the restriction of $\prod_{i \neq j}(T - \lambda_i I_V)$ to V_j is invertible. Since $\prod_{i=1}^{k}(T - \lambda_i I_V) = 0$, we see that $T - \lambda_j I_V \equiv 0$ on V_j, so T is just scalar multiplication by λ_j on V_j. □

Note: Most authors define a semisimple linear operator as one which is diagonalizable. But our definition allows us more flexibility, as the following proposition shows.

Proposition 1.9.9. *Suppose that $T \in \mathcal{L}(V)$ is semisimple, and that W is a subspace of V invariant under T. Then the restriction $T|_W$ is semisimple.*

Proof. Let $P_{\min}(z)$ denote the minimal polynomial of T. We have $P_{\min}(T|_W) = P_{\min}(T)|_W = 0$, so the minimal polynomial of $T|_W$ divides $P_{\min}(z)$. This minimal poynomial must then be of the form $\prod_{i \in J}(z - \lambda_i)$, where J is a subset of the set of eigenvalues of T. □

Consider now the Jordan matrix of the linear operator T on $\mathbb{C}^6$ in Exercise 1.9.4. Let S be the semisimple linear operator on C^6 whose matrix with respect to this Jordan basis of T is

$$\begin{pmatrix} \lambda_1 & & & & & 0 \\ & \lambda_1 & & & \ddots & \\ & & \lambda_1 & & & \\ & \ddots & & \lambda_2 & & \\ & & & & \lambda_2 & \\ 0 & & & & & \lambda_2 \end{pmatrix}$$

and let N be the nilpotent operator with matrix

$$\begin{pmatrix} 0 & 1 & & & & 0 \\ & 0 & 1 & & \ddots & \\ & & 0 & & & \\ & \ddots & & 0 & 1 & \\ & & & & 0 & \\ 0 & & & & & 0 \end{pmatrix}.$$

Then $T = S + N$, and it is easy to check that S and N commute.

Using its Jordan matrix, it is easy to see that, in fact, any linear operator T on a complex vector space V has a *Jordan-Chevalley decomposition* $T = S + N$, where S is semisimple and N is nilpotent, and S and N commute. We will now obtain this decomposition abstractly, without the benefit of the Jordan matrix. This approach will show that S and N satisfy a few additional properties, given in Theorem 1.9.14 below.

Lemma 1.9.10. *(The Chinese Remainder Theorem) Suppose that $p_1(z), \ldots, p_m(z)$ are nonconstant polynomials which are pairwise relatively prime. If $r_1(z), \ldots, r_m(z)$ are any polynomials, then then is a polynomial $P(z)$ such that $P(z) \equiv r_j(z) \mod p_j(z)$, for all j.*

Proof. For each j, let $Q_j(z) = \prod_{i \neq j} p_i(z)$. Then there exist polynomials $A_j(z)$ and $B_j(z)$ such that $A_j(z)p_j(z) + B_j(z)Q_j(z) = 1$. Now put

$$P(z) = \sum_{i=1}^{m} r_i(z)B_i(z)Q_i(z).$$

For $i \neq j$, $p_j(z)$ divides $Q_i(z)$, so

$$\begin{aligned} P(z) &\equiv \sum_{i=1}^{m} r_i(z)B_i(z)Q_i(z) && \mod p_j(z) \\ &\equiv r_j(z)B_j(z)Q_j(z) && \mod p_j(z) \\ &\equiv r_j(z)\left(1 - A_j(z)p_j(z)\right) && \mod p_j(z) \\ &\equiv r_j(z) && \mod p_j(z). \end{aligned}$$

□

The Chinese Remainder Theorem, properly formulated, holds for all principal ideal domains, and in fact for all commutative rings with identity. The original form of the theorem, as it pertains to the integers, appeared in a third-century AD book by the mathematician Sun Tzu ("Master Sun," also Sunzi) (孙子) [Dau85] (not the Sun Tzu who wrote *The Art of War*).

Proposition 1.9.11. *Suppose that S_1 and S_2 are two diagonalizable linear operators on V. Then $S_1S_2 = S_2S_1$ if and only if S_1 and S_2 are simultaneously diagonalizable; that is, if and only if there is a basis of V for which the matrices of S_1 and S_2 are both diagonal.*

Proof. Since diagonal matrices of the same size commute, it is clear that if S_1 and S_2 are simultaneously diagonalizable, then they commute.

Conversely, let us assume that S_1 and S_2 are diagonalizable linear operators such that $S_1S_2 = S_2S_1$. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of S_1, with respective eigenspaces $V_1, \ldots, V_k$. Then $V = \bigoplus_{i=1}^{k} V_i$. Since S_1 and S_2 commute, each eigenspace V_i is invariant under S_2. Then by Lemma 1.9.9, the restriction $S_2|_{V_i}$ is diagonalizable. Choose a basis of V_i for which the matrix of $S_2|_{V_i}$ is diagonal. Then, combining the bases of the V_i, we obtain a basis of V for which the matrices of S_1 and S_2 are both diagonal. □

In particular, this proposition says that $S_1 + S_2$ must be semisimple!

Exercise 1.9.12. Show that if $\{S_1, \ldots, S_m\}$ are pairwise commuting semisimple elements of $\mathcal{L}(V)$, then there exists a basis of V for which the matrices of $\{S_1, \ldots, S_m\}$ are all diagonal.

Lemma 1.9.13. *Let N_1 and N_2 be commuting nilpotent linear operators on a vector space V. Then $N_1 + N_2$ is nilpotent.*

Proof. Assume that $N_1^{m_1} = 0$ and $N_2^{m_2} = 0$. Since N_1 and N_2 commute, we can apply the binomial theorem to obtain, for any $m \in \mathbb{N}$,

$$(N_1 + N_2)^m = \sum_{k=0}^{m} \binom{m}{k} N_1^k N_2^{m-k}.$$

It follows immediately that $(N_1 + N_2)^{m_1+m_2} = 0$. □

Theorem 1.9.14. *(The Jordan-Chevalley Decomposition) Let V be a complex vector space, and let $T \in \mathcal{L}(V)$. Then there exists a polynomial $p(z)$ such that if $q(z) = z - p(z)$, the following properties hold:*

1. *$S := p(T)$ is semisimple and $N := q(T)$ is nilpotent;*
2. *Any linear operator which commutes with T must commute with both S and N;*
3. *If S' and N' are commuting semisimple and nilpotent operators, respectively, such that $T = S' + N'$, the $S' = S$ and $N' = N$; and*
4. *If $A \subset B \subset V$ are subspaces, and if $T : B \to A$, then so do S and N.*

Proof. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T, and assume that the minimal polynomial of T is $P_{\min}(z) = \prod_{i=1}^{k} (z - \lambda_i)^{r_i}$. Now according to the Chinese Remainder Theorem, there exists a polynomial $p(z)$ such that, for each i,

$$p(z) \equiv \lambda_i \mod (z - \lambda_i)^{r_i} \tag{1.28}$$

$$p(z) \equiv 0 \mod z. \tag{1.29}$$

(In case one of the λ_i's equals 0, then (1.28) implies (1.29), so the second condition above is superfluous. Condition (1.29) is really only needed to prove the technical conclusion (4) above.)

Let V_i be the generalized eigenspace of T corresponding to λ_i. Then we have the direct decomposition

$$V = \bigoplus_{i=1}^{k} V_i.$$

Each V_i is invariant under T, hence is invariant under $p(T)$. Now by Exercise 1.9.5, $(T - \lambda_i I_V)^{r_i}$ vanishes on V_i, so by the relation (1.28) above, we see that

the operator $p(T) - \lambda_i I_V$ is identically 0 on V_i. Thus $p(T)v = \lambda_i v$ for all $v \in V_i$, and it follows that $p(T)$ is semisimple on V.

For any $v \in V_i$, we also have $q(T)v = (T - p(T))(v) = (T - \lambda_i I_V)v$, so (since $(T - \lambda_i I_V)^{r_i} = 0$ on V_i), we see that $q(T)$ is nilpotent on V_i, with $q(T)^{r_i} \equiv 0$ on V_i. Putting $R = \max_{1 \leq i \leq k} r_i$, we have $q(T)^R = 0$ on V, so $q(T)$ is a nilpotent operator.

Since $S = p(T)$ and $N = q(T)$ are polynomials in T, they commute, and in addition, any linear operator which commutes with T must commute with S and N. By 1.29, $p(z)$ and $q(z)$ have constant term equal to 0, so clearly S and N satisfy statement (4) above.

The only thing left to prove is the uniqueness statement (3). So let S' and N' be commuting semisimple and nilpotent linear operators, respectively, on V, such that $T = S' + N'$. Then S' and N' commute with T, and so must commute with both S and N. We then have

$$S - S' = N' - N.$$

Since S and S' are commuting semisimple operators, the left hand side above is a semisimple operator by Lemma 1.9.11. On the other hand, since N and N' commute, the right hand side above is a nilpotent operator, by Lemma 1.9.13. The only eigenvalue of $S - S'$ is therefore 0, whence $S - S' = 0$. Therefore, $N' - N = 0$. □

1.10 Symmetric Bilinear Forms

Let V be a vector space over $\mathbb{F}$. A *bilinear form* on V is a map

$$\begin{aligned} \langle\ ,\ \rangle : V \times V &\to \mathbb{F} \\ (v, w) &\mapsto \langle v, w \rangle. \end{aligned} \tag{1.30}$$

which is linear in each of its two arguments:

$$\begin{aligned} \langle \alpha v + \beta v', w \rangle &= \alpha \langle v, w \rangle + \beta \langle v', w \rangle \\ \langle v, \alpha w + \beta w' \rangle &= \alpha \langle v, w \rangle + \beta \langle v, w' \rangle, \end{aligned}$$

for all $v, v', w, w' \in V$ and all $\alpha, \beta \in \mathbb{F}$.

Example 1.10.1. The dot product on $\mathbb{R}^n$ is a bilinear form. More generally, an inner product on a real vector space V is a bilinear form.

Example 1.10.2. Let A be any $n \times n$ matrix over $\mathbb{F}$. Using the matrix A, we can define a bilinear form on $\mathbb{F}^n$ by putting

$$\langle x, y \rangle = {}^t x A x \qquad \text{for all } x, y \in \mathbb{F}^n$$

As a special case, when $A = I_n$ and $\mathbb{F} = \mathbb{R}$, we obtain the dot product on $\mathbb{R}^n$.

Let $\langle\ ,\ \rangle$ be a bilinear form on V. Fix a basis $B = (v_1, \ldots, v_n)$ of V. The *matrix of* $\langle\ ,\ \rangle$ *with respect to* B is the $n \times n$ matrix A whose (i,j) entry is $a_{ij} = \langle v_i, v_j \rangle$. This matrix A completely determines the bilinear form, since each vector in V is a unique linear combination of the basis vectors in B:

$$v = \sum_{i=1}^{n} x_i v_i \text{ and } w = \sum_{j=1}^{n} y_j v_j \implies \langle v, w \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i a_{ij} y_j \tag{1.31}$$

Given the basis B of V, we have a *coordinate map* $[\,\cdot\,]_B$ *from* V *onto* $\mathbb{F}^n$ *with respect to* B: namely, if $v \in V$, then

$$[v]_B = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \iff v = \sum_{i=1}^{n} x_i v_i.$$

The coordinate map $v \mapsto [v]_B$ is a linear isomorphism from V onto $\mathbb{F}^n$.

Exercise 1.10.3. If $T \in \mathcal{L}(V)$ with matrix $M_{B,B}(T)$ with respect to B, show that $[Tv]_B = M_{B,B}(T)[v]_B$, for all $v \in V$.

Now again let $\langle\ ,\ \rangle$ be a bilinear form on V, and let A be its matrix with respect to the basis B of V. Let $v = \sum_{i=1}^{n} x_i v_i$ and $w = \sum_{j=1}^{n} y_j v_j$ be vectors in V. Then by (1.30), we have

$$\begin{aligned} \langle v, w \rangle &= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i a_{ij} y_j \\ &= \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} a_{11} & & a_{1n} \\ & \cdots & \\ a_{n1} & & a_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\ &= {}^t[v]_B \, A \, [w]_B \end{aligned} \tag{1.32}$$

Thus Example 1.10.2 essentially gives us *all* bilinear forms on V, once we've fixed a basis B of V.

A bilinear form $\langle\ ,\ \rangle$ is called *nondegenerate* if, whenever v is a nonzero vector in V, there is a $w \in V$ such that $\langle v, w \rangle \neq 0$. The choice of the vector w, which is necessarily nonzero, depends on the vector v.

Exercise 1.10.4. Show that $\langle\ ,\ \rangle$ is nondegenerate if and only if, whenever w is a nonzero vector in V, there is a vector $v \in V$ such that $\langle v, w \rangle \neq 0$.

Theorem 1.10.5. *Let* $\langle\ ,\ \rangle$ *be a bilinear form on* V*, and let* A *be its matrix with respect to a given basis* B *of* V*. Then* $\langle\ ,\ \rangle$ *is nondegenerate if and only if* A *is nonsingular.*

Proof. Suppose that A is nonsingular. Let v be a nonzero vector in V. Then $[v]_B$ is a nonzero vector in $\mathbb{F}^n$. Since A is nonsingular, its rows are linearly independent, so ${}^t[v]_B A$ is a nonzero row matrix. Hence there exists an element $y \in \mathbb{F}^n$ such that ${}^t[v]_B A\, y \neq 0$. If we let $w \in V$ be the vector such that $[w]_B = y$, then according to (1.32), we have $\langle v, w\rangle \neq 0$. Thus $\langle\ ,\ \rangle$ is nondegenerate.

Suppose next that A is singular. Then its rows are linearly dependent, so there is an $x \neq 0$ in $\mathbb{F}^n$ such that ${}^t x\, A = 0$. Let v be the vector in V such that $[v]_B = x$. Then according to (1.32), we get $\langle v, w\rangle = 0$ for all $w \in V$. This shows that $\langle\ ,\ \rangle$ is not nondegenerate; i.e., is *degenerate*. □

Example 1.10.6. Let

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

Then the bilinear form on $\mathbb{R}^2$ given by $(x, y) = {}^t x\, A\, y$ is nondegenerate.

Example 1.10.7. Let J_n be the $2n \times 2n$ matrix which is given in block form as

$$J_n = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}, \tag{1.33}$$

where the "0" in the matrix above refers to the zero $n \times n$ matrix. Note that J_n is nonsingular, with inverse $J_n^{-1} = -J_n$. J_n gives rise to a nondegenerate symmetric bilinear form on $\mathbb{F}^{2n}$ given by

$$\langle x, y\rangle = {}^t x\, J_n\, y \qquad \text{for all } x, y \in \mathbb{F}^{2n}$$

When $\mathbb{F} = \mathbb{R}$, we call this form the *standard symplectic form on* $\mathbb{R}^{2n}$.

Let A be a square matrix with entries in $\mathbb{F}$. A is said to be *symmetric* if ${}^t A = A$. If $A = (a_{ij})$, this is equivalent to the condition that $a_{ij} = a_{ji}$ for all i and j. A is called *skew-symmetric* if ${}^t A = -A$. This is equivalent to the condition $a_{ij} = -a_{ji}$ for all i, j. Note that the diagonal entries of a skew-symmetric matrix are all 0.

A *symmetric bilinear form* on a vector space V is a bilinear form $\langle\ ,\ \rangle$ on V such that $\langle v, w\rangle = \langle w, v\rangle$ for all v, $w \in V$. The dot product on $\mathbb{R}^n$, or more generally, any inner product on a real vector space, is an example of a (nondegenerate) symmetric bilinear form.

Again fix a basis $B = (v_1, \ldots, v_n)$ of V, and let A be the matrix of $\langle\ ,\ \rangle$ with respect to B. If $\langle\ ,\ \rangle$ is symmetric, then A is a symmetric $n \times n$ matrix: $a_{ij} = \langle v_i, v_j\rangle = \langle v_j, v_i\rangle = a_{ji}$.

Conversely, it is an easy calculation using (1.31) to show that if A is a symmetric matrix, then $\langle\ ,\ \rangle$ is a symmetric bilinear form on V.

Suppose that $\langle\ ,\ \rangle$ is a symmetric bilinear form on a vector space V over $\mathbb{F}$. For any subspace W of V, the *orthogonal complement of* W is the set

$$W^\perp = \{v \in V \mid \langle v, w\rangle = 0 \text{ for all } w \in W\}.$$

It is easy to see that $W^\perp$ is a subspace of V. Note that $V^\perp = \{v \in V \,|\, \langle v, v' \rangle = 0 \text{ for all } v' \in V\}$, so $\langle\ ,\ \rangle$ is nondegenerate if and only if $V^\perp = \{0\}$.

For any $v \in V$, we let f_v be the linear functional on V given by $f_v(v') = \langle v, v' \rangle$, for all $v' \in V$.

Proposition 1.10.8. *Suppose that $\langle\ ,\ \rangle$ is nondegenerate. If W is a subspace of V, then the map $v \mapsto f_v|_W$ is a linear map of V onto the dual space W^*, with kernel $W^\perp$.*

Proof. The map $f : V \to V^*$ given by $v \mapsto f_v$ is easily seen to be linear. Its kernel is $V^\perp = \{0\}$, since $\langle\ ,\ \rangle$ is nondegenerate. Since $\dim V = \dim V^*$, f is onto, and we conclude that any element of V^* is of the form f_v, for a unique $v \in V$.

Next we prove that any linear functional on W can be extended to a linear functional on V. To be precise, suppose that $g \in W^*$. Choose any subspace U of V complementary to W, so that $V = W \oplus U$. Then define the function G on V by $G(w + u) = g(w)$ for all $w \in W$ and all $u \in U$. G is a well-defined linear functional on V such that $G|_W = g$.

The restriction map $f \mapsto f|W$ is a linear map from V^* to W^*, and the above shows that it is surjective. Since the map $v \mapsto f_v$ is a linear bijection from V onto V^*, we see that the composition $v \mapsto f_v|_W$ is a surjective linear map from V onto W^*. The kernel of this map is clearly $W^\perp$. □

This proposition implies that

$$\dim W^\perp = \dim V - \dim W^* = \dim V - \dim W. \tag{1.34}$$

1.11 Inner Products and Adjoints

Let V be a vector space over $\mathbb{F}$, where as usual $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$. Recall that an *inner product* on V is a map $\langle\ ,\ \rangle : V \times V \to \mathbb{F}$ such that

(i) $\langle au + bv, w \rangle = a \langle u, w \rangle + b \langle v, w \rangle$ (Linearity in the First Argument)

(ii) $\langle w, v \rangle = \overline{\langle v, w \rangle}$ (Conjugate Symmetry)

(iii) $\langle v, v \rangle \geq 0$ (Positivity)

(iv) $\langle v, v \rangle = 0$ only if $v = 0$ (Definiteness)

for all u, v, $w \in V$ and all a, $b \in \mathbb{F}$. An *inner product space* is any vector space over $\mathbb{F}$ equipped with a given inner product.

The *norm* of a vector $v \in V$ is

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

Then $\|v\| \geq 0$ and (i) and (iv) imply that $\|v\| = 0$ if and only if $v = 0$.

If $\mathbb{F} = \mathbb{R}$, then $\langle\ ,\ \rangle$ is a nondegenerate symmetric bilinear form on V. If $\mathbb{F} = \mathbb{C}$, then any map from $V \times V$ to $\mathbb{C}$ satisfying (i) and (ii) is called a *Hermitian form*.

We identify $\mathbb{F}^n$ with the vector space $\mathbb{F}^{n\times 1}$ of all $n \times 1$ column matrices with entries in $\mathbb{F}$. Then $\mathbb{C}^n$ is equipped with the *standard inner products*, which for $\mathbb{C}^n$ is given by

$$\langle z, w\rangle = {}^t z\, \overline{w} = \sum_{j=1}^{n} z_j\, \overline{w}_j, \tag{1.35}$$

if

$$z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \qquad \text{and} \qquad w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$$

The standard inner product on $\mathbb{R}^n$ just the restriction of the above to $\mathbb{R}^n \times \mathbb{R}^n$:

$$(x, y) = \sum_{j=1}^{n} x_j\, y_j.$$

This is of course just the usual dot product on $\mathbb{R}^n$.

There are numerous examples of inner product spaces. Here are but two of them:

Example 1.11.1. Let $T = \{z \in \mathbb{C} \,|\, |z| = 1\}$ be the unit circle in the complex plane, and let $C(T)$ denote the complex vector space of all complex-valued continuous functions on T. Then $C(T)$ can be equipped with the inner product

$$\langle f, g\rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(e^{it})\, \overline{g}(e^{it})\, dt \qquad (f, g \in C(T)).$$

Example 1.11.2. Let $\mathbb{F}^{m\times n}$ denote the vector space (over $\mathbb{F}$) of all $m \times n$ matrices with entries in $\mathbb{F}$. Then the *Hilbert-Schmidt inner product* on $\mathbb{F}^{m\times n}$ is given by

$$\langle S, T\rangle = \mathrm{tr}\,(S\, {}^t\overline{T})$$

If $S = (s_{jk})$ and $T = (t_{jk})$, then we have $\langle S, T\rangle = \sum_{j=1}^{m} \sum_{k=1}^{n} s_{jk}\, \overline{t}_{jk}$, the standard inner product on $\mathbb{F}^{mn}$. Notice that this inner product generalizes the standard inner product (1.35) on $\mathbb{F}^n$. The *Hilbert-Schmidt norm* of $S \in \mathbb{F}^{m\times n}$ is

$$\left\{ \sum_{j=1}^{m} \sum_{k=1}^{n} |s_{jk}|^2 \right\}^{1/2}.$$

Theorem 1.11.3. *(The Cauchy-Schwartz Inequality.) Let $\langle\ ,\ \rangle$ be an inner product on the vector space V over $\mathbb{F}$. Then*

$$|\langle u, v\rangle| \le \|u\|\,\|v\| \tag{1.36}$$

for all u, $v \in V$. Equality holds if and only if one of the two vectors is a multiple of the other.

Proof. There is a number $c \in \mathbb{F}$ with $|c| = 1$ such that $c\,\langle u, v\rangle = |\langle u, v\rangle|$. (In particular $c = \pm 1$ in case $\mathbb{F} = \mathbb{R}$.) For any $t \in \mathbb{R}$, we have

$$\begin{aligned} 0 &\le \|t(cu) + v\|^2 \\ &= \langle t(cu) + v, t(cu) + v\rangle \\ &= t^2\,\|cu\|^2 + t\,\langle cu, v\rangle + t\,\langle v, cu\rangle + \|v\|^2 \\ &= t^2\,\|u\|^2 + 2t\ (\mathrm{Re}\,(c\,\langle u, v\rangle)) + \|v\|^2 \\ &= t^2\,\|u\|^2 + 2t\,|\langle u, v\rangle| + \|v\|^2 \end{aligned}$$

The quadratic expression above in t has nonpositive discriminant, and hence $4|\langle u, v\rangle|^2 - 4\|u\|^2\,\|v\|^2 \le 0$, giving (1.36).

Next let us examine when (1.36) becomes an equality. Certainly it happens when $u = 0$ or $v = 0$, so let us assume that $u \ne 0$ and $v \ne 0$. If equality still occurs, then the discriminant $4|\langle u, v\rangle|^2 - 4\|u\|^2\,\|v\|^2$ is zero, so there is a unique value of t such that the quadratic expression above equals 0. For this t, we have $t(cu) + v = 0$, so v is a multiple of u. Conversely, if v is a multiple of u, (1.36) is easily seen to be an equality. □

Theorem 1.11.4. *(The Triangle Inequality) For any vectors u and v in an inner product space V, we have*

$$\|u + v\| \le \|u\| + \|v\|, \tag{1.37}$$

with equality if and only if one of the vectors is a nonnegative multiple of the other.

Proof. By (1.36), we have

$$\begin{aligned} \|u + v\|^2 &= \langle u + v, u + v\rangle \\ &= \|u\|^2 + 2\,\mathrm{Re}\,(\langle u, v\rangle) + \|v\|^2 \\ &\le \|u\|^2 + 2\,\|u\|\,\|v\| + \|v\|^2 \\ &= (\|u\| + \|v\|)^2\,, \end{aligned}$$

proving (1.37). Equality in (1.37) will occur if and only if $\mathrm{Re}\,(\langle u, v\rangle) = \|u\|\,\|v\|$; that is, if and only if $\langle u, v\rangle = \|u\|\,\|v\|$. Hence one of the vectors is a multiple of the other, say $v = cu$. This then gives $\overline{c}\,\|u\|^2 = |c|\,\|u\|^2$. Assuming that $u \ne 0$ (the case $u = 0$ giving $v = 0$, so is trivial), we obtain $\overline{c} = |c|$, so $c \ge 0$. □

From basic linear algebra we recall that a collection $\mathcal{U}$ of vectors in V is *orthonormal* provided that for any u, $v \in \mathcal{U}$ we have

$$\langle u, v\rangle = \begin{cases} 0 & \text{if } u \neq v \\ 1 & \text{if } u = v. \end{cases}$$

An orthonormal set is always linearly independent: if $u_1, \ldots, u_m \in \mathcal{U}$ and $c_1 u_1 + \cdots + c_m u_m = 0$, then for each j, we have $c_j = \langle c_1 u_1 + \cdots + c_m u_m, u_j\rangle = 0$.

The standard basis $e_1, \ldots, e_n$ of $\mathbb{F}^n$ is an orthonormal basis with respect to its standard inner product. In the infinite-dimensional inner product space $C(T)$, the Fourier exponentials

$$f_k(e^{i\theta}) = \frac{1}{\sqrt{2\pi}}\, e^{ik\theta}, \qquad\qquad (e^{i\theta} \in T)$$

for all $k \in \mathbb{Z}$, form an orthonormal set.

In $\mathbb{R}^2$, all orthonormal bases are of the form

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

and in $\mathbb{C}^2$ they are of the form

$$\begin{pmatrix} a \\ b \end{pmatrix}, \qquad c \begin{pmatrix} -\overline{b} \\ \overline{a} \end{pmatrix},$$

where $|a|^2 + |b|^2 = 1$ and $|c| = 1$.

Suppose that $v_1, v_2, \ldots$ is a linearly independent set in an inner product space V. The *Gram-Schmidt process* extracts an orthonormal set $u_1, u_2, \ldots$ out of this with the property that

$$\operatorname{span}(v_1, \ldots, v_k) = \operatorname{span}(u_1, \ldots, u_k) \tag{1.38}$$

for all $k = 1, 2, \ldots$. The vectors $u_1, u_2, \ldots$ are defined inductively as follows: put $u_1 = v_1/\|v_1\|$. Suppose that $k > 1$ and that orthonormal vectors $u_1, \ldots, u_{k-1}$ have been obtained so that

$$\operatorname{span}(v_1, \ldots, v_{k-1}) = \operatorname{span}(u_1, \ldots, u_{k-1}).$$

Let

$$w_k = v_k - \sum_{j=1}^{k-1} \langle v_k, u_j\rangle\, u_j$$

Now $v_k \notin \operatorname{span}(v_1, \ldots, v_{k-1}) = \operatorname{span}(u_1, \ldots, u_{k-1})$, so $w_k \neq 0$. The definition of w_k above shows that $\langle w_k, u_j\rangle = 0$ for $1 \leq j \leq k-1$. If we let $u_k = w_k/\|w_k\|$, then $\{u_1, \ldots, u_k\}$ is orthonormal (hence linearly independent) and satisfies (1.38).

If V is finite-dimensional and has basis $(v_1, \dots, v_n)$, the Gram-Schmidt process yields an orthonormal basis $(u_1, \dots, u_n)$ of V.

For any nonempty subset A of an inner product space V, the set

$$A^\perp = \{v \in V \mid \langle v, w\rangle = 0 \text{ for all } w \in A\}$$

is a subspace of V, called the *orthogonal complement* of A.

If V is finite-dimensional and W is a subspace of V, then $W^\perp$ is a subspace of V and $\dim W^\perp = \dim V - \dim W$. To see this, start with a basis $(v_1, \dots, v_n)$ of V such that $(v_1, \dots, v_k)$ is a basis of W, and apply the Gram-Schmidt process. If $(u_1, \dots, u_n)$ is the resulting orthonormal basis of V, then $(u_1, \dots, u_k)$ is a basis of W and it is easy to see that $(u_{k+1}, \dots, u_n)$ is a basis of $W^\perp$. This also shows that V is an *orthogonal direct sum*

$$V = W \oplus W^\perp.$$

Finally, it is not hard to see from the construction above that $(W^\perp)^\perp = W$. This can also be proved by noting that $W \subset (W^\perp)^\perp$ and that $\dim W = \dim V - \dim W^\perp = \dim(W^\perp)^\perp$.

Proposition 1.11.5. *Let V be a finite-dimensional vector space and let $f \in V^*$. Then there is a unique vector $w \in V$ such that $f(v) = \langle v, w\rangle$ for all $v \in V$.*

Proof. The uniqueness of w is straightforward: if $\langle v, w\rangle = \langle v, w'\rangle$ for all $v \in V$, then $\langle v, w - w'\rangle = 0$ for all $v \in V$ so $\langle w - w', w - w'\rangle = 0$, and hence $w = w'$.

For the existence of w, we can assume that $f \neq 0$. Then $f(V) = \mathbb{F}$, so by the Rank-Nullity Theorem (Theorem1.2.1), $\dim(\ker f) = \dim V - 1$. This implies that $(\ker f)^\perp$ is one-dimensional. Let w_0 be any unit vector in $(\ker f)^\perp$, and then let $w = \overline{f(w_0)}\, w_0$. Now $v - \langle v, w_0\rangle\, w_0$ is orthogonal to w_0, so $v - \langle v, w_0\rangle\, w_0 \in \ker f$. Hence $f(v - \langle v, w_0\rangle\, w_0) = 0$, and this untangles to $f(v) = \langle v, w\rangle$, as desired. □

Note that if $\mathbb{F} = \mathbb{R}$, then the proof is easier: for any $v \in V$, define $f_v \in V^*$ by $f_v(w) = \langle w, v\rangle$ for all $w \in V$. Then the map $v \mapsto f_v$ is a linear map from V into V^* with kernel $\{0\}$ (because the inner product is positive definite, hence nondegenerate), so it must be a linear bijection.

Exercise 1.11.6. Suppose that W is a subspace of a finite-dimensional inner product space V. Since $V = W \oplus W^\perp$, there are unique maps $P\colon V \to W$ and $Q\colon V \to W^\perp$ such that $v = P(v) + Q(v)$ for any $v \in V$. P is called the *orthogonal projection* of V onto W. The we have the *Pythagorean Theorem*

$$\|v\|^2 = \|Pv\|^2 + \|Qv\|^2 \qquad (v \in V).$$

Prove that for any $v \in V$,

$$\|Qv\| = \|Pv - v\| \leq \|w - v\| \qquad (w \in W)$$

with equality if and only if $w = P(v)$. Thus $P(v)$ is the point in W closest to v.

Theorem 1.11.7. *Let V and W be finite-dimensional inner product spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(V, W)$. Then there exists a unique linear map $T^* \in \mathcal{L}(W, V)$ such that*

$$\langle T(v), w\rangle = \langle v, T^*(w)\rangle \tag{1.39}$$

for all $v \in V$, $w \in W$.

(In (1.39) there is a slight abuse of notation, in which $\langle\ ,\ \rangle$ denotes the inner product both in V and in W.)

Proof. For fixed any $w \in W$, the function $f_w(v) = \langle Tv, w\rangle$ is a linear functional on V. By Proposition 1.11.5 there exists a unique vector $v' \in V$ such that $f_w(v) = \langle v, v'\rangle$ for all $v \in V$. The vector v' of course depends on w, so we write $v' = T^*(w)$. Thus $\langle T(v), w\rangle = \langle v, T^*(w)\rangle$ for all $v \in V$ and $w \in W$.

The uniqueness of $T^*(w)$ for each w implies that T^* preserves scalar multiplication: if $w \in W$ and $\lambda \in \mathbb{F}$, we have

$$\langle v, T^*(\lambda w)\rangle = \langle T(v), \lambda w\rangle = \overline{\lambda}\,\langle Tv, w\rangle = \overline{\lambda}\,\langle v, T^* w\rangle = \langle v, \lambda\, T^*(w)\rangle,$$

and hence $T^*(\lambda w) = \lambda\, T^*(w)$.

A similar calculation using the uniqueness of each $T^*(w)$ shows that T^* is additive: $T^*(w_1 + w_2) = T^*(w_1) + T^*(w_2)$ for all $w_1, w_2 \in W$. This shows that $T^* \in \mathcal{L}(W^*, V^*)$. □

The linear map $T^*\colon W \to V$ is called the *adjoint* of T. The relation (1.39) shows that $(T_1 + T_2)^* = T_1^* + T_2^*$ if T_1 and T_2 are in $\mathcal{L}(V, W)$. The same relation shows that $(\lambda T)^* = \overline{\lambda}\, T^*$, for $\lambda \in \mathbb{F}$, $T \in \mathcal{L}(V, W)$.

Finally, suppose $T \in \mathcal{L}(V, W)$, $S \in \mathcal{L}(W, U)$, where V, W, and U are finite-dimensional inner product spaces. Applying the relation (1.39) twice shows that $(ST)^* = T^*\, S^*$.

Let $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_m)$ be orthonormal bases of V and W, respectively, let $T \in \mathcal{L}(V, W)$, and suppose that the $m \times n$ matrix $A = (T_{jk})$ is the matrix of T with respect to these bases. Then ${}^t\overline{A}$ is the matrix of $T^* \in \mathcal{L}(W, V)$ with respect to these bases. In fact,

$$T^*_{kj} = \langle T^*(w_j), v_k\rangle = \langle w_j, T(v_k)\rangle = \overline{T}_{jk}.$$

Theorem 1.11.8. *Let V and W be finite-dimensional inner product spaces over $\mathbb{F}$, and let $T \in \mathcal{L}(V, W)$. Then*

(a). $(T^*)^* = T$.

(b). $T(V)^\perp = \ker(T^*)$.

Proof. The first assertion is an immediate consequence of (1.39), since for any $v \in V$, $w \in W$, we have $\langle Tv, w\rangle = \langle v, T^*(w)\rangle = \langle (T^*)^*(v), w\rangle$.

For the second assertion, we note that for any $w \in W$,

$$\begin{aligned} w \in T(V)^\perp &\iff \langle w, T(v)\rangle = 0 \text{ for all } v \in V \\ &\iff \langle T^*(w), v\rangle = 0 \text{ for all } v \in V \\ &\iff T^*(w) = 0 \\ &\iff w \in \ker(T^*). \end{aligned}$$

□

The conclusions in Theorem 1.11.8 easily imply the following facts:

(i). $(\ker(T^*))^\perp = T(V)$.

(ii). $(\ker T)^\perp = T^*(W)$.

(iii). $\ker T = (T^*(W))^\perp$.

Exercise 1.11.9. Let $T \in \mathcal{L}(V)$. Prove that T is one-to-one if and only if T^* is onto. Then prove that T is invertible if and only if T^* is invertible, and that $(T^{-1})^* = (T^*)^{-1}$.

Exercise 1.11.10. Let W be a subspace of a finite-dimensional inner product space V, and let $P\colon V \to W$ be the orthogonal projection of V onto W (in accordance with the orthogonal direct sum $V = W \oplus W^\perp$). Prove that $P^2 = P$and that $P^* = P$. Conversely, show that any $P \in \mathcal{L}(V)$ satisfying $P^2 = P$ and $P^* = P$ is an orthogonal projection of V onto an appropriate subspace.

1.12 Diagonalizability of Normal Operators

Let V be a finite-dimensional inner product space over $\mathbb{F}$, and let $T \in \mathcal{L}(V)$. It is important to determine sufficient conditions under which T is diagonalizable; that is to say, semisimple. This amounts to determining when there exists a basis of V consisting of eigenvectors of T. On the matrix side, we want to determine sufficient conditions under which an $n \times n$ matrix A is conjugate to a diagonal matrix. We already know one such sufficient condition: if A has n distinct eigenvalues in $\mathbb{F}$, then A is diagonalizable.

Throughout this section we will assume that V is a finite-dimensional inner product space, with inner product $\langle\ ,\ \rangle$. If $T \in \mathcal{L}(V)$, then $T^* \in \mathcal{L}(V)$.

Definition 1.12.1. An operator $T \in \mathcal{L}(V)$ is said to be

1. *normal* if $T^*T = TT^*$;
2. *self-adjoint* if $T = T^*$;
3. *unitary* if $\|Tv\| = \|v\|$ for all $v \in V$. We also say that T is *isometric*.

A self-adjoint operator is clearly normal. A unitary linear operator T preserves the inner product: for any $v, w \in V$, the relations $\|T(v+w)\|^2 = \|v+w\|^2$ and $\|T(v+i\,w)\|^2 = \|v+i\,w\|^2$ imply that $\langle Tv, Tw\rangle = \langle v, w\rangle$. Conversely, it is obvious that any $T \in \mathcal{L}(V)$ preserving the inner product is unitary. Finally, it clear that $\langle Tv, Tw\rangle = \langle v, w\rangle$ for all v and w if and only if $T^*T = I_V$. Thus a unitary operator is normal.

An unitary linear operator on a real vector space is called an *orthogonal operator*. (The word *unitary* is in fact often reserved for operators on complex spaces.) A self-adjoint linear operator on a complex inner product space is often called a *Hermitian operator*, and a self-adjoint linear operator on a real inner product space is also often called a *symmetric operator*).

The definitions above have matrix counterparts. Let $A \in \mathfrak{gl}(n, \mathbb{F})$. We say that A is

1. *normal* if ${}^t\overline{A}\,A = A\,{}^t\overline{A}$;
2. *symmetric* if ${}^tA = A$;
3. *Hermitian* if ${}^t\overline{A} = A$;
4. *unitary* if ${}^t\overline{A}\,A = I_n$;
5. *orthogonal* if ${}^tA\,A = I_n$.

Normal linear operators correspond to normal matrices and vice versa. Specifically, if $T \in \mathcal{L}(V)$ is normal and $B = (u_1, \ldots, u_n)$ is an orthonormal basis of V, then the matrix A of T with respect to B is normal. This follows immediately from the fact that composition of linear operators corresponds to matrix multiplication ((1.6)) and the fact that the matrix of T^* with respect to the basis B is ${}^t\overline{A}$.

Conversely, if A is a normal matrix then the linear operator $z \mapsto Az$ is a normal linear operator on $\mathbb{C}^n$. This is because the adjoint of this operator (with respect to the standard inner product on $\mathbb{C}^n$) is the linear operator $w \mapsto {}^t\overline{A}\,w$.

In a similar fashion, it can be seen that self-adjoint operators are represented by Hermitian matrices, and unitary linear operators by unitary matrices. In the case of linear operators on real inner product spaces, the correspondence

is between self-adjoint operators and real symmetric matrices, and between unitary operators and orthogonal matrices.

Exercise 1.12.2. Prove that any unitary linear operator T on a complex finite-dimensional inner product space V is invertible. Then prove that the set of all unitary linear operators on V forms a group, with multiplication defined as composition of operators.

The group of all unitary linear operators on a finite-dimensional complex inner product space V is called the *unitary group* of V, and is denoted by $U(V)$. If $V = \mathbb{C}^n$, then the unitary group on V can be identified with the group of unitary $n \times n$ matrices, which we denote by $U(n)$. The group of all unitary (= orthogonal) linear operators on a finite-dimensional real inner product space V is denoted $O(V)$; the group of orthogonal real $n \times n$ matrices is called the *orthogonal group*, and is denoted by $O(n)$.

Exercise 1.12.3. Prove that a complex $n \times n$ matrix A is unitary if and only if its columns form an orthonormal basis of $\mathbb{C}^n$. Similarly, prove that an $n \times n$ real matrix is orthogonal if and only if its columns form an orthonormal basis of $\mathbb{R}^n$. What can you say about the rows of these matrices?

Exercise 1.12.4. Let V be a complex inner product space. If T is a unitary linear operator on V, prove that $|\det T| = 1$.

Ths subgroup of $\mathrm{U}(V)$ consisting of all $T \in \mathrm{U}(V)$ such that $\det T = 1$ is called the *special unitary group* on V, and is denoted by $\mathfrak{su}(V)$. The special unitary group on $\mathbb{C}^n$ is denoted by $\mathrm{SU}(n)$. It consists of all $n \times n$ complex matrices A such that ${}^t\overline{A}\,A = I_n$ and $\det A = 1$.

Exercise 1.12.5. (a). Prove that if $T \in \mathcal{L}(V)$ is self-adjoint, then any eigenvalue of T is real.

(b). Prove that if λ is an eigenvalue of a unitary operator $S \in \mathcal{L}(V)$, then $|\lambda| = 1$.

(c). Suppose that u and v are eigenvectors of a self-adjoint operator $T \in \mathcal{L}(V)$ corresponding to different eigenvalues. Prove that u and v are orthogonal.

Our goal is to prove that normal linear operators on complex inner product spaces are diagonalizable. This involves proving a series of lemmas, the first of which is the following.

Lemma 1.12.6. *Let $T \in \mathcal{L}(V)$ be self-adjoint. Then $T = 0$ if and only if $\langle Tv, v\rangle = 0$ for all $v \in V$.*

Proof. The "only if" part being trivial, let us suppose that $T \in \mathcal{L}(V)$ is self-adjoint and satisfies

$$\langle Tv, v\rangle = 0 \qquad \text{for all } v \in V.$$

Replacing v by $v + w$ in the relation above yields $\langle T(v), w\rangle + \langle T(w), v\rangle = 0$; since T is self-adjoint, this gives

$$\operatorname{Re}\langle Tv, w\rangle = 0 \qquad \text{for all } v,\, w \in V.$$

Replacing w above by iw, we then obtain

$$\operatorname{Im}\langle Tv, w\rangle = 0 \qquad \text{for all } v,\, w \in V,$$

from which we conclude that $\langle Tv, w\rangle = 0$ for all $v, w \in V$. It follows that $T = 0$. □

Lemma 1.12.7. *Let $T \in \mathcal{L}(V)$. Then T is normal if and only if $\|T^*(v)\| = \|T(v)\|$ for all $v \in V$.*

Proof. Note that $T^*T - TT^*$ is self-adjoint. Hence by Lemma 1.12.6,

$$\begin{aligned}
\|T^*(v)\| = \|T(v)\| \text{ for all } v \in V &\iff \langle T^*(v), T^*(v)\rangle = \langle T(v), T(v)\rangle \text{ for all } v \in V \\
&\iff \langle TT^*(v), v\rangle = \langle T^*T(v), v\rangle \text{ for all } v \in V \\
&\iff \langle TT^*(v) - T^*T(v), v\rangle \text{ for all } v \in V \\
&\iff TT^* - T^*T = 0 \\
&\iff T \text{ is normal.}
\end{aligned}$$

□

Corollary 1.12.8. *Let $T \in \mathcal{L}(V)$ be normal. Then $\ker T = \ker T^*$.*

In fact $v \in \ker T \iff \|T(v)\| = 0 \iff \|T^*(v)\| = 0 \iff v \in \ker T^*$.

Exercise 1.12.9. Suppose that $T \in \mathcal{L}(V)$ is normal. Prove that $\ker T^2 = \ker T$. (*Hint:* T^*T is self-adjoint.) Then prove that $\ker T^m = \ker T$ for all positive integers m.

Lemma 1.12.10. *Let $T \in \mathcal{L}(V)$ be a normal operator. Then any eigenvector of T is also an eigenvector of T^*.*

Proof. Suppose that $T(v) = \lambda\, v$. Then $v \in \ker(T - \lambda I_V)$. Now $(T - \lambda I_V)^* = T^* - \overline{\lambda} I_V$, and it is clear from the normality of T that $T - \lambda I_V$ is also normal. By Corollary 1.12.8 it follows that $v \in \ker(T^* - \overline{\lambda} I_V)$, and hence $T^*(v) = \overline{\lambda}\, v$. □

Lemma 1.12.11. *Let $T \in \mathcal{L}(V)$ be a normal operator. If a subspace U of V is invariant under T, then $U^\perp$ is invariant under T^*.*

Proof. Obvious, since for any $u \in U$ and $v \in U^\perp$, we have $\langle u, T^*v\rangle = \langle T(u), v\rangle = 0$. □

Theorem 1.12.12. *(The Spectral Theorem for Normal Operators.) Let V be a complex inner product space, and let $T \in \mathcal{L}(V)$ be a normal operator. Then there exists an orthonormal basis of V consisting of eigenvectors of T.*

Proof. The proof is by induction on $\dim V$, the case $\dim V = 1$ being obvious.

For the induction step, let $n > 1$ and assume that the conclusion of the theorem holds for all normal operators on all complex inner product spaces of dimension $< n$.

Suppose that $\dim V = n$ and T is a normal linear operator on V. Since V is a complex vector space, T has an eigenvalue λ. Let v be an eigenvector corresponding to λ. Then according to Lemma 1.12.10, the one-dimensional subspace $U = \mathbb{C}\, v$ is invariant under both T and T^*. Lemma 1.12.11 then shows that $U^\perp$ is also invariant under T^* and T^*. Let S be the restriction of T to $U^\perp$. Then the adjoint S^* of S is the restriction of T^* to $U^\perp$. It follows that S is a normal operator on $U^\perp$. Since $U^\perp$ is $(n-1)$-dimensional, the induction hypothesis guarantees an orthonormal basis $(v_1, \ldots, v_{n-1})$ of $U^\perp$ consisting of eigenvectors of S. Since these are also eigenvectors of T, we see that the orthonormal basis $(v_1, \ldots, v_{n-1}, v)$ of V consists of eigenvectors of T, completing the induction step. □

The theorem above asserts that normal operators in complex inner product spaces are semisimple. We can of course restate the result in matrix form.

Corollary 1.12.13. *Let A be an $n \times n$ complex normal matrix. Then there exists a unitary matrix U such that $U^{-1}AU$ is a diagonal matrix. Up to a permutation of the diagonal entries, this diagonal matrix is unique.*

Proof. The uniqueness assertion is clear, since A and $U^{-1}AU$ have the same characteristic polynomial.

As for the first assertion, if we endow $\mathbb{C}^n$ with the standard inner product (1.35), then the linear operator T on $\mathbb{C}^n$ given by $T(z) = Az$ (i.e., matrix multiplication) is normal. By the preceding theorem, there is an orthonormal basis $(u_1, \ldots, u_n)$ of $\mathbb{C}^n$ consisting of eigenvectors of T. Then we have $A\, u_j = \lambda_j\, u_j$ for $j = 1, \ldots, n$. Let U be the $n \times n$ matrix whose columns (in order) are $u_1, \ldots, u_n$, and let D be the $n \times n$ diagonal matrix whose diagonal entries (in order) are $\lambda_1, \ldots, \lambda_n$. Since the columns of U are orthonormal, U is a unitary matrix (see Exercise 1.12.3), and we also have

$$AU = UD.$$

Hence $D = U^{-1}\, A\, U$. □

Exercise 1.12.14. (*An Alternative proof of Theorem 1.12.12.*) Let T be a normal operator on a finite-dimensional complex inner product space V. Prove the

following assertions to show that V is an orthogonal direct sum of eigenspaces of T.

(a). For any complex number λ, prove that the operator $T - \lambda I_V$ is normal.

(b). By Part (a) above and Exercise 1.12.9, $\ker(T - \lambda I_V)^m = \ker(T - \lambda I_V)$. Use this and Theorem 1.7.7 to prove that V is a direct sum of eigenspaces of T.

(c). Use Lemma 1.12.10 to prove that the eigenspaces you obtained in Part (b) are mutually orthogonal.

(d). Use Part (c) to prove that V has an orthonormal basis consisting of eigenvectors of T.

The proof of Theorem 1.12.12 relied on the existence of at least one eigenvalue, which is of course guaranteed for a linear operator on a finite-dimensional *complex* vector space. For normal operators on real inner product spaces, diagonalizability is not guaranteed. For instance, the matrix

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

is orthogonal, so the corresponding linear operator $x \mapsto A\,x$ on $\mathbb{R}^2$ is orthogonal, hence is a normal operator. This linear operator corresponds to a 90° counterclockwise rotation on $\mathbb{R}^2$, so it has no eigenvectors in $\mathbb{R}^2$. Hence it is not semisimple, and the matrix A is not diagonalizable (or more precisely, is not conjugate to a diagonal matrix).

On the other hand, there are normal operators on real inner product spaces that are diagonalizable; namely, the self-adjoint ones. To prove this, we will need the following lemma.

Lemma 1.12.15. *Let V be a finite-dimensional real inner product space, and let $T \in \mathcal{L}(V)$ be self-adjoint. Suppose that b and c are any real numbers such that $b^2 - 4c < 0$. Then the linear operator $T^2 + b\,T + c\,I_V$ is invertible.*

Proof. It is enough to prove that $T^2 + b\,T + c\,I_V$ is injective. Let u be any unit vector in V. Then by the Cauchy-Schwartz Inequality (1.36),

$$\begin{aligned} \langle (T^2 + b\,T + c\,I_V)(u), u\rangle &= \langle T(u), T(u)\rangle + b\,\langle T(u), u\rangle + c\,\langle u, u\rangle \\ &\geq \|T(u)\|^2 - |b|\,\|T(u)\| + c \\ &> 0. \end{aligned}$$

The last inequality holds because $b^2 - 4c < 0$. Now if $T^2 + b\,T + c\,I_V$ were *not* injective, then there would be a unit vector u in its kernel, which would contradict the inequality above. This of course proves the lemma. □

Theorem 1.12.16. *(The Spectral Theorem for Self-Adjoint Operators.) Let V be a real inner product space, and let $T \in \mathcal{L}(V)$ be self-adjoint. Then T is semisimple; in fact, there is an orthonormal basis of V consisting of eigenvectors of T.*

Proof. The proof is by induction on $\dim V$. The case $\dim V = 1$ is of course trivial.

For the induction step, let $n > 1$, assume that the conclusion holds for all self-adjoint linear operators on real inner product spaces of dimension $< n$, let $\dim V = n$, and assume $T \in \mathcal{L}(V)$ is self-adjoint.

To show that T satisfies the assertion, the key is to prove that T has at least one real eigenvalue λ. Assuming that this real eigenvalue λ exists, we can complete the proof as follows. Let $U = \ker(T - \lambda\, I_V)$ be the eigenspace corresponding to λ, and let $W = U^{\perp}$ be its orthogonal complement. Then U is T-invariant, as is W, because for any $u \in U$ and $w \in W$, we have

$$\langle u, T(w)\rangle = \langle T(u), w\rangle = 0,$$

which shows that $T(w) \in W$. Let S denote the restriction of T to W. Then $S \in \mathcal{L}(W)$ is self-adjoint, because T is. Since $\dim W < \dim V$, we can apply the induction hypothesis to produce an orthonormal basis of W consisting of eigenvectors of S, hence of T. To this basis we tack on an orthonormal basis of U. Since V is the orthogonal direct sum $U \oplus W$ the vectors from these two bases combine to form an orthonormal basis of V consisting of eigenvectors of T.

So all depends on our being able to produce a real eigenvalue of T. For this, fix any vector $v \neq 0$ in V. Since $n = \dim V$, the vectors v, $T(v)$, $T^2(v), \ldots, T^n(v)$ are linearly dependent, so there exist real scalars a_0, a_1 $\ldots$, a_n, not all 0, such that $a_0\, v + a_1\, T(v) + \cdots + a_n\, T^n(v) = 0$. We write this as

$$p(T)\,(v) = 0, \tag{1.40}$$

where $p(x) = a_0 + a_1\, x + \cdots + a_n\, x^n$. The polynomial $p(x)$ clearly has degree ≥ 1. Since it has real coefficients, it has a factorization into either linear or irreducible quadratic factors, or both. Now if there are *no* linear factors in the factorization of $p(x)$, we would have

$$p(x) = a \cdot \prod_{k=1}^{l} (x^2 + b_k\, x + c_k),$$

where $a \neq 0$ and $b_j^2 - 4c_j < 0$ for all j. The relation (1.40) then becomes

$$a \cdot \left(\prod_{k=1}^{m} (T^2 + b_k\, T + c_k\, I_V) \right)(v) = 0.$$

But by Lemma 1.12.15, the operators $T^2 + b_k\, T + c_k\, I_V$ are all invertible, so the left hand side above cannot vanish. This contradiction shows that $p(x)$ has linear factors, and we have

$$p(x) = a \cdot \prod_{j=1}^{l} (x + r_j) \cdot p_1(x) \qquad\qquad (a \neq 0)$$

where $p_1(x)$ is a product of irreducible quadratic factors, or just the number 1, in case there are no such factors. Hence

$$a \cdot \left(\prod_{j=1}^{l} (T + r_j\, I_V) \right) p_1(T)(v) = 0$$

Since $p_1(T)$ is invertible, we can apply $p_1(T)^{-1}$ to both sides above to obtain

$$\prod_{j=1}^{l} (T + r_j\, I_V)\,(v) = 0.$$

The operator $\prod_{j=1}^{l} (T + r_j\, I_V)$ is therefore not invertible, so one of its factors is not invertible. Rearranging the factors if necessary, we can assume that $T + r_1\, I_V$ is not invertible. Then $\lambda = -r_1$ is an eigenvalue of T, completing the proof of the theorem. □

Remark 1.12.17. In the proof above, instead of fixing a vector $v \in V$, we could have just as well used the Cayley-Hamilton Theorem for T to obtain $p(T) = 0$, where $p(x)$ is the characteristic polynomial of T. (Note that according to Remark 1.6.4, a linear operator on a real vector spaces also satisfies its characteristic polynomial.) By the same argument employed above, it will be impossible for $p(x)$ to lack any linear factors.

Corollary 1.12.18. *Let A be an $n \times n$ symmetric matrix with real entries. Then there is a real orthogonal $n \times n$ matrix U such that $U^{-1}\, A\, U$ is a diagonal matrix. Up to a permutation of the diagonal entries, this diagonal matrix is unique.*

The proof is the same as that of Corollary 1.12.13.

Exercise 1.12.19. Let V be a finite-dimensional inner product space, and suppose that Q is a Hermitian form on V; that is to say, $Q : V \times V \to \mathbb{C}$, with

(i). $Q(av_1 + bv_2, w) = a\, Q(v_1, w) + b\, Q(v_2, w)$, and

(ii). $Q(w, v) = \overline{Q(v, w)}$

for all v, v_1, v_2, $w \in V$ and all a, $b \in \mathbb{F}$. (When $\mathbb{F} = \mathbb{R}$, this means that Q is a symmetric bilinear form on V.) Prove that there is a self-adjoint operator

$T \in \mathcal{L}(V)$ such that $Q(v,w) = \langle Tv, w\rangle$ for all $v, w \in V$. Then prove that there is an orthonormal basis $(u_1, \ldots, u_n)$ of V and real numbers $\lambda_1, \ldots, \lambda_n$ such that

$$Q(v,v) = \sum_{j=1}^{n} \lambda_j \, |c_j|^2,$$

for $v = \sum c_j \, u_j \in V$. In particular, if Q is an inner product, then there are n vectors that are simultaneously orthogonal for $\langle \ , \ \rangle$ and Q.

The spectral theorems allow us to express normal or self-adjoint operators in terms of *orthogonal projections*. If U is a subspace of an inner product space V, then V is the orthogonal direct sum $V = U \oplus U^{\perp}$. Every $v \in V$ can be written in a unique way as $v = u + w$, where $u \in U$ and $w \in U^{\perp}$, and we put $P_U(v) = u$. (See Exercise 1.11.6.) As in that exercise, the map $P_U : V \to U$ is called the orthogonal projection of V onto U. Note that $P_U + P_{U^{\perp}} = I_V$. The following properties of P_U are also easy to verify:

(a) $P_U^2 = P_U$.

(b) P_U is self-adjoint.

(c) $P_U \, P_{U'} = 0$ iff $U \perp U'$.

(d) U is invariant under an operator $T \in \mathcal{L}(V)$ iff $P_U \, T = T \, P_U$.

A collection $P_1, \ldots, P_m$ of orthogonal projections of V onto nonzero subspaces is called a *resolution of the identity* in V if the following conditions are met:

(i) $P_i \, P_j = 0$ if $i \neq j$

(ii) $P_1 + \cdots + P_m = I_V$.

If $P_1, \ldots, P_m$ is a resolution of the identity, then V is the orthogonal direct sum $V = P_1(V) \oplus \cdots \oplus P_m(V)$. Conversely, any decomposition of V as an orthogonal direct sum gives rise to a resolution of the identity.

The spectral theorems then say that if an inner product space V is complex and $T \in \mathcal{L}(V)$ is normal, or if V is real and T is self-adjoint, then there are *distinct* scalars $\lambda_1, \ldots, \lambda_m$ and a resolution of the identity $P_1, \ldots, P_m$ in V such that

$$T = \lambda_1 \, P_1 + \cdots + \lambda_m \, P_m. \tag{1.41}$$

The converse also holds, and is easy to prove. The linear combination on the right hand side is called the *spectral resolution of* T. Since the eigenvalues (and corresponding eigenspaces) of T are uniquely determined by T, it is clear that the spectral resolution of T is unique.

Note that if T has spectral resolution (1.41), then for any polynomial $p(\lambda)$, we have

$$p(T) = p(\lambda_1)\, P_1 + \cdots + p(\lambda_m)\, P_m.$$

In fact, if $f : \mathbb{F} \to \mathbb{F}$ is any function, it makes sense to define the linear operator $f(T)$ on V by putting

$$f(T) = f(\lambda_1)\, P_1 + \cdots + f(\lambda_m)\, P_m.$$

For instance,

$$\sin T = \sin \lambda_1\, P_1 + \cdots + \sin \lambda_m\, P_m.$$

The map $f \mapsto f(T)$ is then a ring homomorphism from the ring of $\mathbb{F}$-valued functions on $\mathbb{F}$ into the ring $\mathcal{L}(V)$. We will not develop this any further, except to point out that, in the right setting, this can be extended to the infinite-dimensional case. See, for example, Rudin's book [Rud91].

Exercise 1.12.20. Suppose that $T \in \mathcal{L}(V)$ has the spectral resolution (1.41). Prove that $S \in \mathcal{L}(V)$ commutes with T if and only if S commutes with all the projections P_j.

1.13 Positive Operators and Polar and Singular Value Decompositions

Let V be an inner product space over $\mathbb{F}$. An operator $S \in \mathcal{L}(V)$ is said to be *positive* (or *positive definite*) provided that P is self-adjoint and $\langle Sv, v\rangle > 0$ for all nonzero vectors $v \in V$. The condition that S is positive is often denoted by $S > 0$. (If $\langle Sv, v\rangle \geq 0$ for all v, we say that S is *nonnegative* or *positive semidefinite*, and denote this by $S \geq 0$.)

Actually, when V is a *complex* inner product space, any linear operator S such that $\langle Sv, v\rangle \geq 0$ for all $v \in V$ is self-adjoint. To see this, note that for any v and w in V, the conditions $\langle S(v+w), v+w\rangle \geq 0$ and $\langle S(v+i\,w), v+i\,w\rangle \geq 0$ imply that

$$\langle Sv, w\rangle + \overline{\langle v, Sw\rangle}$$

and

$$i\left(-\langle Sv, w\rangle + \overline{\langle v, Sw\rangle}\right)$$

are both real. This in turn implies that the real and imaginary parts of $\langle Sv, w\rangle$ and $\langle v, Sw\rangle$ coincide, whence $\langle Sv, w\rangle = \langle v, Sw\rangle$.

Exercise 1.13.1. Give an example of a linear operator T on $\mathbb{R}^2$ such that $\langle Tv, v\rangle \geq 0$ for all $v \in \mathbb{R}^2$, but which is not self-adjoint.

If $T \in \mathcal{L}(V)$, then T^*T is nonnegative: it is clearly self-adjoint and $\langle T^*Tv, v\rangle = \langle Tv, Tv\rangle \geq 0$ for all $v \in V$. It is positive if and only if T is invertible.

It is clear that a positive operator S has positive eigenvalues. The Spectral Theorem also shows that a nonnegative operator is positive if and only if it is invertible.

Roughly speaking, nonnegative operators generalize the nonnegative real numbers, positive operators generalize the positive real numbers, self-adjoint operators generalize the real numbers, the adjoint map $T \mapsto T^*$ generalizes complex conjugation, and the relation $T^*T \geq 0$ generalizes the condition $\overline{z}\, z \geq 0$.

The fact that any nonnegative real number has a unique nonegative square root also generalizes in the present setting.

Let V be a vector space over $\mathbb{F}$ and let $T \in \mathcal{L}(V)$. We say that an operator $S \in \mathcal{L}(V)$ is a *square root* of T provided that $S^2 = T$. It is not hard to see that a linear operator (even the identity) can have many square roots.

Theorem 1.13.2. *Every nonnegative linear operator S on an inner product space has a unique nonnegative square root. We denote this square root by $\sqrt{S}$.*

Proof. The proof is very easy. If S is a nonnegative operator, then it has a spectral resolution

$$S = \lambda_1 P_1 + \cdots + \lambda_m P_m,$$

where $\lambda_j \geq 0$ for all j. The operator $R = \sqrt{\lambda_1}\, P_1 + \cdots + \sqrt{\lambda_m}\, P_m$ is then a nonnegative square root of S.

For the uniqueness, suppose that R_0 is a nonnegative square root of S. Then R_0 has a spectral resolution

$$R_0 = \mu_1 Q_1 + \cdots + \mu_l Q_l,$$

where $\mu_k \geq 0$ for all k. We then have

$$R_0^2 = \mu_1^2 Q_1 + \cdots + \mu_l^2 Q_m.$$

Since the μ_j are distinct and nonnegative, so are the μ_j^2, and so the right hand side above must be the spectral resolution of S. In particular, $m = l$, the resolutions of the identity $Q_1, \ldots, Q_m$ and $P_1, \ldots, P_m$ are the same, and with an appropriate reordering of the indices, we must have $\mu_j = \sqrt{\lambda_j}$ for all j.

□

If S is a positive operator on V, the construction in the proof above allows us to define S^t for any real number t:

$$S^t = \lambda_1^t P_1 + \cdots + \lambda_m^t P_m.$$

The operators S^t (for $t \in \mathbb{R}$) form a group called the *one-parameter group containing* S, and the map $t \mapsto S^t$ is a homomorphism from the additive group of $\mathbb{R}$ onto this one-parameter group.

Exercise 1.13.3. Suppose that P is a positive operator on V and $T \in \mathcal{L}(V)$ commutes with P: $TP = PT$. Prove that T commutes with P^t for all real numbers t.

Exercise 1.13.4. A real $n \times n$ matrix A is called *positive definite* provided that the bilinear form

$$Q(x, y) = \langle Ax, y \rangle$$

is an inner product on $\mathbb{R}^n$. (Here $\langle x, y \rangle = {}^t x\, y$ is the standard inner product on $\mathbb{R}^n$, where we identify $\mathbb{R}^n$ with the vector space $\mathbb{R}^{n\times 1}$ of $n \times 1$ column matrices.) Let $A = (a_{ij})_{1\le i,j\le n}$. If A is positive definite, prove that A is symmetric and that $\Delta_k(A) > 0$ for $1 \le k \le n$, where $\Delta_k(A)$ is the determinant of the "top left" $k \times k$ submatrix $(a_{ij})_{1\le i,j\le k}$ of A.

Conversely, prove that if A is symmetric and $\Delta_k(A) > 0$ for $1 \le k \le n$, then A is positive definite.

Exercise 1.13.5. (The norm of a linear operator.) Let V be an inner product space over $\mathbb{F}$ and $T \in \mathcal{L}(V)$. We define the *norm* of T to be the nonnegative number

$$\|T\| \colon = \sup_{v\neq 0} \frac{\|Tv\|}{\|v\|}.$$

It is clear that $\|T\| = \sup\{\|Tv\| \mid \|v\| = 1\}$. Prove that $\|T\| = \sqrt{\Lambda}$, where Λ is the largest eigenvalue of the nonnegative operator T^*T.

Let V and W be inner product spaces of dimensions m and n, respectively, and let $T \in \mathcal{L}(V, W)$. For simplicity, we'll assume that T is of *full rank*. This means that either T is injective or T is surjective. (Of course, the two conditions are equivalent in case $m = n$.)

A linear map is of full rank if and only if its matrix, with respect to any choice of bases of V and W, has maximum rank. Since the subset of $\mathbb{F}^{m\times n}$ consisting of all matrices of maximum rank is dense and open, we see that "most" elements of $\mathcal{L}(V, W)$ are of full rank.

We now consider a factorization of T which is useful in signal processing and other applications. A linear map $T \in \mathcal{L}(V, W)$ is said to be a *linear isometry* from V into W if T preserves inner products: $\langle Tu, Tv \rangle = \langle u, v \rangle$ for all u, $v \in V$. A linear isometry is easily seen to be injective. Note that a linear isometry from V into itself is a unitary linear operator.

Theorem 1.13.6. *(The Singular Value Decomposition for Injective Linear Maps.) Suppose that $T \in \mathcal{L}(V, W)$ is injective. Then T has a unique factorization*

$$T = F\, P, \tag{1.42}$$

where $P \in \mathcal{L}(V)$ is a positive operator and $F \in \mathcal{L}(V, W)$ is a linear isometry.

Remark 1.13.7. The factorization (1.42) is called the *singular value decomposition* of T. By the Spectral Theorem, Theorem 1.13.6 implies that there is an orthonormal basis $(v_1, \ldots, v_n)$ of V - namely, an orthonormal eigenbasis of P - which is mapped by T into an orthogonal set in W.

In the case when $W = V$, the factorization above is called the *polar decomposition.*

Proof. First we observe that T^*T is a positive operator on V: it is certainly self-adjoint, and for any nonzero vector $v \in V$, we have

$$\langle T^*Tv, v\rangle = \langle Tv, Tv\rangle = \|Tv\|^2 > 0.$$

Let $P = \sqrt{T^*T}$, and let $F = TP^{-1}$. Then P is a positive operator on V and $F \in \mathcal{L}(V, W)$ is a *linear isometry* of V into W. The latter fact follows from

$$F^*F = (TP^{-1})^*\,(TP^{-1}) = P^{-1}T^*TP^{-1} = P^{-1} \cdot P^2 \cdot P^{-1} = I_V.$$

We have thus shown that T has the factorization (1.42). This factorization is unique because if $T = F_1\,P_1$, where F_1 is a linear isometry from V into W and $P_1 \in \mathcal{L}(V)$ is a positive operator, then $T^*T = P_1\,F_1^*\,F_1\,P_1 = P_1^2$. But we also have $T^*T = P^2$. By the uniqueness of positive square roots, it follows that $P_1 = P$, and hence $F_1 = TP_1^{-1} = TP^{-1} = F$. □

Exercise 1.13.8. Prove that if T maps an orthonormal basis $(v_1, \ldots, v_n)$ of V into an orthogonal set in W, then each v_j must be an eigenvector of T^*T. Then prove that the basis $(v_1, \ldots, v_n)$ is unique up to the choice of an orthonormal basis of each eigenspace of T^*T.

Exercise 1.13.9. (Singular Value Decomposition for Matrices) Assume that $n \geq k$. An $n \times k$ matrix Y is called a *Stiefel matrix* if ${}^t\overline{Y}\,Y = I_k$. (Thus the columns of a Stiefel matrix form an orthonormal set in $\mathbb{F}^n$.) Let A be any $n \times k$ matrix of rank k

(a) Prove that A has a unique factorization

$$A = Y\,B,$$

where Y is an $n \times k$ Stiefel matrix and B is a positive definite Hermitian $k \times k$ matrix.

(b) Prove that A has a factorization

$$A = F\,D\,U$$

where F is an $n \times k$ Stiefel matrix, D is a diagonal matrix with positive diagonal entries, and U is a $k \times k$ Stiefel matrix (that is, a unitary or orthogonal matrix). Prove that D is unique up to a permutation of the diagonal entries.

Let us now consider the case when $T \in \mathcal{L}(V, W)$ is surjective. Then according to Theorem 1.11.8, $T^* \in \mathcal{L}(W, V)$ is injective. An argument similar to that in the proof of Theorem 1.13.6 yields the following result.

Theorem 1.13.10. *(The Singular Value Decomposition for Surjective Linear Maps.) Suppose that $T \in \mathcal{L}(V, W)$ is surjective. Then T has a unique factorization*

$$T = P\,F \tag{1.43}$$

where P is a positive linear operator on W and $F \in \mathcal{L}(V, W)$ satisfies $F\,F^* = I_W$.

In particular, F^* is an isometric linear map from W into V. Since $P^2 = T\,T^*$, T^* maps an orthonormal eigenbasis of W for P into an orthogonal basis of $T^*(W) \subset V$. The linear operator F^*F is the orthogonal projection of V onto $T^*(W)$. If we denote this orthogonal projection by E, then we have

$$T = P\,F = P\,(F\,F^*)\,F = P\,F\,(F^*\,F) = P\,F\,E$$

Thus there is an orthonormal basis of $(\ker T)^\perp (= T^*(W))$ which T maps into an orthogonal basis of W.

Exercise 1.13.11. Prove all of the assertions above.

Exercise 1.13.12. Let $n \geq k$ and suppose that F is an $n \times k$ Stiefel matrix. Prove that the linear operator on $\mathbb{F}^n$ given by $x \mapsto F\,F^*\,x$ is the orthogonal projection of $\mathbb{F}^n$ onto the linear span of the columns of F.