CHAPTER 8

Operators on Complex Vector Spaces

In this chapter we delve deeper into the structure of operators on complex vector spaces. An inner product does not help with this material, so we return to the general setting of a finite-dimensional vector space (as opposed to the more specialized context of an inner-product space). Thus our assumptions for this chapter are as follows:

Recall that **F** denotes **R** or **C**. Also, V is a finite-dimensional, nonzero vector space over **F**.

Some of the results in this chapter are valid on real vector spaces, so we have not assumed that V is a complex vector space. Most of the results in this chapter that are proved only for complex vector spaces have analogous results on real vector spaces that are proved in the next chapter. We deal with complex vector spaces first because the proofs on complex vector spaces are often simpler than the analogous proofs on real vector spaces.

Generalized Eigenvectors

Unfortunately some operators do not have enough eigenvectors to lead to a good description. Thus in this section we introduce the concept of generalized eigenvectors, which will play a major role in our description of the structure of an operator.

To understand why we need more than eigenvectors, let's examine the question of describing an operator by decomposing its domain into invariant subspaces. Fix $T \in \mathcal{L}(V)$. We seek to describe T by finding a "nice" direct sum decomposition

8.1 $$V = U_1 \oplus \cdots \oplus U_m,$$

where each U_j is a subspace of V invariant under T. The simplest possible nonzero invariant subspaces are one-dimensional. A decomposition 8.1 where each U_j is a one-dimensional subspace of V invariant under T is possible if and only if V has a basis consisting of eigenvectors of T (see 5.21). This happens if and only if V has the decomposition

8.2 $$V = \operatorname{null}(T - \lambda_1 I) \oplus \cdots \oplus \operatorname{null}(T - \lambda_m I),$$

where $\lambda_1, \dots, \lambda_m$ are the distinct eigenvalues of T (see 5.21).

In the last chapter we showed that a decomposition of the form 8.2 holds for every self-adjoint operator on an inner-product space (see 7.14). Sadly, a decomposition of the form 8.2 may not hold for more general operators, even on a complex vector space. An example was given by the operator in 5.19, which does not have enough eigenvectors for 8.2 to hold. Generalized eigenvectors, which we now introduce, will remedy this situation. Our main goal in this chapter is to show that if V is a complex vector space and $T \in \mathcal{L}(V)$, then

$$V = \operatorname{null}(T - \lambda_1 I)^{\dim V} \oplus \cdots \oplus \operatorname{null}(T - \lambda_m I)^{\dim V},$$

where $\lambda_1, \dots, \lambda_m$ are the distinct eigenvalues of T (see 8.23).

Suppose $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T. A vector $v \in V$ is called a ***generalized eigenvector*** of T corresponding to λ if

8.3 $$(T - \lambda I)^j v = 0$$

for some positive integer j. Note that every eigenvector of T is a generalized eigenvector of T (take $j = 1$ in the equation above), but the converse is not true. For example, if $T \in \mathcal{L}(\mathbf{C}^3)$ is defined by

$$T(z_1, z_2, z_3) = (z_2, 0, z_3),$$

then $T^2(z_1, z_2, 0) = 0$ for all $z_1, z_2 \in \mathbf{C}$. Hence every element of $\mathbf{C}^3$ whose last coordinate equals 0 is a generalized eigenvector of T. As you should verify,

$$\mathbf{C}^3 = \{(z_1, z_2, 0) : z_1, z_2 \in \mathbf{C}\} \oplus \{(0, 0, z_3) : z_3 \in \mathbf{C}\},$$

where the first subspace on the right equals the set of generalized eigenvectors for this operator corresponding to the eigenvalue 0 and the second subspace on the right equals the set of generalized eigenvectors corresponding to the eigenvalue 1. Later in this chapter we will prove that a decomposition using generalized eigenvectors exists for every operator on a complex vector space (see 8.23).

Though j is allowed to be an arbitrary integer in the definition of a generalized eigenvector, we will soon see that every generalized eigenvector satisfies an equation of the form 8.3 with j equal to the dimension of V. To prove this, we now turn to a study of null spaces of powers of an operator.

Note that we do not define the concept of a generalized eigenvalue because this would not lead to anything new. Reason: if $(T - \lambda I)^j$ is not injective for some positive integer j, then $T - \lambda I$ is not injective, and hence λ is an eigenvalue of T.

Suppose $T \in \mathcal{L}(V)$ and k is a nonnegative integer. If $T^k v = 0$, then $T^{k+1}v = T(T^k v) = T(0) = 0$. Thus $\operatorname{null} T^k \subset \operatorname{null} T^{k+1}$. In other words, we have

8.4 $$\{0\} = \operatorname{null} T^0 \subset \operatorname{null} T^1 \subset \cdots \subset \operatorname{null} T^k \subset \operatorname{null} T^{k+1} \subset \cdots.$$

The next proposition says that once two consecutive terms in this sequence of subspaces are equal, then all later terms in the sequence are equal.

8.5 Proposition: *If $T \in \mathcal{L}(V)$ and m is a nonnegative integer such that* $\operatorname{null} T^m = \operatorname{null} T^{m+1}$, *then*

$$\operatorname{null} T^0 \subset \operatorname{null} T^1 \subset \cdots \subset \operatorname{null} T^m = \operatorname{null} T^{m+1} = \operatorname{null} T^{m+2} = \cdots.$$

PROOF: Suppose $T \in \mathcal{L}(V)$ and m is a nonnegative integer such that $\operatorname{null} T^m = \operatorname{null} T^{m+1}$. Let k be a positive integer. We want to prove that

$$\operatorname{null} T^{m+k} = \operatorname{null} T^{m+k+1}.$$

We already know that $\operatorname{null} T^{m+k} \subset \operatorname{null} T^{m+k+1}$. To prove the inclusion in the other direction, suppose that $v \in \operatorname{null} T^{m+k+1}$. Then

$$0 = T^{m+k+1}v = T^{m+1}(T^k v).$$

Hence

$$T^k v \in \operatorname{null} T^{m+1} = \operatorname{null} T^m.$$

Thus $0 = T^m(T^k v) = T^{m+k}v$, which means that $v \in \operatorname{null} T^{m+k}$. This implies that $\operatorname{null} T^{m+k+1} \subset \operatorname{null} T^{m+k}$, completing the proof. ■

The proposition above raises the question of whether there must exist a nonnegative integer m such that $\operatorname{null} T^m = \operatorname{null} T^{m+1}$. The proposition below shows that this equality holds at least when m equals the dimension of the vector space on which T operates.

8.6 Proposition: *If $T \in \mathcal{L}(V)$, then*

$$\operatorname{null} T^{\dim V} = \operatorname{null} T^{\dim V+1} = \operatorname{null} T^{\dim V+2} = \cdots.$$

PROOF: Suppose $T \in \mathcal{L}(V)$. To get our desired conclusion, we need only prove that $\operatorname{null} T^{\dim V} = \operatorname{null} T^{\dim V+1}$ (by 8.5). Suppose this is not true. Then, by 8.5, we have

$$\{0\} = \operatorname{null} T^0 \subsetneq \operatorname{null} T^1 \subsetneq \cdots \subsetneq \operatorname{null} T^{\dim V} \subsetneq \operatorname{null} T^{\dim V+1},$$

where the symbol $\subsetneq$ means "contained in but not equal to". At each of the strict inclusions in the chain above, the dimension must increase by at least 1. Thus $\dim \operatorname{null} T^{\dim V+1} \geq \dim V + 1$, a contradiction because a subspace of V cannot have a larger dimension than $\dim V$. ■

Now we have the promised description of generalized eigenvectors.

This corollary implies that the set of generalized eigenvectors of $T \in \mathcal{L}(V)$ corresponding to an eigenvalue λ is a subspace of V.

8.7 Corollary: *Suppose $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T. Then the set of generalized eigenvectors of T corresponding to λ equals $\operatorname{null}(T - \lambda I)^{\dim V}$.*

PROOF: If $v \in \operatorname{null}(T - \lambda I)^{\dim V}$, then clearly v is a generalized eigenvector of T corresponding to λ (by the definition of generalized eigenvector).

Conversely, suppose that $v \in V$ is a generalized eigenvector of T corresponding to λ. Thus there is a positive integer j such that

$$v \in \operatorname{null}(T - \lambda I)^j.$$

From 8.5 and 8.6 (with $T - \lambda I$ replacing T), we get $v \in \operatorname{null}(T - \lambda I)^{\dim V}$, as desired. ■

An operator is called ***nilpotent*** if some power of it equals 0. For example, the operator $N \in \mathcal{L}(\mathbf{F}^4)$ defined by

$$N(z_1, z_2, z_3, z_4) = (z_3, z_4, 0, 0)$$

is nilpotent because $N^2 = 0$. As another example, the operator of differentiation on $\mathcal{P}_m(\mathbf{R})$ is nilpotent because the $(m+1)^{\text{st}}$ derivative of any polynomial of degree at most m equals 0. Note that on this space of dimension $m+1$, we need to raise the nilpotent operator to the power $m+1$ to get 0. The next corollary shows that we never need to use a power higher than the dimension of the space.

*The Latin word **nil** means nothing or zero; the Latin word **potent** means power. Thus **nilpotent** literally means zero power.*

8.8 Corollary: *Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then $N^{\dim V} = 0$.*

PROOF: Because N is nilpotent, every vector in V is a generalized eigenvector corresponding to the eigenvalue 0. Thus from 8.7 we see that $\operatorname{null} N^{\dim V} = V$, as desired. ■

Having dealt with null spaces of powers of operators, we now turn our attention to ranges. Suppose $T \in \mathcal{L}(V)$ and k is a nonnegative integer. If $w \in \operatorname{range} T^{k+1}$, then there exists $v \in V$ with

$$w = T^{k+1}v = T^k(Tv) \in \operatorname{range} T^k.$$

Thus $\operatorname{range} T^{k+1} \subset \operatorname{range} T^k$. In other words, we have

$$V = \operatorname{range} T^0 \supset \operatorname{range} T^1 \supset \cdots \supset \operatorname{range} T^k \supset \operatorname{range} T^{k+1} \supset \cdots .$$

These inclusions go in the opposite direction from the corresponding inclusions for null spaces (8.4).

The proposition below shows that the inclusions above become equalities once the power reaches the dimension of V.

8.9 Proposition: *If $T \in \mathcal{L}(V)$, then*

$$\operatorname{range} T^{\dim V} = \operatorname{range} T^{\dim V+1} = \operatorname{range} T^{\dim V+2} = \cdots .$$

PROOF: We could prove this from scratch, but instead let's make use of the corresponding result already proved for null spaces. Suppose $m > \dim V$. Then

$$\begin{aligned} \dim \operatorname{range} T^m &= \dim V - \dim \operatorname{null} T^m \\ &= \dim V - \dim \operatorname{null} T^{\dim V} \\ &= \dim \operatorname{range} T^{\dim V}, \end{aligned}$$

where the first and third equalities come from 3.4 and the second equality comes from 8.6. We already know that range $T^{\dim V} \supset$ range T^m. We just showed that $\dim \operatorname{range} T^{\dim V} = \dim \operatorname{range} T^m$, so this implies that range $T^{\dim V}$ = range T^m, as desired. ■

The Characteristic Polynomial

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. We know that V has a basis with respect to which T has an upper-triangular matrix (see 5.13). In general, this matrix is not unique—V may have many different bases with respect to which T has an upper-triangular matrix, and with respect to these different bases we may get different upper-triangular matrices. However, the diagonal of any such matrix must contain precisely the eigenvalues of T (see 5.18). Thus if T has $\dim V$ distinct eigenvalues, then each one must appear exactly once on the diagonal of any upper-triangular matrix of T.

What if T has fewer than $\dim V$ distinct eigenvalues, as can easily happen? Then each eigenvalue must appear at least once on the diagonal of any upper-triangular matrix of T, but some of them must be repeated. Could the number of times that a particular eigenvalue is repeated depend on which basis of V we choose?

If T happens to have a diagonal matrix A with respect to some basis, then λ appears on the diagonal of A precisely $\dim \operatorname{null}(T - \lambda I)$ times, as you should verify.

You might guess that a number λ appears on the diagonal of an upper-triangular matrix of T precisely $\dim \operatorname{null}(T - \lambda I)$ times. In general, this is false. For example, consider the operator on $\mathbf{C}^2$ whose matrix with respect to the standard basis is the upper-triangular matrix

$$\begin{bmatrix} 5 & 1 \\ 0 & 5 \end{bmatrix}.$$

For this operator, $\dim \operatorname{null}(T - 5I) = 1$ but 5 appears on the diagonal twice. Note, however, that $\dim \operatorname{null}(T - 5I)^2 = 2$ for this operator. This example illustrates the general situation—a number λ appears on the diagonal of an upper-triangular matrix of T precisely $\dim \operatorname{null}(T - \lambda I)^{\dim V}$ times, as we will show in the following theorem. Because $\operatorname{null}(T - \lambda I)^{\dim V}$ depends only on T and λ and not on a choice of basis, this implies that the number of times an eigenvalue is repeated on the diagonal of an upper-triangular matrix of T is independent of which particular basis we choose. This result will be our key tool in analyzing the structure of an operator on a complex vector space.

8.10 Theorem: *Let $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. Then for every basis of V with respect to which T has an upper-triangular matrix, λ appears on the diagonal of the matrix of T precisely* $\dim \operatorname{null}(T - \lambda I)^{\dim V}$ *times.*

PROOF: We will assume, without loss of generality, that $\lambda = 0$ (once the theorem is proved in this case, the general case is obtained by replacing T with $T - \lambda I$).

For convenience let $n = \dim V$. We will prove this theorem by induction on n. Clearly the desired result holds if $n = 1$. Thus we can assume that $n > 1$ and that the desired result holds on spaces of dimension $n - 1$.

Suppose $(v_1, \dots, v_n)$ is a basis of V with respect to which T has an upper-triangular matrix

Recall that an asterisk is often used in matrices to denote entries that we do not know or care about.

8.11
$$\begin{bmatrix} \lambda_1 & & & * \\ & \ddots & & \\ & & \lambda_{n-1} & \\ 0 & & & \lambda_n \end{bmatrix}.$$

Let $U = \operatorname{span}(v_1, \dots, v_{n-1})$. Clearly U is invariant under T (see 5.12), and the matrix of $T|_U$ with respect to the basis $(v_1, \dots, v_{n-1})$ is

8.12
$$\begin{bmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_{n-1} \end{bmatrix}.$$

Thus, by our induction hypothesis, 0 appears on the diagonal of 8.12 $\dim \operatorname{null}(T|_U)^{n-1}$ times. We know that $\operatorname{null}(T|_U)^{n-1} = \operatorname{null}(T|_U)^n$ (because U has dimension $n - 1$; see 8.6). Hence

8.13 0 appears on the diagonal of 8.12 $\dim \operatorname{null}(T|_U)^n$ times.

The proof breaks into two cases, depending on whether $\lambda_n = 0$. First consider the case where $\lambda_n \neq 0$. We will show that in this case

8.14
$$\operatorname{null} T^n \subset U.$$

Once this has been verified, we will know that $\operatorname{null} T^n = \operatorname{null}(T|_U)^n$, and hence 8.13 will tell us that 0 appears on the diagonal of 8.11 exactly $\dim \operatorname{null} T^n$ times, completing the proof in the case where $\lambda_n \neq 0$.

Because $\mathcal{M}(T)$ is given by 8.11, we have

$$\mathcal{M}(T^n) = \mathcal{M}(T)^n = \begin{bmatrix} {\lambda_1}^n & & & * \\ & \ddots & & \\ & & {\lambda_{n-1}}^n & \\ 0 & & & {\lambda_n}^n \end{bmatrix}.$$

This shows that

$$T^n v_n = u + {\lambda_n}^n v_n$$

for some $u \in U$. To prove 8.14 (still assuming that $\lambda_n \neq 0$), suppose $v \in \operatorname{null} T^n$. We can write v in the form

$$v = \tilde{u} + a v_n,$$

where $\tilde{u} \in U$ and $a \in \mathbf{F}$. Thus

$$0 = T^n v = T^n \tilde{u} + a T^n v_n = T^n \tilde{u} + a u + a {\lambda_n}^n v_n.$$

Because $T^n \tilde{u}$ and au are in U and $v_n \notin U$, this implies that $a{\lambda_n}^n = 0$. However, $\lambda_n \neq 0$, so $a = 0$. Thus $v = \tilde{u} \in U$, completing the proof of 8.14.

Now consider the case where $\lambda_n = 0$. In this case we will show that

8.15 $$\dim \operatorname{null} T^n = \dim \operatorname{null} (T|_U)^n + 1,$$

which along with 8.13 will complete the proof when $\lambda_n = 0$.

Using the formula for the dimension of the sum of two subspaces (2.18), we have

$$\begin{aligned} \dim \operatorname{null} T^n &= \dim(U \cap \operatorname{null} T^n) + \dim(U + \operatorname{null} T^n) - \dim U \\ &= \dim \operatorname{null}(T|_U)^n + \dim(U + \operatorname{null} T^n) - (n-1). \end{aligned}$$

Suppose we can prove that $\operatorname{null} T^n$ contains a vector not in U. Then

$$n = \dim V \geq \dim(U + \operatorname{null} T^n) > \dim U = n - 1,$$

which implies that $\dim(U + \operatorname{null} T^n) = n$, which when combined with the formula above for $\dim \operatorname{null} T^n$ gives 8.15, as desired. Thus to complete the proof, we need only show that $\operatorname{null} T^n$ contains a vector not in U.

Let's think about how we might find a vector in $\operatorname{null} T^n$ that is not in U. We might try a vector of the form

$$u - v_n,$$

where $u \in U$. At least we are guaranteed that any such vector is not in U. Can we choose $u \in U$ such that the vector above is in $\operatorname{null} T^n$? Let's compute:

$$T^n(u - v_n) = T^n u - T^n v_n.$$

To make the above vector equal 0, we must choose (if possible) $u \in U$ such that $T^n u = T^n v_n$. We can do this if $T^n v_n \in \operatorname{range}(T|_U)^n$. Because 8.11 is the matrix of T with respect to $(v_1, \dots, v_n)$, we see that $Tv_n \in U$ (recall that we are considering the case where $\lambda_n = 0$). Thus

$$T^n v_n = T^{n-1}(Tv_n) \in \operatorname{range}(T|_U)^{n-1} = \operatorname{range}(T|_U)^n,$$

where the last equality comes from 8.9. In other words, we can indeed choose $u \in U$ such that $u - v_n \in \operatorname{null} T^n$, completing the proof. ■

Suppose $T \in \mathcal{L}(V)$. The ***multiplicity*** of an eigenvalue λ of T is defined to be the dimension of the subspace of generalized eigenvectors corresponding to λ. In other words, the multiplicity of an eigenvalue λ of T equals $\dim \operatorname{null}(T - \lambda I)^{\dim V}$. If T has an upper-triangular matrix with respect to some basis of V (as always happens when $\mathbf{F} = \mathbf{C}$), then the multiplicity of λ is simply the number of times λ appears on the diagonal of this matrix (by the last theorem).

Our definition of multiplicity has a clear connection with the geometric behavior of T. Most texts define multiplicity in terms of the multiplicity of the roots of a certain polynomial defined by determinants. These two definitions turn out to be equivalent.

As an example of multiplicity, consider the operator $T \in \mathcal{L}(\mathbf{F}^3)$ defined by

8.16 $$T(z_1, z_2, z_3) = (0, z_1, 5z_3).$$

You should verify that 0 is an eigenvalue of T with multiplicity 2, that 5 is an eigenvalue of T with multiplicity 1, and that T has no additional eigenvalues. As another example, if $T \in \mathcal{L}(\mathbf{F}^3)$ is the operator whose matrix is

8.17 $$\begin{bmatrix} 6 & 7 & 7 \\ 0 & 6 & 7 \\ 0 & 0 & 7 \end{bmatrix},$$

then 6 is an eigenvalue of T with multiplicity 2 and 7 is an eigenvalue of T with multiplicity 1 (this follows from the last theorem).

In each of the examples above, the sum of the multiplicities of the eigenvalues of T equals 3, which is the dimension of the domain of T. The next proposition shows that this always happens on a complex vector space.

8.18 Proposition: *If V is a complex vector space and $T \in \mathcal{L}(V)$, then the sum of the multiplicities of all the eigenvalues of T equals* $\dim V$.

PROOF: Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then there is a basis of V with respect to which the matrix of T is upper triangular (by 5.13). The multiplicity of λ equals the number of times λ appears on the diagonal of this matrix (from 8.10). Because the diagonal of this matrix has length $\dim V$, the sum of the multiplicities of all the eigenvalues of T must equal $\dim V$. ■

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ denote the distinct eigenvalues of T. Let d_j denote the multiplicity of λ_j as an eigenvalue of T. The polynomial

$$(z - \lambda_1)^{d_1} \dots (z - \lambda_m)^{d_m}$$

Most texts define the characteristic polynomial using determinants. The approach taken here, which is considerably simpler, leads to an easy proof of the Cayley-Hamilton theorem.

is called the ***characteristic polynomial*** of T. Note that the degree of the characteristic polynomial of T equals $\dim V$ (from 8.18). Obviously the roots of the characteristic polynomial of T equal the eigenvalues of T. As an example, the characteristic polynomial of the operator $T \in \mathcal{L}(\mathbf{C}^3)$ defined by 8.16 equals $z^2(z-5)$.

Here is another description of the characteristic polynomial of an operator on a complex vector space. Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Consider any basis of V with respect to which T has an upper-triangular matrix of the form

8.19 $$\mathcal{M}(T) = \begin{bmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}.$$

Then the characteristic polynomial of T is given by

$$(z - \lambda_1) \dots (z - \lambda_n);$$

this follows immediately from 8.10. As an example of this procedure, if $T \in \mathcal{L}(\mathbf{C}^3)$ is the operator whose matrix is given by 8.17, then the characteristic polynomial of T equals $(z-6)^2(z-7)$.

In the next chapter we will define the characteristic polynomial of an operator on a real vector space and prove that the next result also holds for real vector spaces.

8.20 Cayley-Hamilton Theorem: *Suppose that V is a complex vector space and $T \in \mathcal{L}(V)$. Let q denote the characteristic polynomial of T. Then $q(T) = 0$.*

The English mathematician Arthur Cayley published three mathematics papers before he completed his undergraduate degree in 1842. The Irish mathematician William Hamilton was made a professor in 1827 when he was 22 years old and still an undergraduate!

PROOF: Suppose $(v_1, \dots, v_n)$ is a basis of V with respect to which the matrix of T has the upper-triangular form 8.19. To prove that $q(T) = 0$, we need only show that $q(T)v_j = 0$ for $j = 1, \dots, n$. To do this, it suffices to show that

8.21 $$(T - \lambda_1 I) \dots (T - \lambda_j I)v_j = 0$$

for $j = 1, \dots, n$.

We will prove 8.21 by induction on j. To get started, suppose $j = 1$. Because $\mathcal{M}(T, (v_1, \dots, v_n))$ is given by 8.19, we have $Tv_1 = \lambda_1 v_1$, giving 8.21 when $j = 1$.

Now suppose that $1 < j \le n$ and that

$$\begin{aligned} 0 &= (T - \lambda_1 I)v_1 \\ &= (T - \lambda_1 I)(T - \lambda_2 I)v_2 \\ &\;\;\vdots \\ &= (T - \lambda_1 I) \dots (T - \lambda_{j-1} I)v_{j-1}. \end{aligned}$$

Because $\mathcal{M}(T, (v_1, \dots, v_n))$ is given by 8.19, we see that

$$(T - \lambda_j I)v_j \in \operatorname{span}(v_1, \dots, v_{j-1}).$$

Thus, by our induction hypothesis, $(T - \lambda_1 I) \dots (T - \lambda_{j-1} I)$ applied to $(T - \lambda_j I)v_j$ gives 0. In other words, 8.21 holds, completing the proof. ■

Decomposition of an Operator

We saw earlier that the domain of an operator might not decompose into invariant subspaces consisting of eigenvectors of the operator, even on a complex vector space. In this section we will see that every operator on a complex vector space has enough generalized eigenvectors to provide a decomposition.

We observed earlier that if $T \in \mathcal{L}(V)$, then $\operatorname{null} T$ is invariant under T. Now we show that the null space of any polynomial of T is also invariant under T.

8.22 Proposition: *If $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbf{F})$, then $\operatorname{null} p(T)$ is invariant under T.*

PROOF: Suppose $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbf{F})$. Let $v \in \operatorname{null} p(T)$. Then $p(T)v = 0$. Thus

$$(p(T))(Tv) = T(p(T)v) = T(0) = 0,$$

and hence $Tv \in \operatorname{null} p(T)$. Thus $\operatorname{null} p(T)$ is invariant under T, as desired. ■

The following major structure theorem shows that every operator on a complex vector space can be thought of as composed of pieces, each of which is a nilpotent operator plus a scalar multiple of the identity. Actually we have already done all the hard work, so at this point the proof is easy.

8.23 Theorem: *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T, and let $U_1, \dots, U_m$ be the corresponding subspaces of generalized eigenvectors. Then*

(a) $V = U_1 \oplus \cdots \oplus U_m$;

(b) *each U_j is invariant under T;*

(c) *each $(T - \lambda_j I)|_{U_j}$ is nilpotent.*

PROOF: Note that $U_j = \operatorname{null}(T - \lambda_j I)^{\dim V}$ for each j (by 8.7). From 8.22 (with $p(z) = (z - \lambda_j)^{\dim V}$), we get (b). Obviously (c) follows from the definitions.

To prove (a), recall that the multiplicity of λ_j as an eigenvalue of T is defined to be $\dim U_j$. The sum of these multiplicities equals $\dim V$ (see 8.18); thus

8.24 $$\dim V = \dim U_1 + \cdots + \dim U_m.$$

Let $U = U_1 + \cdots + U_m$. Clearly U is invariant under T. Thus we can define $S \in \mathcal{L}(U)$ by

$$S = T|_U.$$

Note that S has the same eigenvalues, with the same multiplicities, as T because all the generalized eigenvectors of T are in U, the domain of S. Thus applying 8.18 to S, we get

$$\dim U = \dim U_1 + \cdots + \dim U_m.$$

This equation, along with 8.24, shows that $\dim V = \dim U$. Because U is a subspace of V, this implies that $V = U$. In other words,

$$V = U_1 + \cdots + U_m.$$

This equation, along with 8.24, allows us to use 2.19 to conclude that (a) holds, completing the proof. ■

As we know, an operator on a complex vector space may not have enough eigenvectors to form a basis for the domain. The next result shows that on a complex vector space there are enough generalized eigenvectors to do this.

8.25 Corollary: *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then there is a basis of V consisting of generalized eigenvectors of T.*

PROOF: Choose a basis for each U_j in 8.23. Put all these bases together to form a basis of V consisting of generalized eigenvectors of T. ■

Given an operator T on V, we want to find a basis of V so that the matrix of T with respect to this basis is as simple as possible, meaning that the matrix contains many 0's. We begin by showing that if N is nilpotent, we can choose a basis of V such that the matrix of N with respect to this basis has more than half of its entries equal to 0.

8.26 Lemma: *Suppose N is a nilpotent operator on V. Then there is a basis of V with respect to which the matrix of N has the form*

If V is complex vector space, a proof of this lemma follows easily from Exercise 6 in this chapter, 5.13, and 5.18. But the proof given here uses simpler ideas than needed to prove 5.13, and it works for both real and complex vector spaces.

8.27
$$\begin{bmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{bmatrix};$$

here all entries on and below the diagonal are 0's.

PROOF: First choose a basis of $\operatorname{null} N$. Then extend this to a basis of $\operatorname{null} N^2$. Then extend to a basis of $\operatorname{null} N^3$. Continue in this fashion, eventually getting a basis of V (because $\operatorname{null} N^m = V$ for m sufficiently large).

Now let's think about the matrix of N with respect to this basis. The first column, and perhaps additional columns at the beginning, consists of all 0's because the corresponding basis vectors are in $\operatorname{null} N$. The next set of columns comes from basis vectors in $\operatorname{null} N^2$. Applying N to any such vector, we get a vector in $\operatorname{null} N$; in other words, we get a vector that is a linear combination of the previous basis vectors. Thus all nonzero entries in these columns must lie above the diagonal. The next set of columns come from basis vectors in $\operatorname{null} N^3$. Applying N to any such vector, we get a vector in $\operatorname{null} N^2$; in other words, we get a vector that is a linear combination of the previous basis vectors. Thus, once again, all nonzero entries in these columns must lie above the diagonal. Continue in this fashion to complete the proof. ■

Note that in the next theorem we get many more zeros in the matrix of T than are needed to make it upper triangular.

8.28 Theorem: *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. Then there is a basis of V with respect to which T has a block diagonal matrix of the form*

$$\begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{bmatrix},$$

where each A_j is an upper-triangular matrix of the form

8.29
$$A_j = \begin{bmatrix} \lambda_j & & * \\ & \ddots & \\ 0 & & \lambda_j \end{bmatrix}.$$

PROOF: For $j = 1, \dots, m$, let U_j denote the subspace of generalized eigenvectors of T corresponding to λ_j. Thus $(T - \lambda_j I)|_{U_j}$ is nilpotent (see 8.23(c)). For each j, choose a basis of U_j such that the matrix of $(T - \lambda_j I)|_{U_j}$ with respect to this basis is as in 8.26. Thus the matrix of $T|_{U_j}$ with respect to this basis will look like 8.29. Putting the bases for the U_j's together gives a basis for V (by 8.23(a)). The matrix of T with respect to this basis has the desired form. ■

Square Roots

Recall that a square root of an operator $T \in \mathcal{L}(V)$ is an operator $S \in \mathcal{L}(V)$ such that $S^2 = T$. As an application of the main structure theorem from the last section, in this section we will show that every invertible operator on a complex vector space has a square root.

Every complex number has a square root, but not every operator on a complex vector space has a square root. An example of an operator on $\mathbf{C}^3$ that has no square root is given in Exercise 4 in this chapter. The noninvertibility of that particular operator is no accident, as we will soon see. We begin by showing that the identity plus a nilpotent operator always has a square root.

8.30 Lemma: *Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then $I + N$ has a square root.*

PROOF: Consider the Taylor series for the function $\sqrt{1+x}$:

8.31 $$\sqrt{1+x} = 1 + a_1x + a_2x^2 + \cdots .$$

Because $a_1 = 1/2$, this formula shows that $1 + x/2$ is a good estimate for $\sqrt{1+x}$ when x is small.

We will not find an explicit formula for all the coefficients or worry about whether the infinite sum converges because we are using this equation only as motivation, not as a formal part of the proof.

Because N is nilpotent, $N^m = 0$ for some positive integer m. In 8.31, suppose we replace x with N and 1 with I. Then the infinite sum on the right side becomes a finite sum (because $N^j = 0$ for all $j \geq m$). In other words, we guess that there is a square root of $I + N$ of the form

$$I + a_1N + a_2N^2 + \cdots + a_{m-1}N^{m-1}.$$

Having made this guess, we can try to choose $a_1, a_2, \ldots, a_{m-1}$ so that the operator above has its square equal to $I + N$. Now

$$\begin{aligned}(I{+}a_1N + a_2N^2 + a_3N^3 + \cdots + a_{m-1}N^{m-1})^2 \\ = I + 2a_1N + (2a_2 + {a_1}^2)N^2 + (2a_3 + 2a_1a_2)N^3 + \cdots \\ + (2a_{m-1} + \text{terms involving } a_1, \ldots, a_{m-2})N^{m-1}.\end{aligned}$$

We want the right side of the equation above to equal $I + N$. Hence choose a_1 so that $2a_1 = 1$ (thus $a_1 = 1/2$). Next, choose a_2 so that $2a_2 + {a_1}^2 = 0$ (thus $a_2 = -1/8$). Then choose a_3 so that the coefficient of N^3 on the right side of the equation above equals 0 (thus $a_3 = 1/16$).

Continue in this fashion for $j = 4, \dots, m-1$, at each step solving for a_j so that the coefficient of N^j on the right side of the equation above equals 0. Actually we do not care about the explicit formula for the a_j's. We need only know that some choice of the a_j's gives a square root of $I + N$. ■

The previous lemma is valid on real and complex vector spaces. However, the next result holds only on complex vector spaces.

On real vector spaces there exist invertible operators that have no square roots. For example, the operator of multiplication by -1 on **R** *has no square root because no real number has its square equal to -1.*

8.32 Theorem: *Suppose V is a complex vector space. If $T \in \mathcal{L}(V)$ is invertible, then T has a square root.*

PROOF: Suppose $T \in \mathcal{L}(V)$ is invertible. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T, and let $U_1, \dots, U_m$ be the corresponding subspaces of generalized eigenvectors. For each j, there exists a nilpotent operator $N_j \in \mathcal{L}(U_j)$ such that $T|_{U_j} = \lambda_j I + N_j$ (see 8.23(c)). Because T is invertible, none of the λ_j's equals 0, so we can write

$$T|_{U_j} = \lambda_j (I + \frac{N_j}{\lambda_j})$$

for each j. Clearly N_j/λ_j is nilpotent, and so $I + N_j/\lambda_j$ has a square root (by 8.30). Multiplying a square root of the complex number λ_j by a square root of $I + N_j/\lambda_j$, we obtain a square root S_j of $T|_{U_j}$.

A typical vector $v \in V$ can be written uniquely in the form

$$v = u_1 + \cdots + u_m,$$

where each $u_j \in U_j$ (see 8.23). Using this decomposition, define an operator $S \in \mathcal{L}(V)$ by

$$Sv = S_1 u_1 + \cdots + S_m u_m.$$

You should verify that this operator S is a square root of T, completing the proof. ■

By imitating the techniques in this section, you should be able to prove that if V is a complex vector space and $T \in \mathcal{L}(V)$ is invertible, then T has a k^{th}-root for every positive integer k.

The Minimal Polynomial

As we will soon see, given an operator on a finite-dimensional vector space, there is a unique monic polynomial of smallest degree that when applied to the operator gives 0. This polynomial is called the minimal polynomial of the operator and is the focus of attention in this section.

*A **monic polynomial** is a polynomial whose highest degree coefficient equals* 1*. For example,* $2 + 3z^2 + z^8$ *is a monic polynomial.*

Suppose $T \in \mathcal{L}(V)$, where $\dim V = n$. Then

$$(I, T, T^2, \dots, T^{n^2})$$

cannot be linearly independent in $\mathcal{L}(V)$ because $\mathcal{L}(V)$ has dimension n^2 (see 3.20) and we have $n^2 + 1$ operators. Let m be the smallest positive integer such that

8.33 $$(I, T, T^2, \dots, T^m)$$

is linearly dependent. The linear dependence lemma (2.4) implies that one of the operators in the list above is a linear combination of the previous ones. Because m was chosen to be the smallest positive integer such that 8.33 is linearly dependent, we conclude that T^m is a linear combination of $(I, T, T^2, \dots, T^{m-1})$. Thus there exist scalars $a_0, a_1, a_2, \dots, a_{m-1} \in \mathbf{F}$ such that

$$a_0 I + a_1 T + a_2 T^2 + \cdots + a_{m-1} T^{m-1} + T^m = 0.$$

The choice of scalars $a_0, a_1, a_2, \dots, a_{m-1} \in \mathbf{F}$ above is unique because two different such choices would contradict our choice of m (subtracting two different equations of the form above, we would have a linearly dependent list shorter than 8.33). The polynomial

$$a_0 + a_1 z + a_2 z^2 + \cdots + a_{m-1} z^{m-1} + z^m$$

is called the ***minimal polynomial*** of T. It is the monic polynomial $p \in \mathcal{P}(\mathbf{F})$ of smallest degree such that $p(T) = 0$.

For example, the minimal polynomial of the identity operator I is $z - 1$. The minimal polynomial of the operator on $\mathbf{F}^2$ whose matrix equals $\begin{bmatrix} 4 & 1 \\ 0 & 5 \end{bmatrix}$ is $20 - 9z + z^2$, as you should verify.

Clearly the degree of the minimal polynomial of each operator on V is at most $(\dim V)^2$. The Cayley-Hamilton theorem (8.20) tells us that if V is a complex vector space, then the minimal polynomial of each operator on V has degree at most $\dim V$. This remarkable improvement also holds on real vector spaces, as we will see in the next chapter.

A polynomial $p \in \mathcal{P}(\mathbf{F})$ is said to ***divide*** a polynomial $q \in \mathcal{P}(\mathbf{F})$ if there exists a polynomial $s \in \mathcal{P}(\mathbf{F})$ such that $q = sp$. In other words, p divides q if we can take the remainder r in 4.6 to be 0. For example, the polynomial $(1 + 3z)^2$ divides $5 + 32z + 57z^2 + 18z^3$ because $5 + 32z + 57z^2 + 18z^3 = (2z + 5)(1 + 3z)^2$. Obviously every nonzero constant polynomial divides every polynomial.

Note that $(z - \lambda)$ divides a polynomial q if and only if λ is a root of q. This follows immediately from 4.1.

The next result completely characterizes the polynomials that when applied to an operator give the 0 operator.

8.34 Theorem: *Let $T \in \mathcal{L}(V)$ and let $q \in \mathcal{P}(\mathbf{F})$. Then $q(T) = 0$ if and only if the minimal polynomial of T divides q.*

PROOF: Let p denote the minimal polynomial of T.

First we prove the easy direction. Suppose that p divides q. Thus there exists a polynomial $s \in \mathcal{P}(\mathbf{F})$ such that $q = sp$. We have

$$q(T) = s(T)p(T) = s(T)0 = 0,$$

as desired.

To prove the other direction, suppose that $q(T) = 0$. By the division algorithm (4.5), there exist polynomials $s, r \in \mathcal{P}(\mathbf{F})$ such that

8.35 $$q = sp + r$$

and $\deg r < \deg p$. We have

$$0 = q(T) = s(T)p(T) + r(T) = r(T).$$

Because p is the minimal polynomial of T and $\deg r < \deg p$, the equation above implies that $r = 0$. Thus 8.35 becomes the equation $q = sp$, and hence p divides q, as desired. ■

Now we describe the eigenvalues of an operator in terms of its minimal polynomial.

8.36 Theorem: *Let $T \in \mathcal{L}(V)$. Then the roots of the minimal polynomial of T are precisely the eigenvalues of T.*

PROOF: Let

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_{m-1} z^{m-1} + z^m$$

be the minimal polynomial of T.

First suppose that $\lambda \in \mathbf{F}$ is a root of p. Then the minimal polynomial of T can be written in the form

$$p(z) = (z - \lambda) q(z),$$

where q is a monic polynomial with coefficients in $\mathbf{F}$ (see 4.1). Because $p(T) = 0$, we have

$$0 = (T - \lambda I)(q(T)v)$$

for all $v \in V$. Because the degree of q is less than the degree of the minimal polynomial p, there must exist at least one vector $v \in V$ such that $q(T)v \neq 0$. The equation above thus implies that λ is an eigenvalue of T, as desired.

To prove the other direction, now suppose that $\lambda \in \mathbf{F}$ is an eigenvalue of T. Let v be a nonzero vector in V such that $Tv = \lambda v$. Repeated applications of T to both sides of this equation show that $T^j v = \lambda^j v$ for every nonnegative integer j. Thus

$$\begin{aligned} 0 = p(T)v &= (a_0 + a_1 T + a_2 T^2 + \cdots + a_{m-1} T^{m-1} + T^m)v \\ &= (a_0 + a_1 \lambda + a_2 \lambda^2 + \cdots + a_{m-1} \lambda^{m-1} + \lambda^m)v \\ &= p(\lambda)v. \end{aligned}$$

Because $v \neq 0$, the equation above implies that $p(\lambda) = 0$, as desired. ■

Suppose we are given, in concrete form, the matrix (with respect to some basis) of some operator $T \in \mathcal{L}(V)$. To find the minimal polynomial of T, consider

$$(\mathcal{M}(I), \mathcal{M}(T), \mathcal{M}(T)^2, \ldots, \mathcal{M}(T)^m)$$

for $m = 1, 2, \ldots$ until this list is linearly dependent. Then find the scalars $a_0, a_1, a_2, \ldots, a_{m-1} \in \mathbf{F}$ such that

$$a_0 \mathcal{M}(I) + a_1 \mathcal{M}(T) + a_2 \mathcal{M}(T)^2 + \cdots + a_{m-1} \mathcal{M}(T)^{m-1} + \mathcal{M}(T)^m = 0.$$

You can think of this as a system of $(\dim V)^2$ equations in m variables $a_0, a_1, \ldots, a_{m-1}$.

The scalars $a_0, a_1, a_2, \ldots, a_{m-1}, 1$ will then be the coefficients of the minimal polynomial of T. All this can be computed using a familiar process such as Gaussian elimination.

For example, consider the operator T on $\mathbf{C}^5$ whose matrix is given by

8.37
$$\begin{bmatrix} 0 & 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Because of the large number of 0's in this matrix, Gaussian elimination is not needed here. Simply compute powers of $\mathcal{M}(T)$ and notice that there is no linear dependence until the fifth power. Do the computations and you will see that the minimal polynomial of T equals

8.38
$$z^5 - 6z + 3.$$

Now what about the eigenvalues of this particular operator? From 8.36, we see that the eigenvalues of T equal the solutions to the equation

$$z^5 - 6z + 3 = 0.$$

Unfortunately no solution to this equation can be computed using rational numbers, arbitrary roots of rational numbers, and the usual rules of arithmetic (a proof of this would take us considerably beyond linear algebra). Thus we cannot find an exact expression for any eigenvalues of T in any familiar form, though numeric techniques can give good approximations for the eigenvalues of T. The numeric techniques, which we will not discuss here, show that the eigenvalues for this particular operator are approximately

$$-1.67, \quad 0.51, \quad 1.40, \quad -0.12 + 1.59i, \quad -0.12 - 1.59i.$$

Note that the nonreal eigenvalues occur as a pair, with each the complex conjugate of the other, as expected for the roots of a polynomial with real coefficients (see 4.10).

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. The Cayley-Hamilton theorem (8.20) and 8.34 imply that the minimal polynomial of T divides the characteristic polynomial of T. Both these polynomials are monic. Thus if the minimal polynomial of T has degree $\dim V$, then it must equal the characteristic polynomial of T. For example, if T is the operator on $\mathbf{C}^5$ whose matrix is given by 8.37, then the characteristic polynomial of T, as well as the minimal polynomial of T, is given by 8.38.

Jordan Form

We know that if V is a complex vector space, then for every $T \in \mathcal{L}(V)$ there is a basis of V with respect to which T has a nice upper-triangular matrix (see 8.28). In this section we will see that we can do even better—there is a basis of V with respect to which the matrix of T contains zeros everywhere except possibly on the diagonal and the line directly above the diagonal.

We begin by describing the nilpotent operators. Consider, for example, the nilpotent operator $N \in \mathcal{L}(\mathbf{F}^n)$ defined by

$$N(z_1, \dots, z_n) = (0, z_1, \dots, z_{n-1}).$$

If $v = (1, 0, \dots, 0)$, then clearly $(v, Nv, \dots, N^{n-1}v)$ is a basis of $\mathbf{F}^n$ and $(N^{n-1}v)$ is a basis of null N, which has dimension 1.

As another example, consider the nilpotent operator $N \in \mathcal{L}(\mathbf{F}^5)$ defined by

8.39 $$N(z_1, z_2, z_3, z_4, z_5) = (0, z_1, z_2, 0, z_4).$$

Unlike the nilpotent operator discussed in the previous paragraph, for this nilpotent operator there does not exist a vector $v \in \mathbf{F}^5$ such that $(v, Nv, N^2v, N^3v, N^4v)$ is a basis of $\mathbf{F}^5$. However, if $v_1 = (1, 0, 0, 0, 0)$ and $v_2 = (0, 0, 0, 1, 0)$, then $(v_1, Nv_1, N^2v_1, v_2, Nv_2)$ is a basis of $\mathbf{F}^5$ and (N^2v_1, Nv_2) is a basis of null N, which has dimension 2.

Suppose $N \in \mathcal{L}(V)$ is nilpotent. For each nonzero vector $v \in V$, let $m(v)$ denote the largest nonnegative integer such that $N^{m(v)}v \neq 0$. For example, if $N \in \mathcal{L}(\mathbf{F}^5)$ is defined by 8.39, then $m(1, 0, 0, 0, 0) = 2$.

Obviously $m(v)$ depends on N as well as on v, but the choice of N will be clear from the context.

The lemma below shows that every nilpotent operator $N \in \mathcal{L}(V)$ behaves similarly to the example defined by 8.39, in the sense that there is a finite collection of vectors $v_1, \dots, v_k \in V$ such that the nonzero vectors of the form $N^j v_r$ form a basis of V; here r varies from 1 to k and j varies from 0 to $m(v_r)$.

8.40 Lemma: *If $N \in \mathcal{L}(V)$ is nilpotent, then there exist vectors $v_1, \dots, v_k \in V$ such that*

(a) $(v_1, Nv_1, \dots, N^{m(v_1)}v_1, \dots, v_k, Nv_k, \dots, N^{m(v_k)}v_k)$ *is a basis of V;*

(b) $(N^{m(v_1)}v_1, \dots, N^{m(v_k)}v_k)$ *is a basis of* null N.

PROOF: Suppose N is nilpotent. Then N is not injective and thus $\dim \operatorname{range} N < \dim V$ (see 3.21). By induction on the dimension of V, we can assume that the lemma holds on all vector spaces of smaller dimension. Using $\operatorname{range} N$ in place of V and $N|_{\operatorname{range} N}$ in place of N, we thus have vectors $u_1, \dots, u_j \in \operatorname{range} N$ such that

(i) $(u_1, Nu_1, \dots, N^{m(u_1)}u_1, \dots, u_j, Nu_j, \dots, N^{m(u_j)}u_j)$ is a basis of $\operatorname{range} N$;

(ii) $(N^{m(u_1)}u_1, \dots, N^{m(u_j)}u_j)$ is a basis of $\operatorname{null} N \cap \operatorname{range} N$.

Because each $u_r \in \operatorname{range} N$, we can choose $v_1, \dots, v_j \in V$ such that $Nv_r = u_r$ for each r. Note that $m(v_r) = m(u_r) + 1$ for each r.

The existence of a subspace W with this property follows from 2.13.

Let W be a subspace of $\operatorname{null} N$ such that

8.41 $$\operatorname{null} N = (\operatorname{null} N \cap \operatorname{range} N) \oplus W$$

and choose a basis of W, which we will label $(v_{j+1}, \dots, v_k)$. Because $v_{j+1}, \dots, v_k \in \operatorname{null} N$, we have $m(v_{j+1}) = \cdots = m(v_k) = 0$.

Having constructed $v_1, \dots, v_k$, we now need to show that (a) and (b) hold. We begin by showing that the alleged basis in (a) is linearly independent. To do this, suppose

8.42 $$0 = \sum_{r=1}^{k} \sum_{s=0}^{m(v_r)} a_{r,s} N^s(v_r),$$

where each $a_{r,s} \in \mathbf{F}$. Applying N to both sides of the equation above, we get

$$\begin{aligned} 0 &= \sum_{r=1}^{k} \sum_{s=0}^{m(v_r)} a_{r,s} N^{s+1}(v_r) \\ &= \sum_{r=1}^{j} \sum_{s=0}^{m(u_r)} a_{r,s} N^s(u_r). \end{aligned}$$

The last equation, along with (i), implies that $a_{r,s} = 0$ for $1 \le r \le j$, $0 \le s \le m(v_r) - 1$. Thus 8.42 reduces to the equation

$$\begin{aligned} 0 = {} & a_{1,m(v_1)} N^{m(v_1)} v_1 + \cdots + a_{j,m(v_j)} N^{m(v_j)} v_j \\ & + a_{j+1,0} v_{j+1} + \cdots + a_{k,0} v_k. \end{aligned}$$

The terms on the first line on the right are all in null $N \cap$ range N; the terms on the second line are all in W. Thus the last equation and 8.41 imply that

8.43
$$\begin{aligned} 0 &= a_{1,m(v_1)}N^{m(v_1)}v_1 + \cdots + a_{j,m(v_j)}N^{m(v_j)}v_j \\ &= a_{1,m(v_1)}N^{m(u_1)}u_1 + \cdots + a_{j,m(v_j)}N^{m(u_j)}u_j \end{aligned}$$

and

8.44
$$0 = a_{j+1,0}v_{j+1} + \cdots + a_{k,0}v_k.$$

Now 8.43 and (ii) imply that $a_{1,m(v_1)} = \cdots = a_{j,m(v_j)} = 0$. Because $(v_{j+1}, \ldots, v_k)$ is a basis of W, 8.44 implies that $a_{j+1,0} = \cdots = a_{k,0} = 0$. Thus all the a's equal 0, and hence the list of vectors in (a) is linearly independent.

Clearly (ii) implies that $\dim(\operatorname{null} N \cap \operatorname{range} N) = j$. Along with 8.41, this implies that

8.45
$$\dim \operatorname{null} N = k.$$

Clearly (i) implies that

8.46
$$\begin{aligned} \dim \operatorname{range} N &= \sum_{r=0}^{j} (m(u_r) + 1) \\ &= \sum_{r=0}^{j} m(v_r). \end{aligned}$$

The list of vectors in (a) has length

$$\begin{aligned} \sum_{r=0}^{k} (m(v_r) + 1) &= k + \sum_{r=0}^{j} m(v_r) \\ &= \dim \operatorname{null} N + \dim \operatorname{range} N \\ &= \dim V, \end{aligned}$$

where the second equality comes from 8.45 and 8.46, and the third equality comes from 3.4. The last equation shows that the list of vectors in (a) has length $\dim V$; because this list is linearly independent, it is a basis of V (see 2.17), completing the proof of (a).

Finally, note that

$$(N^{m(v_1)}v_1, \ldots, N^{m(v_k)}v_k) = (N^{m(u_1)}u_1, \ldots, N^{m(u_j)}u_j, v_{j+1}, \ldots, v_k).$$

Now (ii) and 8.41 show that the last list above is a basis of null N, completing the proof of (b). ■

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

$$\begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{bmatrix},$$

where each A_j is an upper-triangular matrix of the form

$$A_j = \begin{bmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{bmatrix}.$$

To understand why each λ_j must be an eigenvalue of T, see 5.18.

In each A_j, the diagonal is filled with some eigenvalue λ_j of T, the line directly above the diagonal is filled with 1's, and all other entries are 0 (A_j may be just a 1-by-1 block consisting of just some eigenvalue).

Because there exist operators on real vector spaces that have no eigenvalues, there exist operators on real vector spaces for which there is no corresponding Jordan basis. Thus the hypothesis that V is a complex vector space is required for the next result, even though the previous lemma holds on both real and complex vector spaces.

The French mathematician Camille Jordan first published a proof of this theorem in 1870.

8.47 Theorem: *Suppose V is a complex vector space. If $T \in \mathcal{L}(V)$, then there is a basis of V that is a Jordan basis for T.*

PROOF: First consider a nilpotent operator $N \in \mathcal{L}(V)$ and the vectors $v_1, \dots, v_k \in V$ given by 8.40. For each j, note that N sends the first vector in the list $(N^{m(v_j)}v_j, \dots, Nv_j, v_j)$ to 0 and that N sends each vector in this list other than the first vector to the previous vector. In other words, if we reverse the order of the basis given by 8.40(a), then we obtain a basis of V with respect to which N has a block diagonal matrix, where each matrix on the diagonal has the form

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

Thus the theorem holds for nilpotent operators.

Now suppose $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T, with $U_1, \ldots, U_m$ the corresponding subspaces of generalized eigenvectors. We have

$$V = U_1 \oplus \cdots \oplus U_m,$$

where each $(T - \lambda_j I)|_{U_j}$ is nilpotent (see 8.23). By the previous paragraph, there is a basis of each U_j that is a Jordan basis for $(T - \lambda_j I)|_{U_j}$. Putting these bases together gives a basis of V that is a Jordan basis for T. ■

Exercises

1. Define $T \in \mathcal{L}(\mathbf{C}^2)$ by
$$T(w, z) = (z, 0).$$
Find all generalized eigenvectors of T.

2. Define $T \in \mathcal{L}(\mathbf{C}^2)$ by
$$T(w, z) = (-z, w).$$
Find all generalized eigenvectors of T.

3. Suppose $T \in \mathcal{L}(V)$, m is a positive integer, and $v \in V$ is such that $T^{m-1}v \neq 0$ but $T^m v = 0$. Prove that
$$(v, Tv, T^2 v, \dots, T^{m-1}v)$$
is linearly independent.

4. Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is defined by $T(z_1, z_2, z_3) = (z_2, z_3, 0)$. Prove that T has no square root. More precisely, prove that there does not exist $S \in \mathcal{L}(\mathbf{C}^3)$ such that $S^2 = T$.

5. Suppose $S, T \in \mathcal{L}(V)$. Prove that if ST is nilpotent, then TS is nilpotent.

6. Suppose $N \in \mathcal{L}(V)$ is nilpotent. Prove (without using 8.26) that 0 is the only eigenvalue of N.

7. Suppose V is an inner-product space. Prove that if $N \in \mathcal{L}(V)$ is self-adjoint and nilpotent, then $N = 0$.

8. Suppose $N \in \mathcal{L}(V)$ is such that $\operatorname{null} N^{\dim V - 1} \neq \operatorname{null} N^{\dim V}$. Prove that N is nilpotent and that
$$\dim \operatorname{null} N^j = j$$
for every integer j with $0 \leq j \leq \dim V$.

9. Suppose $T \in \mathcal{L}(V)$ and m is a nonnegative integer such that
$$\operatorname{range} T^m = \operatorname{range} T^{m+1}.$$
Prove that $\operatorname{range} T^k = \operatorname{range} T^m$ for all $k > m$.

10. Prove or give a counterexample: if $T \in \mathcal{L}(V)$, then

$$V = \operatorname{null} T \oplus \operatorname{range} T.$$

11. Prove that if $T \in \mathcal{L}(V)$, then

$$V = \operatorname{null} T^n \oplus \operatorname{range} T^n,$$

where $n = \dim V$.

12. Suppose V is a complex vector space, $N \in \mathcal{L}(V)$, and 0 is the only eigenvalue of N. Prove that N is nilpotent. Give an example to show that this is not necessarily true on a real vector space.

13. Suppose that V is a complex vector space with $\dim V = n$ and $T \in \mathcal{L}(V)$ is such that

$$\operatorname{null} T^{n-2} \neq \operatorname{null} T^{n-1}.$$

Prove that T has at most two distinct eigenvalues.

14. Give an example of an operator on $\mathbf{C}^4$ whose characteristic polynomial equals $(z-7)^2(z-8)^2$.

15. Suppose V is a complex vector space. Suppose $T \in \mathcal{L}(V)$ is such that 5 and 6 are eigenvalues of T and that T has no other eigenvalues. Prove that

$$(T-5I)^{n-1}(T-6I)^{n-1} = 0,$$

where $n = \dim V$.

16. Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Prove that V has a basis consisting of eigenvectors of T if and only if every generalized eigenvector of T is an eigenvector of T.

For complex vector spaces, this exercise adds another equivalence to the list given by 5.21.

17. Suppose V is an inner-product space and $N \in \mathcal{L}(V)$ is nilpotent. Prove that there exists an orthonormal basis of V with respect to which N has an upper-triangular matrix.

18. Define $N \in \mathcal{L}(\mathbf{F}^5)$ by

$$N(x_1, x_2, x_3, x_4, x_5) = (2x_2, 3x_3, -x_4, 4x_5, 0).$$

Find a square root of $I + N$.

19. Prove that if V is a complex vector space, then every invertible operator on V has a cube root.

20. Suppose $T \in \mathcal{L}(V)$ is invertible. Prove that there exists a polynomial $p \in \mathcal{P}(\mathbf{F})$ such that $T^{-1} = p(T)$.

21. Give an example of an operator on $\mathbf{C}^3$ whose minimal polynomial equals z^2.

22. Give an example of an operator on $\mathbf{C}^4$ whose minimal polynomial equals $z(z-1)^2$.

For complex vector spaces, this exercise adds another equivalence to the list given by 5.21.

23. Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Prove that V has a basis consisting of eigenvectors of T if and only if the minimal polynomial of T has no repeated roots.

24. Suppose V is an inner-product space. Prove that if $T \in \mathcal{L}(V)$ is normal, then the minimal polynomial of T has no repeated roots.

25. Suppose $T \in \mathcal{L}(V)$ and $v \in V$. Let p be the monic polynomial of smallest degree such that

$$p(T)v = 0.$$

Prove that p divides the minimal polynomial of T.

26. Give an example of an operator on $\mathbf{C}^4$ whose characteristic and minimal polynomials both equal $z(z-1)^2(z-3)$.

27. Give an example of an operator on $\mathbf{C}^4$ whose characteristic polynomial equals $z(z-1)^2(z-3)$ and whose minimal polynomial equals $z(z-1)(z-3)$.

This exercise shows that every monic polynomial is the characteristic polynomial of some operator.

28. Suppose $a_0, \dots, a_{n-1} \in \mathbf{C}$. Find the minimal and characteristic polynomials of the operator on $\mathbf{C}^n$ whose matrix (with respect to the standard basis) is

$$\begin{bmatrix} 0 & & & & & -a_0 \\ 1 & 0 & & & & -a_1 \\ & 1 & \ddots & & & -a_2 \\ & & \ddots & & & \vdots \\ & & & & 0 & -a_{n-2} \\ & & & & 1 & -a_{n-1} \end{bmatrix}.$$

29. Suppose $N \in \mathcal{L}(V)$ is nilpotent. Prove that the minimal polynomial of N is z^{m+1}, where m is the length of the longest consecutive string of 1's that appears on the line directly above the diagonal in the matrix of N with respect to any Jordan basis for N.

30. Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Prove that there does not exist a direct sum decomposition of V into two proper subspaces invariant under T if and only if the minimal polynomial of T is of the form $(z - \lambda)^{\dim V}$ for some $\lambda \in \mathbf{C}$.

31. Suppose $T \in \mathcal{L}(V)$ and $(v_1, \ldots, v_n)$ is a basis of V that is a Jordan basis for T. Describe the matrix of T with respect to the basis $(v_n, \ldots, v_1)$ obtained by reversing the order of the v's.