Chapter 8

Operators on Complex Vector Spaces

8.A Generalized Eigenvectors and Nilpotent Operators

10/22:

- In this chapter, we will assume that V is a finite-dimensional *nonzero* vector space over $\mathbb{F}$ (just to avoid dealing with some trivialities).
- Null spaces and powers of an operator.

Theorem 8.1. *Suppose $T \in \mathcal{L}(V)$. Then*

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

Proof. We induct on the exponent k of T. For the base case $k = 0$, suppose $v \in \operatorname{null} T^0$. Then $v \in \operatorname{null} I$ since $T^0 = I$ by definition. It follows that

$$0 = Iv = v$$

so $\{0\} = \operatorname{null} T^0$, as desired. Now suppose inductively that we have proven the claim for k; we now wish to show that $\operatorname{null} T^k \subset \operatorname{null} T^{k+1}$. Suppose $v \in \operatorname{null} T^k$. Then $T^k v = 0$. It follows that

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

so $v \in T^{k+1}$, as desired. ■

Theorem 8.2. *Let $T \in \mathcal{L}(V)$, and suppose m is a nonnegative integer such that* $\operatorname{null} T^m = \operatorname{null} T^{m+1}$. *Then*

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

Proof. We induct on k, defined as follows. For the base case $k = 0$, we have that

$$\operatorname{null} T^{m+0} = \operatorname{null} T^m = \operatorname{null} T^{m+1} = \operatorname{null} T^{m+0+1}$$

by hypothesis, as desired. Now suppose inductively that we have proven that $\operatorname{null} T^{m+k-1} = \operatorname{null} T^{m+k}$; we now wish to show that $\operatorname{null} T^{m+k} = \operatorname{null} T^{m+k+1}$. By Theorem 8.1, we have that $\operatorname{null} T^{m+k} \subset \operatorname{null} T^{m+k+1}$. On the other hand, suppose that $v \in \operatorname{null} T^{m+k+1}$. Then

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

But this implies that $T^k v \in \operatorname{null} T^{m+1} = \operatorname{null} T^m$ by hypothesis. Therefore,

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

so $v \in \operatorname{null} T^{m+k}$, as desired. ■

- Theorem 8.2 raises the question how to characterize/define/find nonnegative integers m such that the null space stops growing. We tackle begin to tackle this question with the following.

  Theorem 8.3. *Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then*

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

  Proof. To prove the claim, Theorem 8.2 tells us that we need only verify that $\operatorname{null} T^n = \operatorname{null} T^{n+1}$. Suppose for the sake of contradiction that $\operatorname{null} T^n \neq \operatorname{null} T^{n+1}$. Then by Theorem 8.2, we cannot have $\operatorname{null} T^k = \operatorname{null} T^{k+1}$ for any $0 \leq k \leq n$. However, by Theorem 8.1, we must still have that $\operatorname{null} T^k \subset \operatorname{null} T^{k+1}$ for each $k = 1, \dots, n$. Combining the last two results, we must have the following.

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

  At each of these strict inclusions, the dimension from the previous to the next null space must increase by at least one. Thus, $\dim \operatorname{null} T^{n+1} \geq n+1$. But since $\operatorname{null} T^{n+1} \subset V$, Theorem 2.11 asserts that $\dim \operatorname{null} T^{n+1} \leq n$, so we have that

  $$n + 1 \leq \dim \operatorname{null} T^{n+1} \leq n$$

  a contradiction. ∎

- While it is not true that $V = \operatorname{null} T \oplus \operatorname{range} T$ for each $T \in \mathcal{L}(V)$, we can prove the following related theorem.

  Theorem 8.4. *Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then*

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

  Proof. To prove that $V = \operatorname{null} T^n \oplus \operatorname{range} T^n$, it will suffice to show that $(\operatorname{null} T^n) \cap (\operatorname{range} T^n) = \{0\}$ and that $\dim(\operatorname{null} T^n \oplus \operatorname{range} T^n) = \dim V$ (see Exercise 2.C.1). Let's begin.

  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 $v = T^n u$. Combining these results reveals that

  $$T^{2n} u = T^n v = 0$$

  so $u \in \operatorname{null} T^{2n} = \operatorname{null} T^n$ by Theorem 8.3. Therefore, $v = T^n u = 0$, as desired.

  As to the other equality, we have that

  $$\begin{aligned} \dim(\operatorname{null} T^n \oplus \operatorname{range} T^n) &= \dim \operatorname{null} T^n + \dim \operatorname{range} T^n && \text{Theorem 3.21} \\ &= \dim V && \text{Fundamental Theorem of Linear Maps} \end{aligned}$$

  as desired. ∎

- Although many operators can be described by their eigenvectors, not all can. Thus, we introduce the following more general descriptor.

- **Generalized eigenvector** (of $T \in \mathcal{L}(V)$): A nonzero vector $v \in V$ such that

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

  for some positive integer j, where λ is an eigenvalue of T.

  - Although this definition lets j be arbitrary, we will soon prove that if $j = \dim V$, every generalized eigenvector satisfies the above equation.
  - Note that we do not define generalized eigenvalues because generalized eigenvectors still pertain to the original set of eigenvalues.

- Every eigenvector of T is a generalized eigenvector of T (take $j = 1$ in the definition).
- **Generalized eigenspace** (of $T \in \mathcal{L}(V)$ and λ): The set of all generalized eigenvectors of T corresponding to λ, and the 0 vector. *Denoted by* $\boldsymbol{G(\lambda, T)}$.
- Since every eigenvector of T is a generalized eigenvector of T, we have that $E(\lambda, T) \subset G(\lambda, T)$.
- We now characterize generalized eigenspaces.

Theorem 8.5. *Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. Then $G(\lambda, T) = \text{null}(T - \lambda I)^{\dim V}$.*

Proof. Suppose first that $v \in (T - \lambda I)^{\dim V}$. Then by the definition of $G(\lambda, T)$, $v \in G(\lambda, T)$, as desired.

Now suppose that $v \in G(\lambda, T)$. Then $(T - \lambda I)^j v = 0$ for some positive integer j. Thus, $v \in \text{null}(T - \lambda I)^j$. We divide into two cases ($j < \dim V$ and $j \geq \dim V$). If $j < \dim V$, then by Theorem 8.1, $v \in \text{null}(T - \lambda I)^j \subset \text{null}(T - \lambda I)^{\dim V}$, as desired. On the other hand, if $j \geq \dim V$, then by Theorem 8.3 $v \in \text{null}(T - \lambda I)^j = \text{null}(T - \lambda I)^{\dim V}$, as desired. ■

- We now prove an analogous result to Theorem 5.2 for generalized eigenvectors.

Theorem 8.6. *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 $a_1, \dots, a_m \in \mathbb{F}$ are numbers such that

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

We will prove that each $a_j = 0$ one at a time. Let's begin.

Let $j \in \{1, \dots, n\}$ be arbitrary, and let k be the largest nonnegative integer such that $(T - \lambda_j I)^k v_j \neq 0$. Let

$$w = (T - \lambda_j I)^k v_j$$

Then by the definition of k,

$$\begin{aligned}(T - \lambda_j I)w = (T - \lambda_j I)^{k+1} v_1 = 0 \\ Tw = \lambda_j w\end{aligned}$$

It follows that for any $\lambda \in \mathbb{F}$, $(T - \lambda I)w = (\lambda_j - \lambda)w$, which in turn implies that

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

for any $\lambda \in \mathbb{F}$ where $n = \dim V$. Thus, we have that

$$\begin{aligned}(T - \lambda_j I)^k \prod_{\substack{i=1 \\ i \neq j}}^m (T - \lambda_i I)^n 0 &= (T - \lambda_j I)^k \prod_{\substack{i=1 \\ i \neq j}}^m (T - \lambda_i I)^n (a_1 v_1 + \cdots + a_m v_m) \\ 0 &= a_1 (T - \lambda_j I)^k \prod_{\substack{i=1 \\ i \neq j}}^m (T - \lambda_i I)^n v_1 + \cdots + a_{j-1} (T - \lambda_j I)^k \prod_{\substack{i=1 \\ i \neq j}}^m (T - \lambda_i I)^n v_{j-1} \\ &\quad + a_j (T - \lambda_j I)^k \prod_{\substack{i=1 \\ i \neq j}}^m (T - \lambda_i I)^n v_j \\ &\quad + a_{j+1} (T - \lambda_j I)^k \prod_{\substack{i=1 \\ i \neq j}}^m (T - \lambda_i I)^n v_{j+1} + \cdots + a_m (T - \lambda_j I)^k \prod_{\substack{i=1 \\ i \neq j}}^m (T - \lambda_i I)^n v_m\end{aligned}$$

$$
\begin{aligned}
&= a_1(T-\lambda_j I)^k \left(\prod_{\substack{i=1\\ i\neq j,1}}^{m}(T-\lambda_i I)^n\right)(T-\lambda_1 I)^n v_1 \\
&\quad + \cdots + a_{j-1}(T-\lambda_j I)^k \left(\prod_{\substack{i=1\\ i\neq j,j-1}}^{m}(T-\lambda_i I)^n\right)(T-\lambda_{j-1} I)^n v_{j-1} \\
&\quad + a_j \left(\prod_{\substack{i=1\\ i\neq j}}^{m}(T-\lambda_i I)^n\right)(T-\lambda_j I)^k v_j \\
&\quad + a_{j+1}(T-\lambda_j I)^k \left(\prod_{\substack{i=1\\ i\neq j,j+1}}^{m}(T-\lambda_i I)^n\right)(T-\lambda_{j+1} I)^n v_{j+1} \\
&\quad + \cdots + a_m(T-\lambda_j I)^k \left(\prod_{\substack{i=1\\ i\neq j,m}}^{m}(T-\lambda_i I)^n\right)(T-\lambda_m I)^n v_m && \text{Theorem 5.4} \\
&= a_j \prod_{\substack{i=1\\ i\neq j}}^{m}(T-\lambda_i I)^n w && \text{Theorem 8.5} \\
&= a_j \prod_{\substack{i=1\\ i\neq j}}^{m}(\lambda_j-\lambda_i)^n w
\end{aligned}
$$

so $a_j = 0$, as desired. ■

- **Nilpotent** (operator): An operator T such that $T^j = 0$ for some positive integer j.
- We now show that we never need to raise a nilpotent operator to a $j > \dim V$ to make it equal to zero.

  Theorem 8.7. *Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then $N^{\dim V} = 0$.*

  Proof. Since N is nilpotent, we know that there exists a nonnegative integer j such that

  $$(N - 0I)^j v = N^j v = 0 = 0v$$

  for any $v \in V$. Thus, $G(0, N) = V$. It follows by Theorem 8.5 that $V = G(0, N) = \operatorname{null}(N - 0I)^{\dim V} = \operatorname{null} N^{\dim V}$. Consequently, for any $v \in V$, $N^{\dim V} v = 0$, so $N^{\dim V} = 0$, as desired. ■

- We now show that if N is nilpotent, there exists a basis of V such that $\mathcal{M}(N)$ is more than half zeroes.

  Theorem 8.8. *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*

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

  , i.e., where all entries on and below the diagonal are zeroes.

  Proof. First choose a basis of $\operatorname{null} N$. Then extend this to a basis of $\operatorname{null} N^2$, then to a basis of $\operatorname{null} N^3$, on and on up until we have extended it to a basis $v_1, \dots, v_n$ of $\operatorname{null} N^{\dim V}$ (which, incidentally, will be a basis of V since $\operatorname{null} N^{\dim V} = V$ by Theorem 8.7). We will prove that $\mathcal{M}(N, (v_1, \dots, v_n))$ has the desired form.

Let k be the smallest positive integer such that $v_1 \in \operatorname{null} N^k$. Then $0 = N^k v_1 = N^{k-1} N v_1$, so $N v_1 \in \operatorname{null} N^{k-1} = \{0\}$ by the condition on k. It follows that $N v_1 = 0$, so since $v_1, \dots, v_n$ is linearly independent (as a basis), $\mathcal{M}(N, (v_1, \dots, v_n))_{\cdot,1} = \mathcal{M}(N v_1)$ has only zero entries. Apply the same argument to any other vector in $\operatorname{null} N^k$, getting all zero columns for some number of columns. Having done this, move onto the first vector in the basis that is not in $\operatorname{null} N^k$. Let this vector be v_i. Then in a similar fashion to before, $N v_i \in \operatorname{null} N^k$, so $N v_i$ is a linear combination of all vectors before v_i. Thus, all nonzero entries in $\mathcal{M}(()N, (v_1, \dots, v_n))_{\dots,i} = \mathcal{M}(()N v_i)$ are above the diagonal. We continue in this fashion for the whole basis. ∎

Exercises

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

Proof. Suppose for the sake of contradiction that $\lambda \neq 0$ is an eigenvalue of N with corresponding eigenvector v. Then

$$0 = 0v = N^{\dim V} v = \lambda^{\dim V} v \neq 0$$

a contradiction. ∎

8.B Decomposition of an Operator

10/23:

- We are going to build up in this section to a proof that while not every operator's domain can be decomposed into eigenspaces, every operator's domain can be decomposed into generalized eigenspaces.
- We first show that the null and rance spaces of every polynomial of an operator T are invariant under T.

Theorem 8.9. *Suppose $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbb{F})$. Then $\operatorname{null} p(T)$ and $\operatorname{range} p(T)$ are invariant under T.*

Proof. To prove that $\operatorname{null} p(T)$ is invariant under T, it will suffice to show that for any $v \in \operatorname{null} p(T)$, $Tv \in \operatorname{null} p(T)$. Let $v \in \operatorname{null} p(T)$ be arbitrary. Then $(p(T))v = 0$. It follows that

$$\begin{aligned}
(p(T))(Tv) &= (p(T)T)v \\
&= (Tp(T))v && \text{Theorem 5.4} \\
&= T(p(T)v) \\
&= T(0) \\
&= 0 && \text{Theorem 3.2}
\end{aligned}$$

Therefore, $Tv \in \operatorname{null} p(T)$, as desired.

To prove that $\operatorname{range} p(T)$ is invariant under T, it will suffice to show that for any $v \in \operatorname{range} p(T)$, $Tv \in \operatorname{range} p(T)$. Let $v \in \operatorname{range} p(T)$ ve arbitrary. Then there exists $u \in V$ such that $p(T)u = v$. It follows that

$$\begin{aligned}
Tv &= T(p(T)u) \\
&= p(T)(Tu) && \text{Theorem 5.4}
\end{aligned}$$

Therefore, $Tv \in \operatorname{range} p(T)$, as desired. ∎

- We now prove the main result we've been working up to. It 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" (Axler, 2015, p. 252).

Theorem 8.10. *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*) $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$.

Proof. We induct on $n = \dim V$. For the base case $n = 1$, since we know that there exists *an* eigenvalue of T (see Theorem 5.5) to which there can only correspond one set of eigenvector of T (because V is one-dimensional), we have that $V = G(\lambda_1, T)$, as desired. Now suppose using strong induction that we have proven the claim for all dimensions strictly less than n; we now seek to prove it for n. Let's begin.

By Theorem 5.5, T has an eigenvalue λ_1. Thus,

$$\begin{aligned} V &= \operatorname{null}(T - \lambda_1 I)^n \oplus \operatorname{range}(T - \lambda_1 I)^n && \text{Theorem 8.4} \\ &= G(\lambda_1, T) \oplus U && \text{Theorem 8.5} \end{aligned}$$

where we let $U = \operatorname{range}(T - \lambda_1 I)^n$. Under this definition, we have by Theorem 8.9 with $p(z) = (z - \lambda_1)^n$ that U is invariant under T. Additionally, since λ_1 is an eigenvalue of T (thus with some corresponding generalized eigenvector), $G(\lambda_1, T) \neq \{0\}$. Therefore, $\dim U < n$ by Theorem 3.21.

Consider $T|_U$. 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 elements of $G(\lambda_1, T)$. Thus, the eigenvalues of $T|_U$ are exactly $\lambda_1, \dots, \lambda_m$.

Having established that $\dim U < n$ and $\lambda_2, \dots, \lambda_m$ are the distinct eigenvalues of $T|_U$, we have by the induction hypothesis that $U = G(\lambda_2, T|_U) \oplus \cdots + G(\lambda_m, T|_U)$. Consequently, all that's left is to show that $G(\lambda_k, T|_U) = G(\lambda_k, T)$ for each $k = 2, \dots, m$. Clearly, $G(\lambda_k, T|_U) \subset G(\lambda_k, T)$. In the other direction, suppose that $v \in G(\lambda_k, T)$. As an element of $V = G(\lambda_1, T) \oplus U$, we have that $v = v_1 + u$ where $v_1 \in G(\lambda_1, T)$ and $u \in U$. However, since v and v_1 correspond to different eigenvalues, we have by Theorem 8.6 that they are linearly independent. Thus, $v_1 = 0$. Therefore, $v = u \in U$, so $v \in G(\lambda_k, T|_U)$, as desired. ∎

(*b*) *Each $G(\lambda_j, T)$ is invariant under T.*

Proof. By Theorem 8.5, $G(\lambda_j, T) = \operatorname{null}(T - \lambda_j I)^{\dim V}$. By Theorem 8.9, if $p(z) = (z - \lambda_j)^{\dim V}$, then $\operatorname{null} p(T) = \operatorname{null}(T - \lambda_j I)^{\dim V}$ is invariant under T. Therefore, $G(\lambda_j, T)$ is invariant under T, as desired. ∎

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

Proof. Let $v \in G(\lambda_j, T)$. Then

$$\begin{aligned} ((T - \lambda_j I)|_{G(\lambda_j, T)})^{\dim V} v &= (T - \lambda_j I)^{\dim V} v \\ &= 0 && \text{Theorem 8.5} \end{aligned}$$

Thus, since there is a positive integer power of $(T - \lambda_j I)|_{G(\lambda_j, T)}$ such that $(T - \lambda_j I)|_{G(\lambda_j, T)} = 0$, $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent, as desired. ∎

- We now show that while T may not have enough eigenvectors to form an eigenbasis, T always has enough generalized eigenvectors to form a basis.

Theorem 8.11. *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 $G(\lambda_j, T)$ proven to compose V in Theorem 8.10a. Concatenate these bases to form a basis of V consisting of generalized eigenvectors of V. ∎

- **Multiplicity** (of an eigenvalue λ of $T \in \mathcal{L}(V)$): The dimension of $G(\lambda, T)$. *Also known as* **algebraic multiplicity**.

- We now prove an obvious consequence of the definition of multiplicity.

Theorem 8.12. *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. We have from Theorem 8.10a that

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

Thus, by Theorem 3.21,

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

Therefore, by the definition of the multiplicity of an eigenvalue of T, the above equation proves the desired result. ■

- **Geometric multiplicity** (of an eigenvalue λ of $T \in \mathcal{L}(V)$): The dimension of $E(\lambda, T)$.
- **Block diagonal matrix**: 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.

- We now prove the results we have proven before, but in matrix form.

Theorem 8.13. *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, with multiplicities $d_1, \dots, d_m$. 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. Let $j \in \{1, \dots, m\}$ be arbitrary. By Theorem 8.10c, $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent. Thus, by Theorem 8.8, we can choose a basis of $G(\lambda_j, T)$ (which will be of length d_j) such that $\mathcal{M}((T - \lambda_j I)|_{G(\lambda_j, T)})$ with respect to this basis has all zeroes on and below the diagonal. It follows that $\mathcal{M}(T|_{G(\lambda_j, T)}) = \mathcal{M}((T - \lambda_j I)|_{G(\lambda_j, T)}) + \mathcal{M}(\lambda_j I|_{G(\lambda_j, T)})$ will have the necessary form to be an A_j. Therefore, since concatenating the bases of each $G(\lambda_j, T)$ gives a basis of V by Theorem 8.10a, $\mathcal{M}(T)$ with respect to this basis will have the desired form. ■

- We now harness the power of some of our newer theorems to prove some further results about square roots.

Theorem 8.14. *Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then $I + N$ has a square root.*

Proof. Consider the Taylor series expansion of $\sqrt{I+N}$, as if I were the number 1 and N were some number x. We would have

$$\sqrt{I+N} = I + a_1 N + a_2 N^2 + \cdots$$

for some set of coefficients $a_1, a_2, \dots$. Now while an infinite sum of operators cannot be an operator, this sum can: Since N is nilpotent, $N^m = 0$ for some positive integer m, so every term of degree $j \geq m$ is zero and the sum is finite. Indeed, for the sum above to be a square root of $I+N$, we need only require that

$$\begin{aligned}
I + N &= \left(\sqrt{I+N}\right)^2 \\
&= \left(I + a_1 N + \cdots + a_{m-1} N^{m-1}\right)^2 \\
&= I + 2a_1 N + (2a_2 + a_1^2)N^2 + (2a_3 + 2a_1 a_2)N^3 + \cdots + cN^{m-1}
\end{aligned}$$

where c stands in for a much more complicated coefficient in terms of $a_1, \dots, a_{m-1}$. To do so, simply choose a_1 such that $2a_1 = 1$ (i.e., choose $a_1 = 1/2$), choose a_2 such that $2a_2 + a_1^2 = 0$ (i.e., choose $a_2 = -1/8$), choose a_3 such that $2a_3 + 2a_1 a_2 = 0$, and on and on. ∎

Theorem 8.15. *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. Let $j \in \{1, \dots, m\}$ be arbitrary. Since $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent (see Theorem 8.10c), we have that there exists a nilpotent operator $N_j \in \mathcal{L}(G(\lambda_j, T))$ such that $T|_{G(\lambda_j,T)} = \lambda_j I + N_j$. Additionally, we know that $\lambda_j \neq 0$: Since T is an operator on a finite-dimensional complex vector space, Theorem 5.7 implies that T has an upper-triangular matrix with respect to some basis of V; since T is invertible, Theorem 5.8 implies all of the diagonal entries of this matrix are nonzero; since Theorem 5.9 implies that the eigenvalues of T are exactly the diagonal entries of this matrix, we know that they are nonzero, i.e., in particular, $\lambda_j \neq 0$. Thus, since we can divide by λ_j, we have that

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

By Theorem 8.14, $I + N_j/\lambda_j$ has a square root. Naturally, this operator times a square root of λ_j is a square root R_j of $T|_{G(\lambda_j,T)}$. This combined with the fact that any $v \in V$ can be written uniquely in the form $u_1 + \cdots + u_m$ where $u_i \in G(\lambda_i, T)$ for each $i = 1, \dots, m$ allows us to define the operator $R \in \mathcal{L}(V)$ by

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

To prove that $R^2 = T$, it will suffice to show that $R^2 v = Tv$ for all $v \in V$. Let $v \in V$ be arbitrary. Then it follows from all of our definitions that

$$\begin{aligned}
R^2 v &= R_1^2 u_1 + \cdots + R_m^2 u_m \\
&= \lambda_1 \left(I + \frac{N_1}{\lambda_1}\right) u_1 + \cdots + \lambda_m \left(I + \frac{N_m}{\lambda_m}\right) u_m \\
&= T|_{G(\lambda_1,T)} u_1 + \cdots + T|_{G(\lambda_m,T)} u_m \\
&= Tu_1 + \cdots + Tu_m \\
&= Tv
\end{aligned}$$

as desired. ∎

- Note that the techniques in this section can be adapted 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

10/25: **9** Suppose A and B are block diagonal matrices of the form

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

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

$$AB = \begin{pmatrix} A_1B_1 & & 0 \\ & \ddots & \\ 0 & & A_mB_m \end{pmatrix}$$

Proof. Suppose that A, B are $n \times n$ matrices. By the definition of matrix multiplication, if $1 \leq i, k \leq \dim A_1 = \dim B_1$ we have that

$$\begin{aligned} (AB)_{i,k} &= \sum_{r=1}^{n} A_{i,r}B_{r,k} \\ &= \sum_{r=1}^{\dim A_1} A_{i,r}B_{r,k} + \sum_{\dim A_1+1}^{n} 0 \cdot 0 \\ &= \sum_{r=1}^{\dim A_1} A_{1_{i,r}}B_{2_{r,k}} \\ &= (A_1B_1)_{i,k} \end{aligned}$$

We can do something similar for every other submatrix along the diagonal. ■

8.C Characteristic and Minimal Polynomials

10/23:

- **Characteristic polynomial** (of $T \in \mathcal{L}(V)$, V complex): The polynomial

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

  where $\lambda_1, \dots, \lambda_m$ are the distinct eigenvalues of T with corresponding multiplicities $d_1, \dots, d_m$.
- We first show some elementary properties of the characteristic polynomial.

  Theorem 8.16. *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then*

  (a) *The characteristic polynomial of T has degree* $\dim V$.

  Proof. This follows from the definition of the characteristic polynomial and Theorem 8.12. ■

  (b) *The zeroes of the characteristic polynomial of T are the eigenvalues of T.*

  Proof. This follows from the definition of the characteristic polynomial. ■

- The simple definition (as opposed to the determinant-based definition) of the characteristic polynomial given here affords a simple proof of the Complex Cayley-Hamilton Theorem.

  Theorem 8.17 (Complex Cayley-Hamilton Theorem[1])**.** *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let q denote the characteristic polynomial of T. Then $q(T) = 0$.*

[1] This result is named for English mathematician Arthur Cayley and Irish mathematician William Rowan Hamilton, both of whom found great success even before the completion of their undergraduate degrees.

Proof. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. To prove that $q(T) = 0$, it will suffice to show that $q(T)v = 0$ for all $v \in V$. Let $v \in V$ be arbitrary. Then by Theorem 8.10a, $v = u_1 + \cdots + u_m$ where each $u_j \in G(\lambda_j, T)$. Additionally, Theorems 8.10c and 8.7 assert that each $(T - \lambda_j I)^{d_j}|_{G(\lambda_j,T)} = 0$. Therefore, we have that

$$q(T)v = \prod_{i=1}^{m} (T - \lambda_i I)^{d_i} (u_1 + \cdots + u_m) = 0$$

since after distributing the operator to each term in the sum, we can restrict the domain of each exponential to $G(\lambda_j, T)$ and commute the $(T - \lambda_j I)^{d_j}|_{G(\lambda_j,T)}$ term to be applied first (by Theorem 5.4). ■

- **Monic polynomial**: A polynomial whose highest-degree coefficient equals 1.
- We now prove that we can associate a unique **minimal polynomial** with each operator T.

Theorem 8.18. *Suppose $T \in \mathcal{L}(V)$. Then there is a unique monic polynomial p of smallest degree such that $p(T) = 0$.*

Proof. We first show that such a polynomial exists; then we prove its uniqueness. Let's begin.

Let $n = \dim V$. Consider the list of operators $I, T, T^2, \dots, T^{n^2} \in \mathcal{L}(V)$. Since $\dim \mathcal{L}(V) = n^2$ (see Theorem 3.15) and the length of this list is $n^2 + 1 > n^2$, Theorem 2.3 implies that this list is linearly dependent. Let m be the smallest positive integer such that

$$I, T, T^2, \dots, T^m$$

is linearly dependent.

By the Linear Dependence Lemma, one of the operators in the above list is a linear combination of the previous ones. By the choice of m, we know that this operator is T^m. Thus, there exist scalars $a_0, \dots, a_{m-1} \in \mathbb{F}$ such that

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

Therefore, if we consider the monic polynomial $p \in \mathcal{P}(\mathbb{F})$ defined by

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

we know by the above that $p(T) = 0$.

Suppose for the sake of contradiction that there exists a monic polynomial $q \in \mathcal{P}(\mathbb{F})$ of degree less than or equal to m such that $q(T) = 0$. If $\deg q < m$, then $I, T, T^2, \dots, T^{\deg q}$ will be linearly independent, so $q(z) = 0$. But this implies the highest degree coefficient of q is not 1, so q is not monic, a contradiction. On the other hand, if $\deg q = m$, then we have that $(p - q)(T) = 0$ as well. However, since both p and z have a $1z^m$ term that cancels in $p - q$, $\deg(p - q) < m$. Thus, we can reach the same contradiction in the other case. ■

- **Minimal polynomial** (of T): The unique monic polynomial p of smallest degree such that $p(T) = 0$.
  - By the proof of Theorem 8.18, the degree of the minimal polynomial of each operator on V is at most $(\dim V)^2$.
  - By the Complex Cayley-Hamilton Theorem, the degree of the minimal polynomial of each operator on V *complex* is at most $\dim V$.
  - The minimal polynomial can be computed by considering a homogeneous system of equations

    $$a_0 \mathcal{M}(I) + a_1 \mathcal{M}(T) + \cdots + a_m \mathcal{M}(T)^m = 0$$

    with $(\dim V)^2$ equations in $a_0, \dots, a_m$ for successive values of m until a solution exists. This solution would give the coefficients of the minimal polynomial.

- We now characterize all polynomials that when applied to an operator give the 0 operator.

  Theorem 8.19. *Suppose $T \in \mathcal{L}(V)$ and $q \in \mathcal{P}(\mathbb{F})$. 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.

  Suppose first that $q(T) = 0$. By the Division Algorithm for Polynomials, we have that $q = ps + r$ where $\deg r < \deg p$. Thus,
  $$0 = q(T) = p(T)s(T) + r(T) = r(T)$$
  It follows that $r = 0$ (otherwise, since $\deg r < \deg p$, r divided by the coefficient of the highest-order term would be a monic polynomial that when applied to T of degree less than minimal polynomial, a contradiction). Therefore, q is a polynomial multiple of the minimal polynomial of T, as desired.

  Now suppose that q is a polynomial multiple of the minimal polynomial of T. The $q = ps$ for some $s \in \mathcal{P}(\mathbb{F})$. It follows that
  $$q(T) = p(T)s(T) = 0s(T) = 0$$
  as desired. ■

- We can now apply our discussion of the minimal polynomial back to the characteristic polynomial.

  Theorem 8.20. *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Then the characteristic polynomial of T is a polynomial multiple of the minimal polynomial of T.*

  Proof. Let q denote the characteristic polynomial of T. By the Complex Cayley-Hamilton Theorem, $q(T) = 0$. Therefore, by Theorem 8.19, $q(T)$ is a multiple of the minimal polynomial of T, as desired. ■

- We now show that like the eigenvalues of T are the zeroes of the characteristic polynomial, they are the zeroes of the minimal polynomial[2].

  Theorem 8.21. *Let $T \in \mathcal{L}(V)$. Then the zeroes 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. We will first show that every zero of p is an eigenvalue of T. Then, we will show that every eigenvalue of T is a zero of p. Let's begin.

  Suppose first that $\lambda \in \mathbb{F}$ satisfies $p(\lambda) = 0$. Thus, by Theorem 4.3, there exists a polynomial $q \in \mathcal{P}(\mathbb{F})$ such that $p(z) = (z - \lambda)q(z)$, where q must be a monic polynomial since p is one. Consequently, since $p(T) = 0$, we have that
  $$0 = (T - \lambda I)q(T)v$$
  for all $v \in V$. In particular, we must have $q(T)v \neq 0$ for some $v \in V$ (otherwise $q(T)$ with $\deg q < \deg p$ would be the minimal polynomial of T). Therefore, λ is an eigenvalue of T with corresponding eigenvector $q(T)v$ for this v, as desired.

  Now suppose that $\lambda \in \mathbb{F}$ is an eigenvalue of T. Then there exists a nonzero $v \in V$ such that $Tv = \lambda v$. It follows that $T^j v = \lambda^j v$ for every nonnegative integer j. Thus, we have that
  $$\begin{aligned} 0 &= p(T)v \\ &= (a_0 + a_1 T + a_2 T^2 + \cdots + a_{m-1} T^{m-1} + T^m)v \\ &= (a_0 + a_1 \lambda + a_2 \lambda^2 + \cdots + a_{m-1} \lambda^{m-1} + \lambda^m)v \\ &= p(\lambda)v \end{aligned}$$
  But since $v \neq 0$, we must have $p(\lambda) = 0$, as desired. ■

- Axler (2015) gives some examples of how the previous results can be applied to tangible problems.

[2]It would appear then that the minimal polynomial can be written in the form $(z - \lambda_1) \cdots (z - \lambda_m)$.

8.D Jordan Form

- Theorem 8.13 got us to a pretty nice form for every operator T. We now build up to an even nicer one.
- We first show that every nilpotent operator has a corresponding basis consisting of certain powers of N applied to a select number of vectors.

Theorem 8.22. *Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then there exist vectors $v_1, \dots, v_n \in V$ and nonnegative integers $m_1, \dots, m_n$ such that*

(*a*) $N^{m_1}v_1, \dots, Nv_1, v_1, \dots, N^{m_n}v_n, \dots, Nv_n, v_n$ *is a basis of* V.

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

Proof. We induct on $n = \dim V$. For the base case $n = 1$, 0 is the only nilpotent operator on V. Thus, choose any nonzero v_1 and let $m_1 = 0$.

Now suppose using strong induction that we have proven the claim for all vector spaces of dimension less than n; we now seek to prove the claim for n. First off, note that N is not injective: If j is the smallest nonnegative integer such that of $N^j = 0$, then there exists $v \in V$ such that $N^{j-1}v \neq 0$; it follows that although $v \neq 2v$ and hence $N^{j-1}v \neq 2N^{j-1}v = N^{j-1}2v$, $N(N^{j-1}v) = 0$ and $N(N^{j-1}(2v)) = 2N(N^{j-1}v) = 2 \cdot 0 = 0$. It follows by Theorem 3.18 that N is not surjective. Thus, $\operatorname{range} N \neq V$, so we must have $\dim \operatorname{range} N < \dim V$.

Consider $N|_{\operatorname{range} N} \in \mathcal{L}(\operatorname{range} N)$, to which we can apply our inductive hypothesis by the previous result. Doing so, we find that there exist vectors $v_1, \dots, v_n \in \operatorname{range} N$ and nonnegative integers $m_1, \dots, m_n$ such that

$$N^{m_1}v_1, \dots, Nv_1, v_1, \dots, N^{m_n}v_n, \dots, Nv_n, v_n \tag{8.1}$$

is a basis of $\operatorname{range} N$ and

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

Let $j \in \{1, \dots, n\}$ be arbitrary. Since $v_j \in \operatorname{range} N$, there exists $u_j \in V$ such that $Nu_j = v_j$. It follows that $N^{k+1}u_j = N^k v_j$ for each j and every nonnegative integer k. We now seek to prove that

$$N^{m_1+1}u_1, \dots, Nu_1, u_1, \dots, N^{m_n+1}u_n, \dots, Nu_n, u_n \tag{8.2}$$

is a linearly independent list of vectors in V. To do so, consider a linear combination of the vectors in List 8.2 that is equal to zero. Applying N to said linear combination gives a linear combination of List 8.1 that is equal to zero with related coefficients (since $Nu_j = v_j$, $N(N^k u_j) = N^k v_j$ for all $1 \leq k \leq m_j$, and $N(N^{m_j+1}u_j) = N^{m_j+1}v_j = 0$ for each $j = 1, \dots, n$). But since List 8.1 is linearly independent as a basis, all coefficients are zero with the possible exception of those of the vectors $N^{m_1+1}u_1, \dots, N^{m_n+1}u_n$, since those vectors go to zero as described above. However, once again, the linear independence of List 8.1, to which each of these vectors belongs in the form $N^{m_j}v_j$, implies that their coefficients are equal to zero as well. Thus, we have proven that List 8.2 is linearly independent, as desired.

Using Theorem 2.8, extend List 8.2 to a basis

$$N^{m_1+1}u_1, \dots, Nu_1, u_1, \dots, N^{m_n+1}u_n, \dots, Nu_n, u_n, w_1, \dots, w_p \tag{8.3}$$

of V. Let $j \in \{1, \dots, p\}$ be arbitrary. We know that $Nw_j \in \operatorname{range} N$. Thus, Nw_j is in the span of List 8.1. But since each vector in List 8.1 equals N applied to a vector in List 8.2, we have that $Nw_j = Nx_j$ for some x_j in the span of List 8.2. Let

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

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

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

spans V since its span contains each x_j (in the span of the List 8.2 part) and each u_{n+j} and hence each w_j (and because List 8.3 spans V). Thus, since the above spanning list has the same length as the basis of V in List 8.3, Theorem 2.13 implies that it is a basis of V. It clearly has the required form (choose $m_j = 0$ for $j = n+1, \dots, n+p$). Additionally, we have $Nu_{n+j} = Nw_j - Nx_j = 0$ for each $j = 1, \dots, n$, verifying part (b). ■

- **Jordan basis** (for $T \in \mathcal{L}(V)$): A basis of V such that 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
$$A_j = \begin{pmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{pmatrix}$$

- We are now ready for the main result.

Theorem 8.23 (Jordan Form[3]). *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. We divide into two cases (T is nilpotent and T is not nilpotent), using one to prove the other. Suppose first that T be a nilpotent operator on V. Consider the vectors $v_1, \dots, v_n \in V$ associated with it by Theorem 8.22. Let $j \in \{1, \dots, n\}$ be arbitrary. Notice that N sends the first vector in the list $N^{m_j} v_j, \dots, Nv_j, v_j$ to 0 and every other vector to the previous one in the list. Thus, applying this observation to all j, we realize that the matrix of N with respect to the basis given by Theorem 8.22 is block diagonal with each matrix on the diagonal having the form
$$\begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix}$$
Combining this result with Exercise 8.A.1 proves that the desired result holds for nilpotent operators.

Now suppose that T is not a nilpotent operator. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. By Theorem 8.10c, each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent. Thus, by the proof of the first case, there exists a Jordan basis for each $(T - \lambda_j I)|_{G(\lambda_j, T)}$. It follows since $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$ (see Theorem 8.10a) that concatenating all of these bases gives a basis of V that is a Jordan basis for T. ■

[3] This result is named for French mathematician Camille Jordan, who published the first proof of this theorem.