CHAPTER

8

Hypatia, the 5th century Egyptian mathematician and philosopher, as envisioned around 1900 by Alfred Seifert.

Operators on Complex Vector Spaces

In this chapter we delve deeper into the structure of operators, with most of the attention 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. To avoid some trivialities, we will assume that $V \neq \{0\}$. Thus our assumptions for this chapter are as follows:

8.1 Notation $\mathbf{F}$, V

- $\mathbf{F}$ denotes $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a finite-dimensional nonzero vector space over $\mathbf{F}$.

LEARNING OBJECTIVES FOR THIS CHAPTER

- generalized eigenvectors and generalized eigenspaces
- characteristic polynomial and the Cayley–Hamilton Theorem
- decomposition of an operator
- minimal polynomial
- Jordan Form

S. Axler, *Linear Algebra Done Right*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-3-319-11080-6_8

8.A Generalized Eigenvectors and Nilpotent Operators

Null Spaces of Powers of an Operator

We begin this chapter with a study of null spaces of powers of an operator.

8.2 Sequence of increasing null spaces

Suppose $T \in \mathcal{L}(V)$. Then

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

Proof Suppose k is a nonnegative integer and $v \in \operatorname{null} T^k$. Then $T^k v = 0$, and hence $T^{k+1}v = T(T^k v) = T(0) = 0$. Thus $v \in \operatorname{null} T^{k+1}$. Hence $\operatorname{null} T^k \subset \operatorname{null} T^{k+1}$, as desired. ■

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

8.3 Equality in the sequence of null spaces

Suppose $T \in \mathcal{L}(V)$. Suppose m is a nonnegative integer such that $\operatorname{null} T^m = \operatorname{null} T^{m+1}$. Then

$$\operatorname{null} T^m = \operatorname{null} T^{m+1} = \operatorname{null} T^{m+2} = \operatorname{null} T^{m+3} = \cdots.$$

Proof 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 from 8.2 that $\operatorname{null} T^{m+k} \subset \operatorname{null} T^{m+k+1}$.

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

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

Hence

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

Thus $T^{m+k}v = T^m(T^k v) = 0$, 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 exists 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.

8.4 Null spaces stop growing

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

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

Proof We need only prove that $\operatorname{null} T^n = \operatorname{null} T^{n+1}$ (by 8.3). Suppose this is not true. Then, by 8.2 and 8.3, we have

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

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

Unfortunately, it is not true that $V = \operatorname{null} T \oplus \operatorname{range} T$ for each $T \in \mathcal{L}(V)$. However, the following result is a useful substitute.

8.5 V is the direct sum of $\operatorname{null} T^{\dim V}$ and $\operatorname{range} T^{\dim V}$

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

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

Proof First we show that

8.6 $$(\operatorname{null} T^n) \cap (\operatorname{range} T^n) = \{0\}.$$

Suppose $v \in (\operatorname{null} T^n) \cap (\operatorname{range} T^n)$. Then $T^n v = 0$, and there exists $u \in V$ such that $v = T^n u$. Applying T^n to both sides of the last equation shows that $T^n v = T^{2n} u$. Hence $T^{2n} u = 0$, which implies that $T^n u = 0$ (by 8.4). Thus $v = T^n u = 0$, completing the proof of 8.6.

Now 8.6 implies that $\operatorname{null} T^n + \operatorname{range} T^n$ is a direct sum (by 1.45). Also,

$$\dim(\operatorname{null} T^n \oplus \operatorname{range} T^n) = \dim \operatorname{null} T^n + \dim \operatorname{range} T^n = \dim V,$$

where the first equality above comes from 3.78 and the second equality comes from the Fundamental Theorem of Linear Maps (3.22). The equation above implies that $\operatorname{null} T^n \oplus \operatorname{range} T^n = V$, as desired. ■

8.7 Example Suppose $T \in \mathcal{L}(\mathbf{F}^3)$ is defined by

$$T(z_1, z_2, z_3) = (4z_2, 0, 5z_3).$$

For this operator, $\operatorname{null} T + \operatorname{range} T$ is not a direct sum of subspaces, because $\operatorname{null} T = \{(z_1, 0, 0) : z_1 \in \mathbf{F}\}$ and $\operatorname{range} T = \{(z_1, 0, z_3) : z_1, z_3 \in \mathbf{F}\}$. Thus $\operatorname{null} T \cap \operatorname{range} T \neq \{0\}$ and hence $\operatorname{null} T + \operatorname{range} T$ is not a direct sum. Also note that $\operatorname{null} T + \operatorname{range} T \neq \mathbf{F}^3$.

However, we have $T^3(z_1, z_2, z_3) = (0, 0, 125z_3)$. Thus we see that $\operatorname{null} T^3 = \{(z_1, z_2, 0) : z_1, z_2 \in \mathbf{F}\}$ and $\operatorname{range} T^3 = \{(0, 0, z_3) : z_3 \in \mathbf{F}\}$. Hence $\mathbf{F}^3 = \operatorname{null} T^3 \oplus \operatorname{range} T^3$.

Generalized Eigenvectors

Unfortunately, some operators do not have enough eigenvectors to lead to a good description. Thus in this subsection 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

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

where each U_j is a subspace of V invariant under T. The simplest possible nonzero invariant subspaces are 1-dimensional. A decomposition as above where each U_j is a 1-dimensional subspace of V invariant under T is possible if and only if V has a basis consisting of eigenvectors of T (see 5.41). This happens if and only if V has an eigenspace decomposition

8.8
$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T),$$

where $\lambda_1, \dots, \lambda_m$ are the distinct eigenvalues of T (see 5.41).

The Spectral Theorem in the previous chapter shows that if V is an inner product space, then a decomposition of the form 8.8 holds for every normal operator if $\mathbf{F} = \mathbf{C}$ and for every self-adjoint operator if $\mathbf{F} = \mathbf{R}$ because operators of those types have enough eigenvectors to form a basis of V (see 7.24 and 7.29).

Sadly, a decomposition of the form 8.8 may not hold for more general operators, even on a complex vector space. An example was given by the operator in 5.43, which does not have enough eigenvectors for 8.8 to hold. Generalized eigenvectors and generalized eigenspaces, which we now introduce, will remedy this situation.

8.9 Definition ***generalized eigenvector***

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 $v \neq 0$ and

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

for some positive integer j.

Although j is allowed to be an arbitrary integer in the equation

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

in the definition of a generalized eigenvector, we will soon prove that every generalized eigenvector satisfies this equation with $j = \dim V$.

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.

8.10 Definition ***generalized eigenspace,*** $G(\lambda, T)$

Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. The ***generalized eigenspace*** of T corresponding to λ, denoted $G(\lambda, T)$, is defined to be the set of all generalized eigenvectors of T corresponding to λ, along with the 0 vector.

Because every eigenvector of T is a generalized eigenvector of T (take $j = 1$ in the definition of generalized eigenvector), each eigenspace is contained in the corresponding generalized eigenspace. In other words, if $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$, then

$$E(\lambda, T) \subset G(\lambda, T).$$

The next result implies that if $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$, then $G(\lambda, T)$ is a subspace of V (because the null space of each linear map on V is a subspace of V).

8.11 Description of generalized eigenspaces

Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. Then $G(\lambda, T) = \operatorname{null}(T - \lambda I)^{\dim V}$.

Proof Suppose $v \in \operatorname{null}(T - \lambda I)^{\dim V}$. The definitions imply $v \in G(\lambda, T)$. Thus $G(\lambda, T) \supset \operatorname{null}(T - \lambda I)^{\dim V}$.

Conversely, suppose $v \in G(\lambda, T)$. Thus there is a positive integer j such that

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

From 8.2 and 8.4 (with $T - \lambda I$ replacing T), we get $v \in \operatorname{null}(T - \lambda I)^{\dim V}$. Thus $G(\lambda, T) \subset \operatorname{null}(T - \lambda I)^{\dim V}$, completing the proof. ■

8.12 **Example** Define $T \in \mathcal{L}(\mathbf{C}^3)$ by

$$T(z_1, z_2, z_3) = (4z_2, 0, 5z_3).$$

(a) Find all eigenvalues of T, the corresponding eigenspaces, and the corresponding generalized eigenspaces.

(b) Show that $\mathbf{C}^3$ is the direct sum of generalized eigenspaces corresponding to the distinct eigenvalues of T.

Solution

(a) A routine use of the definition of eigenvalue shows that the eigenvalues of T are 0 and 5. The corresponding eigenspaces are easily seen to be $E(0, T) = \{(z_1, 0, 0) : z_1 \in \mathbf{C}\}$ and $E(5, T) = \{(0, 0, z_3) : z_3 \in \mathbf{C}\}$.

Note that this operator T does not have enough eigenvectors to span its domain $\mathbf{C}^3$.

We have $T^3(z_1, z_2, z_3) = (0, 0, 125z_3)$ for all $z_1, z_2, z_3 \in \mathbf{C}$. Thus 8.11 implies that $G(0, T) = \{(z_1, z_2, 0) : z_1, z_2 \in \mathbf{C}\}$.

We have $(T - 5I)^3(z_1, z_2, z_3) = (-125z_1 + 300z_2, -125z_2, 0)$. Thus 8.11 implies that $G(5, T) = \{(0, 0, z_3) : z_3 \in \mathbf{C}\}$.

(b) The results in part (a) show that $\mathbf{C}^3 = G(0, T) \oplus G(5, T)$.

One of our major goals in this chapter is to show that the result in part (b) of the example above holds in general for operators on finite-dimensional complex vector spaces; we will do this in 8.21.

We saw earlier (5.10) that eigenvectors corresponding to distinct eigenvalues are linearly independent. Now we prove a similar result for generalized eigenvectors.

8.13 Linearly independent generalized eigenvectors

Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \dots, \lambda_m$ are distinct eigenvalues of T and $v_1, \dots, v_m$ are corresponding generalized eigenvectors. Then $v_1, \dots, v_m$ is linearly independent.

Proof Suppose $a_1, \dots, a_m$ are complex numbers such that

8.14 $$0 = a_1 v_1 + \cdots + a_m v_m.$$

Let k be the largest nonnegative integer such that $(T - \lambda_1 I)^k v_1 \neq 0$. Let

$$w = (T - \lambda_1 I)^k v_1.$$

Thus

$$(T - \lambda_1 I)w = (T - \lambda_1 I)^{k+1} w = 0,$$

and hence $Tw = \lambda_1 w$. Thus $(T - \lambda I)w = (\lambda_1 - \lambda)w$ for every $\lambda \in \mathbf{F}$ and hence

8.15 $$(T - \lambda I)^n w = (\lambda_1 - \lambda)^n w$$

for every $\lambda \in \mathbf{F}$, where $n = \dim V$.

Apply the operator

$$(T - \lambda_1 I)^k (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n$$

to both sides of 8.14, getting

$$\begin{aligned} 0 &= a_1 (T - \lambda_1 I)^k (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n v_1 \\ &= a_1 (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n w \\ &= a_1 (\lambda_1 - \lambda_2)^n \cdots (\lambda_1 - \lambda_m)^n w, \end{aligned}$$

where we have used 8.11 to get the first equation above and 8.15 to get the last equation above.

The equation above implies that $a_1 = 0$. In a similar fashion, $a_j = 0$ for each j, which implies that $v_1, \dots, v_m$ is linearly independent. ■

Nilpotent Operators

8.16 Definition ***nilpotent***

An operator is called ***nilpotent*** if some power of it equals 0.

8.17 Example ***nilpotent operators***

(a) 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 $N^2 = 0$.

(b) The operator of differentiation on $\mathcal{P}_m(\mathbf{R})$ is nilpotent because the $(m+1)^{\text{st}}$ derivative of every 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 the 0 operator.

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

The next result shows that we never need to use a power higher than the dimension of the space.

8.18 Nilpotent operator raised to dimension of domain is 0

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

Proof Because N is nilpotent, $G(0, N) = V$. Thus 8.11 implies that $\operatorname{null} N^{\dim V} = V$, as desired. ■

Given an operator T on V, we want to find a basis of V such that the matrix of T with respect to this basis is as simple as possible, meaning that the matrix contains many 0's.

If V is a complex vector space, a proof of the next result follows easily from Exercise 7, 5.27, and 5.32. But the proof given here uses simpler ideas than needed to prove 5.27, and it works for both real and complex vector spaces.

The next result shows that if N is nilpotent, then 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. Later in this chapter we will do even better.

8.19 Matrix of a nilpotent operator

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

$$\begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix};$$

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

Proof First choose a basis of null N. Then extend this to a basis of null N^2. Then extend to a basis of null N^3. Continue in this fashion, eventually getting a basis of V (because 8.18 states that null $N^{\dim V} = V$).

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 null N. The next set of columns comes from basis vectors in null N^2. Applying N to any such vector, we get a vector in 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 lie above the diagonal. The next set of columns comes from basis vectors in null N^3. Applying N to any such vector, we get a vector in 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 lie above the diagonal. Continue in this fashion to complete the proof. ■

EXERCISES 8.A

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 the generalized eigenspaces corresponding to the distinct eigenvalues of T.

3 Suppose $T \in \mathcal{L}(V)$ is invertible. Prove that $G(\lambda, T) = G(\frac{1}{\lambda}, T^{-1})$ for every $\lambda \in \mathbf{F}$ with $\lambda \neq 0$.

4 Suppose $T \in \mathcal{L}(V)$ and $\alpha, \beta \in \mathbf{F}$ with $\alpha \neq \beta$. Prove that

$$G(\alpha, T) \cap G(\beta, T) = \{0\}.$$

5 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

$$v, Tv, T^2 v, \ldots, T^{m-1} v$$

is linearly independent.

6 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 $S^2 = T$.

7 Suppose $N \in \mathcal{L}(V)$ is nilpotent. Prove that 0 is the only eigenvalue of N.

8 Prove or give a counterexample: The set of nilpotent operators on V is a subspace of $\mathcal{L}(V)$.

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

10 Suppose that $T \in \mathcal{L}(V)$ is not nilpotent. Let $n = \dim V$. Show that $V = \operatorname{null} T^{n-1} \oplus \operatorname{range} T^{n-1}$.

11 Prove or give a counterexample: If V is a complex vector space and $\dim V = n$ and $T \in \mathcal{L}(V)$, then T^n is diagonalizable.

12 Suppose $N \in \mathcal{L}(V)$ and there exists a basis of V with respect to which N has an upper-triangular matrix with only 0's on the diagonal. Prove that N is nilpotent.

13 Suppose V is an inner product space and $N \in \mathcal{L}(V)$ is normal and nilpotent. Prove that $N = 0$.

14 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.
[*If* $F = \mathbf{C}$, *then the result above follows from Schur's Theorem* (*6.38*) *without the hypothesis that* N *is nilpotent. Thus the exercise above needs to be proved only when* $\mathbf{F} = \mathbf{R}$.]

15 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 $0 \leq j \leq \dim V$.

16 Suppose $T \in \mathcal{L}(V)$. Show that

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

17 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 $k > m$.

18 Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Prove that

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

19 Suppose $T \in \mathcal{L}(V)$ and m is a nonnegative integer. Prove that

$$\operatorname{null} T^m = \operatorname{null} T^{m+1} \quad \text{if and only if} \quad \operatorname{range} T^m = \operatorname{range} T^{m+1}.$$

20 Suppose $T \in \mathcal{L}(\mathbf{C}^5)$ is such that $\operatorname{range} T^4 \neq \operatorname{range} T^5$. Prove that T is nilpotent.

21 Find a vector space W and $T \in \mathcal{L}(W)$ such that $\operatorname{null} T^k \subsetneq \operatorname{null} T^{k+1}$ and $\operatorname{range} T^k \supsetneq \operatorname{range} T^{k+1}$ for every positive integer k.

8.B Decomposition of an Operator

Description of Operators on Complex Vector Spaces

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

We observed earlier that if $T \in \mathcal{L}(V)$, then null T and range T are invariant under T [see 5.3, parts (c) and (d)]. Now we show that the null space and the range of each polynomial of T is also invariant under T.

8.20 The null space and range of $p(T)$ are invariant under T

Suppose $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbf{F})$. Then null $p(T)$ and range $p(T)$ are invariant under T.

Proof Suppose $v \in \operatorname{null} p(T)$. Then $p(T)v = 0$. Thus

$$\big((p(T)\big)(Tv) = T\big(p(T)v\big) = T(0) = 0.$$

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

Suppose $v \in \operatorname{range} p(T)$. Then there exists $u \in V$ such that $v = p(T)u$. Thus

$$Tv = T\big(p(T)u\big) = p(T)(Tu).$$

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

The following major result 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 the hard work in our discussion of the generalized eigenspaces $G(\lambda, T)$, so at this point the proof is easy.

8.21 Description of operators on complex vector spaces

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

(a) $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$;

(b) each $G(\lambda_j, T)$ is invariant under T;

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

Proof Let $n = \dim V$. Recall that $G(\lambda_j, T) = \operatorname{null}(T - \lambda_j I)^n$ for each j (by 8.11). From 8.20 [with $p(z) = (z - \lambda_j)^n$], we get (b). Obviously (c) follows from the definitions.

We will prove (a) by induction on n. To get started, note that the desired result holds if $n = 1$. Thus we can assume that $n > 1$ and that the desired result holds on all vector spaces of smaller dimension.

Because V is a complex vector space, T has an eigenvalue (see 5.21); thus $m \geq 1$. Applying 8.5 to $T - \lambda_1 I$ shows that

8.22 $$V = G(\lambda_1, T) \oplus U,$$

where $U = \operatorname{range}(T - \lambda_1 I)^n$. Using 8.20 [with $p(z) = (z - \lambda_1)^n$], we see that U is invariant under T. Because $G(\lambda_1, T) \neq \{0\}$, we have $\dim U < n$. Thus we can apply our induction hypothesis to $T|_U$.

None of the generalized eigenvectors of $T|_U$ correspond to the eigenvalue λ_1, because all generalized eigenvectors of T corresponding to λ_1 are in $G(\lambda_1, T)$. Thus each eigenvalue of $T|_U$ is in $\{\lambda_2, \ldots, \lambda_m\}$.

By our induction hypothesis, $U = G(\lambda_2, T|_U) \oplus \cdots \oplus G(\lambda_m, T|_U)$. Combining this information with 8.22 will complete the proof if we can show that $G(\lambda_k, T|_U) = G(\lambda_k, T)$ for $k = 2, \ldots, m$.

Thus fix $k \in \{2, \ldots, m\}$. The inclusion $G(\lambda_k, T|_U) \subset G(\lambda_k, T)$ is clear.

To prove the inclusion in the other direction, suppose $v \in G(\lambda_k, T)$. By 8.22, we can write $v = v_1 + u$, where $v_1 \in G(\lambda_1, T)$ and $u \in U$. Our induction hypothesis implies that

$$u = v_2 + \cdots + v_m,$$

where each v_j is in $G(\lambda_j, T|_U)$, which is a subset of $G(\lambda_j, T)$. Thus

$$v = v_1 + v_2 + \cdots + v_m,$$

Because generalized eigenvectors corresponding to distinct eigenvalues are linearly independent (see 8.13), the equation above implies that each v_j equals 0 except possibly when $j = k$. In particular, $v_1 = 0$ and thus $v = u \in U$. Because $v \in U$, we can conclude that $v \in G(\lambda_k, T|_U)$, completing the proof. ∎

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

8.23 A basis of generalized eigenvectors

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 of each $G(\lambda_j, T)$ in 8.21. Put all these bases together to form a basis of V consisting of generalized eigenvectors of T. ■

Multiplicity of an Eigenvalue

If V is a complex vector space and $T \in \mathcal{L}(V)$, then the decomposition of V provided by 8.21 can be a powerful tool. The dimensions of the subspaces involved in this decomposition are sufficiently important to get a name.

8.24 Definition ***multiplicity***

- Suppose $T \in \mathcal{L}(V)$. The ***multiplicity*** of an eigenvalue λ of T is defined to be the dimension of the corresponding generalized eigenspace $G(\lambda, T)$.
- In other words, the multiplicity of an eigenvalue λ of T equals $\dim \operatorname{null}(T - \lambda I)^{\dim V}$.

The second bullet point above is justified by 8.11.

8.25 Example Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is defined by

$$T(z_1, z_2, z_3) = (6z_1 + 3z_2 + 4z_3, 6z_2 + 2z_3, 7z_3).$$

The matrix of T (with respect to the standard basis) is

$$\begin{pmatrix} 6 & 3 & 4 \\ 0 & 6 & 2 \\ 0 & 0 & 7 \end{pmatrix}.$$

The eigenvalues of T are 6 and 7, as follows from 5.32. You can verify that the generalized eigenspaces of T are as follows:

$$G(6, T) = \operatorname{span}\big((1, 0, 0), (0, 1, 0)\big) \quad \text{and} \quad G(7, T) = \operatorname{span}\big((10, 2, 1)\big).$$

Thus the eigenvalue 6 has multiplicity 2 and the eigenvalue 7 has multiplicity 1.

The direct sum $\mathbf{C}^3 = G(6, T) \oplus G(7, T)$ is the decomposition promised by 8.21. A basis of $\mathbf{C}^3$ consisting of generalized eigenvectors of T, as promised by 8.23, is

$$(1, 0, 0), (0, 1, 0), (10, 2, 1).$$

In Example 8.25, the sum of the multiplicities of the eigenvalues of T equals 3, which is the dimension of the domain of T. The next result shows that this always happens on a complex vector space.

8.26 Sum of the multiplicities equals $\dim V$

Suppose 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 The desired result follows from 8.21 and the obvious formula for the dimension of a direct sum (see 3.78 or Exercise 16 in Section 2.C). ■

The terms ***algebraic multiplicity*** and ***geometric multiplicity*** are used in some books. In case you encounter this terminology, be aware that the algebraic multiplicity is the same as the multiplicity defined here and the geometric multiplicity is the dimension of the corresponding eigenspace. In other words, if $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T, then

$$\text{algebraic multiplicity of } \lambda = \dim \operatorname{null}(T - \lambda I)^{\dim V} = \dim G(\lambda, T),$$
$$\text{geometric multiplicity of } \lambda = \dim \operatorname{null}(T - \lambda I) = \dim E(\lambda, T).$$

Note that as defined above, the algebraic multiplicity also has a geometric meaning as the dimension of a certain null space. The definition of multiplicity given here is cleaner than the traditional definition that involves determinants; 10.25 implies that these definitions are equivalent.

Block Diagonal Matrices

To interpret our results in matrix form, we make the following definition, generalizing the notion of a diagonal matrix.

Often we can understand a matrix better by thinking of it as composed of smaller matrices.

If each matrix A_j in the definition below is a 1-by-1 matrix, then we actually have a diagonal matrix.

8.27 Definition ***block diagonal matrix***

A ***block diagonal matrix*** is a square matrix of the form

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

where $A_1, \dots, A_m$ are square matrices lying along the diagonal and all the other entries of the matrix equal 0.

8.28 Example The 5-by-5 matrix

$$A = \begin{pmatrix} (4) & 0 \quad 0 & 0 \quad 0 \\ 0 & \begin{pmatrix} 2 & -3 \\ 0 & 2 \end{pmatrix} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} \\ \begin{matrix} 0 \\ 0 \end{matrix} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \begin{pmatrix} 1 & 7 \\ 0 & 1 \end{pmatrix} \end{pmatrix}$$

is a block diagonal matrix with

$$A = \begin{pmatrix} A_1 & & 0 \\ & A_2 & \\ 0 & & A_3 \end{pmatrix},$$

where

$$A_1 = \begin{pmatrix} 4 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 2 & -3 \\ 0 & 2 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 1 & 7 \\ 0 & 1 \end{pmatrix}.$$

Here the inner matrices in the 5-by-5 matrix above are blocked off to show how we can think of it as a block diagonal matrix.

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

8.29 Block diagonal matrix with upper-triangular blocks

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T, with multiplicities $d_1, \ldots, d_m$. Then there is a basis of V with respect to which T has a block diagonal matrix of the form

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

where each A_j is a d_j-by-d_j upper-triangular matrix of the form

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

Proof Each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent [see 8.21(c)]. For each j, choose a basis of $G(\lambda_j, T)$, which is a vector space with dimension d_j, such that the matrix of $(T - \lambda_j I)|_{G(\lambda_j, T)}$ with respect to this basis is as in 8.19. Thus the matrix of $T|_{G(\lambda_j, T)}$, which equals $(T - \lambda_j I)|_{G(\lambda_j, T)} + \lambda_j I|_{G(\lambda_j, T)}$, with respect to this basis will look like the desired form shown above for A_j.

Putting the bases of the $G(\lambda_j, T)$'s together gives a basis of V [by 8.21(a)]. The matrix of T with respect to this basis has the desired form. ■

The 5-by-5 matrix in 8.28 is of the form promised by 8.29, with each of the blocks itself an upper-triangular matrix that is constant along the diagonal of the block. If T is an operator on a 5-dimensional vector space whose matrix is as in 8.28, then the eigenvalues of T are 4, 2, 1 (as follows from 5.32), with multiplicities 1, 2, 2.

8.30 **Example** Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is defined by

$$T(z_1, z_2, z_3) = (6z_1 + 3z_2 + 4z_3, 6z_2 + 2z_3, 7z_3).$$

The matrix of T (with respect to the standard basis) is

$$\begin{pmatrix} 6 & 3 & 4 \\ 0 & 6 & 2 \\ 0 & 0 & 7 \end{pmatrix},$$

which is an upper-triangular matrix but is not of the form promised by 8.29.

As we saw in Example 8.25, the eigenvalues of T are 6 and 7 and the corresponding generalized eigenspaces are

$$G(6, T) = \operatorname{span}\big((1, 0, 0), (0, 1, 0)\big) \quad \text{and} \quad G(7, T) = \operatorname{span}\big((10, 2, 1)\big).$$

We also saw that a basis of $\mathbf{C}^3$ consisting of generalized eigenvectors of T is

$$(1, 0, 0), (0, 1, 0), (10, 2, 1).$$

The matrix of T with respect to this basis is

$$\left(\begin{matrix} \begin{pmatrix} 6 & 3 \\ 0 & 6 \end{pmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \begin{pmatrix} 7 \end{pmatrix} \end{matrix}\right),$$

which is a matrix of the block diagonal form promised by 8.29.

When we discuss the Jordan Form in Section 8.D, we will see that we can find a basis with respect to which an operator T has a matrix with even more 0's than promised by 8.29. However, 8.29 and its equivalent companion 8.21 are already quite powerful. For example, in the next subsection we will use 8.21 to show that every invertible operator on a complex vector space has a square root.

Square Roots

Recall that a square root of an operator $T \in \mathcal{L}(V)$ is an operator $R \in \mathcal{L}(V)$ such that $R^2 = T$ (see 7.33). Every complex number has a square root, but not every operator on a complex vector space has a square root. For example, the operator on $\mathbf{C}^3$ in Exercise 6 in Section 8.A has no square root. The noninvertibility of that operator is no accident, as we will soon see. We begin by showing that the identity plus any nilpotent operator has a square root.

8.31 Identity plus nilpotent has a square root

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

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

8.32 $$\sqrt{1+x} = 1 + a_1 x + a_2 x^2 + \cdots.$$

Because $a_1 = 1/2$, the formula above 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 the coefficients or worry about whether the infinite sum converges because we will use this equation only as motivation.

Because N is nilpotent, $N^m = 0$ for some positive integer m. In 8.32, 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

$$I + a_1 N + a_2 N^2 + \cdots + a_{m-1} N^{m-1}.$$

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

$$\begin{aligned}(I + a_1 N &+ a_2 N^2 + a_3 N^3 + \cdots + a_{m-1} N^{m-1})^2 \\ &= I + 2a_1 N + (2a_2 + {a_1}^2) N^2 + (2a_3 + 2a_1 a_2) N^3 + \cdots \\ &\quad + (2a_{m-1} + \text{terms involving } a_1, \dots, a_{m-2}) N^{m-1}.\end{aligned}$$

We want the right side of the equation above to equal $I + N$. Hence choose a_1 such that $2a_1 = 1$ (thus $a_1 = 1/2$). Next, choose a_2 such that $2a_2 + {a_1}^2 = 0$ (thus $a_2 = -1/8$). Then choose a_3 such 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. For example, the operator of multiplication by -1 on the 1-dimensional real vector space $\mathbf{R}$ has no square root.

8.33 Over $\mathbf{C}$, invertible operators have square roots

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$ is invertible. Then T has a square root.

Proof Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. For each j, there exists a nilpotent operator $N_j \in \mathcal{L}\big(G(\lambda_j, T)\big)$ such that $T|_{G(\lambda_j,T)} = \lambda_j I + N_j$ [see 8.21(c)]. Because T is invertible, none of the λ_j's equals 0, so we can write

$$T|_{G(\lambda_j,T)} = \lambda_j\Big(I + \frac{N_j}{\lambda_j}\Big)$$

for each j. Clearly N_j/λ_j is nilpotent, and so $I + N_j/\lambda_j$ has a square root (by 8.31). 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 R_j of $T|_{G(\lambda_j,T)}$.

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

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

where each u_j is in $G(\lambda_j, T)$ (see 8.21). Using this decomposition, define an operator $R \in \mathcal{L}(V)$ by

$$Rv = R_1 u_1 + \cdots + R_m u_m.$$

You should verify that this operator R 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.

EXERCISES 8.B

1 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.

2 Give an example of an operator T on a finite-dimensional real vector space such that 0 is the only eigenvalue of T but T is not nilpotent.

3 Suppose $T \in \mathcal{L}(V)$. Suppose $S \in \mathcal{L}(V)$ is invertible. Prove that T and $S^{-1}TS$ have the same eigenvalues with the same multiplicities.

4 Suppose V is an n-dimensional complex vector space and T is an operator on V such that $\operatorname{null} T^{n-2} \neq \operatorname{null} T^{n-1}$. Prove that T has at most two distinct eigenvalues.

5 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* $\mathbf{F} = \mathbf{C}$, *the exercise above adds an equivalence to the list in 5.41.*]

6 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 $I + N$.

7 Suppose V is a complex vector space. Prove that every invertible operator on V has a cube root.

8 Suppose $T \in \mathcal{L}(V)$ and 3 and 8 are eigenvalues of T. Let $n = \dim V$. Prove that $V = (\operatorname{null} T^{n-2}) \oplus (\operatorname{range} T^{n-2})$.

9 Suppose A and B are block diagonal matrices of the form

$$A = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix}, \quad B = \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_m \end{pmatrix},$$

where A_j has the same size as B_j for $j = 1, \ldots, m$. Show that AB is a block diagonal matrix of the form

$$AB = \begin{pmatrix} A_1 B_1 & & 0 \\ & \ddots & \\ 0 & & A_m B_m \end{pmatrix}.$$

10 Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Prove that there exist $D, N \in \mathcal{L}(V)$ such that $T = D + N$, the operator D is diagonalizable, N is nilpotent, and $DN = ND$.

11 Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. Prove that for every basis of V with respect to which T has an upper-triangular matrix, the number of times that λ appears on the diagonal of the matrix of T equals the multiplicity of λ as an eigenvalue of T.

8.C Characteristic and Minimal Polynomials

The Cayley–Hamilton Theorem

The next definition associates a polynomial with each operator on V if $\mathbf{F} = \mathbf{C}$. For $\mathbf{F} = \mathbf{R}$, the corresponding definition will be given in the next chapter.

8.34 Definition ***characteristic polynomial***

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, with multiplicities $d_1, \dots, d_m$. The polynomial

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

is called the ***characteristic polynomial*** of T.

8.35 Example Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is defined as in Example 8.25. Because the eigenvalues of T are 6, with multiplicity 2, and 7, with multiplicity 1, we see that the characteristic polynomial of T is $(z - 6)^2(z - 7)$.

8.36 Degree and zeros of characteristic polynomial

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then

(a) the characteristic polynomial of T has degree $\dim V$;

(b) the zeros of the characteristic polynomial of T are the eigenvalues of T.

Proof Clearly part (a) follows from 8.26 and part (b) follows from the definition of the characteristic polynomial. ■

Most texts define the characteristic polynomial using determinants (the two definitions are equivalent by 10.25). The approach taken here, which is considerably simpler, leads to the following easy proof of the Cayley–Hamilton Theorem. In the next chapter, we will see that this result also holds on real vector spaces (see 9.24).

8.37 Cayley–Hamilton Theorem

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let q denote the characteristic polynomial of T. Then $q(T) = 0$.

English mathematician Arthur Cayley (1821–1895) published three math papers before completing his undergraduate degree in 1842. Irish mathematician William Rowan Hamilton (1805–1865) was made a professor in 1827 when he was 22 years old and still an undergraduate!

Proof Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of the operator T, and let $d_1, \dots, d_m$ be the dimensions of the corresponding generalized eigenspaces $G(\lambda_1, T), \dots, G(\lambda_m, T)$. For each $j \in \{1, \dots, m\}$, we know that $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent. Thus we have $(T - \lambda_j I)^{d_j}|_{G(\lambda_j, T)} = 0$ (by 8.18).

Every vector in V is a sum of vectors in $G(\lambda_1, T), \dots, G(\lambda_m, T)$ (by 8.21). Thus to prove that $q(T) = 0$, we need only show that $q(T)|_{G(\lambda_j, T)} = 0$ for each j.

Thus fix $j \in \{1, \dots, m\}$. We have

$$q(T) = (T - \lambda_1 I)^{d_1} \cdots (T - \lambda_m I)^{d_m}.$$

The operators on the right side of the equation above all commute, so we can move the factor $(T - \lambda_j I)^{d_j}$ to be the last term in the expression on the right. Because $(T - \lambda_j I)^{d_j}|_{G(\lambda_j, T)} = 0$, we conclude that $q(T)|_{G(\lambda_j, T)} = 0$, as desired. ■

The Minimal Polynomial

In this subsection we introduce another important polynomial associated with each operator. We begin with the following definition.

8.38 Definition ***monic polynomial***

A ***monic polynomial*** is a polynomial whose highest-degree coefficient equals 1.

8.39 Example The polynomial $2 + 9z^2 + z^7$ is a monic polynomial of degree 7.

8.40 Minimal polynomial

Suppose $T \in \mathcal{L}(V)$. Then there is a unique monic polynomial p of smallest degree such that $p(T) = 0$.

Proof Let $n = \dim V$. Then the list

$$I, T, T^2, \ldots, T^{n^2}$$

is not linearly independent in $\mathcal{L}(V)$, because the vector space $\mathcal{L}(V)$ has dimension n^2 (see 3.61) and we have a list of length $n^2 + 1$. Let m be the smallest positive integer such that the list

8.41 $$I, T, T^2, \ldots, T^m$$

is linearly dependent. The Linear Dependence Lemma (2.21) 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 the list above is linearly dependent, we conclude that T^m is a linear combination of $I, T, T^2, \ldots, T^{m-1}$. Thus there exist scalars $a_0, a_1, a_2, \ldots, a_{m-1} \in \mathbf{F}$ such that

8.42 $$a_0 I + a_1 T + a_2 T^2 + \cdots + a_{m-1} T^{m-1} + T^m = 0.$$

Define a monic polynomial $p \in \mathcal{P}(\mathbf{F})$ by

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

Then 8.42 implies that $p(T) = 0$.

To prove the uniqueness part of the result, note that the choice of m implies that no monic polynomial $q \in \mathcal{P}(\mathbf{F})$ with degree smaller than m can satisfy $q(T) = 0$. Suppose $q \in \mathcal{P}(\mathbf{F})$ is a monic polynomial with degree m and $q(T) = 0$. Then $(p - q)(T) = 0$ and $\deg(p - q) < m$. The choice of m now implies that $q = p$, completing the proof. ■

The last result justifies the following definition.

8.43 Definition ***minimal polynomial***

Suppose $T \in \mathcal{L}(V)$. Then the ***minimal polynomial*** of T is the unique monic polynomial p of smallest degree such that $p(T) = 0$.

The proof of the last result shows that the degree of the minimal polynomial of each operator on V is at most $(\dim V)^2$. The Cayley–Hamilton Theorem (8.37) 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.

Suppose you are given the matrix (with respect to some basis) of an operator $T \in \mathcal{L}(V)$. You could program a computer to find the minimal polynomial of T as follows: Consider the system of linear equations

8.44 $$a_0\mathcal{M}(I) + a_1\mathcal{M}(T) + \cdots + a_{m-1}\mathcal{M}(T)^{m-1} = -\mathcal{M}(T)^m$$

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

for successive values of $m = 1, 2, \ldots$ until this system of equations has a solution $a_0, a_1, a_2, \ldots, a_{m-1}$. 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 and fast (for a computer) process such as Gaussian elimination.

8.45 **Example** Let T be the operator on $\mathbf{C}^5$ whose matrix (with respect to the standard basis) is

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

Find the minimal polynomial of T.

Solution 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 then you will notice that there is clearly no solution to 8.44 until $m = 5$. Do the computations and you will see that the minimal polynomial of T equals $z^5 - 6z + 3$.

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

8.46 $q(T) = 0$ implies q is a multiple of the minimal polynomial

Suppose $T \in \mathcal{L}(V)$ and $q \in \mathcal{P}(\mathbf{F})$. Then $q(T) = 0$ if and only if q is a polynomial multiple of the minimal polynomial of T.

Proof Let p denote the minimal polynomial of T.

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

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

as desired.

To prove the other direction, now suppose $q(T) = 0$. By the Division Algorithm for Polynomials (4.8), there exist polynomials $s, r \in \mathcal{P}(\mathbf{F})$ such that

8.47 $$q = ps + r$$

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

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

The equation above implies that $r = 0$ (otherwise, dividing r by its highest-degree coefficient would produce a monic polynomial that when applied to T gives 0; this polynomial would have a smaller degree than the minimal polynomial, which would be a contradiction). Thus 8.47 becomes the equation $q = ps$. Hence q is a polynomial multiple of p, as desired. ■

The next result is stated only for complex vector spaces, because we have not yet defined the characteristic polynomial when $\mathbf{F} = \mathbf{R}$. However, the result also holds for real vector spaces, as we will see in the next chapter.

8.48 Characteristic polynomial is a multiple of minimal polynomial

Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Then the characteristic polynomial of T is a polynomial multiple of the minimal polynomial of T.

Proof The desired result follows immediately from the Cayley–Hamilton Theorem (8.37) and 8.46. ■

We know (at least when $\mathbf{F} = \mathbf{C}$) that the zeros of the characteristic polynomial of T are the eigenvalues of T (see 8.36). Now we show that the minimal polynomial has the same zeros (although the multiplicities of these zeros may differ).

8.49 Eigenvalues are the zeros of the minimal polynomial

Let $T \in \mathcal{L}(V)$. Then the zeros 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 $\lambda \in \mathbf{F}$ is a zero of p. Then p can be written in the form

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

where q is a monic polynomial with coefficients in $\mathbf{F}$ (see 4.11). 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 exists 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 $\lambda \in \mathbf{F}$ is an eigenvalue of T. Thus there exists $v \in V$ with $v \neq 0$ 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

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

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

The next three examples show how our results can be useful in finding minimal polynomials and in understanding why eigenvalues of some operators cannot be exactly computed.

8.50 **Example** Find the minimal polynomial of the operator $T \in \mathcal{L}(\mathbf{C}^3)$ in Example 8.30.

Solution In Example 8.30 we noted that the eigenvalues of T are 6 and 7. Thus by 8.49, the minimal polynomial of T is a polynomial multiple of $(z - 6)(z - 7)$.

In Example 8.35, we saw that the characteristic polynomial of T is $(z - 6)^2(z - 7)$. Thus by 8.48 and the paragraph above, the minimal polynomial of T is either $(z - 6)(z - 7)$ or $(z - 6)^2(z - 7)$. A simple computation shows that

$$(T - 6I)(T - 7I) \neq 0.$$

Thus the minimal polynomial of T is $(z - 6)^2(z - 7)$.

8.51 Example Find the minimal polynomial of the operator $T \in \mathcal{L}(\mathbf{C}^3)$ defined by $T(z_1, z_2, z_3) = (6z_1, 6z_2, 7z_3)$.

Solution It is easy to see that for this operator T, the eigenvalues of T are 6 and 7, and the characteristic polynomial of T is $(z-6)^2(z-7)$.

Thus as in the previous example, the minimal polynomial of T is either $(z-6)(z-7)$ or $(z-6)^2(z-7)$. A simple computation shows that $(T-6I)(T-7I) = 0$. Thus the minimal polynomial of T is $(z-6)(z-7)$.

8.52 Example What are the eigenvalues of the operator in Example 8.45?

Solution From 8.49 and the solution to Example 8.45, 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, 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 eigenvalue of T in any familiar form, although 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

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

The nonreal eigenvalues occur as a pair, with each the complex conjugate of the other, as expected for a polynomial with real coefficients (see 4.15).

EXERCISES 8.C

1 Suppose $T \in \mathcal{L}(\mathbf{C}^4)$ is such that the eigenvalues of T are 3, 5, 8. Prove that $(T-3I)^2(T-5I)^2(T-8I)^2 = 0$.

2 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 $n = \dim V$.

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

4 Give an example of an operator on $\mathbf{C}^4$ whose characteristic polynomial equals $(z-1)(z-5)^3$ and whose minimal polynomial equals $(z-1)(z-5)^2$.

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

6 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 $z(z-1)(z-3)$.

7 Suppose V is a complex vector space. Suppose $T \in \mathcal{L}(V)$ is such that $P^2 = P$. Prove that the characteristic polynomial of P is $z^m(z-1)^n$, where $m = \dim \operatorname{null} P$ and $n = \dim \operatorname{range} P$.

8 Suppose $T \in \mathcal{L}(V)$. Prove that T is invertible if and only if the constant term in the minimal polynomial of T is nonzero.

9 Suppose $T \in \mathcal{L}(V)$ has minimal polynomial $4+5z-6z^2-7z^3+2z^4+z^5$. Find the minimal polynomial of T^{-1}.

10 Suppose V is a complex vector space and $T \in \mathcal{L}(V)$ is invertible. Let p denote the characteristic polynomial of T and let q denote the characteristic polynomial of T^{-1}. Prove that

$$q(z) = \frac{1}{p(0)} z^{\dim V} p\left(\frac{1}{z}\right)$$

for all nonzero $z \in \mathbf{C}$.

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

12 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 zeros.
[*For complex vector spaces, the exercise above adds another equivalence to the list given by 5.41.*]

13 Suppose V is an inner product space and $T \in \mathcal{L}(V)$ is normal. Prove that the minimal polynomial of T has no repeated zeros.

14 Suppose V is a complex inner product space and $S \in \mathcal{L}(V)$ is an isometry. Prove that the constant term in the characteristic polynomial of S has absolute value 1.

15 Suppose $T \in \mathcal{L}(V)$ and $v \in V$.

(a) Prove that there exists a unique monic polynomial p of smallest degree such that $p(T)v = 0$.

(b) Prove that p divides the minimal polynomial of T.

16 Suppose V is an inner product space and $T \in \mathcal{L}(V)$. Suppose

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

is the minimal polynomial of T. Prove that

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

is the minimal polynomial of T^*.

17 Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Suppose the minimal polynomial of T has degree $\dim V$. Prove that the characteristic polynomial of T equals the minimal polynomial of T.

18 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

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

[*The exercise above shows that every monic polynomial is the characteristic polynomial of some operator.*]

19 Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Suppose that with respect to some basis of V the matrix of T is upper triangular, with $\lambda_1, \dots, \lambda_n$ on the diagonal of this matrix. Prove that the characteristic polynomial of T is $(z - \lambda_1) \cdots (z - \lambda_n)$.

20 Suppose V is a complex vector space and $V_1, \dots, V_m$ are nonzero subspaces of V such that $V = V_1 \oplus \cdots \oplus V_m$. Suppose $T \in \mathcal{L}(V)$ and each V_j is invariant under T. For each j, let p_j denote the characteristic polynomial of $T|_{V_j}$. Prove that the characteristic polynomial of T equals $p_1 \cdots p_m$.

8.D 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 8.29). 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 0's everywhere except possibly on the diagonal and the line directly above the diagonal.

We begin by looking at two examples of nilpotent operators.

8.53 Example Let $N \in \mathcal{L}(\mathbf{F}^4)$ be the nilpotent operator defined by

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

If $v = (1, 0, 0, 0)$, then $N^3 v, N^2 v, N v, v$ is a basis of $\mathbf{F}^4$. The matrix of N with respect to this basis is

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

The next example of a nilpotent operator has more complicated behavior than the example above.

8.54 Example Let $N \in \mathcal{L}(\mathbf{F}^6)$ be the nilpotent operator defined by

$$N(z_1, z_2, z_3, z_4, z_5, z_6) = (0, z_1, z_2, 0, z_4, 0).$$

Unlike the nice behavior of the nilpotent operator of the previous example, for this nilpotent operator there does not exist a vector $v \in \mathbf{F}^6$ such that $N^5 v, N^4 v, N^3 v, N^2 v, N v, v$ is a basis of $\mathbf{F}^6$. However, if we take $v_1 = (1, 0, 0, 0, 0, 0)$, $v_2 = (0, 0, 0, 1, 0, 0)$, and $v_3 = (0, 0, 0, 0, 0, 1)$, then $N^2 v_1, N v_1, v_1, N v_2, v_2, v_3$ is a basis of $\mathbf{F}^6$. The matrix of N with respect to this basis is

$$\left(\begin{array}{cccccc} \multicolumn{3}{c}{\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ 0 & 0 & 0 & \multicolumn{2}{c}{\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}} & \begin{matrix} 0 \\ 0 \end{matrix} \\ 0 & 0 & 0 & 0 & 0 & (0) \end{array}\right).$$

Here the inner matrices are blocked off to show that we can think of the 6-by-6 matrix above as a block diagonal matrix consisting of a 3-by-3 block with 1's on the line above the diagonal and 0's elsewhere, a 2-by-2 block with 1 above the diagonal and 0's elsewhere, and a 1-by-1 block containing 0.

Our next result shows that every nilpotent operator $N \in \mathcal{L}(V)$ behaves similarly to the previous example. Specifically, there is a finite collection of vectors $v_1, \ldots, v_n \in V$ such that there is a basis of V consisting of the vectors of the form $N^k v_j$, as j varies from 1 to n and k varies (in reverse order) from 0 to the largest nonnegative integer m_j such that $N^{m_j} v_j \neq 0$. For the matrix interpretation of the next result, see the first part of the proof of 8.60.

8.55 Basis corresponding to a nilpotent operator

Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then there exist vectors $v_1, \ldots, v_n \in V$ and nonnegative integers $m_1, \ldots, m_n$ such that

(a) $N^{m_1} v_1, \ldots, N v_1, v_1, \ldots, N^{m_n} v_n, \ldots, N v_n, v_n$ is a basis of V;

(b) $N^{m_1+1} v_1 = \cdots = N^{m_n+1} v_n = 0$.

Proof We will prove this result by induction on $\dim V$. To get started, note that the desired result obviously holds if $\dim V = 1$ (in that case, the only nilpotent operator is the 0 operator, so take v_1 to be any nonzero vector and $m_1 = 0$). Now assume that $\dim V > 1$ and that the desired result holds on all vector spaces of smaller dimension.

Because N is nilpotent, N is not injective. Thus N is not surjective (by 3.69) and hence range N is a subspace of V that has a smaller dimension than V. Thus we can apply our induction hypothesis to the restriction operator $N|_{\text{range } N} \in \mathcal{L}(\text{range } N)$. [We can ignore the trivial case range $N = \{0\}$, because in that case N is the 0 operator and we can choose $v_1, \ldots, v_n$ to be any basis of V and $m_1 = \cdots = m_n = 0$ to get the desired result.]

By our induction hypothesis applied to $N|_{\text{range } N}$, there exist vectors $v_1, \ldots, v_n \in \text{range } N$ and nonnegative integers $m_1, \ldots, m_n$ such that

8.56 $$N^{m_1} v_1, \ldots, N v_1, v_1, \ldots, N^{m_n} v_n, \ldots, N v_n, v_n$$

is a basis of range N and

$$N^{m_1+1} v_1 = \cdots = N^{m_n+1} v_n = 0.$$

Because each v_j is in range N, for each j there exists $u_j \in V$ such that $v_j = N u_j$. Thus $N^{k+1} u_j = N^k v_j$ for each j and each nonnegative integer k. We now claim that

8.57 $$N^{m_1+1} u_1, \ldots, N u_1, u_1, \ldots, N^{m_n+1} u_n, \ldots, N u_n, u_n$$

is a linearly independent list of vectors in V. To verify this claim, suppose that some linear combination of 8.57 equals 0. Applying N to that linear combination, we get a linear combination of 8.56 equal to 0. However, the list 8.56 is linearly independent, and hence all the coefficients in our original linear combination of 8.57 equal 0 except possibly the coefficients of the vectors

$$N^{m_1+1}u_1, \dots, N^{m_n+1}u_n,$$

which equal the vectors

$$N^{m_1}v_1, \dots, N^{m_n}v_n.$$

Again using the linear independence of the list 8.56, we conclude that those coefficients also equal 0, completing our proof that the list 8.57 is linearly independent.

Now extend 8.57 to a basis

8.58 $$N^{m_1+1}u_1, \dots, Nu_1, u_1, \dots, N^{m_n+1}u_n, \dots, Nu_n, u_n, w_1, \dots, w_p$$

of V (which is possible by 2.33). Each Nw_j is in range N and hence is in the span of 8.56. Each vector in the list 8.56 equals N applied to some vector in the list 8.57. Thus there exists x_j in the span of 8.57 such that $Nw_j = Nx_j$. Now let

$$u_{n+j} = w_j - x_j.$$

Then $Nu_{n+j} = 0$. Furthermore,

$$N^{m_1+1}u_1, \dots, Nu_1, u_1, \dots, N^{m_n+1}u_n, \dots, Nu_n, u_n, u_{n+1}, \dots, u_{n+p}$$

spans V because its span contains each x_j and each u_{n+j} and hence each w_j (and because 8.58 spans V).

Thus the spanning list above is a basis of V because it has the same length as the basis 8.58 (where we have used 2.42). This basis has the required form, completing the proof. ■

French mathematician Camille Jordan (1838–1922) first published a proof of 8.60 in 1870.

In the next definition, the diagonal of each A_j is filled with some eigenvalue λ_j of T, the line directly above the diagonal of A_j is filled with 1's, and all other entries in A_j are 0 (to understand why each λ_j is an eigenvalue of T, see 5.32). The λ_j's need not be distinct. Also, A_j may be a 1-by-1 matrix (λ_j) containing just an eigenvalue of T.

8.59 Definition ***Jordan basis***

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

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_p \end{pmatrix},$$

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

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

8.60 Jordan Form

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 $v_1, \dots, v_n \in V$ given by 8.55. For each j, note that N sends the first vector in the list $N^{m_j} v_j, \dots, N v_j, v_j$ to 0 and that N sends each vector in this list other than the first vector to the previous vector. In other words, 8.55 gives a basis of V with respect to which N has a block diagonal matrix, where each matrix on the diagonal has the form

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

Thus the desired result holds for nilpotent operators.

Now suppose $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. We have the generalized eigenspace decomposition

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T),$$

where each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent (see 8.21). Thus some basis of each $G(\lambda_j, T)$ is a Jordan basis for $(T - \lambda_j I)|_{G(\lambda_m, T)}$ (see previous paragraph). Put these bases together to get a basis of V that is a Jordan basis for T. ■

EXERCISES 8.D

1 Find the characteristic polynomial and the minimal polynomial of the operator N in Example 8.53.

2 Find the characteristic polynomial and the minimal polynomial of the operator N in Example 8.54.

3 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.

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

5 Suppose $T \in \mathcal{L}(V)$ and $v_1, \dots, v_n$ is a basis of V that is a Jordan basis for T. Describe the matrix of T^2 with respect to this basis.

6 Suppose $N \in \mathcal{L}(V)$ is nilpotent and $v_1, \dots, v_n$ and $m_1, \dots, m_n$ are as in 8.55. Prove that $N^{m_1} v_1, \dots, N^{m_n} v_n$ is a basis of null N.
[*The exercise above implies that n, which equals* $\dim \operatorname{null} N$, *depends only on N and not on the specific Jordan basis chosen for N.*]

7 Suppose $p, q \in \mathcal{P}(\mathbf{C})$ are monic polynomials with the same zeros and q is a polynomial multiple of p. Prove that there exists $T \in \mathcal{L}(C^{\deg q})$ such that the characteristic polynomial of T is q and the minimal polynomial of T is p.

8 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 $\lambda \in \mathbf{C}$.