# V. Theory of a Single Linear Transformation, 211-247

DOI: 10.3792/euclid/9781429799980-5

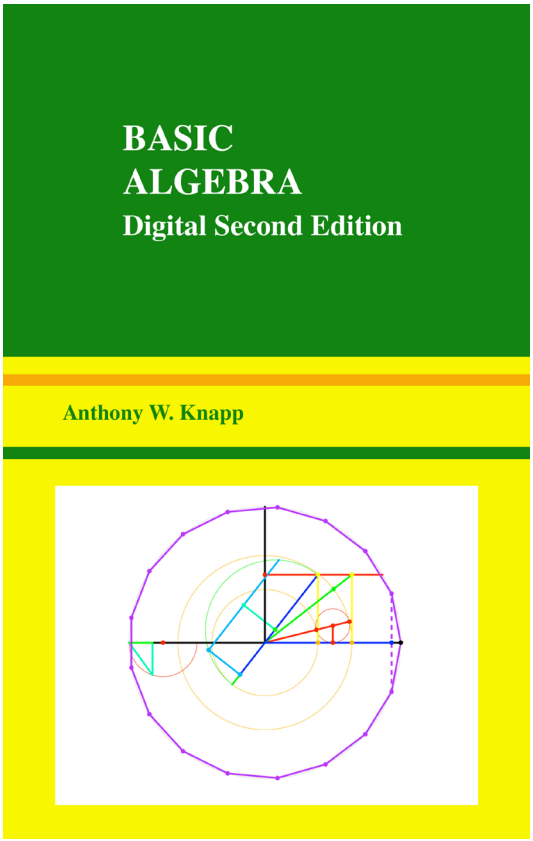
from

***Basic Algebra***

***Digital Second Edition***

Anthony W. Knapp

Full Book DOI: 10.3792/euclid/9781429799980
ISBN: 978-1-4297-9998-0



Anthony W. Knapp
81 Upper Sheep Pasture Road
East Setauket, N.Y. 11733–1729, U.S.A.
Email to: aknapp@math.stonybrook.edu
Homepage: www.math.stonybrook.edu/~aknapp

Title: Basic Algebra
Cover: Construction of a regular heptadecagon, the steps shown in color sequence; see page 505.

Mathematics Subject Classification (2010): 15–01, 20–01, 13–01, 12–01, 16–01, 08–01, 18A05, 68P30.

First Edition, ISBN-13 978-0-8176-3248-9

Published by Birkhäuser Boston

Digital Second Edition, not to be sold, no ISBN

Published by the Author

# CHAPTER V

# Theory of a Single Linear Transformation

**Abstract.** This goal of this chapter is to find finitely many canonical representatives of each similarity class of square matrices with entries in a field and correspondingly of each isomorphism class of linear maps from a finite-dimensional vector space to itself.

Section 1 frames the problem in more detail. Section 2 develops the theory of determinants over a commutative ring with identity in order to be able to work easily with characteristic polynomials $\det(XI - A)$. The discussion is built around the principle of "permanence of identities," which allows for passage from certain identities with integer coefficients to identities with coefficients in the ring in question.

Section 3 introduces the minimal polynomial of a square matrix or linear map. The Cayley–Hamilton Theorem establishes that such a matrix satisfies its characteristic equation, and it follows that the minimal polynomial divides the characteristic polynomial. It is proved that a matrix is similar to a diagonal matrix if and only if its minimal polynomial is the product of distinct factors of degree 1. In combination with the fact that two diagonal matrices are similar if and only if their diagonal entries are permutations of one another, this result solves the canonical-form problem for matrices whose minimal polynomial is the product of distinct factors of degree 1.

Section 4 introduces general projection operators from a vector space to itself and relates them to vector-space direct-sum decompositions with finitely many summands. The summands of a direct-sum decomposition are invariant under a linear map if and only if the linear map commutes with each of the projections associated to the direct-sum decomposition.

Section 5 concerns the Primary Decomposition Theorem, whose subject is the operation of a linear map $L : V \to V$ with $V$ finite-dimensional. The statement is that if $L$ has minimal polynomial $P_1(X)^{l_1} \cdots P_k(X)^{l_k}$ with the $P_j(X)$ distinct monic prime, then $V$ has a unique direct-sum decomposition in which the respective summands are the kernels of the linear maps $P_j(L)^{l_j}$, and moreover the minimal polynomial of the restriction of $L$ to the $j^{\text{th}}$ summand is $P_j(X)^{l_j}$.

Sections 6–7 concern Jordan canonical form. For the case that the prime factors of the minimal polynomial of a square matrix all have degree 1, the main theorem gives a canonical form under similarity, saying that a given matrix is similar to one in "Jordan form" and that the Jordan form is completely determined up to permutation of the constituent blocks. The theorem applies to all square matrices if the field is algebraically closed, as is the case for $\mathbb{C}$. The theorem is stated and proved in Section 6, and Section 7 shows how to make computations in two different ways.

## 1. Introduction

This chapter will work with vector spaces over a common field of "scalars," which will be called $\mathbb{K}$. As was observed near the end of Section IV.5, all the results

concerning vector spaces in Chapter II remain valid when the scalars are taken from $\mathbb{K}$ rather than just $\mathbb{Q}$ or $\mathbb{R}$ or $\mathbb{C}$. The ring of polynomials in one indeterminate $X$ over $\mathbb{K}$ will be denoted by $\mathbb{K}[X]$.

For the field $\mathbb{C}$ of complex numbers, every nonconstant polynomial in $\mathbb{C}[X]$ has a root, according to the Fundamental Theorem of Algebra (Theorem 1.18). Because of this fact some results in this chapter will take an especially simple form when $\mathbb{K} = \mathbb{C}$, and this simple form will persist for any field with this same property. Accordingly, we make a definition. Let us say that a field $\mathbb{K}$ is **algebraically closed** if every nonconstant polynomial in $\mathbb{K}[X]$ has a root. We shall work hard in Chapter IX to obtain examples of algebraically closed fields beyond $\mathbb{K} = \mathbb{C}$, but let us mention now what a few of them are.

EXAMPLES.

(1) The subset of $\mathbb{C}$ of all roots of polynomials with rational coefficients is an algebraically closed field.

(2) For each prime $p$, we have seen that any finite field of characteristic $p$ has $p^n$ elements for some $n$. It turns out that there is one and only one field of $p^n$ elements, up to isomorphism, for each $n$. If we align them suitably for fixed $p$ and take their union on $n$, then the result is an algebraically closed field.

(3) If $\mathbb{K}$ is any field, then there exists an algebraically closed field having $\mathbb{K}$ as a subfield. We shall prove this existence in Chapter IX by means of Zermelo's Well-Ordering Theorem (which appears in Section A5 of the appendix).

The general problem to be addressed in this chapter is to find "canonical forms" for linear maps from finite-dimensional vector spaces to themselves, special ways of realizing the linear maps that bring out some of their properties. Let us phrase a specific problem of this kind completely in terms of linear algebra at first. Then we can rephrase it in terms of a combination of linear algebra and group theory, and we shall see how it fits into a more general context.

In terms of matrices, the specific problem is to find a way of deciding whether two square matrices represent the same linear map in different bases. We know from Proposition 2.17 that if $L : V \to V$ is linear on the finite-dimensional vector space $V$ and if $A$ is the matrix of $L$ relative to a particular ordered basis in domain and range, then the matrix $B$ of $L$ in another ordered basis is of the form $B = C^{-1}AC$ for some invertible matrix $C$, i.e., $A$ and $B$ are similar.[1] Thus one kind of solution to the problem would be to specify one representative of each similarity class of square matrices. But this is not a convenient kind of answer to look for; in fact, the matrices $A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ are similar via

[1] A square matrix $A$ with a two-sided inverse is sometimes said to be **nonsingular**. A square matrix with no inverse is then said to be **singular**.

$C = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, but there is no particular reason to prefer one of $A$ or $B$ to the other. Thus a "canonical form" for detecting similarity will allow more than one representative of each similarity class (but typically only finitely many such representatives), and a supplementary statement will tell us when two such are similar.

So far, the best information that we have about solving this problem concerning square matrices comes from Section II.8. In that section the discussion of eigenvalues gave us some necessary conditions for similarity, but we did not obtain a useful necessary and sufficient condition.

In terms of linear maps, what we seek for a linear $L : V \to V$ is to use the geometry of $L$ to construct an ordered basis of $V$ such that $L$ acts in a particularly simple way on that ordered basis. Ideally the description of how $L$ acts on the ordered basis is to be detailed enough so that the matrix of $L$ in that ordered basis is completely determined by the description, even though the ordered basis may not be determined by it. For example, if $L$ were to have a basis of eigenvectors, then the description could be that "$L$ has an ordered basis of eigenvectors with eigenvalues $x_1, \ldots, x_n$." In any ordered basis with this property, the matrix of $L$ would then be diagonal with diagonal entries $x_1, \ldots, x_n$.

Suppose then that we have this kind of detailed description of how a linear map $L$ acts on some ordered basis. To what extent is $L$ completely determined? The answer is that $L$ is determined up to an isomorphism of the underlying vector space. In fact, suppose that $L$ and $M$ are linear maps from $V$ to itself such that $\begin{pmatrix} L \\ \Gamma\Gamma \end{pmatrix} = A = \begin{pmatrix} M \\ \Delta\Delta \end{pmatrix}$ for some ordered bases $\Gamma$ and $\Delta$. Then

$$\begin{aligned} \begin{pmatrix} L \\ \Gamma\Gamma \end{pmatrix} = A = \begin{pmatrix} M \\ \Delta\Delta \end{pmatrix} &= \begin{pmatrix} I \\ \Delta\Gamma \end{pmatrix} \begin{pmatrix} M \\ \Gamma\Gamma \end{pmatrix} \begin{pmatrix} I \\ \Gamma\Delta \end{pmatrix} \\ &= \begin{pmatrix} S \\ \Gamma\Gamma \end{pmatrix}^{-1} \begin{pmatrix} M \\ \Gamma\Gamma \end{pmatrix} \begin{pmatrix} S \\ \Gamma\Gamma \end{pmatrix} = \begin{pmatrix} S^{-1}MS \\ \Gamma\Gamma \end{pmatrix}, \end{aligned}$$

where $S : V \to V$ is the invertible linear map defined by $\begin{pmatrix} S \\ \Gamma\Gamma \end{pmatrix} = \begin{pmatrix} I \\ \Gamma\Delta \end{pmatrix}$. Hence $L = S^{-1}MS$ and $SL = MS$. In other words, if we think of having two copies of $V$, one called $V_1$ and the other called $V_2$, that are isomorphic via $S : V_1 \to V_2$, then the effect of $M$ in $V_2$ corresponds under $S$ to the effect of $L$ in $V_1$. In this sense, $L$ is determined up to an isomorphism of $V$.

Thus we are looking for a geometric description that determines linear maps up to isomorphism. Two linear maps $L$ and $M$ that are related in this way have $L = S^{-1}MS$ for some invertible linear map $S$. Passing to matrices with respect to some basis, we see that the matrices of $L$ and $M$ are to be similar. Consequently our two problems, one to characterize similarity for matrices and the other to characterize isomorphism for linear maps, come to the same thing.

These two problems have an interpretation in terms of group theory. In the case of $n$-by-$n$ matrices, the group $\mathrm{GL}(n, \mathbb{K})$ of invertible matrices acts on the set of all square matrices of size $n$ by conjugation via $(g, x) \mapsto gxg^{-1}$; the similarity classes are exactly the orbits of this group action, and the canonical form is to single out finitely many representatives from each orbit. In the case of linear maps, the group $\mathrm{GL}(V)$ of invertible linear maps on the finite-dimensional vector space $V$ acts by conjugation on the set of all linear maps from $V$ into itself; the isomorphism classes of linear maps on $V$ are the orbits, and the canonical form is to single out finitely many representatives from each orbit.

The above problem, whether for matrices or for linear maps, does not have a unique acceptable solution. Nevertheless, the text of this chapter will ultimately concentrate on one such solution, known as the "Jordan canonical form."

Now that we have brought group theory into the statement of the problem, we can put matters in a more general context: The situation is that some "important" group $G$ acts in an important way on an "interesting" vector space of matrices. The **canonical-form problem** for this situation is to single out finitely many representatives of each orbit and give a way of deciding, in terms of these representatives, whether two of the given matrices lie in the same orbit. We shall not pursue the more general problem in the text at this time. However, Problem 1 at the end of the chapter addresses one version beyond the one concerning similarity: to find a canonical form for the action of $\mathrm{GL}(m, \mathbb{K}) \times \mathrm{GL}(n, \mathbb{K})$ on $m$-by-$n$ matrices by $((g, h), x) = gxh^{-1}$. Some other groups that are important in this sense, besides products of general linear groups, are introduced in Chapter VI, and a problem at the end of Chapter VI reinterprets two theorems of that chapter as further canonical-form theorems under the action of a general linear group.

Let us return to the canonical-form problems for similarity of matrices and isomorphism of linear maps. The basic tool in studying these problems is the characteristic polynomial of a matrix or a linear map, as in Chapter II. However, we subtly used a special feature of $\mathbb{Q}$ and $\mathbb{R}$ and $\mathbb{C}$ in working with characteristic polynomials in Chapter II: we passed back and forth between the characteristic polynomial $\det(\lambda I - A)$ as a polynomial in one indeterminate (defined by its expression after expanding it out) and as a polynomial function of $\lambda$, defined for each value of $\lambda$ in $\mathbb{Q}$ or $\mathbb{R}$ or $\mathbb{C}$, one value at a time. This passage was legitimate because the homomorphism of the ring of polynomials in one indeterminate over a field to the ring of polynomial functions is one-one when the field is infinite, by Proposition 4.28c or Corollary 1.14. Some care is required, however, in working with general fields, and we begin by supplying the necessary details for justifying manipulations with determinants in a more general setting than earlier. The end result will be that the characteristic polynomial is a polynomial in one indeterminate, and we shall henceforth call that indeterminate $X$, rather than $\lambda$, so as to emphasize this point of view.

## 2. Determinants over Commutative Rings with Identity

Throughout this section let $R$ be a commutative ring with identity. The main case of interest for us at this time will be that $R = \mathbb{K}[X]$ is the polynomial ring in one indeterminate $X$ over a field $\mathbb{K}$.

The set of $n$-by-$n$ matrices with entries in $R$ is an abelian group under entry-by-entry addition, and matrix multiplication makes it into a ring with identity. Following tradition, we shall usually write $M_n(R)$ rather than $M_{nn}(R)$ for this ring. In this section we shall define a determinant function $\det : M_n(R) \to R$ and establish some of its properties. For the case that $R$ is a field, some of our earlier proofs concerning determinants used vector-space concepts—bases, dimensions, and so forth—and these are not available for general $R$. Yet most of the properties of determinants remain valid for general $R$ because of a phenomenon known as **permanence of identities**. We shall not try to state a general theorem about this principle but instead will be content to observe a pattern in how the relevant identities are proved.

If $A$ is in $M_n(R)$, we define its **determinant** to be

$$\det A = \sum_{\sigma \in \mathfrak{S}_n} (\operatorname{sgn} \sigma) A_{1\sigma(1)} A_{2\sigma(2)} \cdots A_{n\sigma(n)},$$

in effect converting into a definition the formula obtained in Theorem 2.34d when $R$ is a field.

A sample of the kind of identity we have in mind is the formula

$$\det(AB) = \det A \det B \qquad \text{for } A \text{ and } B \text{ in } M_n(R).$$

The key is that this formula says that two polynomials in $2n^2$ variables, with integer coefficients, are equal whenever arbitrary members of $R$ are substituted for the variables. Thus let us introduce $2n^2$ indeterminates $X_{11}, X_{12}, \ldots, X_{nn}$ and $Y_{11}, Y_{12}, \ldots, Y_{nn}$ to correspond to these variables. Forming the commutative ring $S = \mathbb{Z}[X_{11}, X_{12}, \ldots, X_{nn}, Y_{11}, Y_{12}, \ldots, Y_{nn}]$, we assemble the matrices $X = [X_{ij}]$, $Y = [Y_{ij}]$, and $XY = \left[\sum_k X_{ik} Y_{kj}\right]$ in $M_n(S)$. Consider the two members of $S$ given by

$$\begin{aligned} &\det X \det Y \\ &= \Big( \sum_{\sigma \in \mathfrak{S}_n} (\operatorname{sgn} \sigma) X_{1\sigma(1)} X_{2\sigma(2)} \cdots X_{n\sigma(n)} \Big) \Big( \sum_{\sigma \in \mathfrak{S}_n} (\operatorname{sgn} \sigma) Y_{1\sigma(1)} Y_{2\sigma(2)} \cdots Y_{n\sigma(n)} \Big) \end{aligned}$$

and

$$\det(XY) = \sum_{\sigma \in \mathfrak{S}_n} (\operatorname{sgn} \sigma) (XY)_{1\sigma(1)} (XY)_{2\sigma(2)} \cdots (XY)_{n\sigma(n)},$$

where $(XY)_{ij} = \sum_k X_{ik}Y_{kj}$. If we fix arbitrary elements $x_{11}, x_{12}, \ldots, x_{nn}$ and $y_{11}, y_{12}, \ldots, y_{nn}$ of $\mathbb{Z}$, then Proposition 4.30 gives us a unique substitution homomorphism $\Psi : S \to \mathbb{Z}$ such that $\Psi(1) = 1$, $\Psi(X_{ij}) = x_{ij}$, and $\Psi(Y_{ij}) = y_{ij}$ for all $i$ and $j$. Writing $x = [x_{ij}]$ and $y = [y_{ij}]$ and using that matrices with integer entries have $\det(xy) = \det x \det y$ because $\mathbb{Z}$ is a subset of the field $\mathbb{Q}$, we see that $\Psi(\det(XY)) = \Psi(\det X \det Y)$ for each choice of $x$ and $y$. Since $\mathbb{Z}$ is an infinite integral domain and since $x$ and $y$ are arbitrary, Corollary 4.32 allows us to deduce that

$$\det(XY) = \det X \det Y$$

as an equality in $S$.

Now we pass from an identity in $S$ to an identity in $R$. Let $1_R$ be the identity in $R$. Proposition 4.19 gives us a unique homomorphism of rings $\varphi_1 : \mathbb{Z} \to R$ such that $\varphi_1(1) = 1_R$. If we fix arbitrary elements $A_{11}, A_{12}, \ldots, A_{nn}$ and $B_{11}, B_{12}, \ldots, B_{nn}$ of $R$, then Proposition 4.30 gives us a unique substitution homomorphism $\Phi : S \to R$ such that $\Phi(1) = \varphi_1(1) = 1_R$, $\Phi(X_{ij}) = A_{ij}$ for all $i$ and $j$, and $\Phi(Y_{ij}) = B_{ij}$ for all $i$ and $j$. Applying $\Phi$ to the equality $\det(XY) = \det X \det Y$, we obtain the identity we sought, namely

$$\det(AB) = \det A \det B \qquad \text{for } A \text{ and } B \text{ in } M_n(R).$$

**Proposition 5.1.** If $R$ is a commutative ring with identity, then the determinant function $\det : M_n(R) \to R$ has the following properties:

(a) $\det(AB) = \det A \det B$,
(b) $\det I = 1$,
(c) $\det A^t = \det A$,
(d) $\det C = \det A + \det B$ if $A$, $B$, and $C$ match in all rows but the $j^{\text{th}}$ and if the $j^{\text{th}}$ row of $C$ is the sum of the $j^{\text{th}}$ rows of $A$ and $B$,
(e) $\det B = r \det A$ if $A$ and $B$ match in all rows but the $j^{\text{th}}$ and if the $j^{\text{th}}$ row of $B$ is equal entry by entry to $r$ times the $j^{\text{th}}$ row of $A$ for some $r$ in $R$,
(f) $\det A = 0$ if $A$ has two equal rows,
(g) $\det \begin{pmatrix} A & B \\ 0 & D \end{pmatrix} = \det A \det D$ if $A$ is in $M_k(R)$, $D$ is in $M_l(R)$, and $k + l = n$.

REMARKS. Properties (d), (e), and (f) imply that usual steps in manipulating determinants by row reduction continue to be valid.

PROOF. Part (a) was proved above, and parts (c) through (f) may be proved in the same way from the corresponding facts about integer matrices in Section II.7. Part (b) is immediate from the definition.

For (g), we first prove the result when the entries are in $\mathbb{Q}$, and then we argue in the same way as with (a) above. When the entries are in $\mathbb{Q}$, row reduction of $D$ allows us to reduce to the case either that $D$ has a row of 0's or that $D$

is the identity. If $D$ has a row of 0's, then $\det \begin{pmatrix} A & B \\ 0 & D \end{pmatrix}$ and $\det A \det D$ are both 0 and hence are equal. If $D$ is the identity, then further row reduction shows that $\det \begin{pmatrix} A & B \\ 0 & I \end{pmatrix} = \det \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix}$, and the right side equals $\det A = \det A \det I$, as required. □

**Proposition 5.2** (expansion in cofactors). Let $R$ be a commutative ring with identity, let $A$ be in $M_n(R)$, and let $\widehat{A_{ij}}$ be the member of $M_{n-1}(R)$ obtained by deleting the $i^{\text{th}}$ row and the $j^{\text{th}}$ column from $A$. Then

(a) for any $j$, $\det A = \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \det \widehat{A_{ij}}$, i.e., $\det A$ may be calculated by "expansion in cofactors" about the $j^{\text{th}}$ column,

(b) for any $i$, $\det A = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \det \widehat{A_{ij}}$, i.e., $\det A$ may be calculated by "expansion in cofactors" about the $i^{\text{th}}$ row.

PROOF. This may be derived in the same way from Proposition 2.36 by using the principle of permanence of identities. □

**Corollary 5.3** (Vandermonde matrix and determinant). If $r_1, \dots, r_n$ lie in a commutative ring $R$ with identity, then

$$\det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ r_1 & r_2 & \cdots & r_n \\ r_1^2 & r_2^2 & \cdots & r_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ r_1^{n-1} & r_2^{n-1} & \cdots & r_n^{n-1} \end{pmatrix} = \prod_{j>i} (r_j - r_i).$$

PROOF. The derivation of this from Proposition 5.2 is the same as the derivation of Corollary 2.37 from Proposition 2.35. □

**Proposition 5.4** (Cramer's rule). Let $R$ be a commutative ring with identity, let $A$ be in $M_n(R)$, and define $A^{\text{adj}}$ in $M_n(R)$ to be the classical adjoint of $A$, namely the matrix with entries $A^{\text{adj}}_{ij} = (-1)^{i+j} \det \widehat{A_{ji}}$, where $\widehat{A_{kl}}$ defined as in the statement of Proposition 5.2. Then $AA^{\text{adj}} = A^{\text{adj}}A = (\det A)I$.

PROOF. This may be derived from Proposition 2.38 in the same way as for Propositions 5.1 and 5.2 using the principle of permanence of identities. □

**Corollary 5.5.** Let $R$ be a commutative ring with identity, and let $A$ be in $M_n(R)$. If $\det A$ is a unit in $R$, then $A$ has a two-sided inverse in $M_n(R)$. Conversely if $A$ has a one-sided inverse in $M_n(R)$, then $\det A$ is a unit in $R$.

REMARK. If $R$ is a field, then $A$ and any associated linear map are often called **nonsingular** if invertible, **singular** otherwise. When $R$ is not a field, terminology varies for what to call a noninvertible matrix whose determinant is not 0.

PROOF. If $\det A$ is a unit in $R$, let $r$ be its multiplicative inverse. Then Proposition 5.4 shows that $r A^{\mathrm{adj}}$ is a two-sided inverse of $A$. Conversely if $A$ has, say, a left inverse $B$, then $BA = I$ implies $(\det B)(\det A) = \det I = 1$, and $\det B$ is an inverse for $\det A$. A similar argument applies if $A$ has a right inverse. □

## 3. Characteristic and Minimal Polynomials

Again let $\mathbb{K}$ be a field. If $A$ is in $M_n(\mathbb{K})$, the **characteristic polynomial** of $A$ is defined to be the member of the ring $\mathbb{K}[X]$ of polynomials in one indeterminate $X$ given by $F(X) = \det(XI - A)$. The material of Section 2 shows that $F(X)$ is well defined, being the determinant of a member of $M_n(\mathbb{K}[X])$. It is apparent from the definition of determinant in Section 2 that $F(X)$ is a monic polynomial of degree $n$ with coefficient $-\operatorname{Tr} A = -\sum_{j=1}^{n} A_{jj}$ for $X^{n-1}$. Evaluating $F(X)$ at 0, we see that the constant term is $(-1)^n \det A$.

Since the determinant of a product in $M_n(\mathbb{K}[X])$ is the product of the determinants (Proposition 5.1a) and since $C^{-1}(XI - A)C = XI - C^{-1}AC$, we have

$$\det(XI - C^{-1}AC) = (\det C)^{-1} \det(XI - A)(\det C) = \det(XI - A).$$

Thus similar matrices have equal characteristic polynomials. If $V$ is an $n$-dimensional vector space over $\mathbb{K}$ and $L : V \to V$ is linear, then the matrices of $L$ in any two ordered bases of $V$ (the domain basis being assumed equal to the range basis) are similar, and their characteristic polynomials are the same. Consequently we can define the **characteristic polynomial** of $L$ to be the characteristic polynomial of any matrix of $L$.

The development of characteristic polynomials has thus be redone in a way that is valid over any field $\mathbb{K}$ without making use of the ring homomorphism from polynomials in one indeterminate over $\mathbb{K}$ to polynomial functions from $\mathbb{K}$ into itself. The discussion in Section II.8 of eigenvectors and eigenvalues for members $A$ of $M_n(\mathbb{K})$ and for linear maps $L : V \to V$ with $V$ finite-dimensional over $\mathbb{K}$ is now meaningful, and there is no need to repeat it.

In particular, the eigenvalues of $A$ and $L$ are exactly the roots of their characteristic polynomial, no matter what $\mathbb{K}$ is. If $\mathbb{K}$ is algebraically closed, then the characteristic polynomial has a root, and consequently $A$ and $L$ each have at least one eigenvalue.

If $L : V \to V$ is linear and $V$ is finite-dimensional, then a vector subspace $U$ of $V$ is said to be **invariant** under $L$ if $L(U) \subseteq U$. In this case $L\big|_U$ is a well-defined linear map from $U$ to itself. Since $L(U) \subseteq U$, Proposition 2.25 shows that $L : V \to V$ factors through $V/U$ as a linear map $\overline{L} : V/U \to V/U$. We shall use this construction, the existence of eigenvalues in the algebraically closed case, and an induction to prove the following.

**Proposition 5.6.** If $\mathbb{K}$ is an algebraically closed field, if $V$ is a finite-dimensional vector space over $\mathbb{K}$, and if $L : V \to V$ is linear, then $V$ has an ordered basis in which the matrix of $L$ is upper triangular. Consequently any member of $M_n(\mathbb{K})$ is similar to an upper triangular matrix.

REMARKS. For an upper triangular matrix $A = \begin{pmatrix} c_1 & & * \\ & \ddots & \\ 0 & & c_n \end{pmatrix}$ in $M_n(\mathbb{K})$, the characteristic polynomial is $\prod_{j=1}^{n} (X - c_j)$ because the only nonzero term in the definition of $\det(XI - A)$ is the one corresponding to the identity permutation. Triangular form is not yet the canonical form we seek for a square matrix because a particular square matrix may be similar to infinitely many matrices in triangular form.

PROOF. We proceed by induction on $n = \dim V$, with the base case $n = 1$ being clear. Suppose that the result holds for all linear maps from spaces of dimension $< n$ to themselves. Given $L : V \to V$ with $\dim V = n$, let $v_1$ be an eigenvector of $L$. This exists by the remarks before the proposition since $\mathbb{K}$ is algebraically closed. Let $U$ be the vector subspace $\mathbb{K}v_1$. Then $L(U) \subseteq U$, and Proposition 2.25 shows that $L : V \to V$ factors through $V/U$ as a linear map $\overline{L} : V/U \to V/U$. Since $\dim V/U = n - 1$, the inductive hypothesis produces an ordered basis $(\bar{v}_2, \dots, \bar{v}_n)$ of $V/U$ such that the matrix of $\overline{L}$ is upper triangular in this basis. This condition means that $\overline{L}(\bar{v}_j) = \sum_{i=2}^{j} c_{ij} \bar{v}_i$ for $j \geq 2$. Select coset representatives $v_2, \dots, v_n$ of $\bar{v}_2, \dots, \bar{v}_n$ so that $\bar{v}_j = v_j + U$ for $j \geq 2$. Then $L(v_j + U) = \sum_{i=2}^{j} c_{ij}(v_i + U)$ for $j \geq 2$, and hence $L(v_j)$ lies in the coset $\sum_{i=2}^{j} c_{ij} v_i + U$ for $j \geq 2$. For each $j \geq 1$, we then have $L(v_j) = \sum_{i=2}^{j} c_{ij} v_i + c_{1j} v_1$ for some scalar $c_{1j}$, and we see that $(v_1, \dots, v_n)$ is the required ordered basis. □

Let us return to the situation in which $\mathbb{K}$ is any field. For a matrix $A$ in $M_n(\mathbb{K})$ and a polynomial $P$ in $\mathbb{K}[X]$, it is meaningful to form $P(A)$. We can do so by two equivalent methods, both useful. The concrete way of forming $P(A)$ is as $P(A) = c_n A^n + \cdots + c_1 A + c_0 I$ if $P(X) = c_n X^n + \cdots + c_1 X + c_0$. The abstract way is to form the subring $T$ of $M_n(\mathbb{K})$ generated by $\mathbb{K}I$ and $A$. This subring is commutative. We let $\varphi : \mathbb{K} \to T$ be given by $\varphi(c) = cI$. Then the universal mapping property of $\mathbb{K}[X]$ given in Proposition 4.24 produces a unique ring homomorphism $\Phi : \mathbb{K}[X] \to T$ such that $\Phi(c) = cI$ for all $c \in \mathbb{K}$ and $\Phi(X) = A$. The value of $P(A)$ is the element $\Phi(P)$ of $T$.

For $A$ in $M_n(\mathbb{K})$, let us study all polynomials $P$ such that $P(A) = 0$. For any polynomial $P$ and any invertible matrix $C$, we have

$$P(C^{-1}AC) = C^{-1}P(A)C$$

because if $P(X) = c_n X^n + \cdots + c_1 X + c_0$, then

$$\begin{aligned} P(C^{-1}AC) &= c_n(C^{-1}AC)^n + \cdots + c_1 C^{-1}AC + c_0 I \\ &= C^{-1}(c_n A^n + \cdots + c_1 A + c_0 I)C. \end{aligned}$$

Consequently if $P(A) = 0$, then $P(C^{-1}AC) = 0$, and the set of matrices with $P(A) = 0$ is closed under similarity. We shall make use of this observation a little later in this section.

**Proposition 5.7.** If $A$ is in $M_n(\mathbb{K})$, then there exists a nonzero polynomial $P$ in $\mathbb{K}[X]$ such that $P(A) = 0$.

PROOF. The $\mathbb{K}$ vector space $M_n(\mathbb{K})$ has dimension $n^2$. Therefore the $n^2 + 1$ matrices $I, A, A^2, \ldots, A^{n^2}$ are linearly dependent, and we have

$$c_0 + c_1 A + c_2 A^2 + \cdots + c_{n^2} A^{n^2} = 0$$

for some set of scalars not all 0. Then $P(A) = 0$ for the polynomial $P(X) = c_0 + c_1 X + c_2 X^2 + \cdots + c_{n^2} X^{n^2}$; this $P$ is not the 0 polynomial since at least one of the coefficients is not 0. □

ALTERNATIVE PROOF IF $\mathbb{K}$ IS ALGEBRAICALLY CLOSED. Since the set of polynomials $P$ with $P(A) = 0$ depends only on the similarity class of $A$, Proposition 5.6 shows that there is no loss of generality in assuming that $A$ is upper triangular, say of the form $\begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$. Then $A - \lambda_j I$ is upper triangular with 0 in the $j^{\text{th}}$ diagonal entry, and $\prod_{j=1}^n (A - \lambda_j I)$ is upper triangular with 0 in all diagonal entries. Therefore $\left(\prod_{j=1}^n (A - \lambda_j I)\right)^n = 0$. □

With $A$ fixed, we continue to consider the set of all polynomials $P(X)$ such that $P(A) = 0$. Let us think of $P(A)$ as being computed by the abstract procedure described above, namely as the image of $A$ under the ring homomorphism $\Phi : \mathbb{K}[X] \to T$ such that $\Phi(c) = cI$ for all $c \in \mathbb{K}$ and $\Phi(X) = A$, where $T$ is the commutative subring of $M_n(\mathbb{K})$ generated by $\mathbb{K}I$ and $A$. Then the set of all polynomials $P(X)$ with $P(A) = 0$ is the kernel of the ring homomorphism $\Phi$. This set is therefore an ideal, and Proposition 5.7 shows that the ideal is nonzero. We shall apply the following proposition to this ideal.

**Proposition 5.8.** If $I$ is a nonzero ideal in $\mathbb{K}[X]$, then there exists a unique monic polynomial of lowest degree in $I$, and every member of $I$ is the product of this particular polynomial by some other polynomial.

PROOF. Let $B(X)$ be a nonzero member of $I$ of lowest possible degree; adjusting $B$ by a scalar factor, we may assume that $B$ is monic. If $A$ is in $I$, then Proposition 1.12 produces polynomials $Q$ and $R$ such that $A = BQ + R$ and either $R = 0$ or $\deg R < \deg B$. Since $I$ is an ideal, $BQ$ is in $I$ and hence $R = A - BQ$ is in $I$. From minimality of the degree of $B$, we conclude that $R = 0$. Hence $A = BQ$, and $A$ is exhibited as the product of $B$ and some other polynomial $Q$. If $B_1$ is a second monic polynomial of lowest degree in $I$, then we can take $A = B_1$ to see that $B_1 = QB$. Since $\deg B_1 = \deg B$, we conclude that $\deg Q = 0$. Thus $Q$ is a constant polynomial. Comparing the leading coefficients of $B$ and $B_1$, we see that $Q(X) = 1$. □

With $A$ fixed in $M_n(\mathbb{K})$, let us apply Proposition 5.8 to the ideal of all polynomials $P$ in $\mathbb{K}[X]$ with $P(A) = 0$. The unique monic polynomial of lowest degree in this ideal is called the **minimal polynomial** of $A$. Let us try to identify this minimal polynomial.

**Theorem 5.9** (Cayley–Hamilton Theorem). If $A$ is in $M_n(\mathbb{K})$ and if $F(X) = \det(XI - A)$ is its characteristic polynomial, then $F(A) = 0$.

PROOF. Let $T$ be the commutative subring of $M_n(\mathbb{K})$ generated by $\mathbb{K}I$ and $A$, and define a member $B(X)$ of the ring $T[X]$ by $B(X) = XI - A$. The $(i, j)^{\text{th}}$ entry of $B(X)$ is $B_{ij}(X) = \delta_{ij}X - A_{ij}$, and $F(X) = \det B(X)$.

Let $C(X) = B(X)^{\text{adj}}$ denote the classical adjoint of $B(X)$ as a member of $T[X]$; the form of $C(X)$ is given in the statement of Cramer's rule (Proposition 5.4), and that proposition says that

$$B(X)C(X) = (\det B(X))I = F(X)I.$$

The equality in the $(i, j)^{\text{th}}$ entry is the equality $\delta_{ij}F(X) = \sum_j B_{ik}(X)C_{kj}(X)$ of members of $\mathbb{K}[X]$. Application of the substitution homomorphism $X \mapsto A$ gives

$$\delta_{ij}F(A) = \sum_k B_{ik}(A)C_{kj}(A) = \sum_k (\delta_{ik}A - A_{ik}I)C_{kj}(A).$$

Multiplying on the right by the $i^{\text{th}}$ standard basis vector $e_i$ and summing on $i$, we obtain the equality of vectors

$$F(A)e_j = \sum_i \sum_k (\delta_{ik}Ae_i - A_{ik}e_i)C_{kj}(A) = \sum_k C_{kj}(A)\Big(\sum_i (\delta_{ik}Ae_i - A_{ik}e_i)\Big)$$

since $C_{kj}(A)$ is a scalar. But $\sum_i (\delta_{ik}Ae_i - A_{ik}e_i) = Ae_k - \sum_i A_{ik}e_i = 0$ for all $k$, and therefore $F(A)e_j = 0$. Since $j$ is arbitrary, $F(A) = 0$. □

**Corollary 5.10.** If $A$ is in $M_n(\mathbb{K})$, then the minimal polynomial of $A$ divides the characteristic polynomial of $A$.

PROOF. Theorem 5.9 shows that the characteristic polynomial of $A$ lies in the ideal of all polynomials vanishing on $A$. Then the corollary follows from Proposition 5.8. □

For our matrix $A$ in $M_n(\mathbb{K})$, let $F(X)$ be the characteristic polynomial, and let $M(X)$ be the minimal polynomial. By unique factorization (Theorem 1.17), the monic polynomial $F(X)$ has a factorization into powers of distinct prime monic polynomials of the form

$$F(X) = P_1(X)^{k_1} \cdots P_r(X)^{k_r},$$

and this factorization is unique up to the order of the factors. Since $M(X)$ is a monic polynomial dividing $F(X)$, we must have

$$M(X) = P_1(X)^{l_1} \cdots P_r(X)^{l_r}$$

with $l_1 \leq k_1, \ldots, l_r \leq k_r$, by the same argument that deduced Corollary 1.7 from unique factorization in the ring of integers. We shall see shortly that $k_j > 0$ implies $l_j > 0$ if $P_j(X)$ is of degree 1, i.e., if $P_j(X)$ is of the form $X - \lambda_0$; in other words, if $\lambda_0$ is an eigenvalue of $A$, then $X - \lambda_0$ divides its minimal polynomial. We return to this point in a moment. Problem 31 at the end of the chapter will address the same question when $P_j(X)$ has degree $> 1$.

EXAMPLES.

(1) In the 2-by-2 case, $\begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix}$ has minimal polynomial $M(X) = X - c$, and $\begin{pmatrix} c & 1 \\ 0 & c \end{pmatrix}$ has $M(X) = (X - c)^2$. Both matrices have characteristic polynomial $F(X) = (X - c)^2$.

(2) The $k$-by-$k$ matrix

$$\begin{pmatrix} c & 1 & 0 & \cdots & 0 & 0 \\ 0 & c & 1 & \cdots & 0 & 0 \\ & & \ddots & & & \\ 0 & 0 & 0 & \cdots & c & 1 \\ 0 & 0 & 0 & \cdots & 0 & c \end{pmatrix}$$

with $c$ in every diagonal entry, with 1 in every entry just above the diagonal, and with 0 elsewhere has minimal polynomial $M(X) = (X - c)^k$ and characteristic polynomial $F(X) = (X - c)^k$.

(3) If a matrix $A$ is made up exclusively of several blocks of the type in Example 2 with the same $c$ in each case, the $i^{\text{th}}$ block being of size $k_i$, then the minimal polynomial is $M(X) = (X - c)^{\max_i k_i}$, and the characteristic polynomial is $F(X) = (X - c)^{\sum_i k_i}$.

(4) If $A$ is made up exclusively of several blocks as in Example 3 but with $c$ different for each block, then the minimal and characteristic polynomials for $A$ are obtained by multiplying the minimal and characteristic polynomials obtained from Example 3 for the various $c$'s.

To proceed further, let us change our point of view, working with linear maps $L : V \to V$, where $V$ is a finite-dimensional vector space over $\mathbb{K}$. We have already defined the characteristic polynomial of $L$ to be the characteristic polynomial of the matrix of $L$ in any ordered basis; this is well defined because similar matrices have the same characteristic polynomial. In analogous fashion we can define the **minimal polynomial** of $L$ to be the minimal polynomial of the matrix of $L$ in any ordered basis; this is well defined since, as we have seen, the set of polynomials $P$ in one indeterminate with $P(A) = 0$ is the same as the set with $P(C^{-1}AC) = 0$ if $C$ is invertible.

Another way of approaching the matter of the minimal polynomial of $L$ is to define $P(L)$ for any polynomial $P$ in one indeterminate. As with matrices, we can define $P(L)$ either concretely by substituting $L$ for $X$ in the expression for $P(X)$, or we can define $P(L)$ abstractly by appealing to the universal mapping property in Proposition 4.24. For the latter we work with the subring $T'$ of linear maps from $V$ to itself generated by $\mathbb{K}I$ and $L$. This subring is commutative. We let $\varphi : \mathbb{K} \to T'$ be given by $\varphi(c) = cI$, and we use Proposition 4.24 to obtain the unique ring homomorphism $\Phi : \mathbb{K}[X] \to T'$ such that $\Phi(c) = cI$ for all $c \in \mathbb{K}$ and $\Phi(X) = L$. Then $P(L)$ is the element $\Phi(P)$ of $T'$. Once $P(L)$ is defined, we observe that the set of polynomials $P(X)$ such that $P(L) = 0$ is a nonzero ideal in $\mathbb{K}[X]$; Proposition 5.8 yields a unique monic polynomial of lowest degree in this ideal, and that is the minimal polynomial of $L$.

Linear maps enable us to make convenient use of invariant subspaces. Recall from earlier in the section that a vector subspace $U$ of $V$ is said to be invariant under the linear map $L : V \to V$ if $L(U) \subseteq U$; in this case we obtain associated linear maps $L\big|_U : U \to U$ and $\overline{L} : V/U \to V/U$. Relationships among the characteristic polynomials and minimal polynomials of these linear maps are given in the next two propositions.

**Proposition 5.11.** Let $V$ be a finite-dimensional vector space over $\mathbb{K}$, let $L : V \to V$ be linear, let $U$ be a proper nonzero invariant subspace under $L$, and let $\overline{L} : V/U \to V/U$ be the induced linear map on $V/U$. Then the characteristic polynomials of $L$, $L\big|_U$, and $\overline{L}$ are related by

$$\det(XI - L) = \det\left(XI - L\big|_U\right) \det(XI - \overline{L}).$$

PROOF. Let $\Gamma_U = (v_1, \ldots, v_k)$ be an ordered basis of $U$, and extend $\Gamma_U$ to an ordered basis $\Gamma = (v_1, \ldots, v_n)$ of $V$. Then $\overline{\Gamma} = (v_{k+1} + U, \ldots, v_n + U)$

is an ordered basis of $V/U$. Since $U$ is invariant under $L$, the matrix of $L$ in the ordered basis $\Gamma$ is of the form $\begin{pmatrix} A & B \\ 0 & D \end{pmatrix}$, where $A$ is the matrix of $L\big|_U$ in the ordered basis $\Gamma_U$ and $D$ is the matrix of $\overline{L}$ in the ordered basis $\overline{\Gamma}$. Passing to the characteristic polynomials and applying Proposition 5.1g, we obtain the desired conclusion. □

**Proposition 5.12.** Let $V$ be a finite-dimensional vector space over $\mathbb{K}$, let $L : V \to V$ be linear, let $U$ be a proper nonzero invariant subspace under $L$, and let $\overline{L} : V/U \to V/U$ be the induced linear map on $V/U$. Then the minimal polynomials of $L\big|_U$ and $\overline{L}$ divide the minimal polynomial of $L$.

PROOF. Let $N(X)$ be the minimal polynomial of $L\big|_U$. Then $N(X)$ is the unique monic polynomial of lowest degree in the ideal of all polynomials $P(X)$ such that $P(L)u = 0$ for all $u$ in $U$. The minimal polynomial $M(X)$ of $L$ has this property because $M(X)v = 0$ for all $v$ in $V$. Therefore $M(X)$ is in the ideal and is the product of $N(X)$ and some other polynomial.

Among linear maps $S$ from $V$ into $V$ carrying $U$ into itself, the function $S \mapsto \overline{S}$ sending $S$ to the linear map $\overline{S}$ induced on $V/U$ is a homomorphism of rings. It follows that if $P(X)$ is a polynomial with $P(L) = 0$, then $P(\overline{L}) = 0$. Taking $P(X)$ to be the minimal polynomial of $L$, we see that the minimal polynomial of $L$ is in the ideal of polynomials vanishing on $\overline{L}$. Therefore it is the product of the minimal polynomial of $\overline{L}$ and some other polynomial. □

Let us come back to the unproved assertion before the examples—that $k_j > 0$ implies $l_j > 0$ if $P_r(X)$ has degree 1. We prove the linear-function version of this statement as a corollary of Proposition 5.12.

**Corollary 5.13.** If $L : V \to V$ is linear on a finite-dimensional vector space over $\mathbb{K}$ and if a first-degree polynomial $X - \lambda_0$ divides the characteristic polynomial of $L$, then $X - \lambda_0$ divides the minimal polynomial of $L$.

PROOF. If $X - \lambda_0$ divides the characteristic polynomial, then $\lambda_0$ is an eigenvalue of $L$, say with $v$ as an eigenvector. Then $U = \mathbb{K}v$ is an invariant subspace under $L$, and the characteristic and minimal polynomials of $L\big|_U$ are both $X - \lambda_0$. By Proposition 5.12, $X - \lambda_0$ divides the minimal polynomial of $L$. □

**Theorem 5.14.** If $L : V \to V$ is linear on a finite-dimensional vector space over $\mathbb{K}$, then $L$ has a basis of eigenvectors if and only if the minimal polynomial $M(X)$ of $L$ is the product of distinct factors of degree 1; in this case, $M(X)$ equals $(X - \lambda_1) \cdots (X - \lambda_k)$, where $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of $L$. Consequently a matrix $A$ in $M_n(\mathbb{K})$ is similar to a diagonal matrix if and only if its minimal polynomial is the product of distinct factors of degree 1.

PROOF. The easy direction is that $v_1, \ldots, v_n$ are the members of a basis of eigenvectors for $L$ with respective eigenvalues $\mu_1, \ldots, \mu_n$. In this case, let $\lambda_1, \ldots, \lambda_k$ be the distinct members of the set of eigenvalues, with $\mu_i = \lambda_{j(i)}$ for some function $j : \{1, \ldots, n\} \to \{1, \ldots, k\}$. Then $(L - \lambda_j I)(v) = 0$ for $v$ equal to any $v_i$ with $j(i) = j$. Since the linear maps $L - \lambda_j I$ commute as $j$ varies, $\prod_{j=1}^{k} (L - \lambda_j I)(v) = 0$ for $v$ equal to each of $v_1, \ldots, v_n$, hence for all $v$. Therefore the minimal polynomial $M(X)$ of $L$ divides $\prod_{j=1}^{k} (X - \lambda_j)$. On the other hand, Corollary 5.13 shows that the deg $M(X) \geq k$. Hence $M(X) = \prod_{j=1}^{k} (X - \lambda_j)$.

Conversely suppose that $M(X) = \prod_{j=1}^{k} (X - \lambda_j)$ with the $\lambda_j$ distinct. If $S_1$ is the linear map $S_1 = \prod_{j=2}^{k} (L - \lambda_j I)$, then the formula for $M(X)$ shows that $(L - \lambda_1 I) S_1(v) = 0$ for all $v$ in $V$, and hence image $S_1$ is a vector subspace of the eigenspace of $L$ for the eigenvalue $\lambda_1$. If $v$ is in $\ker S_1 \cap$ image $S_1$, we then have $0 = S_1(v) = \prod_{j=2}^{k} (L - \lambda_j I)(v) = \prod_{j=2}^{k} (\lambda_1 - \lambda_j) v$. Since $\lambda_1$ is distinct from $\lambda_2, \ldots, \lambda_k$, we conclude that $v = 0$, hence that $\ker S_1 \cap$ image $S_1 = 0$. Since $\dim \ker S_1 + \dim$ image $S_1 = \dim V$, Corollary 2.29 therefore gives

$$\begin{aligned} \dim V &= \dim \ker S_1 + \dim \text{image } S_1 \\ &= \dim(\ker S_1 + \text{image } S_1) + \dim(\ker S_1 \cap \text{image } S_1) \\ &= \dim(\ker S_1 + \text{image } S_1). \end{aligned}$$

Hence $V = \ker S_1 + \text{image } S_1$. Since $\ker S_1 \cap \text{image } S_1 = 0$, we conclude that $V = \ker S_1 \oplus \text{image } S_1$.

Actually, the same calculation of $S_1(v)$ as above shows that image $S_1$ is the full eigenspace of $L$ for the eigenvalue $\lambda_1$. In fact, if $L(v) = \lambda_1 v$, then $S_1(v) = \prod_{j=2}^{k} (\lambda_1 - \lambda_j) v$, and hence $v$ equals the image under $S_1$ of $\left( \prod_{j=2}^{k} (\lambda_1 - \lambda_j) \right)^{-1} v$.

Next, since $L$ commutes with $S_1$, $\ker S_1$ is an invariant subspace under $L$, and $\lambda_1$ is not an eigenvalue of $L\big|_{\ker S_1}$. Thus $X - \lambda_1$ does not divide the minimal polynomial of $L\big|_{\ker S_1}$. On the other hand, $S_1$ vanishes on the eigenspaces of $L$ for eigenvalues $\lambda_2, \ldots, \lambda_k$, and Corollary 5.13 shows for $j \geq 2$ that $X - \lambda_j$ divides the minimal polynomial of $L\big|_{\ker S_1}$. Taking Proposition 5.12 into account, we conclude that $L\big|_{\ker S_1}$ has minimal polynomial $\prod_{j=2}^{k} (X - \lambda_j)$. We have succeeded in splitting off the eigenspace of $L$ under $\lambda_1$ as a direct summand and reducing the proposition to the case of $k - 1$ eigenvalues. Thus induction shows that $V$ is the direct sum of its eigenspaces for the eigenvalues $\lambda_2, \ldots, \lambda_k$, and $L$ thus has a basis of eigenvectors. □

Theorem 5.14 comes close to solving the canonical-form problem for similarity in the case of one kind of square matrices: if the minimal polynomial of $A$ is the product of distinct factors of degree 1, then $A$ is similar to a diagonal matrix. To

complete the solution for this case, all we have to do is to say when two diagonal matrices are similar to each other; this step is handled by the following easy proposition.

**Proposition 5.15.** Two diagonal matrices $A$ and $A'$ in $M_n(\mathbb{K})$ with respective diagonal entries $d_1, \dots, d_n$ and $d'_1, \dots, d'_n$ are similar if and only if there is a permutation $\sigma$ in $\mathfrak{S}_n$ such that $d'_j = d_{\sigma(j)}$ for all $j$.

PROOF. The respective characteristic polynomials are $\prod_{j=1}^n (X - d_j)$ and $\prod_{j=1}^n (X - d'_j)$. If $A$ and $A'$ are similar, then the characteristic polynomials are equal, and unique factorization (Theorem 1.17) shows that the factors $X - d'_j$ match the factors $X - d_j$ up to order. Conversely if there is a permutation $\sigma$ in $\mathfrak{S}_n$ such that $d'_j = d_{\sigma(j)}$ for all $j$, then the matrix $C$ whose $j^{\text{th}}$ column is $e_{\sigma(j)}$ has the property that $A' = C^{-1}AC$. □

To proceed further with obtaining canonical forms for matrices under similarity and for linear maps under isomorphism, we shall use linear maps in ways that we have not used them before. In particular, it will be convenient to be able to recognize direct-sum decompositions from properties of linear maps. We take up this matter in the next section.

## 4. Projection Operators

In this section we shall see how to recognize direct-sum decompositions of a vector space $V$ from the associated projection operators, and we shall relate these operators to invariant subspaces under a linear map $L : V \to V$.

If $V = U_1 \oplus U_2$, then the function $E_1$ defined by $E_1(u_1 + u_2) = u_1$ when $u_1$ is in $U_1$ and $u_2$ is in $U_2$ is linear, satisfies $E_1^2 = E_1$, and has image $E_1 = U_1$ and $\ker E_1 = U_2$. We call $E_1$ the **projection** of $V$ on $U_1$ along $U_2$. A decomposition of $V$ as the direct sum of two vector spaces, when the first of the two spaces is singled out, therefore determines a projection operator uniquely. A converse is as follows.

**Proposition 5.16.** If $V$ is a vector space and $E_1 : V \to V$ is a linear map such that $E_1^2 = E_1$, then there exists a direct-sum decomposition $V = U_1 \oplus U_2$ such that $E_1$ is the projection of $V$ on $U_1$ along $U_2$. In this case, $(I - E_1)^2 = I - E_1$, and $I - E_1$ is the projection of $V$ on $U_2$ along $U_1$.

PROOF. Define $U_1 = \text{image } E_1$ and $U_2 = \ker E_1$. If $v$ is in image $E_1 \cap \ker E_1$, then $E_1(v) = 0$ since $v$ is in $\ker E_1$ and $v = E_1(w)$ for some $w$ in $V$ since

$v$ is in image $E_1$. Then $0 = E_1(v) = E_1^2(w) = E_1(w) = v$, and therefore image $E_1 \cap \ker E_1 = 0$.

If $v \in V$ is given , write $v = E_1(v) + (I - E_1)(v)$. Then $E_1(v)$ is in image $E_1$, and the computation $E_1(I - E_1)(v) = (E_1 - E_1^2)(v) = (E_1 - E_1)(v) = 0$ shows that $(I - E_1)(v) = 0$. Consequently $V = \text{image } E_1 + \ker E_1$, and we conclude that $V = \text{image } E_1 \oplus \ker E_1$.

Hence $V = U_1 \oplus U_2$, where $U_1 = \text{image } E_1$ and $U_2 = \ker E_1$. In this notation, $E_1$ is 0 on $U_2$. If $v$ is in $U_1$, then $v = E_1(w)$ for some $w$, and we have $v = E_1(w) = E_1^2(w) = E_1(E_1(w)) = E_1(v)$. Thus $E_1$ is the identity on $U_1$ and is the projection as asserted.

For $(I - E_1)^2$, we have $(I - E_1)^2 = I - 2E_1 + E_1^2 = I - 2E_1 + E_1 = I - E_1$, and $I - E_1$ is a projection. It is 1 on $U_2$ and is 0 on $U_1$, hence is the projection of $V$ on $U_2$ along $U_1$. □

Let us generalize these considerations to the situation that $V$ is the direct sum of $r$ vector subspaces. The following facts about the situation in Proposition 5.16, with the definition $E_2 = I - E_1$, are relevant to formulating the generalization:

(i) $E_1$ and $E_2$ have $E_1^2 = E_1$ and $E_2^2 = E_2$,
(ii) $E_1E_2 = E_2E_1 = 0$,
(iii) $E_1 + E_2 = I$.

Suppose that $V = U_1 \oplus \cdots \oplus U_r$. Define $E_j(u_1 + \cdots + u_r) = u_j$. Then $E_j$ is linear from $V$ to itself with $E_j^2 = E_j$, and Proposition 5.16 shows that $E_j$ is the projection of $V$ on $U_j$ along the direct sum of the remaining $U_i$'s. The linear maps $E_1, \ldots, E_r$ then satisfy

(i′) $E_j^2 = E_j$ for $1 \le j \le r$,
(ii′) $E_jE_i = 0$ if $i \ne j$,
(iii′) $E_1 + \cdots + E_r = I$.

A converse is as follows.

**Proposition 5.17.** If $V$ is a vector space and $E_j : V \to V$ for $1 \le j \le r$ are linear maps such that

(a) $E_jE_i = 0$ if $i \ne j$, and
(b) $E_1 + \cdots + E_r = I$,

then $E_j^2 = E_j$ for $1 \le j \le r$ and the vector subspaces $U_j = \text{image } E_j$ have the properties that $V = U_1 \oplus \cdots \oplus U_r$ and that $E_j$ is the projection of $V$ on $U_j$ along the direct sum of all $U_i$ but $U_j$.

PROOF. Multiplying (b) through by $E_j$ on the left and applying (a) to each term on the left side except the $j^{\text{th}}$, we obtain $E_j^2 = E_j$. Therefore, for each $j$, $E_j$ is a projection on $U_j$ along some vector subspace depending on $j$.

If $v$ is in $V$, then (b) gives $v = E_1(v) + \cdots + E_r(v)$ and shows that $V = U_1 + \cdots + U_r$. Suppose that $v$ is in the intersection of $U_j$ with the sum of the other $U_i$'s. Write $v = \sum_{i \neq j} u_i$ with $u_i = E_i(w_i)$ in $U_i$. Applying $E_j$ and using the fact that $v$ is in $U_j$, we obtain $v = E_j(v) = \sum_{i \neq j} E_j E_i(w_i)$. Every term of the right side is 0 by (a), and hence $v = 0$. Thus $V = U_1 \oplus \cdots \oplus U_r$.

Since $E_j E_i = 0$ for $i \neq j$, $E_j$ is 0 on each $U_i$ for $i \neq j$. Therefore the sum of all $U_i$ except $U_j$ is contained in the kernel of $E_j$. Since the image and kernel of $E_j$ intersect in 0, the sum of all $U_i$ except $U_j$ is exactly equal to the kernel of $E_j$. This completes the proof. □

**Proposition 5.18.** Suppose that a vector space $V$ is a direct sum $V = U_1 \oplus \cdots \oplus U_r$ of vector subspaces, that $E_1, \ldots, E_r$ are the corresponding projections, and that $L : V \to V$ is linear. Then all the subspaces $U_j$ are invariant under $L$ if and only if $LE_j = E_j L$ for all $j$.

PROOF. If $L(U_j) \subseteq U_j$ for all $j$, then $i \neq j$ implies $E_i L(U_j) \subseteq E_i(U_j) = 0$ and $LE_i(U_j) = L(0) = 0$. Also, $v \in U_j$ implies $E_j L(v) = L(v) = LE_j(v)$. Hence $E_i L = E_i L$ for all $i$.

Conversely if $E_j L = LE_j$ and if $v$ is in $U_j$, then $E_j L(v) = LE_j(v) = L(v)$ shows that $L(v)$ is in $U_j$. Therefore $L(U_j) \subseteq U_j$ for all $j$. □

## 5. Primary Decomposition

For the case that the minimal polynomial of a linear map $L : V \to V$ is the product of distinct factors of degree 1, Theorem 5.14 showed that $V$ is a direct sum of its eigenspaces. The proof used elementary vector-space techniques from Chapter II but did not take full advantage of the machinery developed in the present chapter for passing back and forth between polynomials in one indeterminate and the values of polynomials on $L$. Let us therefore rework the proof of that proposition, taking into account the discussion of projections in Section 4.

We seek an eigenspace decomposition $V = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_k}$ relative to $L$. Proposition 5.17 suggests looking for the corresponding decomposition of the identity operator as a sum of projections: $I = E_1 + \cdots + E_k$. According to that proposition, we obtain a direct-sum decomposition as soon as we obtain this kind of sum of linear maps such that $E_i E_j = 0$ for $i \neq j$. The $E_j$'s will automatically be projections.

The proof of Theorem 5.14 showed that $S_1 = \prod_{j=2}^{k} (L - \lambda_j I)$ has image equal to the kernel of $L - \lambda_1 I$, i.e., equal to the eigenspace for eigenvalue $\lambda_1$. If $v$ is in this eigenspace, then $S_1(v) = \prod_{j=2}^{k} (\lambda_1 - \lambda_j) v$. Hence $E_1 = c_1 S_1$, where $c_1^{-1} = \prod_{j=2}^{k} (\lambda_1 - \lambda_j)$. The linear map $S_1$ equals $Q_1(L)$, where $Q_1(X) =$

$\prod_{j=2}^{k}(X-\lambda_j)$. Thus $E_1 = c_1 Q_1(L)$. Similar remarks apply to the other eigenspaces, and therefore the required decomposition of the identity operator has to be of the form $I = c_1 Q_1(L) + \cdots + c_k Q_k(L)$ with $c_1, \ldots, c_k$ equal to certain scalars.

The polynomials $Q_1(X), \ldots, Q_l(X)$ are at hand from the start, each containing all but one factor of the minimal polynomial. Moreover, $i \neq j$ implies that

$$Q_i(L)Q_j(L) = \Big(\prod_{l=1}^{k}(L-\lambda_l I)\Big)\Big(\prod_{l\neq i,j}(L-\lambda_l I)\Big).$$

The first factor on the right side is the value of the minimal polynomial of $L$ with $L$ substituted for $X$. Hence the right side is 0, and we see that our linear maps $E_1, \ldots, E_k$ have $E_i E_j = 0$ for $i \neq j$.

As soon as we allow nonconstant coefficients in place of the $c_j$'s in the above argument, we obtain a generalization of Theorem 5.14 to the situation that the minimal polynomial of $L$ is arbitrary. The prime factors of the minimal polynomial need not even be of degree 1. Hence the theorem applies to all $L$'s even if $\mathbb{K}$ is not algebraically closed.

**Theorem 5.19** (Primary Decomposition Theorem). Let $L : V \to V$ be linear on a finite-dimensional vector space over $\mathbb{K}$, and let $M(X) = P_1(X)^{l_1} \cdots P_k(X)^{l_k}$ be the unique factorization of the minimal polynomial $M(X)$ of $L$ into the product of powers of distinct monic prime polynomials $P_j(X)$. Define $U_j = \ker(P_j(L)^{l_j})$ for $1 \leq j \leq k$. Then

(a) $V = U_1 \oplus \cdots \oplus U_k$,
(b) the projection $E_j$ of $V$ on $U_j$ along the sum of the other $U_i$'s is of the form $T_j(L)$ for some polynomial $T_j$,
(c) each vector subspace $U_j$ is invariant under $L$,
(d) any linear map from $V$ to itself that commutes with $L$ carries each $U_j$ into itself,
(e) any vector subspace $W$ invariant under $L$ has the property that

$$W = (W \cap U_1) \oplus \cdots \oplus (W \cap U_k),$$

(f) the minimal polynomial of $L_j = L\big|_{U_j}$ is $P_j(X)^{l_j}$.

REMARKS. The decomposition in (a) is called the **primary decomposition** of $V$ under $L$, and the vector subspaces $U_j$ are called the **primary subspaces** of $V$ under $L$.

PROOF. For $1 \leq j \leq k$, define $Q_j(X) = M(X)/P_j(X)^{l_j}$. The ideal in $\mathbb{K}[X]$ generated by $Q_1(X), \ldots, Q_k(X)$ consists of all products of a single monic

polynomial $D(X)$ by arbitrary polynomials, according to Proposition 5.8, and $D(X)$ has to divide each $Q_j(X)$. Since $Q_j(X) = \prod_{i \neq j} P_i(X)^{l_i}$, $D(X)$ cannot be divisible by any $P_j(X)$, and consequently $D(X) = 1$. Thus there exist polynomials $R_1(X), \ldots, R_k(X)$ such that

$$1 = Q_1(X)R_1(X) + \cdots + Q_k(X)R_k(X).$$

Define $E_j = Q_j(L)R_j(L)$, so that $E_1 + \cdots + E_k = I$. If $i \neq j$, then $Q_i(X)Q_j(X) = M(X)\prod_{r \neq i,j} P_r(X)^{l_r}$. Since $M(L) = 0$, we see that $E_i E_j = 0$.

Proposition 5.17 says that each $E_j$ is a projection. Also, it says that if $U_j$ denotes image $E_j$, then $V = U_1 \oplus \cdots \oplus U_k$, and $E_j$ is the projection on $U_j$ along the sum of the other $U_i$'s. With this definition of the $U_j$'s (rather than the one in the statement of the theorem), we have therefore shown that (a) and (b) hold.

Let us see that conclusions (c), (d), and (e) follow from (b). Conclusion (c) holds by Proposition 5.18 since $L$ commutes with $T_j(L)$ whenever $T_j$ is a polynomial. For (d), if $J : V \to V$ is a linear map commuting with $L$, then $J$ commutes with each $E_j$ since (b) shows that each $E_j$ is of the form $T_j(L)$. From Proposition 5.18 we conclude that each $U_j$ is invariant under $J$. For (e), the subspace $W$ certainly contains $(W \cap U_1) \oplus \cdots \oplus (W \cap U_k)$. For the reverse containment suppose $w$ is in $W$. Since $E_j$ is of the form $T_j(L)$ and since $W$ is invariant under $L$, $E_j(w)$ is in $W$. But also $E_j(w)$ is in $U_j$. Therefore the expansion $w = \sum_j E_j(w)$ exhibits $w$ as the sum of members of the spaces $W \cap U_j$.

Next let us prove that $U_j$, as we have defined it, is given also by the definition in the statement of the theorem. In other words, let us prove that

$$\text{image } E_j = \ker(P_j(L)^{l_j}). \tag{$*$}$$

We need a preliminary fact. The polynomial $P_j(X)^{l_j}$ has the property that $M(X) = P_j(X)^{l_j}Q_j(X)$. Hence $P_j(L)^{l_j}Q_j(L) = M(L) = 0$. Multiplying by $R_j(L)$, we obtain

$$P_j(L)^{l_j}E_j = 0. \tag{$**$}$$

Now suppose that $v$ is in image $E_j$. Then $P_j(L)^{l_j}(v) = P_j(L)^{l_j}E_j(v) = 0$ by ($**$), and hence image $E_j \subseteq \ker(P_j(L)^{l_j})$. For the reverse inclusion, let $v$ be in $\ker(P_j(L)^{l_j})$. For $i \neq j$, $Q_i(X)R_i(X) = \big(\prod_{r \neq i,j} P_r(X)^{l_r}\big)R_i(X)P_j(X)^{l_j}$ and hence

$$E_i(v) = \big(\textstyle\prod_{r \neq i,j} P_r(L)^{l_r}\big)R_i(L)P_j(L)^{l_j}(v) = 0.$$

Writing $v = E_1(v) + \cdots + E_k(v)$, we see that $v = E_j(v)$. Thus $\ker(P_j(L)^{l_j}) \subseteq$ image $E_j$. Therefore ($*$) holds, and $U_j$ is as in the statement of the theorem.

Finally let us prove (f). Let $M_j(X)$ be the minimal polynomial of $L_j = L\big|_{U_j}$. From ($**$) we see that $P_j(L_j)^{l_j} = 0$. Hence $M_j(X)$ divides $P_j(X)^{l_j}$. For the

reverse divisibility we have $M_j(L_j) = 0$. Then certainly $M_j(L_j)Q_j(L_j)R_j(L_j)$, which equals $M_j(L)E_j$ on $U_j$, is 0 on $U_j$. Consider $M_j(L)E_j$ on $U_i = \text{image } E_i$ when $i \neq j$. Since $E_jE_i = 0$, $M_j(L)E_j$ equals 0 on all $U_i$ other than $U_j$. We conclude that $M_j(L)E_j$ equals 0 on $V$, i.e., $M_j(L)Q_j(L)R_j(L) = 0$. Since $M(X)$ is the minimal polynomial of $L$, $M(X)$ divides

$$M_j(X)Q_j(X)R_j(X) = M_j(X)\big(1 - \sum_{i\neq j} Q_i(X)R_i(X)\big), \tag{$\dagger$}$$

and the factor $P_j(X)^{l_j}$ of $M(X)$ must divide the right side of ($\dagger$). On that right side, $P_j(X)^{l_j}$ divides each $Q_i(X)$ with $i \neq j$. Since $P_j(X)$ does not divide 1, $P_j(X)$ does not divide the factor $1 - \sum_{i\neq j} Q_i(X)R_i(X)$. Since $P_j(X)$ is prime, $P_j(X)^{l_j}$ and $1 - \sum_{i\neq j} Q_i(X)R_i(X)$ are relatively prime. We know that $P_j(X)^{l_j}$ divides the product of $M_j(X)$ and $1 - \sum_{i\neq j} Q_i(X)R_i(X)$, and consequently $P_j(X)^{l_j}$ divides $M_j(X)$. This proves the reverse divisibility and completes the proof of (f). □

## 6. Jordan Canonical Form

Now we can return to the canonical-form problem for similarity of square matrices and isomorphism of linear maps from a finite-dimensional vector space to itself. The answer obtained in this section will solve the problem completely if $\mathbb{K}$ is algebraically closed but only partially if $\mathbb{K}$ fails to be algebraically closed. Problems 32–40 at the end of the chapter extend the content of this section to give a complete answer for general $\mathbb{K}$.

The present theorem is most easily stated in terms of matrices. A square matrix is called a **Jordan block** if it is of the form

$$\begin{pmatrix} c & 1 & 0 & 0 & \cdots & 0 & 0 \\ & c & 1 & 0 & \cdots & 0 & 0 \\ & & c & 1 & \cdots & 0 & 0 \\ & & & \ddots & \ddots & \vdots & \vdots \\ & & & & c & 1 & 0 \\ & & & & & c & 1 \\ & & & & & & c \end{pmatrix},$$

of some size and for some $c$ in $\mathbb{K}$, as in Example 2 of Section 3, with 0 everywhere below the diagonal. A square matrix is in **Jordan form**, or **Jordan normal form**, if it is block diagonal and each block is a Jordan block. One can insist on grouping the blocks for which the constant $c$ is the same and arranging the blocks for given $c$ in some order, but these refinements are inessential.

**Theorem 5.20** (Jordan canonical form).

(a) If the field $\mathbb{K}$ is algebraically closed, then every square matrix over $\mathbb{K}$ is similar to a matrix in Jordan form, and two matrices in Jordan form are similar to each other if and only if their Jordan blocks can be permuted so as to match exactly.

(b) For a general field $\mathbb{K}$, a square matrix $A$ is similar to a matrix in Jordan form if and only if each prime factor of its minimal polynomial has degree 1. Two matrices in Jordan form are similar to each other if and only if their Jordan blocks can be permuted so as to match exactly.

The first step in proving existence of a matrix in Jordan form similar to a given matrix is to use the Primary Decomposition Theorem (Theorem 5.19). We think of the matrix $A$ as operating on the space $\mathbb{K}^n$ of column vectors in the usual way. The primary subspaces are uniquely defined vector subspaces of $\mathbb{K}^n$, and we introduce an ordered basis, yet to be specified in full detail, within each primary subspace. The union of these ordered bases gives an ordered basis of $\mathbb{K}^n$, and we change from the standard basis to this one. The result is that the given matrix has been conjugated so that its appearance is block diagonal, each block having minimal polynomial equal to a power of a prime polynomial and the prime polynomials all being different. Let us call these blocks **primary blocks**. The effect of Theorem 5.19 has been to reduce matters to a consideration of each primary block separately. The hypothesis either that $\mathbb{K}$ is algebraically closed or, more generally, that the prime divisors of the minimal polynomial all have degree 1 means that the minimal polynomial of the primary block under study may be taken to be $(X - c)^l$ for some $c$ in $\mathbb{K}$ and some integer $l \geq 1$. In terms of Jordan form, we have isolated, for each $c$ in $\mathbb{K}$, what will turn out to be the subspace of $\mathbb{K}^n$ corresponding to Jordan blocks with $c$ in every diagonal entry.

Let us write $B$ for a primary block with minimal polynomial $(X - c)^l$. We certainly have $(B - cI)^l = 0$, and it follows that the matrix $N = B - cI$ has $N^l = 0$. A matrix $N$ with $N^l = 0$ for some integer $l \geq 0$ is said to be **nilpotent**. To prove the existence part of Theorem 5.20, it is enough to prove the following theorem.

**Theorem 5.21.** For any field $\mathbb{K}$, each nilpotent matrix $N$ in $M_n(\mathbb{K})$ is similar to a matrix in Jordan form.

The proof of Theorem 5.21 and of the uniqueness statements in Theorem 5.20 will occupy the remainder of this section. It is implicit in Theorem 5.21 that a nilpotent matrix in $M_n(\mathbb{K})$ has 0 as a root of its characteristic polynomial with multiplicity $n$, in particular that the only prime polynomials dividing the characteristic polynomial are the ones dividing the minimal polynomial. We

proved such a fact about divisibility earlier for general square matrices when the prime factor has degree 1, but we did not give a proof for general degree. We pause for a moment to give a direct proof in the nilpotent case.

**Lemma 5.22.** If $N$ is a nilpotent matrix in $M_n(K)$, then $N$ has characteristic polynomial $X^n$ and satisfies $N^n = 0$.

PROOF. If $N^l = 0$, then

$$(XI - N)(X^{l-1}I + X^{l-2}N + \cdots + X^2N^{l-3} + XN^{l-2} + N^{l-1}) = X^lI - N^l = X^lI.$$

Taking determinants and using Proposition 5.1 in the ring $R = \mathbb{K}[X]$, we obtain

$$\det(XI - N)\det(\text{other factor}) = \det(X^lI) = X^{ln}.$$

Thus $\det(XI - N)$ divides $X^{ln}$. By unique factorization in $\mathbb{K}[X]$, $\det(XI - N)$ is a constant times a power of $X$. Then we must have $\det(XI - N) = X^n$. Applying the Cayley–Hamilton Theorem (Theorem 5.9), we obtain $N^n = 0$. □

Let us now prove the uniqueness statements in Theorem 5.20; this step will in fact help orient us for the proof of Theorem 5.21. In (b), one thing we are to prove is that if $A$ is similar to a matrix in Jordan form, then every prime polynomial dividing the minimal polynomial has degree 1. Since characteristic and minimal polynomials are unchanged under similarity, we may assume that $A$ is itself in Jordan form. The characteristic and minimal polynomials of $A$ are computed in the four examples of Section 3. Since the minimal polynomial is the product of polynomials of degree 1, the only primes dividing it have degree 1.

In both (a) and (b) of Theorem 5.20, we are to prove that the Jordan form is unique up to permutation of the Jordan blocks. The matrix $A$ determines its characteristic polynomial, which determines the roots of the characteristic polynomial, which are the diagonal entries of the Jordan form. Thus the sizes of the primary blocks within the Jordan form are determined by $A$. Within each primary block, we need to see that the sizes of the various Jordan blocks are completely determined.

Thus we may assume that $N$ is nilpotent and that $C^{-1}NC = J$ is in Jordan form with 0's on the diagonal. Although we shall make statements that apply in all cases, the reader may be helped by referring to the particular matrix $J$ in Figure 5.1 and its powers in Figure 5.2.

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

FIGURE 5.1. Example of a nilpotent matrix in Jordan form.

Each block of the Jordan form $J$ contributes 1 to the dimension of the kernel (or null space really) of $J$ via the first column of the block, and hence

$$\dim(\ker J) = \#\{\text{Jordan blocks in } J\}.$$

In Figure 5.1 this number is 5.

$$J^2 = \begin{pmatrix} 0 & 0 & 1 & 0 & & & & & & & & \\ 0 & 0 & 0 & 1 & & & & & & & & \\ 0 & 0 & 0 & 0 & & & & & & & & \\ 0 & 0 & 0 & 0 & & & & & & & & \\ & & & & 0 & 0 & 1 & & & & & \\ & & & & 0 & 0 & 0 & & & & & \\ & & & & 0 & 0 & 0 & & & & & \\ & & & & & & & 0 & 0 & & & \\ & & & & & & & 0 & 0 & & & \\ & & & & & & & & & 0 & 0 & \\ & & & & & & & & & 0 & 0 & \\ & & & & & & & & & & & 0 \end{pmatrix} \quad \text{and} \quad J^3 = \begin{pmatrix} 0 & 0 & 0 & 1 & & & & & & & & \\ 0 & 0 & 0 & 0 & & & & & & & & \\ 0 & 0 & 0 & 0 & & & & & & & & \\ 0 & 0 & 0 & 0 & & & & & & & & \\ & & & & 0 & 0 & 0 & & & & & \\ & & & & 0 & 0 & 0 & & & & & \\ & & & & 0 & 0 & 0 & & & & & \\ & & & & & & & 0 & 0 & & & \\ & & & & & & & 0 & 0 & & & \\ & & & & & & & & & 0 & 0 & \\ & & & & & & & & & 0 & 0 & \\ & & & & & & & & & & & 0 \end{pmatrix}$$

FIGURE 5.2. Powers of the nilpotent matrix in Figure 5.1.

When $J$ is squared, the 1's in $J$ move up and to the right one more step beyond the diagonal except that blocks of size 2 become 0. When $J$ is cubed, the 1's in $J$ move up and to the right one further step except that blocks of size 3 become 0. Each time $J$ is raised to a new power one higher than before, each block that is nonzero in the old power contributes an additional 1 to the dimension of the kernel. Thus we have

$$\dim(\ker J^2) - \dim(\ker J) = \#\{\text{Jordan blocks of size} \geq 2\}$$

and $$\dim(\ker J^3) - \dim(\ker J^2) = \#\{\text{Jordan blocks of size} \geq 3\};$$

in the general case,

$$\dim(\ker J^k) - \dim(\ker J^{k-1}) = \#\{\text{Jordan blocks of size} \geq k\} \qquad \text{for } k \geq 1.$$

Lemma 5.22 says that $J^k = 0$ when $k$ is $\geq$ the size of $J$, and the differences need not be computed beyond that point.

For Figure 5.2 the values by inspection are $\dim(\ker J^2) = 9$ and $\dim(\ker J^3) = 11$; also $J^4 = 0$ and hence $\dim(\ker J^4) = 12$. The numbers of Jordan blocks of size $\geq k$ for $k = 1, 2, 3, 4$ are 5, 4, 2, 1, and these numbers indeed match the differences $5 - 0,\ 9 - 5,\ 11 - 9,\ 12 - 11$, as predicted by the above formula.

Since $C^{-1}NC = J$, we have $C^{-1}N^kC = J^k$ and $N^kC = CJ^k$. The matrix $C$ is invertible, and therefore $\dim(\ker J^k) = \dim(\ker CJ^k) = \dim(\ker N^kC) = \dim(\ker N^k)$. Hence

$$\dim(\ker N^k) - \dim(\ker N^{k-1}) = \#\{\text{Jordan blocks of size} \geq k\} \qquad \text{for } k \geq 1,$$

and the number of Jordan blocks of each size is uniquely determined by properties of $N$. This completes the proof of all the uniqueness statements in Theorem 5.20.

Now let us turn to the proof of Theorem 5.21, first giving the idea. The argument involves a great many choices, and it may be helpful to understand it in the context of Figures 5.1 and 5.2. Let $\Sigma = (e_1, \ldots, e_{12})$ be the standard ordered basis of $\mathbb{K}^{12}$. The matrix $J$, when operating by multiplication on the left, moves basis vectors to other basis vectors or to 0. Namely,

$$\begin{gathered} Je_1 = 0, \quad Je_2 = e_1, \quad Je_3 = e_2, \quad Je_4 = e_3, \\ Je_5 = 0, \quad Je_6 = e_5, \quad Je_7 = e_6, \\ Je_8 = 0, \quad Je_9 = e_8, \\ Je_{10} = 0, \quad Je_{11} = e_{10}, \\ Je_{12} = 0, \end{gathered}$$

with each line describing what happens for a single Jordan block. Let us think of the given nilpotent matrix $N$ as equal to $\begin{pmatrix} L \\ \Sigma\Sigma \end{pmatrix}$ for some linear map $L$. We want to find a new ordered basis $\Gamma = (v_1, \ldots, v_{12})$ in which the matrix of $L$ is $J$. In the expression $C^{-1}NC = J$, the matrix $C$ equals $\begin{pmatrix} I \\ \Sigma\Gamma \end{pmatrix}$, and its columns are expressions for $v_1, \ldots, v_{12}$ in the basis $\Sigma$, i.e., $Ce_i = v_i$. For each index $i$, we have $Je_i = Je_{i-1}$ or $Je_i = 0$. The formula $NC = CJ$, when applied to $e_i$, therefore says that

$$Nv_i = NCe_i = CJe_i = \begin{cases} Ce_{i-1} = v_{i-1} & \text{if } Je_i = e_{i-1}, \\ 0 & \text{if } Je_i = 0. \end{cases}$$

Thus we are looking for an ordered basis such that $N$ sends each member of the basis either into the previous member or into 0. The procedure in this example

will be to pick out $v_4$ as a vector not annihilated by $N^3$, obtain $v_3, v_2, v_1$, from it by successively applying $N$, pick out $v_7$ as a vector not annihilated by $N^2$ and independent of what has been found, obtain $v_6, v_5$ from it by successively applying $N$, and so on. It is necessary to check that the appropriate linear independence can be maintained, and that step will be what the proof is really about.

The proof of Theorem 5.21 will now be given in the general case. The core of the argument concerns linear maps and appears as three lemmas. Afterward the results of the lemmas will be interpreted in terms of matrices. For all the lemmas let $V$ be an $n$-dimensional vector space over $\mathbb{K}$, and let $N : V \to V$ be linear with $N^n = 0$. Define $K_j = \ker N^j$, so that

$$0 = K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_n = V.$$

**Lemma 5.23.** Suppose $j \geq 1$ and suppose $S_j$ is any vector subspace of $V$ such that $K_{j+1} = K_j \oplus S_j$. Then $N$ is one-one from $S_j$ into $K_j$ and $N(S_j) \cap K_{j-1} = 0$.

PROOF. Since $N(\ker N^{j+1}) \subseteq \ker N^j$, we obtain $N(S_j) \subseteq K_j$; thus $N$ indeed sends $S_j$ into $K_j$. To see that $N$ is one-one from $S_j$ into $K_j$, suppose that $s$ is a member of $S_j$ with $N(s) = 0$. Then $s$ is in $K_1$. Since $j \geq 1$, $K_1 \subseteq K_j$. Thus $s$ is in $K_j$. Since $K_j \cap S_j = 0$, $s$ is 0. Hence $N$ is one-one from $S_j$ into $K_j$. To see that $N(S_j) \cap K_{j-1} = 0$, suppose $s$ is a member of $S_j$ with $N(s)$ in $K_{j-1}$. Then $0 = N^{j-1}(N(s)) = N^j(s)$ shows that $s$ is in $K_j$. Since $K_j \cap S_j = 0$, $s$ equals 0. □

**Lemma 5.24.** Define $U_n = W_n = 0$. For $0 \leq j \leq n-1$, there exist vector subspaces $U_j$ and $W_j$ of $K_{j+1}$ such that

$$K_{j+1} = K_j \oplus U_j \oplus W_j,$$
$$U_j = N(U_{j+1} \oplus W_{j+1}),$$
and
$$N : U_{j+1} \oplus W_{j+1} \to U_j \quad \text{is one-one.}$$

PROOF. Define $U_{n-1} = N(U_n \oplus W_n) = 0$, and let $W_{n-1}$ be a vector subspace such that $V = K_n = K_{n-1} \oplus W_{n-1}$. Put $S_{n-1} = U_{n-1} \oplus W_{n-1}$. Proceeding inductively downward, suppose that $U_n, U_{n-1}, \ldots, U_{j+1}, W_n, W_{n-1}, \ldots, W_{j+1}$ have been defined so that $U_k = N(U_{k+1} \oplus W_{k+1})$, $N : U_{k+1} \oplus W_{k+1} \to U_k$ is one-one, and $K_{k+1} = K_k \oplus U_k \oplus W_k$ whenever $k$ satisfies $j < k \leq n-1$. We put $S_k = U_k \oplus W_k$ for these values of $k$, and then $S_k$ satisfies the hypothesis of Lemma 5.23 whenever $k$ satisfies $j < k \leq n-1$. We now construct $U_j$ and $W_j$. We put $U_j = N(S_{j+1})$. Since $S_{j+1}$ satisfies the hypothesis of Lemma 5.23, we see that $U_j \subseteq K_{j+1}$, $N$ is one-one from $S_{j+1}$ into $U_j$, and $U_j \cap K_j = 0$. Thus we can find a vector subspace $W_j$ with $K_{j+1} = K_j \oplus U_j \oplus W_j$, and the inductive construction is complete. □

**Lemma 5.25.** The vector subspaces of Lemma 5.24 satisfy

$$V = U_0 \oplus W_0 \oplus U_1 \oplus W_1 \oplus \cdots \oplus U_{n-1} \oplus W_{n-1}.$$

PROOF. Iterated use of Lemma 5.24 gives

$$\begin{aligned} V = K_n &= K_{n-1} \oplus (U_{n-1} \oplus W_{n-1}) \\ &= K_{n-2} \oplus (U_{n-2} \oplus W_{n-2}) \oplus (U_{n-1} \oplus W_{n-1}) \\ &= \cdots = K_0 \oplus (U_0 \oplus W_0) \oplus \cdots \oplus (U_{n-1} \oplus W_{n-1}) \\ &= (U_0 \oplus W_0) \oplus \cdots \oplus (U_{n-1} \oplus W_{n-1}), \end{aligned}$$

the last step holding since $K_0 = 0$, $K_0$ being the kernel of the identity function. □

PROOF OF THEOREM 5.21. We regard $N$ as acting on $V = \mathbb{K}^n$ by multiplication on the left, and we describe an ordered basis in which the matrix of $N$ is in Jordan form. For $0 \le j \le n-1$, form a basis of the vector subspace $W_j$ of Lemma 5.24, and let $v^{(j)}$ be a typical member of this basis. Each $v^{(j)}$ will be used as the last basis vector corresponding to a Jordan block of size $j+1$. The full ordered basis for that Jordan block will therefore be $N^j v^{(j)}, N^{j-1} v^{(j)}, \ldots, N v^{(j)}, v^{(j)}$. The theorem will be proved if we show that the union of these sets as $j$ and $v^{(j)}$ vary is a basis of $\mathbb{K}^n$ and that $N^{j+1} v^{(j)} = 0$ for all $j$ and $v^{(j)}$.

From the first conclusion of Lemma 5.24 we see for $j \ge 0$ that $W_j \subseteq K_{j+1}$, and hence $N^{j+1}(W_j) = 0$. Therefore $N^{j+1} v^{(j)} = 0$ for all $j$ and $v^{(j)}$.

Let us prove by induction downward on $j$ that a basis of $U_j \oplus W_j$ consists of all $v^{(j)}$ and all $N^k v^{(j+k)}$ for $k > 0$. The base case of the induction is $j = n-1$, and the statement holds in that case since $U_{n-1} = 0$ and since the vectors $v^{(n-1)}$ form a basis of $W_{n-1}$. The inductive hypothesis is that all $v^{(j+1)}$ and all $N^k v^{(j+1+k)}$ for $k > 0$ together form a basis of $U_{j+1} \oplus W_{j+1}$. The second and third conclusions of Lemma 5.24 together show that all $N v^{(j+1)}$ and all $N^{k+1} v^{(j+1+k)}$ for $k > 0$ together form a basis of $U_j$. In other words, all $N^k v^{(j+k)}$ with $k > 0$ together form a basis of $U_j$. The vectors $v^{(j)}$ by construction form a basis of $W_j$, and $U_j \cap W_j = 0$. Therefore the union of these separate bases is a basis for $U_j \oplus W_j$, and the induction is complete.

Taking the union of the bases of $U_j \oplus W_j$ for all $j$ and applying Lemma 5.25, we see that we have a basis of $V = \mathbb{K}^n$. This shows that the desired set is a basis of $\mathbb{K}^n$ and completes the proof of Theorem 5.21. □

## 7. Computations with Jordan Form

Let us illustrate the computation of Jordan form and the change-of-basis matrix with a few examples. We are given a matrix $A$ and we seek $J$ and $C$ with $J = C^{-1}AC$. We regard $A$ as the matrix of some linear $L$ in the standard ordered basis $\Sigma$, and we regard $J$ as the matrix of $L$ in some other ordered basis $\Gamma$. Then $C = \begin{pmatrix} I \\ \Sigma\Gamma \end{pmatrix}$, and so the columns of $C$ give the members of $\Gamma$ written as ordinary column vectors (in the standard ordered basis).

EXAMPLE 1. This example will be a nilpotent matrix, and we shall compute $J$ and $C$ merely by interpreting the proof of Theorem 5.21 in concrete terms. Let

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

The first step is to compute the characteristic polynomial, which is

$$\det(XI - A) = \det \begin{pmatrix} X+1 & -1 & 0 \\ 1 & X-1 & 0 \\ 1 & -1 & X \end{pmatrix} = X \det \begin{pmatrix} X+1 & -1 \\ 1 & X-1 \end{pmatrix} = X^3.$$

Then $A^3 = 0$ by the Cayley–Hamilton Theorem (Theorem 5.9), and $A$ is indeed nilpotent. The diagonal entries of $J$ are thus all 0, and we have to compute the sizes of the various Jordan blocks. To do so, we compute the dimension of the kernel of each power of $A$. The dimension of the kernel of a matrix equals the number of independent variables when we solve $AX = 0$ by row reduction. With the first power of $A$, the variable $x_1$ is dependent, and $x_2$ and $x_3$ are independent. Also, $A^2 = 0$. Thus

$$\dim(\ker A^0) = 0, \quad \dim(\ker A) = 2, \quad \text{and} \quad \dim(\ker A^2) = 3.$$

Hence

$$\#\{\text{Jordan blocks of size } \geq 1\} = \dim(\ker A) - \dim(\ker A^0) = 2 - 0 = 2,$$
$$\#\{\text{Jordan blocks of size } \geq 2\} = \dim(\ker A^2) - \dim(\ker A) = 3 - 2 = 1.$$

From these equalities we see that one Jordan block has size 2 and the other has size 1. Thus

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

We want to set up vector subspaces as in Lemma 5.24 so that $K_{j+1} = K_j \oplus U_j \oplus W_j$ and $U_j = A(U_{j+1} \oplus W_{j+1})$ for $0 \le j \le 2$. Since $K_3 = K_2$, the equations begin with $K_2 = \cdots$ and are

$$K_2 = K_1 \oplus 0 \oplus W_1, \quad U_0 = A(0 \oplus W_1), \quad K_1 = K_0 \oplus U_0 \oplus W_0.$$

Here $K_2 = \mathbb{K}^3$ and $K_1$ is the subspace of all $X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ such that $AX = 0$. The space $W_1$ is to satisfy $K_2 = K_1 \oplus W_1$, and we see that $W_1$ is 1-dimensional. Let $\{v^{(1)}\}$ be a basis of the 1-dimensional vector subspace $W_1$. Then $U_0$ is 1-dimensional with basis $\{Av^{(1)}\}$. The subspace $K_1$ is 2-dimensional and contains $U_0$. The space $W_0$ is to satisfy $K_1 = U_0 \oplus W_0$, and we see that $W_0$ is 1-dimensional. Let $\{v^{(0)}\}$ be a basis of $W_0$. Then the respective columns of $C$ may be taken to be

$$Av^{(1)}, \quad v^{(1)}, \quad v^{(0)}.$$

Let us compute these vectors.

If we extend a basis of $K_1$ to a basis of $K_2$, then $W_1$ may be taken to be the linear span of the added vector. To obtain a basis of $K_1$, we compute that the reduced row-echelon form of $A$ is $\begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, and the resulting system consists of the single equation $x_1 - x_2 = 0$. Thus $x_1 = x_2$, and

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

The coefficients of $x_2$ and $x_3$ on the right side form a basis of $K_1$, and we are to choose a vector that is not a linear combination of these. Thus we can take $v^{(1)} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ as the basis vector of $W_1$. Then $U_0 = A(W_1)$ has $Av^{(1)} = A\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}$ as a basis, and the basis of $W_0$ may be taken as any vector in $K_1$ but not $U_0$. We can take this basis to consist of $v^{(0)} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

Lining up our three basis vectors as the columns of $C$ gives us $C = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix}$. Computation gives $C^{-1} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & -1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$, and we readily check that $C^{-1}AC = J$.

EXAMPLE 2. We continue with $A$ and $J$ as in Example 1, but we compute the columns of $C$ without directly following the proof of Theorem 5.21. The method starts from the fact that each Jordan block corresponds to a 1-dimensional space of eigenvectors, and then we backtrack to find vectors corresponding to the other

columns. For this particular $A$, we know that the three columns of $C$ are to be of the form $v_1 = Av^{(1)}$, $v_2 = v^{(1)}$, and $v_3 = v^{(0)}$. The vectors $v_1$ and $v_3$ together span the 0 eigenspace of $A$. We find all the 0 eigenvectors, writing them as a two-parameter family. This eigenspace is just $K_1 = \ker A$, and we found in Example 1 that $K_1 = \left\{ \begin{pmatrix} x_2 \\ x_2 \\ x_3 \end{pmatrix} \right\}$. One of these vectors is to be $v_1$, and it has to equal $Av_2$. Thus we solve $Av_2 = \begin{pmatrix} x_2 \\ x_2 \\ x_3 \end{pmatrix}$. Applying the solution procedure yields

$$\left( \begin{array}{ccc|c} 1 & -1 & 0 & -x_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_3 - x_2 \end{array} \right).$$

This system has no solutions unless $x_3 - x_2 = 0$. If we take $x_2 = x_3 = -1$, then we obtain the same first two columns of $C$ as in Example 1, and any vector in $K_1$ independent of $\begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}$ may be taken as the third column.

EXAMPLE 3. Let

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

Direct calculation shows that the characteristic polynomial is $\det(XI - A) = X^3 - 8X^2 + 21X - 18 = (X-2)(X-3)^2$. The possibilities for $J$ are therefore

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix};$$

the first one will be correct if the dimension of the eigenspace for the eigenvalue 3 is 2, and the second one will be correct if that dimension is 1.

The third column of $C$ corresponds to an eigenvector for the eigenvalue 2, hence to a nonzero solution of $(A - 2I)v = 0$. The solutions are $v = k \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, and we can therefore use $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

For the first two columns of $C$, we have to find $\ker(A - 3I)$ no matter which of the methods we use, the one in Example 1 or the one in Example 2. Solving the system of equations, we obtain all vectors in the space $\left\{ z \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$. The dimension of the space is 1, and the second possibility for the Jordan form is the correct one.

Following the method of Example 1 to find the columns of $C$ means that we pick a basis of this kernel and extend it to a basis of $\ker(A - 3I)^2$. A basis of

$\ker(A-3I)$ consists of the vector $\begin{pmatrix}1\\1\\1\end{pmatrix}$. The matrix $(A-3I)^2$ is $\begin{pmatrix}0&0&0\\0&0&0\\0&-1&1\end{pmatrix}$, and the solution procedure leads to the formula

$$\begin{pmatrix}a\\b\\c\end{pmatrix} = a\begin{pmatrix}1\\0\\0\end{pmatrix} + c\begin{pmatrix}0\\1\\1\end{pmatrix}$$

for its kernel. The vector $\begin{pmatrix}1\\1\\1\end{pmatrix}$ arises from $a=1$ and $c=1$. We are to make an independent choice, say $a=1$ and $c=0$. Then the second basis vector to use is $\begin{pmatrix}1\\0\\0\end{pmatrix}$. This becomes the second column of $C$, and the first column then has to be $(A-3I)\begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}-1\\-1\\-1\end{pmatrix}$. The result is that $C = \begin{pmatrix}-1&1&0\\-1&0&0\\-1&0&1\end{pmatrix}$.

Following the method of Example 2 for this example means that we retain the entire kernel of $A-3I$, namely all vectors $v_1 = z\begin{pmatrix}1\\1\\1\end{pmatrix}$, as candidates for the first column of $C$. The second column is to satisfy $(A-3I)v_2 = v_1$. Solving leads to $v_2 = z\begin{pmatrix}-1\\0\\0\end{pmatrix} + c\begin{pmatrix}1\\1\\1\end{pmatrix}$. In contrast to Example 2, there is no potential contradictory equation. So we choose $z$ and then $c$. If we take $z=1$ and $c=0$, we find that the first two columns of $C$ are to be $\begin{pmatrix}1\\1\\1\end{pmatrix}$ and $\begin{pmatrix}-1\\0\\0\end{pmatrix}$. Then $C = \begin{pmatrix}1&-1&0\\1&0&0\\1&0&1\end{pmatrix}$.

For any example in which we can factor the characteristic polynomial exactly, either of the two methods used above will work. The first method appears complicated but uses numbers throughout; it tends to be more efficient with large examples involving high-degree minimal polynomials. The second method appears direct but requires solving equations with symbolic variables; it tends to be more efficient for relatively simple examples.

## 8. Problems

In Problems 1–25 all vector spaces are assumed finite-dimensional, and all linear transformations are assumed defined from such spaces into themselves. Unless information is given to the contrary, the underlying field $\mathbb{K}$ is assumed arbitrary.

1. Let $M_{mn}(\mathbb{C})$ be the vector space of $m$-by-$n$ complex matrices. The group $\mathrm{GL}(m,\mathbb{C}) \times \mathrm{GL}(n,\mathbb{C})$ acts on $M_{mn}(\mathbb{C})$ by $((g,h),x) \mapsto gxh^{-1}$, where $gxh^{-1}$ denotes a matrix product. Do the following:

(a) Verify that this is indeed a group action.
(b) Prove that two members of $M_{mn}(\mathbb{C})$ lie in the same orbit if and only if they have the same rank.
(c) For each possible rank, give an example of a member of $M_{mn}(\mathbb{C})$ with that rank.

2. Prove that a member of $M_n(\mathbb{K})$ is invertible if and only if the constant term of its minimal polynomial is different from 0.

3. Suppose that $L : V \to V$ is a linear map with minimal polynomial $M(X) = P_1(X)^{l_1} \cdots P_k(X)^{l_k}$ and that $V = U \oplus W$ with $U$ and $W$ both invariant under $L$. Let $P_1(X)^{r_1} \cdots P_k(X)^{r_k}$ and $P_1(X)^{s_1} \cdots P_k(X)^{s_k}$ be the respective minimal polynomials of $L\big|_U$ and $L\big|_W$. Prove that $l_j = \max(r_j, s_j)$ for $1 \le j \le k$.

4. (a) If $A$ and $B$ are in $M_n(\mathbb{K})$, if $P(X)$ is a polynomial such that $P(AB) = 0$, and if $Q(X) = XP(X)$, prove that $Q(BA) = 0$.
(b) What can be inferred from (a) about the relationship between the minimal polynomials of $AB$ and of $BA$?

5. (a) Suppose that $D$ and $D'$ are in $M_n(\mathbb{K})$, are similar to diagonal matrices, and have $DD' = D'D$. Prove that there is a matrix $C$ such that $C^{-1}DC$ and $C^{-1}D'C$ are both diagonal.
(b) Give an example of two nilpotent matrices $N$ and $N'$ in $M_n(\mathbb{K})$ with $NN' = N'N$ such that there is no $C$ with $C^{-1}NC$ and $C^{-1}N'C$ both in Jordan form.

6. (a) Prove that the matrix of a projection is similar to a diagonal matrix. What are the eigenvalues?
(b) Give a necessary and sufficient condition for two projections involving the same $V$ to be given by similar matrices.

7. Let $E : V \to V$ and $F : V \to V$ be projections. Prove that $E$ and $F$ have
(a) the same image if and only if $EF = F$ and $FE = E$,
(b) the same kernel if and only if $EF = E$ and $FE = F$.

8. Let $E : V \to V$ and $F : V \to V$ be projections. Prove that $EF$ is a projection if $EF = FE$. Prove or disprove a converse.

9. An **involution** on $V$ is a linear map $U : V \to V$ such that $U^2 = I$. Show that the equation $U = 2E - 1$ establishes a one-one correspondence between all projections $E$ and all involutions $U$.

9A. Explain how the proof of the converse half of Theorem 5.14 greatly simplifies once the Primary Decomposition Theorem (Theorem 5.19) is available.

10. Let $L : V \to V$ be linear. Prove that there exist vector subspaces $U$ and $W$ of $V$ such that
    (i) $V = U \oplus W$,
    (ii) $L(U) \subseteq U$ and $L(W) \subseteq W$,
    (iii) $L$ is nilpotent on $U$,
    (iv) $L$ is nonsingular on $W$.

11. Prove that the vector subspaces $U$ and $W$ in the previous problem are uniquely characterized by (i) through (iv).

12. (Special case of **Jordan–Chevalley decomposition**) Let $L : V \to V$ be a linear map, and suppose that its minimal polynomial is of the form $M(X) = \prod_{j=1}^{k} (X-\lambda_j)^{l_j}$ with the $\lambda_j$ distinct. Let $V = U_1 \oplus \cdots \oplus U_k$ be the corresponding primary decomposition of $V$, and define $D : V \to V$ by $D = \lambda_1 E_1 + \cdots + \lambda_k E_k$, where $E_1, \ldots, E_k$ are the projections associated with the primary decomposition. Finally put $N = L - D$. Prove that
    (a) $L = D + N$,
    (b) $D$ has a basis of eigenvectors,
    (c) $N$ is nilpotent, i.e., has $N^{\dim V} = 0$,
    (d) $DN = ND$.
    (e) $D$ and $N$ are given by unique polynomials in $L$ such that each of the polynomials is equal to 0 or has degree less than the degree of $M(X)$,
    (f) the minimal polynomial of $D$ is $\prod_{j=1}^{k} (X - \lambda_j)$,
    (g) the minimal polynomial of $N$ is $X^{\max l_j}$.

13. (Special case of **Jordan–Chevalley decomposition,** continued) In the previous problem with $L$ given, prove that a decomposition $L = D + N$ is uniquely determined by properties (a) through (d). Avoid using (e) in the argument.

14. (a) Let $N'$ be a nilpotent square matrix of size $n'$. Prove for arbitrary $c \in \mathbb{K}$ that the characteristic polynomial of $N' + cI$ is $(X - c)^{n'}$, and deduce that the only eigenvalue of $N' + cI$ is $c$.
    (b) Let $L = D + N$ be the decomposition in Problems 12 and 13 of a square matrix $L$ of size $n$. Prove that $L$ and $D$ have the same characteristic polynomial.

15. For the complex matrix $A = \begin{pmatrix} -5 & 9 \\ -4 & 7 \end{pmatrix}$, find a Jordan-form matrix $J$ and an invertible matrix $C$ such that $J = C^{-1}AC$.

16. For the complex matrix $A = \begin{pmatrix} 4 & 1 & -1 \\ -8 & -2 & 2 \\ 8 & 2 & -2 \end{pmatrix}$, find a Jordan-form matrix $J$ and an invertible matrix $C$ such that $J = C^{-1}AC$.

17. For the upper triangular matrix

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

find a Jordan-form matrix $J$ and an invertible matrix $C$ such that $J = C^{-1}AC$.

18. (a) For $M_3(\mathbb{C})$, prove that any two matrices with the same minimal polynomial and the same characteristic polynomial must be similar.
    (b) Is the same thing true for $M_4(\mathbb{C})$?
19. Suppose that $\mathbb{K}$ has characteristic 0 and that $J$ is a Jordan block with nonzero eigenvalue and with size $> 1$. Prove that there is no $n \geq 1$ such that $J^n$ is diagonal.
20. Classify up to similarity all members $A$ of $M_n(\mathbb{C})$ with $A^n = I$.
21. How many similarity classes are there of 3-by-3 matrices $A$ with entries in $\mathbb{C}$ such that $A^3 = A$? Explain.
22. Let $n \geq 2$, and let $N$ be a member of $M_n(\mathbb{K})$ with $N^n = 0$ but $N^{n-1} \neq 0$. Prove that there is no $n$-by-$n$ matrix $A$ with $A^2 = N$.
23. For a Jordan block $J$, prove that $J^t$ is similar to $J$.
24. Prove that if $A$ is in $M_n(\mathbb{C})$, then $A^t$ is similar to $A$.
25. Let $N$ be the 2-by-2 matrix $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, and let $A$ and $B$ be the 4-by-4 matrices $A = \begin{pmatrix} N & 0 \\ 0 & N \end{pmatrix}$ and $B = \begin{pmatrix} N & N \\ 0 & N \end{pmatrix}$. Prove that $A$ and $B$ are similar.

Problems 26–31 concern cyclic vectors. Fix a linear map $L : V \to V$ from a finite-dimensional vector space $V$ to itself. For $v$ in $V$, let $\mathcal{P}(v)$ denote the set of all vectors $Q(L)(v)$ in $V$ for $Q(X)$ in $\mathbb{K}[X]$; $\mathcal{P}(v)$ is a vector subspace and is invariant under $L$. If $U$ is an invariant subspace of $V$, we say that $U$ is a **cyclic subspace** if there is some $v$ in $U$ such that $\mathcal{P}(v) = U$; in this case, $v$ is said to be a **cyclic vector** for $U$, and $U$ is called the **cyclic subspace generated** by $v$. For $v$ in $V$, let $\mathcal{I}_v$ be the ideal of all polynomials $Q(X)$ in $\mathbb{K}[X]$ with $Q(L)v = 0$. The **monic generator** of $v$ is the unique monic polynomial $M_v(X)$ such that $M_v(X)$ divides every member of $\mathcal{I}_v$.

26. For $v \in V$, explain why $\mathcal{I}_v$ is nonzero and why $M_v(X)$ therefore exists.
27. For $v \in V$, prove that
    (a) the degree of the monic generator $M_v(X)$ equals the dimension of the cyclic subspace $\mathcal{P}(v)$,
    (b) the vectors $v, L(v), L^2(v), \ldots, L^{\deg M_v - 1}(v)$ form a vector-space basis of $\mathcal{P}(v)$,

(c) the minimal and characteristic polynomials of $L\big|_{\mathcal{P}(v)}$ are both equal to $M_v(X)$.

28. Suppose that $M_v(X) = c_0 + c_1X + \cdots + c_{d-1}X^{d-1} + X^d$. Prove that the matrix of $L\big|_{\mathcal{P}(v)}$ in a suitable ordered basis is

$$\begin{pmatrix} -c_{d-1} & 1 & 0 & \cdots & & \\ -c_{d-2} & 0 & 1 & & & \\ -c_{d-3} & 0 & 0 & & & \\ \vdots & & & \ddots & \ddots & \vdots \\ -c_2 & 0 & 0 & \cdots & 0 & 1\ 0 \\ -c_1 & 0 & 0 & \cdots & 0 & 0\ 1 \\ -c_0 & 0 & 0 & \cdots & & 0\ 0 \end{pmatrix}.$$

29. Suppose that $v$ is in $V$, that $M_v(X)$ is a power of a prime polynomial $P(X)$, and that $Q(X)$ is a nonzero polynomial with $\deg Q(X) < \deg P(X)$. Prove that $\mathcal{P}(Q(L)(v)) = \mathcal{P}(v)$.

30. Let $P(X)$ be a prime polynomial.
    (a) Prove by induction on $\dim V$ that if the minimal polynomial of $L$ is $P(X)$, then the characteristic polynomial of $L$ is a power of $P(X)$.
    (b) Prove by induction on $l$ that if the minimal polynomial of $L$ is $P(X)^l$, then the characteristic polynomial of $L$ is a power of $P(X)$.
    (c) Conclude that if the minimal polynomial of $L$ is a power of $P(X)$, then $\deg P(X)$ divides $\dim V$.

31. Prove that every prime factor of the characteristic polynomial of $L$ divides the minimal polynomial of $L$.

Problems 32–40 continue the study of cyclic vectors begun in Problems 26–31, using the same notation. The goal is to obtain a canonical-form theorem like Theorem 5.20 for $L$ but with no assumption on $\mathbb{K}$ or $P(X)$, namely that each primary subspace for $L$ is the direct sum of cyclic subspaces and the resulting decomposition is unique up to isomorphism. This result and the Fundamental Theorem of Finitely Generated Abelian Groups (Theorem 4.56) will be seen in Chapter VIII to be special cases of a single more general theorem. Still another canonical form for matrices and linear maps is an analog of the result with elementary divisors mentioned in the remarks with Theorem 4.56 and is valid here; it is called **rational canonical form**, but we shall not pursue it until the problems at the end of Chapter VIII. The proof in Problems 32–40 uses ideas similar to those used for Theorem 5.21 except that the hypothesis will now be that the minimal polynomial of $L$ is $P(X)^l$ with $P(X)$ prime, rather than just $X^l$. Define $K_j = \ker(P(L)^j)$ for $j \geq 0$, so that $K_0 = 0$, $K_j \subseteq K_{j+1}$ for all $j$, $K_l = V$, and each $K_j$ is an invariant subspace under $L$. Define $d = \deg P(X)$.

32. Suppose $j \geq 1$, and suppose $S_j$ is any vector subspace of $V$ such that $K_{j+1} = K_j \oplus S_j$. Prove that $P(L)$ is one-one from $S_j$ into $K_j$ and $P(L)(S_j) \cap K_{j-1} = 0$.

33. Define $U_l = W_l = 0$. For $0 \le j \le l-1$, prove that there exist vector subspaces $U_j$ and $W_j$ of $K_{j+1}$ such that
$$K_{j+1} = K_j \oplus U_j \oplus W_j,$$
$$U_j = P(L)(U_{j+1} \oplus W_{j+1}),$$
$$P(L) : U_{j+1} \oplus W_{j+1} \to U_j \quad \text{is one-one.}$$

34. Prove that the vector subspaces of the previous problem satisfy
$$V = U_0 \oplus W_0 \oplus U_1 \oplus W_1 \oplus \cdots \oplus U_{l-1} \oplus W_{l-1}.$$

35. For $v \neq 0$ in $W_j$, prove that the set of all $L^r P(L)^s(v)$ with $0 \le r \le d-1$ and $0 \le s \le j$ is a vector-space basis of $\mathcal{P}(v)$.

36. Going back over the construction in Problem 33, prove that each $W_j$ can be chosen to have a basis consisting of vectors $L^r(v_i^{(j)})$ for $1 \le i \le (\dim W_j)/d$ and $0 \le r \le d-1$.

37. Let the index $i$ used in the previous problem with $j$ be denoted by $i_j$ for $1 \le i_j \le (\dim W_j)/d$. Prove that a vector-space basis of $U_j \oplus W_j$ consists of all $L^r P(L)^k(v_{i_{j+k}}^{(j+k)})$ for $0 \le r \le d-1,\ k \ge 0,\ 1 \le i_{j+k} \le (\dim W_{j+k})/d$.

38. Prove that $V$ is the direct sum of cyclic subspaces under $L$. Prove specifically that each $v_{i_j}^{(j)}$ generates a cyclic subspace and that the sum of all these vector subspaces, with $0 \le j \le l$ and $1 \le i_j \le (\dim W_j)/d$, is a direct sum and equals $V$.

39. In the decomposition of the previous problem, each cyclic subspace generated by some $v_{i_j}^{(j)}$ has minimal polynomial $P(X)^{j+1}$. Prove that
$$\#\left\{\begin{matrix}\text{direct summands with minimal polynomial} \\ P(X)^k \text{ for some } k \ge j+1\end{matrix}\right\} = (\dim K_{j+1} - \dim K_j)/d.$$

40. Prove that the formula of the previous problem persists for any decomposition of $V$ as the direct sum of cyclic subspaces, and conclude from Problem 28 that the decomposition into cyclic subspaces is unique up to isomorphism.

Problems 41–46 concern systems of ordinary differential equations with constant coefficients. The underlying field is taken to be $\mathbb{C}$, and differential calculus is used. For $A$ in $M_n(\mathbb{C})$ and $t$ in $\mathbb{R}$, define $e^{tA} = \sum_{k=0}^{\infty} \frac{t^k A^k}{k!}$. Take for granted that the series defining $e^{tA}$ converges entry by entry, that the series may be differentiated term by term to yield $\frac{d}{dt}(e^{tA}) = Ae^{tA} = e^{tA}A$, and that $e^{sA+tB} = e^{sA}e^{tB}$ if $A$ and $B$ commute.

41. Calculate $e^{tA}$ for $A$ equal to
(a) $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$,

(b) $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$,

(c) the diagonal matrix with diagonal entries $d_1, \ldots, d_n$.

42. (a) Calculate $e^{tJ}$ when $J$ is a nilpotent $n$-by-$n$ Jordan block.

(b) Use (a) to calculate $e^{tJ}$ when $J$ is a general $n$-by-$n$ Jordan block.

43. Let $y_1, \ldots, y_n$ be unknown functions from $\mathbb{R}$ to $\mathbb{C}$, and let $y$ be the vector-valued function formed by arranging $y_1, \ldots, y_n$ in a column. Suppose that $A$ is in $M_n(\mathbb{C})$. Prove for each vector $v \in \mathbb{C}^n$ that $y(t) = e^{tA}v$ is a solution of the system of differential equations $\frac{dy}{dt} = Ay(t)$.

44. With notation as in the previous problem and with $v$ fixed in $\mathbb{C}^n$, use $e^{-tA}y(t)$ to show, for each open interval of $t$'s containing 0, that the only solution of $\frac{dy}{dt} = Ay(t)$ on that interval such that $y(0) = v$ is $y(t) = e^{tA}v$.

45. For $C$ invertible, prove that $e^{tC^{-1}AC} = C^{-1}e^{tA}C$, and deduce a relationship between solutions of $\frac{dy}{dt} = Ay(t)$ and solutions of $\frac{dy}{dt} = (C^{-1}AC)y(t)$.

46. Let $A = \begin{pmatrix} 2 & 1 & 0 \\ -1 & 4 & 0 \\ -1 & 2 & 2 \end{pmatrix}$. Taking into account Example 3 in Section 7 and Problems 42 through 45 above, find all solutions for $t$ in $(-1, 1)$ to the system $\frac{dy}{dt} = Ay(t)$ such that $y(0) = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.