THE MINIMAL POLYNOMIAL AND SOME APPLICATIONS

KEITH CONRAD

1. Introduction

The easiest matrices to compute with are the diagonal ones. The sum and product of diagonal matrices can be computed componentwise along the main diagonal, and taking powers of a diagonal matrix is simple too. All the complications of matrix operations are gone when working only with diagonal matrices. If a matrix A is not diagonal but can be conjugated to a diagonal matrix, say $D := PAP^{-1}$ is diagonal, then $A = P^{-1}DP$ so $A^k = P^{-1}D^kP$ for all integers k, which reduces us to computations with a diagonal matrix. In many applications of linear algebra (*e.g.*, dynamical systems, differential equations, Markov chains, recursive sequences) powers of a matrix are crucial to understanding the situation, so the relevance of knowing when we can conjugate a nondiagonal matrix into a diagonal matrix is clear.

We want look at the coordinate-free formulation of the idea of a diagonal matrix, which will be called a diagonalizable operator. There is a special polynomial, the minimal polynomial (generally not equal to the characteristic polynomial), which will tell us exactly when a linear operator is diagonalizable. The minimal polynomial will also give us information about nilpotent operators (those having a power equal to O).

All linear operators under discussion are understood to be acting on nonzero finite-dimensional vector spaces over a given field F.

2. Diagonalizable Operators

Definition 2.1. We say the linear operator $A\colon V \to V$ is *diagonalizable* when it admits a diagonal matrix representation with respect to some basis of V: there is a basis $\mathcal{B}$ of V such that the matrix $[A]_{\mathcal{B}}$ is diagonal.

Let's translate diagoinalizability into the language of eigenvectors rather than matrices.

Theorem 2.2. *The linear operator $A\colon V \to V$ is diagonalizable if and only if there is a basis of eigenvectors for A in V.*

Proof. Suppose there is a basis $\mathcal{B} = \{e_1, \ldots, e_n\}$ of V in which $[A]_{\mathcal{B}}$ is diagonal:

$$[A]_{\mathcal{B}} = \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & a_n \end{pmatrix}.$$

Then $Ae_i = a_ie_i$ for all i, so each e_i is an eigenvector for A. Conversely, if V has a basis $\{v_1, \ldots, v_n\}$ of eigenvectors of A, with $Av_i = \lambda_i v_i$ for $\lambda_i \in F$, then in this basis the matrix representation of A is $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. □

A basis of eigenvectors for an operator is called an *eigenbasis*.

An example of a linear operator that is *not* diagonalizable over all fields F is $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ acting on F^2. Its only eigenvectors are the vectors $\binom{x}{0}$. There are not enough eigenvectors to form a basis for F^2, so $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ on F^2 does not diagonalize. Remember this example! Since $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$

and $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ have the same characteristic polynomial, and the second matrix is diagonalizable, the characteristic polynomial *doesn't* determine (in general) if an operator is diagonalizable.

Here are the main results we will obtain about diagonalizability:

(1) There are ways of determining if an operator is diagonalizable without having to look explicitly for a basis of eigenvectors.
(2) When F is *algebraically closed*, "most" operators on a finite-dimensional F-vector space are diagonalizable.
(3) There is a polynomial, the minimal polynomial of the operator, which can be used to detect diagonalizability.
(4) If two operators are each diagonalizable, they can be simultaneously diagonalized (*i.e.*, there is a common eigenbasis) precisely when they *commute*.

Let's look at three examples related to diagonalizability over $\mathbf{R}$ and $\mathbf{C}$.

Example 2.3. Let $R = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, the 90-degree rotation matrix acting on $\mathbf{R}^2$. It is not diagonalizable on $\mathbf{R}^2$ since there are no eigenvectors: a rotation in $\mathbf{R}^2$ sends no nonzero vector to a scalar multiple of itself. This geometric reason is complemented by an algebraic reason: the characteristic polynomial $T^2 + 1$ of R has no roots in $\mathbf{R}$, so there are no real eigenvalues and thus no eigenvectors in $\mathbf{R}^2$. However, there are roots $\pm i$ of $T^2 + 1$ in $\mathbf{C}$, and there are eigenvectors of R as an operator on $\mathbf{C}^2$ rather than $\mathbf{R}^2$. Eigenvectors of R in $\mathbf{C}^2$ for the eigenvalues i and $-i$ are $\binom{i}{1}$ and $\binom{-i}{1}$, respectively. In the basis $\mathcal{B} = \{\binom{i}{1}, \binom{-i}{1}\}$, the matrix of R is $[R]_{\mathcal{B}} = \left(\begin{smallmatrix} i & 0 \\ 0 & -i \end{smallmatrix}\right)$, where the first diagonal entry is the eigenvalue of the first basis vector in $\mathcal{B}$ and the second diagonal entry is the eigenvalue of the second basis vector in $\mathcal{B}$. (Review the proof of Theorem 2.2 to see why this relation between the ordering of vectors in an eigenbasis and the ordering of entries in a diagonal matrix always holds.)

Put more concretely, since passing to a new matrix representation of an operator from an old one amounts to conjugating the old matrix representation by the change-of-basis matrix expressing the old basis in terms of the new basis, we must have $\left(\begin{smallmatrix} i & 0 \\ 0 & -i \end{smallmatrix}\right) = PRP^{-1}$ where $P = ([\binom{1}{0}]_{\mathcal{B}}\ [\binom{0}{1}]_{\mathcal{B}}) = \left(\begin{smallmatrix} -i/2 & 1/2 \\ i/2 & 1/2 \end{smallmatrix}\right)$. Verify this P really conjugates R to $\left(\begin{smallmatrix} i & 0 \\ 0 & -i \end{smallmatrix}\right)$. Note $P^{-1} = \left(\begin{smallmatrix} i & -i \\ 1 & 1 \end{smallmatrix}\right)$ has as its columns the eigenvectors of R in the standard basis $\binom{1}{0}, \binom{0}{1}$ of $\mathbf{C}^2$.

Example 2.4. Every $A \in \mathrm{M}_n(\mathbf{R})$ satisfying $A = A^\top$ can be diagonalized over $\mathbf{R}$. This is a significant result, called the real spectral theorem. (Any theorem that gives sufficient conditions under which an operator can be diagonalized is called a spectral theorem, because the eigenvalues of an operator is called its spectrum.) The essential step in the proof of the real spectral theorem is to show that every real symmetric matrix has a real eigenvalue.

Example 2.5. Any $A \in \mathrm{M}_n(\mathbf{C})$ satisfying $A\overline{A}^\top = \overline{A}^\top A$ is diagonalizable in $\mathrm{M}_n(\mathbf{C})$.[1] When A is real, so $\overline{A}^\top = A^\top$, saying $AA^\top = A^\top A$ is weaker than saying $A = A^\top$. In particular, the real matrix $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ commutes with its transpose and thus is diagonalizable over $\mathbf{C}$, but the real spectral theorem does not apply to this matrix and in fact this matrix isn't diagonalizable over $\mathbf{R}$ (it has no real eigenvalues).

3. Distinct Eigenvalues and Diagonalizability

If a linear operator on a finite-dimensional F-vector space is diagonalizable, its eigenvalues all lie in F, since a diagonal matrix representation has the eigenvalues along the diagonal.

[1] A complex square matrix A satisfying $A\overline{A}^\top = \overline{A}^\top A$ is called *normal*, and normal matrices are unitarily diagonalizable: $A = UDU^{-1}$ where D is diagonal and U is unitary, meaning $U\overline{U}^\top = I_n$. While the condition of A being unitarily diagonalizable is equivalent to $A\overline{A}^\top = \overline{A}^\top A$, the condition of being diagonalizable alone is not equivalent to an algebraic identity on complex matrices.

(See the proof of Theorem 2.2.) The converse is false: if all the eigenvalues of an operator are in F this does *not* necessarily mean the operator is diagonalizable. Just think about our basic example $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, whose only eigenvalue is 1. It is a "repeated eigenvalue," in the sense that the characteristic polynomial $(T-1)^2$ has 1 as a repeated root. Imposing an additional condition, that the eigenvalues lie in F and are *simple* roots of the characteristic polynomial, does force diagonalizability. To prove this, we start with a general lemma on eigenvalues and linear independence.

Lemma 3.1. *Eigenvectors for distinct eigenvalues are linearly independent. More precisely, if $A\colon V \to V$ is linear and $v_1, \ldots, v_r$ are eigenvectors of A with distinct eigenvalues $\lambda_1, \ldots, \lambda_r$, the v_i's are linearly independent.*

Proof. This will be an induction on r. The case $r = 1$ is easy. If $r > 1$, suppose there is a linear relation

$$c_1 v_1 + \cdots + c_{r-1} v_{r-1} + c_r v_r = 0 \tag{3.1}$$

with $c_i \in F$. Apply A to both sides: v_i becomes $Av_i = \lambda_i v_i$, so

$$c_1 \lambda_1 v_1 + \cdots + c_{r-1} \lambda_{r-1} v_{r-1} + c_r \lambda_r v_r = 0. \tag{3.2}$$

Multiply the linear relation in (3.1) by λ_r:

$$c_1 \lambda_r v_1 + \cdots + c_{r-1} \lambda_r v_{r-1} + c_r \lambda_r v_r = 0. \tag{3.3}$$

Subtracting (3.3) from (3.2), the last terms on the left cancel:

$$c_1(\lambda_1 - \lambda_r)v_1 + \cdots + c_{r-1}(\lambda_{r-1} - \lambda_r)v_{r-1} = 0.$$

Now we have a linear relation with $r-1$ eigenvectors having distinct eigenvalues. By induction, all the coefficients are 0: $c_i(\lambda_i - \lambda_r) = 0$ for $i = 1, \ldots, r-1$. Since $\lambda_1, \ldots, \lambda_{r-1}, \lambda_r$ are distinct, $\lambda_i - \lambda_r \neq 0$ for $i = 1, \ldots, r-1$. Thus $c_i = 0$ for $i = 1, \ldots, r-1$. Now our original linear relation (3.1) becomes $c_r v_r = 0$. The vector v_r is not 0 (eigenvectors are always nonzero by definition), so $c_r = 0$. □

Theorem 3.2. *A linear operator on V whose characteristic polynomial is a product of linear factors in $F[T]$ with distinct roots is diagonalizable.*

Proof. The assumption is that there are n different eigenvalues in F, where $n = \dim V$. Call them $\lambda_1, \ldots, \lambda_n$. Let e_i be an eigenvector with eigenvalue λ_i. The eigenvalues are distinct, so by Lemma 3.1 the e_i's are linearly independent. Since there are n of these vectors, the e_i's are a basis of V, so the operator admits an eigenbasis and is diagonalizable. □

Example 3.3. The matrix $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ has characteristic polynomial $(T-1)^2$, which has linear factors in $\mathbf{R}[T]$ but the roots are not distinct, so Theorem 3.2 does not say the matrix is diagonalizable in $\mathrm{M}_2(\mathbf{R})$, and in fact it isn't.

Example 3.4. The matrix $\left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ has characteristic polynomial $T^2 - T - 1$, which has 2 different real roots, so the matrix is diagonalizable in $\mathrm{M}_2(\mathbf{R})$.

There are many diagonal matrices with repeated diagonal entries (take the simplest example, I_n!), and their characteristic polynomials have repeated roots. The criterion in Theorem 3.2 will never detect a diagonalizable operator with a repeated eigenvalue, so that criterion is a sufficient but not necessary condition for diagonalizability. In Section 4 we will see a way to detect diagonalizability using a different polynomial than the characteristic polynomial that is both necessary and sufficient.

Exactly how common is it for a characteristic polynomial to have distinct roots (whether or not they lie in F)? Consider 2×2 matrices: the characteristic polynomial $T^2 + bT + c$ has repeated roots if and only if $b^2 - 4c = 0$. A random quadratic will usually satisfy $b^2 - 4c \neq 0$

(you have to be *careful* to arrange things so that $b^2 - 4c$ is 0), so "most" 2×2 matrices have a characteristic polynomial with distinct roots. Similarly, a random $n \times n$ matrix usually has a characteristic polynomial with distinct roots. In particular, over the complex numbers this means a random $n \times n$ complex matrix almost certainly has distinct eigenvalues and therefore (since the eigenvalues lie in $\mathbf{C}$) Theorem 3.2 tells us that a random $n \times n$ complex matrix is diagonalizable. So diagonalizability is the rule rather than the exception over $\mathbf{C}$, or more generally over an algebraically closed field.

4. The Minimal Polynomial

By the Cayley-Hamilton theorem, there is a nonzero monic polynomial that kills a linear operator A: its characteristic polynomial.[2]

Definition 4.1. The nonzero monic polynomial in $F[T]$ that kills A and has least degree is called the *minimal* polynomial of A in $F[T]$.

What this means for a matrix $A \in \mathrm{M}_n(F)$, viewed as an operator on F^n, is that its minimal polynomial is the polynomial $f(T)$ of least degree such that $f(A)$ is the zero matrix.

Example 4.2. Both $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ have the same characteristic polynomial $(T-1)^2$, but $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ has minimal polynomial $T-1$ while $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ has minimal polynomial $(T-1)^2$. No linear polynomial can kill $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ since that would imply $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ is a scalar diagonal matrix.

Theorem 4.3. *The minimal polynomial of a linear operator $A\colon V \to V$ is equal to that of each matrix representation for it.*

Proof. Picking a basis of V lets us identify $\mathrm{End}_F(V)$ and $\mathrm{M}_n(F)$ as F-algebras. If M is the matrix in $\mathrm{M}_n(F)$ corresponding to A under this isomorphism then for each $f(T) \in F[T]$ the matrix representation of $f(A)$ is $f(M)$. Therefore $f(A) = O$ if and only if $f(M) = O$. Using f of least degree in either equation shows A and M have the same minimal polynomial in $F[T]$. □

We will usually denote the minimal polynomial of A as $m_A(T)$. Its characteristic polynomial $\chi_A(T)$ is $\det(TI_n - [A])$ for a matrix representation $[A]$ of A relative to a basis of V.[3]

Theorem 4.4. *Let $A\colon V \to V$ be linear. A polynomial $f(T) \in F[T]$ satisfies $f(A) = O$ if and only if $m_A(T) \mid f(T)$.*

Proof. Suppose $m_A(T) \mid f(T)$, so $f(T) = m_A(T)g(T)$. Since substitution of A for T gives a homomorphism $F[T] \to \mathrm{End}_F(V)$, we have $f(A) = m_A(A)g(A) = O \cdot g(A) = O$.

Conversely, suppose $f(A) = O$. Using polynomial division in $F[T]$, write $f(T) = m_A(T)q(T) + r(T)$ where $q(T), r(T) \in F[T]$ and $r(T) = 0$ or $\deg r < \deg m_A$. Substituting A for T in the polynomials, we have

$$O = m_A(A)q(A) + r(A) = r(A).$$

[2]That $f(A) = O$ for some nonconstant $f(T) \in F[T]$ can be shown without the Cayley-Hamilton theorem: the space $\mathrm{End}_F(V)$ has dimension n^2, so the $n^2 + 1$ operators $I, A, A^2, \ldots, A^{n^2}$ in $\mathrm{End}_F(V)$ must have a non-trivial linear dependence relation, and this gives a polynomial relation on A with coefficients in F of degree $\leq n^2$, which is weaker than what we know from the Cayley-Hamilton theorem.

[3]Often in linear algebra courses, the characteristic polynomial of A is $\det([A] - TI_n)$, not $\det(TI_n - [A])$. These differ at most a sign change: $\det(TI_n - [A]) = \det(-([A] - TI_n)) = (-1)^n \det([A] - TI_n)$, where $n = \dim_F(V)$. In particular, if n is even then both determinants are the same. An advantage of using $\det(TI_n - [A])$ is that it has leading coefficient 1 all the time.

Since $r(T)$ vanishes at A and either $r(T) = 0$ or $r(T)$ has degree less than the degree of the minimal polynomial of A, it must be the case that $r(T) = 0$. Therefore $f(T) = m_A(T)q(T)$, so $m_A(T) \mid f(T)$. □

Theorem 4.4 justifies speaking of *the* minimal polynomial. If two monic polynomials are both of least degree killing A, Theorem 4.4 shows they divide each other, and therefore they are equal (since they are both monic). Minimal polynomials of linear operators need not be irreducible (*e.g.*, $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ has minimal polynomial $(T-1)^2$).

Example 4.5. Write V as a direct sum of subspaces, say $V = U \oplus W$. Let $P_U \colon V \to V$ be projection onto the subspace U from this particular decomposition: $P(u+w) = u$. Since $P(u) = u$, we have $P^2(u+w) = P(u+w)$, so $P^2 = P$. Thus P is killed by the polynomial $T^2 - T = T(T-1)$. If $T^2 - T$ is not the minimal polynomial then by Theorem 4.4 either T or $T-1$ kills P; the first case means $P = O$ (so $U = \{0\}$) and the second case means $P = \mathrm{id}_V$ (so $U = V$). As long as U and W are both nonzero, P is neither O nor id_V and $T^2 - T$ is the minimal polynomial of the projection P.

While all operators on an n-dimensional space have characteristic polynomials of degree n, the degree of the minimal polynomial varies from one operator to the next. Its computation is not as mechanical as the characteristic polynomial, since there isn't a universal formula for the minimal polynomial. Indeed, consider the matrix

$$A_\varepsilon = \begin{pmatrix} 1+\varepsilon & 0 \\ 0 & 1 \end{pmatrix}.$$

For $\varepsilon \neq 0$, A_ε has two different eigenvalues, $1+\varepsilon$ and 1. Therefore the minimal polynomial of A_ε is not of degree 1, so its minimal polynomial must be its characteristic polynomial $T^2 - (2+\varepsilon)T + 1 + \varepsilon$. However, when $\varepsilon = 0$ the matrix A_ε is the identity I_2 with minimal polynomial $T - 1$. When eigenvalue multipicities change, the minimal polynomial changes in a drastic way. It is *not* a continuous function of the matrix.

To compute the minimal polynomial of a linear operator, Theorem 4.4 looks useful. For example, the minimal polynomial divides the characteristic polynomial, since the characteristic polynomial kills the operator by the Cayley–Hamilton theorem.[4]

Example 4.6. Let

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

The characteristic polynomial is $T^3 - 2T^2 + T = T(T-1)^2$. If the characteristic polynomial is not the minimal polynomial then the minimal polynomial divides one of the quadratic factors. There are two of these: $T(T-1)$ and $(T-1)^2$. A calculation shows $A(A - I_3) = O$, so the minimal polynomial divides $T(T-1)$. Since A and $A - I_3$ are not O, the minimal polynomial of A is $T(T-1) = T^2 - T$.

Since the minimal polynomial divides the characteristic polynomial, every root of the minimal polynomial (possibly in an extension of F) is an eigenvalue. The converse is also true:

Theorem 4.7. *Any eigenvalue of a linear operator is a root of its minimal polynomial in $F[T]$, so the minimal polynomial and characteristic polynomial have the same roots.*

[4]In particular, $\deg m_A(T) \leq \dim_F V$. This inequality can also be proved directly by an induction on the dimension, without using the Cayley-Hamilton theorem. See [1].

Proof. The minimal polynomials of a linear operator and its matrix representation in some basis are the same, so we pick bases to work with a matrix A acting on F^n ($n = \dim_F V$). Say λ is an eigenvalue of A, in some extension field E. We want to show $m_A(\lambda) = 0$. There is an eigenvector in E^n for this eigenvalue: $Av = \lambda v$ and $v \neq 0$. Then $A^k v = \lambda^k v$ for all $k \geq 1$, so $f(A)v = f(\lambda)v$ for all $f \in E[T]$. In particular, taking $f(T) = m_A(T)$, $m_A(A) = O$ so $0 = m_A(\lambda)v$. Thus $m_A(\lambda) = 0$. □

Remark 4.8. The proof may look a bit funny: nowhere in the argument did we use minimality of $m_A(T)$. Indeed, we showed every polynomial that kills A has an eigenvalue of A as a root. Since the minimal polynomial for A divides all other polynomials killing A, it is the minimal polynomial of A to which this result has its main use, and that's why we formulated Theorem 4.7 for the minimal polynomial.

Example 4.9. In Example 4.6, $\chi_A(T) = T(T-1)^2$, so the eigenvalues of A are 0 and 1. Theorem 4.7 says $m_A(T)$ has roots 0 and 1, so $m_A(T)$ is divisible by T and $T-1$. Therefore if $m_A \neq \chi_A$ then $m_A = T(T-1)$. The consideration of $(T-1)^2$ in Example 4.6 was unnecessary; it couldn't have worked because $m_A(T)$ must have both 0 and 1 as roots.

Corollary 4.10. *The minimal and characteristic polynomials of a linear operator have the same irreducible factors in $F[T]$.*

Proof. Any irreducible factor of the minimal polynomial is a factor of the characteristic polynomial since the minimal polynomial divides the characteristic polynomial. Conversely, if $\pi(T)$ is an irreducible factor of the characteristic polynomial, a root of it (possibly in some larger field than F) is an eigenvalue and therefore is also a root of the minimal polynomial by Theorem 4.7. Any polynomial in $F[T]$ sharing a root with $\pi(T)$ is divisible by $\pi(T)$, so $\pi(T)$ divides the minimal polynomial. □

If we compare the irreducible factorizations in $F[T]$

$$\chi_A(T) = \pi_1(T)^{e_1} \cdots \pi_k(T)^{e_k}, \quad m_A(T) = \pi_1(T)^{f_1} \cdots \pi_k(T)^{f_k},$$

we have $1 \leq f_k \leq e_k$ since $m_A(T) \mid \chi_A(T)$. Since $e_i \leq n \leq nf_i$, we also get a "reverse" divisibility: $\chi_A(T) \mid m_A(T)^n$ in $F[T]$.

We say a polynomial in $F[T]$ *splits* if it is a product of linear factors in $F[T]$. For instance, $T^2 - 5$ splits in $\mathbf{R}[T]$, but not in $\mathbf{Q}[T]$. The polynomial $(T-1)^2$ splits in every $F[T]$. Any factor of a polynomial in $F[T]$ that splits also splits.

Using the minimal polynomial in place of the characteristic polynomial provides a good criterion for diagonalizability over a field, which is our main result:

Theorem 4.11. *Let $A\colon V \to V$ be a linear operator. Then A is diagonalizable if and only if its minimal polynomial in $F[T]$ splits in $F[T]$ and has distinct roots.*

Theorem 4.11 gives necessary *and* sufficient conditions for diagonalizability, rather than just sufficient conditions as in Theorem 3.2. Because the minimal polynomial is a factor of the characteristic polynomial, Theorem 4.11 implies Theorem 3.2.

Proof. Suppose $m_A(T)$ splits in $F[T]$ with distinct roots. We will show V has a basis of eigenvectors for A, so A is diagonalizable. Let

$$m_A(T) = (T - \lambda_1) \cdots (T - \lambda_r),$$

so the λ_i's are the eigenvalues of A (Theorem 4.7) and by hypothesis they are distinct.

For an eigenvalue λ_i, let

$$E_{\lambda_i} = \{v \in V : Av = \lambda_i v\}$$

be the corresponding eigenspace. We will show

$$V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_r}, \tag{4.1}$$

so using bases from each E_{λ_i} provides an eigenbasis for A. Since eigenvectors with different eigenvalues are linearly independent, it suffices to show

$$V = E_{\lambda_1} + \cdots + E_{\lambda_r},$$

as the sum will then automatically be direct by linear independence.

The way to get the eigenspace components of a vector is to show that it is possible to "project" from V to each eigenspace E_{λ_i} using *polynomials* in the operator A. Specifically, we want to find polynomials $h_1(T), \ldots, h_r(T)$ in $F[T]$ such that

$$1 = h_1(T) + \cdots + h_r(T), \quad h_i(T) \equiv 0 \bmod m_A(T)/(T - \lambda_i). \tag{4.2}$$

The congruence condition implies the polynomial $(T - \lambda_i)h_i(T)$ is divisible by $m_A(T)$, so $(A - \lambda_i)h_i(A)$ acts on V as O. So if we substitute the operator A for T in (4.2) and apply both sides to a vector $v \in V$, we get

$$v = h_1(A)(v) + \cdots + h_r(A)(v), \quad (A - \lambda_i)h_i(A)(v) = 0.$$

The second equation says $h_i(A)(v)$ lies in E_{λ_i} and the first equation says v is a sum of such eigenvectors, hence $V = \sum_{i=1}^r E_{\lambda_i}$.

It remains to find $h_i(T)$'s fitting (4.2). For $1 \leq i \leq r$, let $f_i(T) = m_A(T)/(T - \lambda_i) = \prod_{j \neq i}(T - \lambda_j)$. Since the λ_i's are distinct, the polynomials $f_1(T), \ldots, f_r(T)$ are relatively prime as an r-tuple, so some $F[T]$-linear combination of them is equal to 1:

$$1 = \sum_{i=1}^r g_i(T) f_i(T), \tag{4.3}$$

where $g_i(T) \in F[T]$. Let $h_i(T) = g_i(T)f_i(T)$!

Now assume A is diagonalizable, so all eigenvalues of A are in F and V is the direct sum of the eigenspaces for A. We want to show the minimal polynomial of A in $F[T]$ splits and has distinct roots.

Let $\lambda_1, \ldots, \lambda_r$ be the different eigenvalues of A, so $V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_r}$. We will show $f(T) := (T - \lambda_1) \cdots (T - \lambda_r) \in F[T]$ is the minimal polynomial of A in $F[T]$.

By hypothesis, the eigenvectors of A span V. Let v be an eigenvector, say $Av = \lambda v$. Then $A - \lambda$ kills v. The operators $A - \lambda_i$ commute with each other and one of them kills v, so their product $f(A)$ kills v. Thus $f(A)$ kills the span of the eigenvectors, which is V, so $f(A) = O$. The minimal polynomial is therefore a factor of $f(T)$. At the same time, each root of $f(T)$ is an eigenvalue of A and therefore is a root of the minimal polynomial of A (Theorem 4.7). Since the roots of $f(T)$ each occur once, $f(T)$ must be the minimal polynomial of A. □

Remark 4.12. In (4.3), the polynomial $g_i(T)$ can in fact be taken to be the constant $1/f_i(\lambda_i)$. Indeed, the sum

$$\sum_{i=1}^r \frac{1}{f_i(\lambda_i)} f_i(T)$$

is a polynomial of degree at most $r - 1$ (since each $f_i(T)$ has degree $r - 1$) and at $\lambda_1, \ldots, \lambda_r$ it takes the value 1 (all but one term in the sum vanishes at each λ_i and the remaining term is 1). Taking a value r times when the degree is at most $r - 1$ forces the polynomial to be that constant value.

Corollary 4.13. *Let A be a linear transformation on a complex vector space V such that A^k is diagonalizable and invertible for some positive integer k. Then A is diagonalizable. In particular, if $A^k = \mathrm{id}_V$ for some positive integer k, then A is diagonalizable.*

Proof. Since $\mathbf{C}$ is algebraically closed, all nonconstant polynomials in $\mathbf{C}[T]$ split into linear factors, so Theorem 4.11 says a linear operator on a finite-dimensional complex vector space is diagonalizable if and only if its minimal polynomial in $\mathbf{C}[T]$ has distinct roots.

Since A^k is diagonalizable, the "only if" direction of Theorem 4.11 tells us the minimal polynomial of A^k has the form $(T-\lambda_1)\cdots(T-\lambda_r)$ for distinct λ_i in $\mathbf{C}$. Therefore $f(A) = O$ where

$$f(T) = (T^k - \lambda_1)\cdots(T^k - \lambda_r).$$

The different factors $T^k - \lambda_i$ share no common roots since the λ_i's are distinct, The numbers λ_i are all nonzero, since otherwise A^k would not be invertible, so each $T^k - \lambda_i$ has no repeated roots (a nonzero complex number has k different kth roots). Thus $f(T)$ has distinct roots in $\mathbf{C}$. The minimal polynomial of A is a factor of $f(T)$, so the minimal polynomial of A has distinct roots. Therefore by the "if" direction of Theorem 4.11, A is diagonalizable. □

Earlier (Example 4.5) we saw that the projection from V onto a subspace in a direct sum decomposition is a linear operator P where $P^2 = P$. Now we can prove the converse using Theorem 4.11.

Corollary 4.14. *Any linear operator $A\colon V \to V$ satisfying $A^2 = A$ is the projection of V onto some subspace: there is a decomposition $V = U \oplus W$ such that $A(u + w) = u$.*

Proof. Since A is killed by $T^2 - T$, its minimal polynomial in $F[T]$ is $T^2 - T$, T, or $T - 1$. These all split into linear factors with distinct roots 0 or 1. Thus A is diagonalizable by Theorem 4.11 with eigenvalues 0 or 1 (or both). Let $U = E_1$ be the 1-eigenspace of A and $W = E_0$ be the 0-eigenspace of A, so $V = U \oplus W$ (U or W might be $\{0\}$ if $A = O$ or $A = \mathrm{id}_V$). For $u \in U$ and $w \in W$, $A(u + w) = A(u) + A(w) = 1 \cdot u + 0 \cdot w = u$, so A is projection onto the subspace U. □

Remark 4.15. One does not need all this development about diagonalizability to show A is a projection when $A^2 = A$. For $v \in V$, $v = Av + (v - Av)$ and $A(Av) = Av$ while $A(v - Av) = Av - A^2v = Av - Av = 0$, so A is the projection of V onto the subspace $U := A(V)$ with complementary subspace $W := (\mathrm{id}_V - A)(V)$ on which A acts as O.

Example 4.16. The 3×3 matrix in Example 4.6 satisfies $A^2 = A$, so it must be a projection. Using row reduction, the image of $A\colon F^3 \to F^3$ is the plane spanned by $(1, 0, 1)$ and $(0, 1, 1)$. The kernel of the matrix is the line spanned by $(-1, 1, 1)$, so

$$A(a(1, 0, 1) + b(0, 1, 1) + c(-1, 1, 1)) = a(1, 0, 1) + b(0, 1, 1).$$

So we see that A is projection onto the plane $F(1, 0, 1) + F(0, 1, 1)$ in F^3.

While diagonalizability of A is equivalent to its minimal polynomial splitting in $F[T]$ with distinct roots, is just the splitting of the minimal polynomial in $F[T]$ (perhaps with repeated roots) equivalent to anything interesting?

Theorem 4.17. *A linear operator $A\colon V \to V$ has a basis in which its matrix representation is upper triangular if and only if $m_A(T)$ splits in $F[T]$.*

This is consistent with what we know about $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, which is already upper triangular and is not diagonalizable. Its minimal polynomial is $(T - 1)^2$.

As a special case of Theorem 4.17, every complex square matrix can be conjugated to an upper-triangular form since $\mathbf{C}$ is algebraically closed. We will discuss the proof of Theorem 4.17 after looking at nilpotent operators in Section 6.

Even if the minimal polynomial of an operator $A\colon V \to V$ doesn't split, this polynomial still gives us information about the operator. For example, the ring $F[A] \subset \mathrm{End}_F(V)$ generated over F by A is isomorphic to $F[T]/(m_A(T))$. Indeed, the substitution homomorphism $F[T] \to \mathrm{End}_F(V)$ sending $f(T)$ to $f(A)$ has image $F[A]$ and kernel $(m_A(T))$ by Theorem 4.4, so $F[T]/(m_A(T)) \cong F[A]$.

5. Simultaneous Diagonalizability

Now that we understand when a single linear operator is diagonalizable (if and only if the minimal polynomial splits with distinct roots), we consider *simultaneous* diagonalizability of several linear operators $A_j\colon V \to V$, $j = 1, 2, \ldots, r$. Assuming each A_j has a diagonal matrix representation in some basis, can we find a common basis in which the A_j's are all diagonal matrices? (This possibility is called simultaneous diagonalization.) A necessary constraint is commutativity: every set of diagonal matrices commutes, so *if* the A_j's can be simultaneously diagonalized, they must commute. Happily, this necessary condition is also sufficient, as we will soon see. What is special about commuting operators is they preserve each other's eigenspaces: if $AB = BA$ and $Av = \lambda v$ then $A(Bv) = B(Av) = B(\lambda v) = \lambda(Bv)$, so B sends each vector in the λ-eigenspace of A to another vector in the λ-eigenspace of A. Pay attention to how this is used in the next theorem.

Theorem 5.1. *If $A_1, \ldots, A_r$ are linear operators on V and each A_i is diagonalizable, they are simultaneously diagonalizable if and only if they commute.*

Proof. We already indicated why simultaneously diagonalizable operators have to commute.

To show the converse direction, assume $A_1, \ldots, A_r$ commute and are each diagonalizable. To show they are simultaneously diagonalizable, we induct on the number r of linear operators. The result is clear if $r = 1$, so assume $r \geq 2$. Let

$$E_\lambda = \{v : A_r v = \lambda v\}$$

be an eigenspace for A_r for some eigenvalue λ of A_r. Since A_r is diagonalizable on V, V is the direct sum of the eigenspaces for A_r.

For $v \in E_\lambda$, $A_r(A_i v) = A_i(A_r v) = A_i(\lambda v) = \lambda(A_i v)$, so $A_i v \in E_\lambda$. Thus each A_i *restricts* to a linear operator on the subspace E_λ. The linear operators $A_1|_{E_\lambda}, \ldots, A_{r-1}|_{E_\lambda}$ commute since the A_i's already commuted as operators on V, and these restrictions to E_λ are diagonalizable by Corollary 7.5. There are $r-1$ of them, so induction on r (while quantifying over *all* finite-dimensional vector spaces) implies there is a basis for E_λ consisting of common eigenvectors for $A_1|_{E_\lambda}, \ldots, A_{r-1}|_{E_\lambda}$.[5] The elements of this basis for E_λ are eigenvectors for $A_r|_{E_\lambda}$ as well, since *all* nonzero vectors in E_λ are eigenvectors for A_r. Thus $A_1|_{E_\lambda}, \ldots, A_{r-1}|_{E_\lambda}, A_r|_{E_\lambda}$ are all diagonalizable. The vector space V is the direct sum of the eigenspaces E_λ of A_r, so stringing together common eigenbases of all $A_i|_{E_\lambda}$ as λ runs over the eigenvalues of A_r gives a common eigenbasis of V for all the A_i's. □

Remark 5.2. Theorem 5.1 is *not* saying commuting operators diagonalize! It says commuting diagonalizable operators simultaneously diagonalize. For example, the matrices $\left(\begin{smallmatrix} 1 & a \\ 0 & 1 \end{smallmatrix}\right)$ for all a commute with each other, but none of them are diagonalizable when $a \neq 0$.

Because we are dealing with operators on finite-dimensional spaces, Theorem 5.1 extends to a possibly infinite number of commuting operators, as follows.

Corollary 5.3. *Let $\{A_i\}$ be commuting linear operators on a finite-dimensional vector space V. If each A_i is diagonalizable on V then they are simultaneously diagonalizable.*

[5] This choice of basis for E_λ is not made by A_r, but by the other operators together.

Proof. Let U be the subspace of $\mathrm{End}_F(V)$ spanned by the operators A_i's. Since $\mathrm{End}_F(V)$ is finite-dimensional, its subspace U is finite-dimensional, so U is spanned by a finite number of A_i's, say $A_{i_1}, \ldots, A_{i_r}$. By Theorem 5.1, there is a common eigenbasis of V for $A_{i_1}, \ldots, A_{i_r}$. A common eigenbasis for linear operators is also an eigenbasis for every linear combination of the operators, so this common eigenbasis of $A_{i_1}, \ldots, A_{i_r}$ diagonalizes every element of U, and in particular diagonalizes each A_i. □

Corollary 5.4. *Let A and B be linear operators $V \to V$ that are diagonalizable and commute. Then every operator in the ring $F[A, B]$ is diagonalizable. In particular, $A + B$ and AB are diagonalizable.*

Proof. Since A and B commute, there is a common eigenbasis of V for A and B. Members of this basis are also eigenvectors for each operator in $F[A, B]$, so all these operators are diagonalizable too. □

Without commutativity, diagonalizability need not be preserved by addition and multiplication. For example, the matrices $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 0 & 1 \\ 0 & -1 \end{smallmatrix}\right)$ are both diagonalizable but their sum $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ is not. The reader is invited to find diagonalizable 2×2 matrices with a non-diagonalizable product.

6. Nilpotent Operators

The minimal polynomial classifies not only diagonalizable operators, but also nilpotent operators (those having a power equal to O).

Theorem 6.1. *For a linear operator $N \colon V \to V$, the following are equivalent:*

(1) *N is nilpotent: $N^k = O$ for some $k \geq 1$,*
(2) *$N^n = O$, where $n = \dim V$,*
(3) *the minimal polynomial of N is T^k for some $k \leq n$.*

Proof. We will show (1) $\Rightarrow$ (2) $\Rightarrow$ (3). (That (3) implies (1) is obvious.) If $N^k = O$ for some $k \geq 1$ then the minimal polynomial of N is a factor of T^k, so the minimal polynomial of N is a power of T. The characteristic polynomial is monic of degree n with the same irreducible factors as the minimal polynomial (Corollary 4.10) so $\chi_N(T) = T^n$, which implies $N^n = O$ by Cayley-Hamilton. The minimal polynomial divides the characteristic polynomial, so if $\chi_N(T) = T^n$ then the minimal polynomial is T^k for some $k \leq n$. □

Corollary 6.2. *A linear operator $N \colon V \to V$ is nilpotent if and only if its only eigenvalue in extensions of F is 0.*

Proof. We know N is nilpotent if and only if its characteristic polynomial is T^n, which is equivalent to saying the only eigenvalue of N in extensions of F is 0. □

We can't diagonalize a nilpotent operator except if it is O: a minimal polynomial of the form T^k has distinct roots only when it is T, and the only operator with minimal polynomial T is O. But there is something we can say from Theorem 6.1 about possible matrix representations of a nilpotent operator.

Corollary 6.3. *A nilpotent linear operator $N \colon V \to V$ has a strictly upper triangular matrix representation.*

A strictly upper triangular matrix, like $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, is an upper triangular matrix with 0's along the main diagonal.

Proof. We argue by induction on $\dim V$. When $\dim V = 1$, $N = O$ and the result is easy. If $\dim V > 1$ and $N = O$, the result is still easy. If $\dim V > 1$ and $N \neq O$, then $W := \ker N$ is a proper subspace of V and $W \neq \{0\}$ since N is not injective (a nonzero nilpotent operator on V certainly isn't one-to-one, as $V \neq \{0\}$ from our conventions at the end of the introduction). Since $N(W) = \{0\} \subset W$, N induces linear operators on W and V/W. Since a power of N on V is O, that same power of N is O on W and V/W, so the operators that N induces on W and V/W are both nilpotent. Of course N really acts on W as O, but on V/W all we can say is that N is nilpotent. Since $0 < \dim V/W < \dim V$, by induction there is a basis of V/W with respect to which the operator induced by N on V/W has a strictly upper triangular matrix representation. Lift such a basis of V/W (arbitrarily) to vectors in V (the lifts are automatically linearly independent) and add to this set a basis of W to get a basis of V. Put the basis vectors from W *first* in the ordering of this basis for V. With respect to this choice of basis of V, the matrix representation of N on V has the form $\left(\begin{smallmatrix} O & * \\ O & U \end{smallmatrix}\right)$, where U is the strictly upper triangular (square) matrix for N on V/W. This matrix for N on V is strictly upper triangular, so we are done. □

Now we give the proof of Theorem 4.17: a linear operator has an upper triangular matrix representation if and only if its minimal polynomial splits in $F[T]$.

Proof. If a linear operator has an upper triangular matrix representation, then the characteristic polynomial of the upper triangular matrix splits in $F[T]$, so the minimal polynomial (a factor of that) also splits in $F[T]$.

Conversely, assume $m_A(T)$ splits in $F[T]$, say

$$m_A(T) = (T - \lambda_1)^{e_1} \cdots (T - \lambda_r)^{e_r},$$

where the λ_i's are the distinct roots. Then the polynomials $f_i(T) = m_A(T)/(T - \lambda_i)^{e_i}$ are relatively prime, so arguing as in the proof of Theorem 4.11 (where all the exponents e_i are 1) we get

$$V = \bigoplus \ker((A - \lambda_i)^{e_i}).$$

Let $W_i = \ker((A - \lambda_i)^{e_i})$. Since A commutes with $(A - \lambda_i)^{e_i}$, $A(W_i) \subset W_i$. We will show $A|_{W_i}$ has an upper-triangular matrix representation, and by stringing together bases of the W_i's to get a basis of V we will obtain an upper-triangular matrix representation of A on V.

On W_i, $(A - \lambda_i)^{e_i} = O$, so $A - \lambda_i$ is a nilpotent operator. Write $A = \lambda_i + (A - \lambda_i)$, which expresses A on W_i as the sum of a scaling operator and a nilpotent operator. By what we know about nilpotent operators, there is a basis of W_i with respect to which $A - \lambda_i$ is strictly upper triangular. Now with respect to *every* basis, the scaling operator λ_i is diagonal. So using the basis that makes $A - \lambda_i$ a strictly upper triangular matrix, the matrix for A is the sum of a diagonal and strictly upper triangular matrix, and that's an upper triangular matrix. □

Corollary 6.4. *If $A_1, \ldots, A_r$ are commuting linear operators on V and each A_i is upper triangularizable, they are simultaneously upper triangularizable.*

Unlike Theorem 5.1, the commuting hypothesis used here is far from being necessary: most upper triangular matrices (unlike all diagonal matrices) do not commute with each other! There is a theorem of A. Borel about linear algebraic groups that relaxes the commutativity assumption to a more reasonable hypothesis (solvability, together with some other technical conditions).

Proof. We argue by induction on the dimension of the vector space (*not* on the number of operators, as in the proof of Theorem 5.1). The one-dimensional case is clear. Assume now

$\dim V \geq 2$ and the corollary is known for lower-dimensional spaces. We may assume the A_i's are not all scalar operators on V (otherwise the result is obvious using an arbitrary basis of V). Without loss of generality, let A_r not be a scalar operator.

Since A_r is upper triangularizable on V, its eigenvalues are in F. Let $\lambda \in F$ be an eigenvalue of A_r and set E_λ to be the λ-eigenspace of A_r in V. Then $0 < \dim E_\lambda < \dim V$.

Since the A_i's commute, $A_i(E_\lambda) \subset E_\lambda$ for all i. Moreover, the minimal polynomial of $A_i|_{E_\lambda}$ is a factor of the minimal polynomial of A_i on V, so every $A_i|_{E_\lambda}$ is upper triangularizable by Theorem 4.17. Since the A_i's commute on V, they also commute as operators on E_λ, so by induction on dimension the A_i's are simultaneously upper triangularizable on E_λ. In particular, the first vector in a simultaneous "upper triangular basis" for the A_i's is a common eigenvector of all the A_i's. Call this vector e_1 and set $W = Fe_1$. Then $A_i(W) \subset W$ for all i. The A_i's are all operators on W and thus also on V/W. Let $\overline{A}_i \colon V/W \to V/W$ be the operator A_i induces on V/W. On V/W these operators commute and are upper triangularizable (since their minimal polynomials divide those of the A_i's on V, which split in $F[T]$), so again by induction on dimension the operators $\overline{A}_i$ on V/W are simultaneously upper triangularizable. If we lift a common upper triangular basis for the $\overline{A}_i$'s from V/W to V and tack on e_1 as the first member of a basis for V, we obtain a common upper triangular basis for the A_i's by an argument similar to that in the proof of Corollary 6.3. □

7. Computing the Minimal Polynomial

Finding the minimal polynomial by computing the characteristic polynomial and testing its factors is a "top down" method. There is a method of computing the minimal polynomial on its own terms, without finding it among the factors of the characteristic polynomial. That is, we can perform a calculation from the "bottom up" instead of from the "top down."

Theorem 7.1. *Let $A \colon V \to V$ be linear. Suppose $W_1, \ldots, W_k$ are subspaces of V such that $V = W_1 + \cdots + W_k$, $A(W_i) \subset W_i$ for all i, and the restriction of A to W_i has minimal polynomial $m_i(T)$. Then the minimal polynomial of A on V is $\mathrm{lcm}(m_1, \ldots, m_k)$.*

We are not assuming that the W_i's are linearly independent subspaces (that is, we don't assume $V = \bigoplus_{i=1}^k W_i$), but only that the W_i's add up to V. That $A(W_i) \subset W_i$ is needed to make sense of the restriction of A to W_i as an operator, with its own minimal polynomial.

Proof. Set $f(T) = \mathrm{lcm}(m_1(T), \ldots, m_k(T))$, so $f(T)$ is a multiple of each $m_i(T)$. Write $f(T) = g_i(T)m_i(T)$ for each i. For $v \in W_i$, $f(A)v = g_i(A)(m_i(A)v) = g_i(A)(0) = 0$, so $f(A)$ is zero on each W_i. Since the W_i's add up to V, $f(A) = O$ on V. Therefore $m_A(T) \mid f(T)$.

Next we show $f(T) \mid m_A(T)$. Since $m_A(A) = O$ on V, $m_A(A)$ kills each W_i. Restricting to W_i, $m_A(A)|_{W_i} = m_A(A|_{W_i})$, so $A|_{W_i}$ is annihilated by $m_A(T)$. Therefore $m_A(T)$ is divisible by $m_i(T)$. Since $m_i(T) \mid m_A(T)$ for all i, their least common multiple $f(T)$ divides $m_A(T)$. The polynomials $f(T)$ and $m_A(T)$ are monic and divide each other, so they are equal. □

Example 7.2. On F^3 let

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

Let $W_1 = \{(x, y, 0)\}$ and $W_2 = \{(0, y, z)\}$. These are planes in F^3, and $W_1 + W_2 = F^3$, although $W_1 \cap W_2 \neq \{0\}$ so $W_1 + W_2$ is not a direct sum decomposition of F^3. A computation shows $A(W_1) \subset W_1$ and $A(W_2) \subset W_2$. Using basis $\{(1, 0, 0), (0, 1, 0)\}$ for W_1 and $\{(0, 1, 0), (0, 0, 1)\}$ for W_2, we get the matrix representations $[A|_{W_1}] = \left(\begin{smallmatrix} 2 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ and

$[A|_{W_2}] = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, so $m_1(T) = (T-2)(T-1)$ and $m_2(T) = (T-1)^2$. Therefore $m_A(T) = \mathrm{lcm}(m_1, m_2) = (T-1)^2(T-2)$.

Theorem 7.1 leads to an algorithm for computing the minimal polynomial of a linear operator $A\colon V \to V$. Pick *arbitrarily* $v \neq 0$ in V and consider the sequence of vectors $\{v, A(v), A^2(v), \dots\}$. They span a subspace of V. Call it W, so $W = \{f(A)v : f(T) \in F[T]\}$. The elements of W have the form $f(A)v$ as $f(T)$ runs through $F[T]$. The nice feature of W is that by its very definition $A(W) \subset W$, so A makes sense as a linear operator on W. To determine the minimal polynomial of A on W, find the smallest d such that the vectors $v, A(v), \dots, A^d(v)$ are *linearly dependent*. (Using a basis to turn V into F^n, we could use row reduction to check when linear dependence occurs for the first time. That tells us d.) Note $d \geq 1$ since $v \neq 0$. Since $v, A(v), \dots, A^{d-1}(v)$ are linearly independent, the only linear relation of the form

$$b_{d-1}A^{d-1}(v) + \cdots + b_1A(v) + b_0v = 0$$

with $b_i \in F$ is the one where every $b_i = 0$. Hence for every nonzero polynomial $f(T) \in F[T]$ with degree less than d, $f(A)(v) \neq 0$. Therefore $f(A) \neq O$ as an operator on W, which means the minimal polynomial of A acting on W has degree at least d.

There is a linear dependence relation on the set $\{v, A(v), \dots, A^d(v)\}$, and the coefficient of $A^d(v)$ in the relation must be nonzero since the other vectors are linearly independent. We can scale the coefficient of $A^d(v)$ in such a relation to be 1, say

$$A^d(v) + c_{d-1}A^{d-1}(v) + \cdots + c_1A(v) + c_0 = 0, \tag{7.1}$$

where $c_i \in F$. This tells us the polynomial

$$m(T) := T^d + c_{d-1}T^{d-1} + \cdots + c_1T + c_0$$

satisfies $m(A)(v) = 0$, so for every $f(T) \in F[T]$ we have $m(A)(f(A)v) = f(A)(m(A)v) = f(A)(0) = 0$. Every element of W is $f(A)(v)$ for some $f(T)$, so $m(A)$ kills all of W: $m(T)$ is the minimal polynomial of A acting on W. (Incidentally, this also shows $\dim W = d$ and W has basis $\{v, A(v), \dots, A^{d-1}(v)\}$.)

Set $W_1 = W$ and $m_1(T) = m(T)$. If $W_1 \neq V$, pick a vector $v' \notin W_1$ and run through the same argument for the subspace W_2 of V spanned by the vectors $v', A(v'), A^2(v'), \dots$ to get a minimal polynomial $m_2(T)$ for A on W_2. And so on. Since V is finite-dimensional, eventually we will get a sequence of subspaces $W_1, W_2, \dots, W_k$ where $A(W_i) \subset W_i$ for all i and the W_i's add up to V. The minimal polynomial of A on V is the least common multiple of the $m_i(T)$'s by Theorem 7.1.

Example 7.3. Let

$$A = \begin{pmatrix} 0 & 4 & 1 & -2 \\ -1 & 4 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ -1 & 3 & 0 & 0 \end{pmatrix}.$$

Set $v_1 = (1,0,0,0)$, so $A(v_1) = (0,-1,0,-1)$ and $A^2(v_1) = (-2,-3,0,-3) = 3A(v_1) - 2v_1$. Thus $A^2(v_1) - 3A(v_1) + 2v_1 = 0$, so $m_1(T) = T^2 - 3T + 2$. Let W_1 be the span of v_1 and $A(v_1)$. The vector $v_2 = (0,1,0,0)$ is not in W_1. It turns out v_2, $A(v_2)$, and $A^2(v_2)$ are linearly independent and $A^3(v_2) = 4A^2(v_2) - 5A(v_2) + 2v_2$, so $m_2(T) = T^3 - 4T^2 + 5T - 2$. Let W_2 be the span of $v_2, A(v_2)$, and $A^2(v_2)$. Both W_1 and W_2 are in the subspace of vectors with third component 0, so $W_1 + W_2 \neq F^4$. Take $v_3 = (0,0,1,0)$. A calculation shows the same linear relations hold for A-iterates of v_3 as for v_2 , so $m_3(T) = m_2(T)$. Since $\{v_1, A(v_1), v_2, v_3\}$ is a basis for F^4, the minimal polynomial of A is $\mathrm{lcm}(m_1, m_2, m_3) = T^3 - 4T^2 + 5T - 2$.

The algorithm we described for computing the minimal polynomial of a linear operator used subspaces preserved by the operator. Let's look at such subspaces more closely. If $A\colon V \to V$ is linear, a subspace $W \subset V$ is called *A-stable* when $A(W) \subset W$. For example, an eigenspace of A is A-stable. If $A = \mathrm{id}_V$ then every subspace of V is A-stable. (We don't require $A(W) = W$, only $A(W) \subset W$.) When $A(W) \subset W$, A induces a linear operator on W and on the quotient space V/W. Denote these induced linear maps as A_W and $A_{V/W}$. We look at how each of diagonalizability and nilpotency are related for A, A_W, and $A_{V/W}$ when W is an A-stable subspace. First we see how the minimal polynomials of these three linear maps are related.

Theorem 7.4. *Suppose $A\colon V \to V$ is linear and W is an A-stable subspace of V. The induced linear maps $A_W\colon W \to W$ and $A_{V/W}\colon V/W \to V/W$ have minimal polynomials that are factors of $m_A(T)$. More precisely, their least common multiple is a factor of $m_A(T)$ and their product is divisible by $m_A(T)$.*

Proof. The polynomial $m_{A_W}(T)$ kills W when T is replaced by $A|_W$, and the polynomial $m_{A_{V/W}}(T)$ kills V/W when T is replaced by A acting on V/W (that is, by $A_{V/W}$). Since $m_A(T)$ kills W when T is replaced by A and kills V/W when T is replaced by A, $m_A(T)$ is divisible by both $m_{A_W}(T)$ and by $m_{A_{V/W}}(T)$ and hence by their least common multiple.

For $v \in V$, with its coset in V/W denoted $\overline{v}$, $m_{A_{V/W}}(A)(\overline{v}) = \overline{0}$, so $m_{A_{V/W}}(A)(v) \in W$. Then $m_{A_W}(A)(m_{A_{V/W}}(A)(v)) = 0$. Since this holds for all $v \in V$, $m_{A_W}(A)m_{A_{V/W}}(A) = O$, so $m_{A_W}(T)m_{A_{V/W}}(T)$ kills A as an operator on V. Therefore $m_A(T) \mid m_{A_W}(T)m_{A_{V/W}}(T)$. □

Theorem 7.4 says $m_A(T)$ lies between $\mathrm{lcm}(m_{A_W}(T), m_{A_{V/W}}(T))$ and $m_{A_W}(T)m_{A_{V/W}}(T)$ in the sense of divisibility. It is not always the least common multiple or always the product. When $V = F^2$ and $A = I_2$, $m_A(T) = T - 1 \neq m_{A_W}(T)m_{A_{V/W}}(T)$ for each nonzero proper subspace $W \subset F^2$. When $V = F^2$ and $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, $m_A(T) = (T-1)^2 \neq \mathrm{lcm}(m_{A_W}(T), m_{A_{V/W}}(T))$ using $W = F\binom{1}{0}$.

Corollary 7.5. *Using the notation of Theorem 7.4,*

(1) *A is upper triangularizable if and only if A_W and $A_{V/W}$ are upper triangularizable,*
(2) *if A is diagonalizable, A_W and $A_{V/W}$ are diagonalizable,*
(3) *A is nilpotent if and only if A_W and $A_{V/W}$ are nilpotent.*

Proof. If a polynomial either splits or splits with distinct roots, every factor of it has the same property, so (1) and (2) are immediate from the relations in Theorem 7.4. To prove (3), $m_A(T)$ is a power of T if and only if $m_{A_W}(T)$ and $m_{A_{V/W}}(T)$ are powers of T. □

The converse to part 2 of Corollary 7.5 is false. Consider $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ acting on $V = F^2$. Let $W = F\binom{1}{0}$, so $A(W) = W$. Both W and V/W are 1-dimensional, so all linear operators on them are diagonalizable. Thus A_W and $A_{V/W}$ are diagonalizable but A is not diagonalizable. This is analogous to a product of squarefree integers not being squarefree (such as 6 and 15, which are squarefree while their product 90 is not squarefree).

References

[1] M. D. Burrow, "The Minimal Polynomial of a Linear Transformation," Amer. Math. Monthly **80** (1973), 1129–1131.