LINEAR ALGEBRA

Second Edition

KENNETH HOFFMAN
Professor of Mathematics
Massachusetts Institute of Technology

RAY KUNZE
Professor of Mathematics
University of California, Irvine

PRENTICE-HALL, INC., Englewood Cliffs, New Jersey

PRENTICE-HALL INTERNATIONAL, INC., *London*
PRENTICE-HALL OF AUSTRALIA, PTY. LTD., *Sydney*
PRENTICE-HALL OF CANADA, LTD., *Toronto*
PRENTICE-HALL OF INDIA PRIVATE LIMITED, *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*

Current printing (last digit):

10 9 8 7 6

Library of Congress Catalog Card No. 75-142120
Printed in the United States of America

Preface

Our original purpose in writing this book was to provide a text for the undergraduate linear algebra course at the Massachusetts Institute of Technology. This course was designed for mathematics majors at the junior level, although three-fourths of the students were drawn from other scientific and technological disciplines and ranged from freshmen through graduate students. This description of the M.I.T. audience for the text remains generally accurate today. The ten years since the first edition have seen the proliferation of linear algebra courses throughout the country and have afforded one of the authors the opportunity to teach the basic material to a variety of groups at Brandeis University, Washington University (St. Louis), and the University of California (Irvine).

Our principal aim in revising *Linear Algebra* has been to increase the variety of courses which can easily be taught from it. On one hand, we have structured the chapters, especially the more difficult ones, so that there are several natural stopping points along the way, allowing the instructor in a one-quarter or one-semester course to exercise a considerable amount of choice in the subject matter. On the other hand, we have increased the amount of material in the text, so that it can be used for a rather comprehensive one-year course in linear algebra and even as a reference book for mathematicians.

The major changes have been in our treatments of canonical forms and inner product spaces. In Chapter 6 we no longer begin with the general spatial theory which underlies the theory of canonical forms. We first handle characteristic values in relation to triangulation and diagonalization theorems and then build our way up to the general theory. We have split Chapter 8 so that the basic material on inner product spaces and unitary diagonalization is followed by a Chapter 9 which treats sesqui-linear forms and the more sophisticated properties of normal operators, including normal operators on real inner product spaces.

We have also made a number of small changes and improvements from the first edition. But the basic philosophy behind the text is unchanged.

We have made no particular concession to the fact that the majority of the students may not be primarily interested in mathematics. For we believe a mathematics course should not give science, engineering, or social science students a hodgepodge of techniques, but should provide them with an understanding of basic mathematical concepts.

On the other hand, we have been keenly aware of the wide range of backgrounds which the students may possess and, in particular, of the fact that the students have had very little experience with abstract mathematical reasoning. For this reason, we have avoided the introduction of too many abstract ideas at the very beginning of the book. In addition, we have included an Appendix which presents such basic ideas as set, function, and equivalence relation. We have found it most profitable not to dwell on these ideas independently, but to advise the students to read the Appendix when these ideas arise.

Throughout the book we have included a great variety of examples of the important concepts which occur. The study of such examples is of fundamental importance and tends to minimize the number of students who can repeat definition, theorem, proof in logical order without grasping the meaning of the abstract concepts. The book also contains a wide variety of graded exercises (about six hundred), ranging from routine applications to ones which will extend the very best students. These exercises are intended to be an important part of the text.

Chapter 1 deals with systems of linear equations and their solution by means of elementary row operations on matrices. It has been our practice to spend about six lectures on this material. It provides the student with some picture of the origins of linear algebra and with the computational technique necessary to understand examples of the more abstract ideas occurring in the later chapters. Chapter 2 deals with vector spaces, subspaces, bases, and dimension. Chapter 3 treats linear transformations, their algebra, their representation by matrices, as well as isomorphism, linear functionals, and dual spaces. Chapter 4 defines the algebra of polynomials over a field, the ideals in that algebra, and the prime factorization of a polynomial. It also deals with roots, Taylor's formula, and the Lagrange interpolation formula. Chapter 5 develops determinants of square matrices, the determinant being viewed as an alternating n-linear function of the rows of a matrix, and then proceeds to multilinear functions on modules as well as the Grassman ring. The material on modules places the concept of determinant in a wider and more comprehensive setting than is usually found in elementary textbooks. Chapters 6 and 7 contain a discussion of the concepts which are basic to the analysis of a single linear transformation on a finite-dimensional vector space; the analysis of characteristic (eigen) values, triangulable and diagonalizable transformations; the concepts of the diagonalizable and nilpotent parts of a more general transformation, and the rational and Jordan canonical forms. The primary and cyclic decomposition theorems play a central role, the latter being arrived at through the study of admissible subspaces. Chapter 7 includes a discussion of matrices over a polynomial domain, the computation of invariant factors and elementary divisors of a matrix, and the development of the Smith canonical form. The chapter ends with a discussion of semi-simple operators, to round out the analysis of a single operator. Chapter 8 treats finite-dimensional inner product spaces in some detail. It covers the basic geometry, relating orthogonalization to the idea of 'best approximation to a vector' and leading to the concepts of the orthogonal projection of a vector onto a subspace and the orthogonal complement of a subspace. The chapter treats unitary operators and culminates in the diagonalization of self-adjoint and normal operators. Chapter 9 introduces sesqui-linear forms, relates them to positive and self-adjoint operators on an inner product space, moves on to the spectral theory of normal operators and then to more sophisticated results concerning normal operators on real or complex inner product spaces. Chapter 10 discusses bilinear forms, emphasizing canonical forms for symmetric and skew-symmetric forms, as well as groups preserving non-degenerate forms, especially the orthogonal, unitary, pseudo-orthogonal and Lorentz groups.

We feel that any course which uses this text should cover Chapters 1, 2, and 3 thoroughly, possibly excluding Sections 3.6 and 3.7 which deal with the double dual and the transpose of a linear transformation. Chapters 4 and 5, on polynomials and determinants, may be treated with varying degrees of thoroughness. In fact, polynomial ideals and basic properties of determinants may be covered quite sketchily without serious damage to the flow of the logic in the text; however, our inclination is to deal with these chapters carefully (except the results on modules), because the material illustrates so well the basic ideas of linear algebra. An elementary course may now be concluded nicely with the first four sections of Chapter 6, together with (the new) Chapter 8. If the rational and Jordan forms are to be included, a more extensive coverage of Chapter 6 is necessary.

Our indebtedness remains to those who contributed to the first edition, especially to Professors Harry Furstenberg, Louis Howard, Daniel Kan, Edward Thorp, to Mrs. Judith Bowers, Mrs. Betty Ann (Sargent) Rose and Miss Phyllis Ruby. In addition, we would like to thank the many students and colleagues whose perceptive comments led to this revision, and the staff of Prentice-Hall for their patience in dealing with two authors caught in the throes of academic administration. Lastly, special thanks are due to Mrs. Sophia Koulouras for both her skill and her tireless efforts in typing the revised manuscript.

K. M. H. / R. A. K.

Contents

2. Let a, b, and c be elements of a field F, and let A be the following 3×3 matrix over F:

$$A = \begin{bmatrix} 0 & 0 & c \\ 1 & 0 & b \\ 0 & 1 & a \end{bmatrix}.$$

Prove that the characteristic polynomial for A is $x^3 - ax^2 - bx - c$ and that this is also the minimal polynomial for A.

3. Let A be the 4×4 real matrix

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

Show that the characteristic polynomial for A is $x^2(x - 1)^2$ and that it is also the minimal polynomial.

4. Is the matrix A of Exercise 3 similar over the field of complex numbers to a diagonal matrix?

5. Let V be an n-dimensional vector space and let T be a linear operator on V. Suppose that there exists some positive integer k so that $T^k = 0$. Prove that $T^n = 0$.

6. Find a 3×3 matrix for which the minimal polynomial is x^2.

7. Let n be a positive integer, and let V be the space of polynomials over R which have degree at most n (throw in the 0-polynomial). Let D be the differentiation operator on V. What is the minimal polynomial for D?

8. Let P be the operator on R^2 which projects each vector onto the x-axis, parallel to the y-axis: $P(x, y) = (x, 0)$. Show that P is linear. What is the minimal polynomial for P?

9. Let A be an $n \times n$ matrix with characteristic polynomial

$$f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}.$$

Show that

$$c_1d_1 + \cdots + c_kd_k = \text{trace } (A).$$

10. Let V be the vector space of $n \times n$ matrices over the field F. Let A be a fixed $n \times n$ matrix. Let T be the linear operator on V defined by

$$T(B) = AB.$$

Show that the minimal polynomial for T is the minimal polynomial for A.

11. Let A and B be $n \times n$ matrices over the field F. According to Exercise 9 of Section 6.1, the matrices AB and BA have the same characteristic values. Do they have the same characteristic polynomial? Do they have the same minimal polynomial?

6.4. Invariant Subspaces

In this section, we shall introduce a few concepts which are useful in attempting to analyze a linear operator. We shall use these ideas to obtain characterizations of diagonalizable (and triangulable) operators in terms of their minimal polynomials.

Definition. *Let* V *be a vector space and* T *a linear operator on* V. *If* W *is a subspace of* V, *we say that* W *is* **invariant under** T *if for each vector* α *in* W *the vector* Tα *is in* W, *i.e., if* T(W) *is contained in* W.

EXAMPLE 6. If T is any linear operator on V, then V is invariant under T, as is the zero subspace. The range of T and the null space of T are also invariant under T.

EXAMPLE 7. Let F be a field and let D be the differentiation operator on the space $F[x]$ of polynomials over F. Let n be a positive integer and let W be the subspace of polynomials of degree not greater than n. Then W is invariant under D. This is just another way of saying that D is 'degree decreasing.'

EXAMPLE 8. Here is a very useful generalization of Example 6. Let T be a linear operator on V. Let U be any linear operator on V which commutes with T, i.e., $TU = UT$. Let W be the range of U and let N be the null space of U. Both W and N are invariant under T. If α is in the range of U, say $\alpha = U\beta$, then $T\alpha = T(U\beta) = U(T\beta)$ so that $T\alpha$ is in the range of U. If α is in N, then $U(T\alpha) = T(U\alpha) = T(0) = 0$; hence, $T\alpha$ is in N.

A particular type of operator which commutes with T is an operator $U = g(T)$, where g is a polynomial. For instance, we might have $U = T - cI$, where c is a characteristic value of T. The null space of U is familiar to us. We see that this example includes the (obvious) fact that the space of characteristic vectors of T associated with the characteristic value c is invariant under T.

EXAMPLE 9. Let T be the linear operator on R^2 which is represented in the standard ordered basis by the matrix

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

Then the only subspaces of R^2 which are invariant under T are R^2 and the zero subspace. Any other invariant subspace would necessarily have dimension 1. But, if W is the subspace spanned by some non-zero vector α, the fact that W is invariant under T means that α is a characteristic vector, but A has no real characteristic values.

When the subspace W is invariant under the operator T, then T induces a linear operator T_W on the space W. The linear operator T_W is defined by $T_W(\alpha) = T(\alpha)$, for α in W, but T_W is quite a different object from T since its domain is W not V.

When V is finite-dimensional, the invariance of W under T has a simple matrix interpretation, and perhaps we should mention it at this point. Suppose we choose an ordered basis $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ for V such that $\mathcal{B}' = \{\alpha_1, \ldots, \alpha_r\}$ is an ordered basis for W ($r = \dim W$). Let $A = [T]_{\mathcal{B}}$ so that

$$T\alpha_j = \sum_{i=1}^{n} A_{ij}\alpha_i.$$

Since W is invariant under T, the vector $T\alpha_j$ belongs to W for $j \leq r$. This means that

(6-9) $$T\alpha_j = \sum_{i=1}^{r} A_{ij}\alpha_i, \qquad j \leq r.$$

In other words, $A_{ij} = 0$ if $j \leq r$ and $i > r$.

Schematically, A has the block form

(6-10) $$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$$

where B is an $r \times r$ matrix, C is an $r \times (n - r)$ matrix, and D is an $(n - r) \times (n - r)$ matrix. The reader should note that according to (6-9) the matrix B is precisely the matrix of the induced operator T_W in the ordered basis $\mathcal{B}'$.

Most often, we shall carry out arguments about T and T_W without making use of the block form of the matrix A in (6-10). But we should note how certain relations between T_W and T are apparent from that block form.

Lemma. *Let* W *be an invariant subspace for* T. *The characteristic polynomial for the restriction operator* T_W *divides the characteristic polynomial for* T. *The minimal polynomial for* T_W *divides the minimal polynomial for* T.

Proof. We have

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$$

where $A = [T]_{\mathcal{B}}$ and $B = [T_W]_{\mathcal{B}'}$. Because of the block form of the matrix

$$\det (xI - A) = \det (xI - B) \det (xI - D).$$

That proves the statement about characteristic polynomials. Notice that we used I to represent identity matrices of three different sizes.

The kth power of the matrix A has the block form

$$A^k = \begin{bmatrix} B^k & C_k \\ 0 & D^k \end{bmatrix}$$

where C_k is some $r \times (n - r)$ matrix. Therefore, any polynomial which annihilates A also annihilates B (and D too). So, the minimal polynomial for B divides the minimal polynomial for A. ∎

EXAMPLE 10. Let T be any linear operator on a finite-dimensional space V. Let W be the subspace spanned by *all* of the characteristic vectors of T. Let $c_1, \ldots, c_k$ be the distinct characteristic values of T. For each i, let W_i be the space of characteristic vectors associated with the characteristic value c_i, and let $\mathcal{B}_i$ be an ordered basis for W_i. The lemma before Theorem 2 tells us that $\mathcal{B}' = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$ is an ordered basis for W. In particular,

$$\dim W = \dim W_1 + \cdots + \dim W_k.$$

Let $\mathcal{B}' = \{\alpha_1, \ldots, \alpha_r\}$ so that the first few α's form the basis $\mathcal{B}_1$, the next few $\mathcal{B}_2$, and so on. Then

$$T\alpha_i = t_i\alpha_i, \qquad i = 1, \ldots, r$$

where $(t_1, \ldots, t_r) = (c_1, c_1, \ldots, c_1, \ldots, c_k, c_k, \ldots, c_k)$ with c_i repeated $\dim W_i$ times.

Now W is invariant under T, since for each α in W we have

$$\begin{aligned} \alpha &= x_1\alpha_1 + \cdots + x_r\alpha_r \\ T\alpha &= t_1x_1\alpha_1 + \cdots + t_rx_r\alpha_r. \end{aligned}$$

Choose any other vectors $\alpha_{r+1}, \ldots, \alpha_n$ in V such that $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ is a basis for V. The matrix of T relative to $\mathcal{B}$ has the block form (6-10), and the matrix of the restriction operator T_W relative to the basis $\mathcal{B}'$ is

$$B = \begin{bmatrix} t_1 & 0 & \cdots & 0 \\ 0 & t_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & t_r \end{bmatrix}.$$

The characteristic polynomial of B (i.e., of T_W) is

$$g = (x - c_1)^{e_1} \cdots (x - c_k)^{e_k}$$

where $e_i = \dim W_i$. Furthermore, g divides f, the characteristic polynomial for T. Therefore, the multiplicity of c_i as a root of f is at least $\dim W_i$.

All of this should make Theorem 2 transparent. It merely says that T is diagonalizable if and only if $r = n$, if and only if $e_1 + \cdots + e_k = n$. It does not help us too much with the non-diagonalizable case, since we don't know the matrices C and D of (6-10).

Definition. *Let W be an invariant subspace for T and let α be a vector in V. The* **T-conductor of α into** W *is the set* $S_T(\alpha; W)$, *which consists of all polynomials* g *(over the scalar field) such that* g(T)α *is in* W.

Since the operator T will be fixed throughout most discussions, we shall usually drop the subscript T and write $S(\alpha; W)$. The authors usually call that collection of polynomials the 'stuffer' (*das einstopfende Ideal*). 'Conductor' is the more standard term, preferred by those who envision a less aggressive operator $g(T)$, gently leading the vector α into W. In the special case $W = \{0\}$ the conductor is called the ***T*-annihilator of α.**

Lemma. *If* W *is an invariant subspace for* T, *then* W *is invariant under every polynomial in* T. *Thus, for each* α *in* V, *the conductor* $S(\alpha; W)$ *is an ideal in the polynomial algebra* F[x].

Proof. If β is in W, then $T\beta$ is in W. Consequently, $T(T\beta) = T^2\beta$ is in W. By induction, $T^k\beta$ is in W for each k. Take linear combinations to see that $f(T)\beta$ is in W for every polynomial f.

The definition of $S(\alpha; W)$ makes sense if W is any subset of V. If W is a subspace, then $S(\alpha; W)$ is a subspace of $F[x]$, because

$$(cf + g)(T) = cf(T) + g(T).$$

If W is also invariant under T, let g be a polynomial in $S(\alpha; W)$, i.e., let $g(T)\alpha$ be in W. If f is any polynomial, then $f(T)[g(T)\alpha]$ will be in W. Since

$$(fg)(T) = f(T)g(T)$$

fg is in $S(\alpha; W)$. Thus the conductor absorbs multiplication by any polynomial. ▮

The unique monic generator of the ideal $S(\alpha; W)$ is also called the **T-conductor of α into W** (the **T-annihilator** in case $W = \{0\}$). The T-conductor of α into W is the monic polynomial g of least degree such that $g(T)\alpha$ is in W. A polynomial f is in $S(\alpha; W)$ if and only if g divides f. Note that the conductor $S(\alpha; W)$ always contains the minimal polynomial for T; hence, *every* T-*conductor divides the minimal polynomial for* T.

As the first illustration of how to use the conductor $S(\alpha; W)$, we shall characterize triangulable operators. The linear operator T is called **triangulable** if there is an ordered basis in which T is represented by a triangular matrix.

Lemma. *Let* V *be a finite-dimensional vector space over the field* F. *Let* T *be a linear operator on* V *such that the minimal polynomial for* T *is a product of linear factors*

$$p = (x - c_1)^{r_1} \cdots (x - c_k)^{r_k}, \qquad c_i \textit{ in } F.$$

Let W *be a proper* (W $\neq$ V) *subspace of* V *which is invariant under* T. *There exists a vector* α *in* V *such that*

(a) α *is not in* W;
(b) $(T - cI)\alpha$ *is in* W, *for some characteristic value* c *of the operator* T.

Proof. What (a) and (b) say is that the T-conductor of α into W is a linear polynomial. Let β be any vector in V which is not in W. Let g be the T-conductor of β into W. Then g divides p, the minimal polynomial for T. Since β is not in W, the polynomial g is not constant. Therefore,

$$g = (x - c_1)^{e_1} \cdots (x - c_k)^{e_k}$$

where at least one of the integers e_i is positive. Choose j so that $e_j > 0$. Then $(x - c_j)$ divides g:

$$g = (x - c_j)h.$$

By the definition of g, the vector $\alpha = h(T)\beta$ cannot be in W. But

$$\begin{aligned}(T - c_jI)\alpha &= (T - c_jI)h(T)\beta \\ &= g(T)\beta\end{aligned}$$

is in W. ∎

Theorem 5. *Let V be a finite-dimensional vector space over the field F and let T be a linear operator on V. Then T is triangulable if and only if the minimal polynomial for T is a product of linear polynomials over F.*

Proof. Suppose that the minimal polynomial factors

$$p = (x - c_1)^{r_1} \cdots (x - c_k)^{r_k}.$$

By repeated application of the lemma above, we shall arrive at an ordered basis $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ in which the matrix representing T is upper-triangular:

$$(6\text{-}11) \qquad [T]_{\mathcal{B}} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}.$$

Now (6-11) merely says that

$$(6\text{-}12) \qquad T\alpha_j = a_{1j}\alpha_1 + \cdots + a_{jj}\alpha_j, \qquad 1 \leq j \leq n$$

that is, $T\alpha_j$ is in the subspace spanned by $\alpha_1, \ldots, \alpha_j$. To find $\alpha_1, \ldots, \alpha_n$, we start by applying the lemma to the subspace $W = \{0\}$, to obtain the vector α_1. Then apply the lemma to W_1, the space spanned by α_1, and we obtain α_2. Next apply the lemma to W_2, the space spanned by α_1 and α_2. Continue in that way. One point deserves comment. After $\alpha_1, \ldots, \alpha_i$ have been found, it is the triangular-type relations (6-12) for $j = 1, \ldots, i$ which ensure that the subspace spanned by $\alpha_1, \ldots, \alpha_i$ is invariant under T.

If T is triangulable, it is evident that the characteristic polynomial for T has the form

$$f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}, \qquad c_i \text{ in } F.$$

Just look at the triangular matrix (6-11). The diagonal entries $a_{11}, \ldots, a_{1n}$ are the characteristic values, with c_i repeated d_i times. But, if f can be so factored, so can the minimal polynomial p, because it divides f. ∎

Corollary. *Let F be an algebraically closed field, e.g., the complex number field. Every n × n matrix over F is similar over F to a triangular matrix.*

Theorem 6. *Let V be a finite-dimensional vector space over the field F and let T be a linear operator on V. Then T is diagonalizable if and only if the minimal polynomial for T has the form*

$$p = (x - c_1) \cdots (x - c_k)$$

where $c_1, \ldots, c_k$ *are distinct elements of F.*

Proof. We have noted earlier that, if T is diagonalizable, its minimal polynomial is a product of distinct linear factors (see the discussion prior to Example 4). To prove the converse, let W be the subspace spanned by all of the characteristic vectors of T, and suppose $W \neq V$. By the lemma used in the proof of Theorem 5, there is a vector α not in W and a characteristic value c_j of T such that the vector

$$\beta = (T - c_jI)\alpha$$

lies in W. Since β is in W,

$$\beta = \beta_1 + \cdots + \beta_k$$

where $T\beta_i = c_i\beta_i$, $1 \leq i \leq k$, and therefore the vector

$$h(T)\beta = h(c_1)\beta_1 + \cdots + h(c_k)\beta_k$$

is in W, for every polynomial h.

Now $p = (x - c_j)q$, for some polynomial q. Also

$$q - q(c_j) = (x - c_j)h.$$

We have

$$q(T)\alpha - q(c_j)\alpha = h(T)(T - c_jI)\alpha = h(T)\beta.$$

But $h(T)\beta$ is in W and, since

$$0 = p(T)\alpha = (T - c_jI)q(T)\alpha$$

the vector $q(T)\alpha$ is in W. Therefore, $q(c_j)\alpha$ is in W. Since α is not in W, we have $q(c_j) = 0$. That contradicts the fact that p has distinct roots. ∎

At the end of Section 6.7, we shall give a different proof of Theorem 6. In addition to being an elegant result, Theorem 6 is useful in a computational way. Suppose we have a linear operator T, represented by the matrix A in some ordered basis, and we wish to know if T is diagonalizable. We compute the characteristic polynomial f. *If* we can factor f:

$$f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}$$

we have two different methods for determining whether or not T is diagonalizable. One method is to see whether (for each i) we can find d_i independent characteristic vectors associated with the characteristic value c_i. The other method is to check whether or not $(T - c_1I) \cdots (T - c_kI)$ is the zero operator.

Theorem 5 provides a different proof of the Cayley-Hamilton theorem. That theorem is easy for a triangular matrix. Hence, via Theorem 5, we obtain the result for any matrix over an algebraically closed field. Any field is a subfield of an algebraically closed field. If one knows that result, one obtains a proof of the Cayley-Hamilton theorem for matrices over any field. If we at least admit into our discussion the Fundamental Theorem of Algebra (the complex number field is algebraically closed), then Theorem 5 provides a proof of the Cayley-Hamilton theorem for complex matrices, and that proof is independent of the one which we gave earlier.

Exercises

1. Let T be the linear operator on R^2, the matrix of which in the standard ordered basis is

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

(a) Prove that the only subspaces of R^2 invariant under T are R^2 and the zero subspace.

(b) If U is the linear operator on C^2, the matrix of which in the standard ordered basis is A, show that U has 1-dimensional invariant subspaces.

2. Let W be an invariant subspace for T. Prove that the minimal polynomial for the restriction operator T_W divides the minimal polynomial for T, without referring to matrices.

3. Let c be a characteristic value of T and let W be the space of characteristic vectors associated with the characteristic value c. What is the restriction operator T_W?

4. Let

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

Is A similar over the field of real numbers to a triangular matrix? If so, find such a triangular matrix.

5. Every matrix A such that $A^2 = A$ is similar to a diagonal matrix.

6. Let T be a diagonalizable linear operator on the n-dimensional vector space V, and let W be a subspace which is invariant under T. Prove that the restriction operator T_W is diagonalizable.

7. Let T be a linear operator on a finite-dimensional vector space over the field of complex numbers. Prove that T is diagonalizable if and only if T is annihilated by some polynomial over C which has distinct roots.

8. Let T be a linear operator on V. If every subspace of V is invariant under T, then T is a scalar multiple of the identity operator.

9. Let T be the indefinite integral operator

$$(Tf)(x) = \int_0^x f(t)\, dt$$

on the space of continuous functions on the interval [0, 1]. Is the space of polynomial functions invariant under T? The space of differentiable functions? The space of functions which vanish at $x = \frac{1}{2}$?

10. Let A be a 3×3 matrix with real entries. Prove that, if A is not similar over R to a triangular matrix, then A is similar over C to a diagonal matrix.

11. True or false? If the triangular matrix A is similar to a diagonal matrix, then A is already diagonal.

12. Let T be a linear operator on a finite-dimensional vector space over an algebraically closed field F. Let f be a polynomial over F. Prove that c is a characteristic value of $f(T)$ if and only if $c = f(t)$, where t is a characteristic value of T.

13. Let V be the space of $n \times n$ matrices over F. Let A be a fixed $n \times n$ matrix over F. Let T and U be the linear operators on V defined by

$$T(B) = AB$$
$$U(B) = AB - BA.$$

(a) True or false? If A is diagonalizable (over F), then T is diagonalizable.
(b) True or false? If A is diagonalizable, then U is diagonalizable.

6.5. Simultaneous Triangulation; Simultaneous Diagonalization

Let V be a finite-dimensional space and let $\mathfrak{F}$ be a family of linear operators on V. We ask when we can simultaneously triangulate or diagonalize the operators in $\mathfrak{F}$, i.e., find one basis $\mathfrak{B}$ such that all of the matrices $[T]_{\mathfrak{B}}$, T in $\mathfrak{F}$, are triangular (or diagonal). In the case of diagonalization, it is necessary that $\mathfrak{F}$ be a commuting family of operators: $UT = TU$ for all T, U in $\mathfrak{F}$. That follows from the fact that all diagonal matrices commute. Of course, it is also necessary that each operator in $\mathfrak{F}$ be a diagonalizable operator. In order to simultaneously triangulate, each operator in $\mathfrak{F}$ must be triangulable. It is not necessary that $\mathfrak{F}$ be a commuting family; however, that condition is sufficient for simultaneous triangulation (if each T can be individually triangulated). These results follow from minor variations of the proofs of Theorems 5 and 6.

The subspace W is **invariant under** (the family of operators) $\mathfrak{F}$ if W is invariant under each operator in $\mathfrak{F}$.

Lemma. *Let $\mathfrak{F}$ be a commuting family of triangulable linear operators on V. Let W be a proper subspace of V which is invariant under $\mathfrak{F}$. There exists a vector α in V such that*

(a) *α is not in W;*
(b) *for each T in $\mathfrak{F}$, the vector $T\alpha$ is in the subspace spanned by α and W.*

Proof. It is no loss of generality to assume that $\mathfrak{F}$ contains only a finite number of operators, because of this observation. Let $\{T_1, \ldots, T_r\}$

be a maximal linearly independent subset of $\mathfrak{F}$, i.e., a basis for the subspace spanned by $\mathfrak{F}$. If α is a vector such that (b) holds for each T_i, then (b) will hold for every operator which is a linear combination of $T_1, \ldots, T_r$.

By the lemma before Theorem 5 (this lemma for a single operator), we can find a vector β_1 (not in W) and a scalar c_1 such that $(T_1 - c_1I)\beta_1$ is in W. Let V_1 be the collection of all vectors β in V such that $(T_1 - c_1I)\beta$ is in W. Then V_1 is a subspace of V which is properly larger than W. Furthermore, V_1 is invariant under $\mathfrak{F}$, for this reason. If T commutes with T_1, then

$$(T_1 - c_1I)(T\beta) = T(T_1 - c_1I)\beta.$$

If β is in V_1, then $(T_1 - c_1I)\beta$ is in W. Since W is invariant under each T in $\mathfrak{F}$, we have $T(T_1 - c_1I)\beta$ in W, i.e., $T\beta$ in V_1, for all β in V_1 and all T in $\mathfrak{F}$.

Now W is a proper subspace of V_1. Let U_2 be the linear operator on V_1 obtained by restricting T_2 to the subspace V_1. The minimal polynomial for U_2 divides the minimal polynomial for T_2. Therefore, we may apply the lemma before Theorem 5 to that operator and the invariant subspace W. We obtain a vector β_2 in V_1 (not in W) and a scalar c_2 such that $(T_2 - c_2I)\beta_2$ is in W. Note that

(a) β_2 is not in W;
(b) $(T_1 - c_1I)\beta_2$ is in W;
(c) $(T_2 - c_2I)\beta_2$ is in W.

Let V_2 be the set of all vectors β in V_1 such that $(T_2 - c_2I)\beta$ is in W. Then V_2 is invariant under $\mathfrak{F}$. Apply the lemma before Theorem 5 to U_3, the restriction of T_3 to V_2. If we continue in this way, we shall reach a vector $\alpha = \beta_r$ (not in W) such that $(T_j - c_jI)\alpha$ is in W, $j = 1, \ldots, r$. ∎

Theorem 7. *Let V be a finite-dimensional vector space over the field F. Let $\mathfrak{F}$ be a commuting family of triangulable linear operators on V. There exists an ordered basis for V such that every operator in $\mathfrak{F}$ is represented by a triangular matrix in that basis.*

Proof. Given the lemma which we just proved, this theorem has the same proof as does Theorem 5, if one replaces T by $\mathfrak{F}$. ∎

Corollary. *Let $\mathfrak{F}$ be a commuting family of n × n matrices over an algebraically closed field F. There exists a non-singular n × n matrix P with entries in F such that $P^{-1}AP$ is upper-triangular, for every matrix A in $\mathfrak{F}$.*

Theorem 8. *Let $\mathfrak{F}$ be a commuting family of diagonalizable linear operators on the finite-dimensional vector space V. There exists an ordered basis for V such that every operator in $\mathfrak{F}$ is represented in that basis by a diagonal matrix.*

Proof. We could prove this theorem by adapting the lemma before Theorem 7 to the diagonalizable case, just as we adapted the lemma before Theorem 5 to the diagonalizable case in order to prove Theorem 6. However, at this point it is easier to proceed by induction on the dimension of V.

If $\dim V = 1$, there is nothing to prove. Assume the theorem for vector spaces of dimension less than n, and let V be an n-dimensional space. Choose any T in $\mathfrak{F}$ which is not a scalar multiple of the identity. Let $c_1, \ldots, c_k$ be the distinct characteristic values of T, and (for each i) let W_i be the null space of $T - c_iI$. Fix an index i. Then W_i is invariant under every operator which commutes with T. Let $\mathfrak{F}_i$ be the family of linear operators on W_i obtained by restricting the operators in $\mathfrak{F}$ to the (invariant) subspace W_i. Each operator in $\mathfrak{F}_i$ is diagonalizable, because its minimal polynomial divides the minimal polynomial for the corresponding operator in $\mathfrak{F}$. Since $\dim W_i < \dim V$, the operators in $\mathfrak{F}_i$ can be simultaneously diagonalized. In other words, W_i has a basis $\mathfrak{B}_i$ which consists of vectors which are simultaneously characteristic vectors for every operator in $\mathfrak{F}_i$.

Since T is diagonalizable, the lemma before Theorem 2 tells us that $\mathfrak{B} = (\mathfrak{B}_1, \ldots, \mathfrak{B}_k)$ is a basis for V. That is the basis we seek. ∎

Exercises

1. Find an invertible real matrix P such that $P^{-1}AP$ and $P^{-1}BP$ are both diagonal, where A and B are the real matrices

(a) $$A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & -8 \\ 0 & -1 \end{bmatrix}$$

(b) $$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}.$$

2. Let $\mathfrak{F}$ be a commuting family of 3×3 complex matrices. How many linearly independent matrices can $\mathfrak{F}$ contain? What about the $n \times n$ case?

3. Let T be a linear operator on an n-dimensional space, and suppose that T has n distinct characteristic values. Prove that any linear operator which commutes with T is a polynomial in T.

4. Let A, B, C, and D be $n \times n$ complex matrices which commute. Let E be the $2n \times 2n$ matrix

$$E = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

Prove that $\det E = \det (AD - BC)$.

5. Let F be a field, n a positive integer, and let V be the space of $n \times n$ matrices over F. If A is a fixed $n \times n$ matrix over F, let T_A be the linear operator on V defined by $T_A(B) = AB - BA$. Consider the family of linear operators T_A obtained by letting A vary over all diagonal matrices. Prove that the operators in that family are simultaneously diagonalizable.

6.6. Direct-Sum Decompositions

As we continue with our analysis of a single linear operator, we shall formulate our ideas in a slightly more sophisticated way—less in terms of matrices and more in terms of subspaces. When we began this chapter, we described our goal this way: To find an ordered basis in which the matrix of T assumes an especially simple form. Now, we shall describe our goal as follows: To decompose the underlying space V into a sum of invariant subspaces for T such that the restriction operators on those subspaces are simple.

Definition. *Let* $W_1, \ldots, W_k$ *be subspaces of the vector space* V. *We say that* $W_1, \ldots, W_k$ *are* **independent** *if*

$$\alpha_1 + \cdots + \alpha_k = 0, \qquad \alpha_i \text{ in } W_i$$

implies that each α_i *is* 0.

For $k = 2$, the meaning of independence is $\{0\}$ intersection, i.e., W_1 and W_2 are independent if and only if $W_1 \cap W_2 = \{0\}$. If $k > 2$, the independence of $W_1, \ldots, W_k$ says much more than $W_1 \cap \cdots \cap W_k = \{0\}$. It says that each W_j intersects the sum of the other subspaces W_i only in the zero vector.

The significance of independence is this. Let $W = W_1 + \cdots + W_k$ be the subspace spanned by $W_1, \ldots, W_k$. Each vector α in W can be expressed as a sum

$$\alpha = \alpha_1 + \cdots + \alpha_k, \qquad \alpha_i \text{ in } W_i.$$

If $W_1, \ldots, W_k$ are independent, then that expression for α is unique; for if

$$\alpha = \beta_1 + \cdots + \beta_k, \qquad \beta_i \text{ in } W_i$$

then $0 = (\alpha_1 - \beta_1) + \cdots + (\alpha_k - \beta_k)$, hence $\alpha_i - \beta_i = 0$, $i = 1, \ldots, k$. Thus, when $W_1, \ldots, W_k$ are independent, we can operate with the vectors in W as k-tuples $(\alpha_1, \ldots, \alpha_k)$, α_i in W_i, in the same way as we operate with vectors in R^k as k-tuples of numbers.

Lemma. *Let* V *be a finite-dimensional vector space. Let* $W_1, \ldots, W_k$ *be subspaces of* V *and let* $W = W_1 + \cdots + W_k$. *The following are equivalent.*

(a) $W_1, \ldots, W_k$ *are independent.*

(b) *For each* j, $2 \leq j \leq k$, *we have*

$$W_j \cap (W_1 + \cdots + W_{j-1}) = \{0\}.$$

(c) *If* $\mathcal{B}_i$ *is an ordered basis for* W_i, $1 \leq i \leq k$, *then the sequence* $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$ *is an ordered basis for* W.

Proof. Assume (a). Let α be a vector in the intersection $W_j \cap (W_1 + \cdots + W_{j-1})$. Then there are vectors $\alpha_1, \ldots, \alpha_{j-1}$ with α_i in W_i such that $\alpha = \alpha_1 + \cdots + \alpha_{j-1}$. Since

$$\alpha_1 + \cdots + \alpha_{j-1} + (-\alpha) + 0 + \cdots + 0 = 0$$

and since $W_1, \ldots, W_k$ are independent, it must be that $\alpha_1 = \alpha_2 = \cdots = \alpha_{j-1} = \alpha = 0$.

Now, let us observe that (b) implies (a). Suppose

$$0 = \alpha_1 + \cdots + \alpha_k, \qquad \alpha_i \text{ in } W_i.$$

Let j be the largest integer i such that $\alpha_i \neq 0$. Then

$$0 = \alpha_1 + \cdots + \alpha_j, \qquad \alpha_j \neq 0.$$

Thus $\alpha_j = -\alpha_1 - \cdots - \alpha_{j-1}$ is a non-zero vector in $W_j \cap (W_1 + \cdots + W_{j-1})$.

Now that we know (a) and (b) are the same, let us see why (a) is equivalent to (c). Assume (a). Let $\mathcal{B}_i$ be a basis for W_i, $1 \leq i \leq k$, and let $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$. Any linear relation between the vectors in $\mathcal{B}$ will have the form

$$\beta_1 + \cdots + \beta_k = 0$$

where β_i is some linear combination of the vectors in $\mathcal{B}_i$. Since $W_1, \ldots, W_k$ are independent, each β_i is 0. Since each $\mathcal{B}_i$ is independent, the relation we have between the vectors in $\mathcal{B}$ is the trivial relation.

We relegate the proof that (c) implies (a) to the exercises (Exercise 2). ∎

If any (and hence all) of the conditions of the last lemma hold, we say that the sum $W = W_1 + \cdots + W_k$ is **direct** or that W is the **direct sum** of $W_1, \ldots, W_k$ and we write

$$W = W_1 \oplus \cdots \oplus W_k.$$

In the literature, the reader may find this direct sum referred to as an independent sum or the interior direct sum of $W_1, \ldots, W_k$.

EXAMPLE 11. Let V be a finite-dimensional vector space over the field F and let $\{\alpha_1, \ldots, \alpha_n\}$ be any basis for V. If W_i is the one-dimensional subspace spanned by α_i, then $V = W_1 \oplus \cdots \oplus W_n$.

EXAMPLE 12. Let n be a positive integer and F a subfield of the complex numbers, and let V be the space of all $n \times n$ matrices over F. Let W_1 be the subspace of all **symmetric** matrices, i.e., matrices A such that $A^t = A$. Let W_2 be the subspace of all **skew-symmetric** matrices, i.e., matrices A such that $A^t = -A$. Then $V = W_1 \oplus W_2$. If A is any matrix in V, the unique expression for A as a sum of matrices, one in W_1 and the other in W_2, is

$$\begin{aligned} A &= A_1 + A_2 \\ A_1 &= \tfrac{1}{2}(A + A^t) \\ A_2 &= \tfrac{1}{2}(A - A^t). \end{aligned}$$

EXAMPLE 13. Let T be any linear operator on a finite-dimensional space V. Let $c_1, \dots, c_k$ be the distinct characteristic values of T, and let W_i be the space of characteristic vectors associated with the characteristic value c_i. Then $W_1, \dots, W_k$ are independent. See the lemma before Theorem 2. In particular, if T is diagonalizable, then $V = W_1 \oplus \cdots \oplus W_k$.

Definition. *If V is a vector space, a* **projection** *of V is a linear operator E on V such that* $E^2 = E$.

Suppose that E is a projection. Let R be the range of E and let N be the null space of E.

1. The vector β is in the range R if and only if $E\beta = \beta$. If $\beta = E\alpha$, then $E\beta = E^2\alpha = E\alpha = \beta$. Conversely, if $\beta = E\beta$, then (of course) β is in the range of E.
2. $V = R \oplus N$.
3. The unique expression for α as a sum of vectors in R and N is $\alpha = E\alpha + (\alpha - E\alpha)$.

From (1), (2), (3) it is easy to see the following. If R and N are subspaces of V such that $V = R \oplus N$, there is one and only one projection operator E which has range R and null space N. That operator is called the **projection on R along N**.

Any projection E is (trivially) diagonalizable. If $\{\alpha_1, \dots, \alpha_r\}$ is a basis for R and $\{\alpha_{r+1}, \dots, \alpha_n\}$ a basis for N, then the basis $\mathcal{B} = \{\alpha_1, \dots, \alpha_n\}$ diagonalizes E:

$$[E]_{\mathcal{B}} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

where I is the $r \times r$ identity matrix. That should help explain some of the terminology connected with projections. The reader should look at various cases in the plane R^2 (or 3-space, R^3), to convince himself that the projection on R along N sends each vector into R by projecting it parallel to N.

Projections can be used to describe direct-sum decompositions of the space V. For, suppose $V = W_1 \oplus \cdots \oplus W_k$. For each j we shall define an operator E_j on V. Let α be in V, say $\alpha = \alpha_1 + \cdots + \alpha_k$ with α_i in W_i. Define $E_j\alpha = \alpha_j$. Then E_j is a well-defined rule. It is easy to see that E_j is linear, that the range of E_j is W_j, and that $E_j^2 = E_j$. The null space of E_j is the subspace

$$(W_1 + \cdots + W_{j-1} + W_{j+1} + \cdots + W_k)$$

for, the statement that $E_j\alpha = 0$ simply means $\alpha_j = 0$, i.e., that α is actually a sum of vectors from the spaces W_i with $i \neq j$. In terms of the projections E_j we have

$$\alpha = E_1\alpha + \cdots + E_k\alpha \tag{6-13}$$

for each α in V. What (6-13) says is that

$$I = E_1 + \cdots + E_k.$$

Note also that if $i \neq j$, then $E_iE_j = 0$, because the range of E_j is the subspace W_j which is contained in the null space of E_i. We shall now summarize our findings and state and prove a converse.

Theorem 9. *If* $V = W_1 \oplus \cdots \oplus W_k$, *then there exist* k *linear operators* $E_1, \ldots, E_k$ *on* V *such that*

(i) *each* E_i *is a projection* ($E_i^2 = E_i$);
(ii) $E_iE_j = 0$, *if* $i \neq j$;
(iii) $I = E_1 + \cdots + E_k$;
(iv) *the range of* E_i *is* W_i.

Conversely, if $E_1, \ldots, E_k$ *are* k *linear operators on* V *which satisfy conditions* (i), (ii), *and* (iii), *and if we let* W_i *be the range of* E_i, *then* $V = W_1 \oplus \cdots \oplus W_k$.

Proof. We have only to prove the converse statement. Suppose $E_1, \ldots, E_k$ are linear operators on V which satisfy the first three conditions, and let W_i be the range of E_i. Then certainly

$$V = W_1 + \cdots + W_k;$$

for, by condition (iii) we have

$$\alpha = E_1\alpha + \cdots + E_k\alpha$$

for each α in V, and $E_i\alpha$ is in W_i. This expression for α is unique, because if

$$\alpha = \alpha_1 + \cdots + \alpha_k$$

with α_i in W_i, say $\alpha_i = E_i\beta_i$, then using (i) and (ii) we have

$$\begin{aligned} E_j\alpha &= \sum_{i=1}^{k} E_j\alpha_i \\ &= \sum_{i=1}^{k} E_jE_i\beta_i \\ &= E_j^2\beta_j \\ &= E_j\beta_j \\ &= \alpha_j. \end{aligned}$$

This shows that V is the direct sum of the W_i. ▌

Exercises

1. Let V be a finite-dimensional vector space and let W_1 be any subspace of V. Prove that there is a subspace W_2 of V such that $V = W_1 \oplus W_2$.

2. Let V be a finite-dimensional vector space and let $W_1, \ldots, W_k$ be subspaces of V such that

$$V = W_1 + \cdots + W_k \quad \text{and} \quad \dim V = \dim W_1 + \cdots + \dim W_k.$$

Prove that $V = W_1 \oplus \cdots \oplus W_k$.

3. Find a projection E which projects R^2 onto the subspace spanned by $(1, -1)$ along the subspace spanned by $(1, 2)$.

4. If E_1 and E_2 are projections onto independent subspaces, then $E_1 + E_2$ is a projection. True or false?

5. If E is a projection and f is a polynomial, then $f(E) = aI + bE$. What are a and b in terms of the coefficients of f?

6. True or false? If a diagonalizable operator has only the characteristic values 0 and 1, it is a projection.

7. Prove that if E is the projection on R along N, then $(I - E)$ is the projection on N along R.

8. Let $E_1, \ldots, E_k$ be linear operators on the space V such that $E_1 + \cdots + E_k = I$.

(a) Prove that if $E_iE_j = 0$ for $i \neq j$, then $E_i^2 = E_i$ for each i.

(b) In the case $k = 2$, prove the converse of (a). That is, if $E_1 + E_2 = I$ and $E_1^2 = E_1$, $E_2^2 = E_2$, then $E_1E_2 = 0$.

9. Let V be a real vector space and E an idempotent linear operator on V, i.e., a projection. Prove that $(I + E)$ is invertible. Find $(I + E)^{-1}$.

10. Let F be a subfield of the complex numbers (or, a field of characteristic zero). Let V be a finite-dimensional vector space over F. Suppose that $E_1, \ldots, E_k$ are projections of V and that $E_1 + \cdots + E_k = I$. Prove that $E_iE_j = 0$ for $i \neq j$ (*Hint:* Use the trace function and ask yourself what the trace of a projection is.)

11. Let V be a vector space, let $W_1, \ldots, W_k$ be subspaces of V, and let

$$V_j = W_1 + \cdots + W_{j-1} + W_{j+1} + \cdots + W_k.$$

Suppose that $V = W_1 \oplus \cdots \oplus W_k$. Prove that the dual space V^* has the direct-sum decomposition $V^* = V_1^0 \oplus \cdots \oplus V_k^0$.

6.7. Invariant Direct Sums

We are primarily interested in direct-sum decompositions $V = W_1 \oplus \cdots \oplus W_k$, where each of the subspaces W_i is invariant under some given linear operator T. Given such a decomposition of V, T induces a linear operator T_i on each W_i by restriction. The action of T is then this.

If α is a vector in V, we have unique vectors $\alpha_1, \ldots, \alpha_k$ with α_i in W_i such that

$$\alpha = \alpha_1 + \cdots + \alpha_k$$

and then

$$T\alpha = T_1\alpha_1 + \cdots + T_k\alpha_k.$$

We shall describe this situation by saying that T is the **direct sum** of the operators $T_1, \ldots, T_k$. It must be remembered in using this terminology that the T_i are not linear operators on the space V but on the various subspaces W_i. The fact that $V = W_1 \oplus \cdots \oplus W_k$ enables us to associate with each α in V a unique k-tuple $(\alpha_1, \ldots, \alpha_k)$ of vectors α_i in W_i (by $\alpha = \alpha_1 + \cdots + \alpha_k$) in such a way that we can carry out the linear operations in V by working in the individual subspaces W_i. The fact that each W_i is invariant under T enables us to view the action of T as the independent action of the operators T_i on the subspaces W_i. Our purpose is to study T by finding invariant direct-sum decompositions in which the T_i are operators of an elementary nature.

Before looking at an example, let us note the matrix analogue of this situation. Suppose we select an ordered basis $\mathcal{B}_i$ for each W_i, and let $\mathcal{B}$ be the ordered basis for V consisting of the union of the $\mathcal{B}_i$ arranged in the order $\mathcal{B}_1, \ldots, \mathcal{B}_k$, so that $\mathcal{B}$ is a basis for V. From our discussion concerning the matrix analogue for a single invariant subspace, it is easy to see that if $A = [T]_{\mathcal{B}}$ and $A_i = [T_i]_{\mathcal{B}_i}$, then A has the block form

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & A_k \end{bmatrix}. \tag{6-14}$$

In (6-14), A_i is a $d_i \times d_i$ matrix ($d_i = \dim W_i$), and the 0's are symbols for rectangular blocks of scalar 0's of various sizes. It also seems appropriate to describe (6-14) by saying that A is the **direct sum** of the matrices $A_1, \ldots, A_k$.

Most often, we shall describe the subspace W_i by means of the associated projections E_i (Theorem 9). Therefore, we need to be able to phrase the invariance of the subspaces W_i in terms of the E_i.

Theorem 10. *Let* T *be a linear operator on the space* V, *and let* $W_1, \ldots, W_k$ *and* $E_1, \ldots, E_k$ *be as in Theorem 9. Then a necessary and sufficient condition that each subspace* W_i *be invariant under* T *is that* T *commute with each of the projections* E_i, *i.e.,*

$$TE_i = E_iT, \qquad i = 1, \ldots, k.$$

Proof. Suppose T commutes with each E_i. Let α be in W_j. Then $E_j\alpha = \alpha$, and

$$\begin{aligned} T\alpha &= T(E_j\alpha) \\ &= E_j(T\alpha) \end{aligned}$$

which shows that $T\alpha$ is in the range of E_j, i.e., that W_j is invariant under T.

Assume now that each W_i is invariant under T. We shall show that $TE_j = E_jT$. Let α be any vector in V. Then

$$\begin{aligned} \alpha &= E_1\alpha + \cdots + E_k\alpha \\ T\alpha &= TE_1\alpha + \cdots + TE_k\alpha. \end{aligned}$$

Since $E_i\alpha$ is in W_i, which is invariant under T, we must have $T(E_i\alpha) = E_i\beta_i$ for some vector β_i. Then

$$\begin{aligned} E_jTE_i\alpha &= E_jE_i\beta_i \\ &= \begin{cases} 0, & \text{if} \quad i \neq j \\ E_j\beta_j, & \text{if} \quad i = j. \end{cases} \end{aligned}$$

Thus

$$\begin{aligned} E_jT\alpha &= E_jTE_1\alpha + \cdots + E_jTE_k\alpha \\ &= E_j\beta_j \\ &= TE_j\alpha. \end{aligned}$$

This holds for each α in V, so $E_jT = TE_j$. ▮

We shall now describe a diagonalizable operator T in the language of invariant direct sum decompositions (projections which commute with T). This will be a great help to us in understanding some deeper decomposition theorems later. The reader may feel that the description which we are about to give is rather complicated, in comparison to the matrix formulation or to the simple statement that the characteristic vectors of T span the underlying space. But, he should bear in mind that this is our first glimpse at a very effective method, by means of which various problems concerned with subspaces, bases, matrices, and the like can be reduced to algebraic calculations with linear operators. With a little experience, the efficiency and elegance of this method of reasoning should become apparent.

Theorem 11. *Let* T *be a linear operator on a finite-dimensional space* V. *If* T *is diagonalizable and if* $c_1, \ldots, c_k$ *are the distinct characteristic values of* T, *then there exist linear operators* $E_1, \ldots, E_k$ *on* V *such that*

(i) $T = c_1E_1 + \cdots + c_kE_k$;
(ii) $I = E_1 + \cdots + E_k$;
(iii) $E_iE_j = 0,\ i \neq j$;
(iv) $E_i^2 = E_i$ (E_i *is a projection*);
(v) *the range of* E_i *is the characteristic space for* T *associated with* c_i.

Conversely, if there exist k *distinct scalars* $c_1, \ldots, c_k$ *and* k *non-zero linear operators* $E_1, \ldots, E_k$ *which satisfy conditions* (i), (ii), *and* (iii), *then* T *is diagonalizable,* $c_1, \ldots, c_k$ *are the distinct characteristic values of* T, *and conditions* (iv) *and* (v) *are satisfied also.*

Proof. Suppose that T is diagonalizable, with distinct characteristic values $c_1, \ldots, c_k$. Let W_i be the space of characteristic vectors associated with the characteristic value c_i. As we have seen,

$$V = W_1 \oplus \cdots \oplus W_k.$$

Let $E_1, \ldots, E_k$ be the projections associated with this decomposition, as in Theorem 9. Then (ii), (iii), (iv) and (v) are satisfied. To verify (i), proceed as follows. For each α in V,

$$\alpha = E_1\alpha + \cdots + E_k\alpha$$

and so

$$\begin{aligned} T\alpha &= TE_1\alpha + \cdots + TE_k\alpha \\ &= c_1E_1\alpha + \cdots + c_kE_k\alpha. \end{aligned}$$

In other words, $T = c_1E_1 + \cdots + c_kE_k$.

Now suppose that we are given a linear operator T along with distinct scalars c_i and non-zero operators E_i which satisfy (i), (ii) and (iii). Since $E_iE_j = 0$ when $i \neq j$, we multiply both sides of $I = E_1 + \cdots + E_k$ by E_i and obtain immediately $E_i^2 = E_i$. Multiplying $T = c_1E_1 + \cdots + c_kE_k$ by E_i, we then have $TE_i = c_iE_i$, which shows that any vector in the range of E_i is in the null space of $(T - c_iI)$. Since we have assumed that $E_i \neq 0$, this proves that there is a non-zero vector in the null space of $(T - c_iI)$, i.e., that c_i is a characteristic value of T. Furthermore, the c_i are all of the characteristic values of T; for, if c is any scalar, then

$$T - cI = (c_1 - c)E_1 + \cdots + (c_k - c)E_k$$

so if $(T - cI)\alpha = 0$, we must have $(c_i - c)E_i\alpha = 0$. If α is not the zero vector, then $E_i\alpha \neq 0$ for some i, so that for this i we have $c_i - c = 0$.

Certainly T is diagonalizable, since we have shown that every non-zero vector in the range of E_i is a characteristic vector of T, and the fact that $I = E_1 + \cdots + E_k$ shows that these characteristic vectors span V. All that remains to be demonstrated is that the null space of $(T - c_iI)$ is exactly the range of E_i. But this is clear, because if $T\alpha = c_i\alpha$, then

$$\sum_{j=1}^{k} (c_j - c_i)E_j\alpha = 0$$

hence

$$(c_j - c_i)E_j\alpha = 0 \qquad \text{for each } j$$

and then

$$E_j\alpha = 0, \qquad j \neq i.$$

Since $\alpha = E_1\alpha + \cdots + E_k\alpha$, and $E_j\alpha = 0$ for $j \neq i$, we have $\alpha = E_i\alpha$, which proves that α is in the range of E_i. ∎

One part of Theorem 9 says that for a diagonalizable operator T, the scalars $c_1, \ldots, c_k$ and the operators $E_1, \ldots, E_k$ are uniquely determined by conditions (i), (ii), (iii), the fact that the c_i are distinct, and the fact that the E_i are non-zero. One of the pleasant features of the decomposition $T = c_1E_1 + \cdots + c_kE_k$ is that if g is any polynomial over the field F, then

$$g(T) = g(c_1)E_1 + \cdots + g(c_k)E_k.$$

We leave the details of the proof to the reader. To see how it is proved one need only compute T^r for each positive integer r. For example,

$$\begin{aligned} T^2 &= \sum_{i=1}^{k} c_iE_i \sum_{j=1}^{k} c_jE_j \\ &= \sum_{i=1}^{k} \sum_{j=1}^{k} c_ic_jE_iE_j \\ &= \sum_{i=1}^{k} c_i^2E_i^2 \\ &= \sum_{i=1}^{k} c_i^2E_i. \end{aligned}$$

The reader should compare this with $g(A)$ where A is a diagonal matrix; for then $g(A)$ is simply the diagonal matrix with diagonal entries $g(A_{11})$, $\ldots, g(A_{nn})$.

We should like in particular to note what happens when one applies the Lagrange polynomials corresponding to the scalars $c_1, \ldots, c_k$:

$$p_j = \prod_{i \neq j} \frac{(x - c_i)}{(c_j - c_i)}.$$

We have $p_j(c_i) = \delta_{ij}$, which means that

$$\begin{aligned} p_j(T) &= \sum_{i=1}^{k} \delta_{ij}E_i \\ &= E_j. \end{aligned}$$

Thus the projections E_j not only commute with T but are polynomials in T.

Such calculations with polynomials in T can be used to give an alternative proof of Theorem 6, which characterized diagonalizable operators in terms of their minimal polynomials. The proof is entirely independent of our earlier proof.

If T is diagonalizable, $T = c_1E_1 + \cdots + c_kE_k$, then

$$g(T) = g(c_1)E_1 + \cdots + g(c_k)E_k$$

for every polynomial g. Thus $g(T) = 0$ if and only if $g(c_i) = 0$ for each i. In particular, the minimal polynomial for T is

$$p = (x - c_1) \cdots (x - c_k).$$

Now suppose T is a linear operator with minimal polynomial $p = (x - c_1) \cdots (x - c_k)$, where $c_1, \ldots, c_k$ are distinct elements of the scalar field. We form the Lagrange polynomials

$$p_j = \prod_{i \neq j} \frac{(x - c_i)}{(c_j - c_i)}.$$

We recall from Chapter 4 that $p_j(c_i) = \delta_{ij}$ and for any polynomial g of degree less than or equal to $(k - 1)$ we have

$$g = g(c_1)p_1 + \cdots + g(c_k)p_k.$$

Taking g to be the scalar polynomial 1 and then the polynomial x, we have

$$\begin{aligned} 1 &= p_1 + \cdots + p_k \\ x &= c_1p_1 + \cdots + c_kp_k. \end{aligned} \tag{6-15}$$

(The astute reader will note that the application to x may not be valid because k may be 1. But if $k = 1$, T is a scalar multiple of the identity and hence diagonalizable.) Now let $E_j = p_j(T)$. From (6-15) we have

$$\begin{aligned} I &= E_1 + \cdots + E_k \\ T &= c_1E_1 + \cdots + c_kE_k. \end{aligned} \tag{6-16}$$

Observe that if $i \neq j$, then p_ip_j is divisible by the minimal polynomial p, because p_ip_j contains every $(x - c_r)$ as a factor. Thus

$$E_iE_j = 0, \qquad i \neq j. \tag{6-17}$$

We must note one further thing, namely, that $E_i \neq 0$ for each i. This is because p is the minimal polynomial for T and so we cannot have $p_i(T) = 0$ since p_i has degree less than the degree of p. This last comment, together with (6-16), (6-17), and the fact that the c_i are distinct enables us to apply Theorem 11 to conclude that T is diagonalizable. ▌

Exercises

1. Let E be a projection of V and let T be a linear operator on V. Prove that the range of E is invariant under T if and only if $ETE = TE$. Prove that both the range and null space of E are invariant under T if and only if $ET = TE$.

2. Let T be the linear operator on R^2, the matrix of which in the standard ordered basis is

$$\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}.$$

Let W_1 be the subspace of R^2 spanned by the vector $\epsilon_1 = (1, 0)$.

(a) Prove that W_1 is invariant under T.

(b) Prove that there is no subspace W_2 which is invariant under T and which is complementary to W_1:

$$R^2 = W_1 \oplus W_2.$$

(Compare with Exercise 1 of Section 6.5.)

3. Let T be a linear operator on a finite-dimensional vector space V. Let R be the range of T and let N be the null space of T. Prove that R and N are independent if and only if $V = R \oplus N$.

4. Let T be a linear operator on V. Suppose $V = W_1 \oplus \cdots \oplus W_k$, where each W_i is invariant under T. Let T_i be the induced (restriction) operator on W_i.

(a) Prove that $\det(T) = \det(T_1) \cdots \det(T_k)$.

(b) Prove that the characteristic polynomial for f is the product of the characteristic polynomials for $f_1, \ldots, f_k$.

(c) Prove that the minimal polynomial for T is the least common multiple of the minimal polynomials for $T_1, \ldots, T_k$. (*Hint:* Prove and then use the corresponding facts about direct sums of matrices.)

5. Let T be the diagonalizable linear operator on R^3 which we discussed in Example 3 of Section 6.2. Use the Lagrange polynomials to write the representing matrix A in the form $A = E_1 + 2E_2$, $E_1 + E_2 = I$, $E_1E_2 = 0$.

6. Let A be the 4×4 matrix in Example 6 of Section 6.3. Find matrices E_1, E_2, E_3 such that $A = c_1E_1 + c_2E_2 + c_3E_3$, $E_1 + E_2 + E_3 = I$, and $E_iE_j = 0$, $i \neq j$.

7. In Exercises 5 and 6, notice that (for each i) the space of characteristic vectors associated with the characteristic value c_i is spanned by the column vectors of the various matrices E_j with $j \neq i$. Is that a coincidence?

8. Let T be a linear operator on V which commutes with every projection operator on V. What can you say about T?

9. Let V be the vector space of continuous real-valued functions on the interval $[-1, 1]$ of the real line. Let W_e be the subspace of even functions, $f(-x) = f(x)$, and let W_o be the subspace of odd functions, $f(-x) = -f(x)$.

(a) Show that $V = W_e \oplus W_o$.

(b) If T is the indefinite integral operator

$$(Tf)(x) = \int_0^x f(t)\, dt$$

are W_e and W_o invariant under T?

6.8. The Primary Decomposition Theorem

We are trying to study a linear operator T on the finite-dimensional space V, by decomposing T into a direct sum of operators which are in some sense elementary. We can do this through the characteristic values and vectors of T in certain special cases, i.e., when the minimal polynomial for T factors over the scalar field F into a product of distinct monic polynomials of degree 1. What can we do with the general T? If we try to study T using characteristic values, we are confronted with two problems. First, T may not have a single characteristic value; this is really a deficiency in the scalar field, namely, that it is not algebraically closed. Second, even if the characteristic polynomial factors completely over F into a product of polynomials of degree 1, there may not be enough characteristic vectors for T to span the space V; this is clearly a deficiency in T. The second situation is illustrated by the operator T on F^3 (F any field) represented in the standard basis by

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

The characteristic polynomial for A is $(x - 2)^2(x + 1)$ and this is plainly also the minimal polynomial for A (or for T). Thus T is not diagonalizable. One sees that this happens because the null space of $(T - 2I)$ has dimension 1 only. On the other hand, the null space of $(T + I)$ and the null space of $(T - 2I)^2$ together span V, the former being the subspace spanned by ϵ_3 and the latter the subspace spanned by ϵ_1 and ϵ_2.

This will be more or less our general method for the second problem. If (remember this is an assumption) the minimal polynomial for T decomposes

$$p = (x - c_1)^{r_1} \cdots (x - c_k)^{r_k}$$

where $c_1, \ldots, c_k$ are distinct elements of F, then we shall show that the space V is the direct sum of the null spaces of $(T - c_iI)^{r_i}$, $i = 1, \ldots, k$. The hypothesis about p is equivalent to the fact that T is triangulable (Theorem 5); however, that knowledge will not help us.

The theorem which we prove is more general than what we have described, since it works with the primary decomposition of the minimal polynomial, whether or not the primes which enter are all of first degree. The reader will find it helpful to think of the special case when the primes are of degree 1, and even more particularly, to think of the projection-type proof of Theorem 6, a special case of this theorem.

Theorem 12 (Primary Decomposition Theorem). *Let* T *be a linear operator on the finite-dimensional vector space* V *over the field* F. *Let* p *be the minimal polynomial for* T,

$$p = p_1^{r_1} \cdots p_k^{r_k}$$

where the p_i *are distinct irreducible monic polynomials over* F *and the* r_i *are positive integers. Let* W_i *be the null space of* $p_i(T)^{r_i}$, $i = 1, \ldots, k$. *Then*

(i) $V = W_1 \oplus \cdots \oplus W_k$;
(ii) *each* W_i *is invariant under* T;
(iii) *if* T_i *is the operator induced on* W_i *by* T, *then the minimal polynomial for* T_i *is* $p_i^{r_i}$.

Proof. The idea of the proof is this. If the direct-sum decomposition (i) is valid, how can we get hold of the projections $E_1, \ldots, E_k$ associated with the decomposition? The projection E_i will be the identity on W_i and zero on the other W_j. We shall find a polynomial h_i such that $h_i(T)$ is the identity on W_i and is zero on the other W_j, and so that $h_1(T) + \cdots + h_k(T) = I$, etc.

For each i, let

$$f_i = \frac{p}{p_i^{r_i}} = \prod_{j \neq i} p_j^{r_j}.$$

Since $p_1, \ldots, p_k$ are distinct prime polynomials, the polynomials $f_1, \ldots, f_k$ are relatively prime (Theorem 10, Chapter 4). Thus there are polynomials $g_1, \ldots, g_k$ such that

$$\sum_{i=1}^{n} f_i g_i = 1.$$

Note also that if $i \neq j$, then $f_i f_j$ is divisible by the polynomial p, because $f_i f_j$ contains each $p_m^{r_m}$ as a factor. We shall show that the polynomials $h_i = f_i g_i$ behave in the manner described in the first paragraph of the proof.

Let $E_i = h_i(T) = f_i(T)g_i(T)$. Since $h_1 + \cdots + h_k = 1$ and p divides $f_i f_j$ for $i \neq j$, we have

$$\begin{aligned} E_1 + \cdots + E_k &= I \\ E_i E_j &= 0, \quad \text{if } i \neq j. \end{aligned}$$

Thus the E_i are projections which correspond to some direct-sum decomposition of the space V. We wish to show that the range of E_i is exactly the subspace W_i. It is clear that each vector in the range of E_i is in W_i, for if α is in the range of E_i, then $\alpha = E_i\alpha$ and so

$$\begin{aligned} p_i(T)^{r_i}\alpha &= p_i(T)^{r_i}E_i\alpha \\ &= p_i(T)^{r_i}f_i(T)g_i(T)\alpha \\ &= 0 \end{aligned}$$

because $p^{r_i}f_i g_i$ is divisible by the minimal polynomial p. Conversely, suppose that α is in the null space of $p_i(T)^{r_i}$. If $j \neq i$, then $f_j g_j$ is divisible by $p_i^{r_i}$ and so $f_j(T)g_j(T)\alpha = 0$, i.e., $E_j\alpha = 0$ for $j \neq i$. But then it is immediate that $E_i\alpha = \alpha$, i.e., that α is in the range of E_i. This completes the proof of statement (i).

It is certainly clear that the subspaces W_i are invariant under T. If T_i is the operator induced on W_i by T, then evidently $p_i(T_i)^{r_i} = 0$, because by definition $p_i(T)^{r_i}$ is 0 on the subspace W_i. This shows that the minimal polynomial for T_i divides $p_i^{r_i}$. Conversely, let g be any polynomial such that $g(T_i) = 0$. Then $g(T)f_i(T) = 0$. Thus gf_i is divisible by the minimal polynomial p of T, i.e., $p_i^{r_i}f_i$ divides gf_i. It is easily seen that $p_i^{r_i}$ divides g. Hence the minimal polynomial for T_i is $p_i^{r_i}$. ∎

Corollary. *If* $E_1, \ldots, E_k$ *are the projections associated with the primary decomposition of* T, *then each* E_i *is a polynomial in* T, *and accordingly if a linear operator* U *commutes with* T *then* U *commutes with each of the* E_i, *i.e., each subspace* W_i *is invariant under* U.

In the notation of the proof of Theorem 12, let us take a look at the special case in which the minimal polynomial for T is a product of first-degree polynomials, i.e., the case in which each p_i is of the form $p_i = x - c_i$. Now the range of E_i is the null space W_i of $(T - c_iI)^{r_i}$. Let us put $D = c_1E_1 + \cdots + c_kE_k$. By Theorem 11, D is a diagonalizable operator which we shall call the **diagonalizable part** of T. Let us look at the operator $N = T - D$. Now

$$T = TE_1 + \cdots + TE_k$$
$$D = c_1E_1 + \cdots + c_kE_k$$

so

$$N = (T - c_1I)E_1 + \cdots + (T - c_kI)E_k.$$

The reader should be familiar enough with projections by now so that he sees that

$$N^2 = (T - c_1I)^2E_1 + \cdots + (T - c_kI)^2E_k$$

and in general that

$$N^r = (T - c_1I)^rE_1 + \cdots + (T - c_kI)^rE_k.$$

When $r \geq r_i$ for each i, we shall have $N^r = 0$, because the operator $(T - c_iI)^r$ will then be 0 on the range of E_i.

Definition. *Let* N *be a linear operator on the vector space* V. *We say that* N *is* **nilpotent** *if there is some positive integer* r *such that* $N^r = 0$.

Theorem 13. *Let* T *be a linear operator on the finite-dimensional vector space* V *over the field* F. *Suppose that the minimal polynomial for* T *decomposes over* F *into a product of linear polynomials. Then there is a diagonalizable operator* D *on* V *and a nilpotent operator* N *on* V *such that*

(i) $T = D + N$,
(ii) $DN = ND$.

The diagonalizable operator D *and the nilpotent operator* N *are uniquely determined by* (i) *and* (ii) *and each of them is a polynomial in* T.

Proof. We have just observed that we can write $T = D + N$ where D is diagonalizable and N is nilpotent, and where D and N not only commute but are polynomials in T. Now suppose that we also have $T = D' + N'$ where D' is diagonalizable, N' is nilpotent, and $D'N' = N'D'$. We shall prove that $D = D'$ and $N = N'$.

Since D' and N' commute with one another and $T = D' + N'$, we see that D' and N' commute with T. Thus D' and N' commute with any polynomial in T; hence they commute with D and with N. Now we have

$$D + N = D' + N'$$

or

$$D - D' = N' - N$$

and all four of these operators commute with one another. Since D and D' are both diagonalizable and they commute, they are simultaneously diagonalizable, and $D - D'$ is diagonalizable. Since N and N' are both nilpotent and they commute, the operator $(N' - N)$ is nilpotent; for, using the fact that N and N' commute

$$(N' - N)^r = \sum_{j=0}^{r} \binom{r}{j} (N')^{r-j}(-N)^j$$

and so when r is sufficiently large every term in this expression for $(N' - N)^r$ will be 0. (Actually, a nilpotent operator on an n-dimensional space must have its nth power 0; if we take $r = 2n$ above, that will be large enough. It then follows that $r = n$ is large enough, but this is not obvious from the above expression.) Now $D - D'$ is a diagonalizable operator which is also nilpotent. Such an operator is obviously the zero operator; for since it is nilpotent, the minimal polynomial for this operator is of the form x^r for some $r \leq m$; but then since the operator is diagonalizable, the minimal polynomial cannot have a repeated root; hence $r = 1$ and the minimal polynomial is simply x, which says the operator is 0. Thus we see that $D = D'$ and $N = N'$. ∎

Corollary. *Let* V *be a finite-dimensional vector space over an algebraically closed field* F, *e.g., the field of complex numbers. Then every linear operator* T *on* V *can be written as the sum of a diagonalizable operator* D *and a nilpotent operator* N *which commute. These operators* D *and* N *are unique and each is a polynomial in* T.

From these results, one sees that the study of linear operators on vector spaces over an algebraically closed field is essentially reduced to the study of nilpotent operators. For vector spaces over non-algebraically closed fields, we still need to find some substitute for characteristic values and vectors. It is a very interesting fact that these two problems can be handled simultaneously and this is what we shall do in the next chapter.

In concluding this section, we should like to give an example which illustrates some of the ideas of the primary decomposition theorem. We have chosen to give it at the end of the section since it deals with differential equations and thus is not purely linear algebra.

EXAMPLE 14. In the primary decomposition theorem, it is not necessary that the vector space V be finite dimensional, nor is it necessary for parts (i) and (ii) that p be the minimal polynomial for T. If T is a linear operator on an arbitrary vector space and *if* there is a monic polynomial p such that $p(T) = 0$, then parts (i) and (ii) of Theorem 12 are valid for T with the proof which we gave.

Let n be a positive integer and let V be the space of all n times continuously differentiable functions f on the real line which satisfy the differential equation

$$\frac{d^n f}{dt^n} + a_{n-1}\frac{d^{n-1}f}{dt^{n-1}} + \cdots + a_1\frac{df}{dt} + a_0 f = 0 \tag{6-18}$$

where $a_0, \ldots, a_{n-1}$ are some fixed constants. If C_n denotes the space of n times continuously differentiable functions, then the space V of solutions of this differential equation is a subspace of C_n. If D denotes the differentiation operator and p is the polynomial

$$p = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

then V is the null space of the operator $p(D)$, because (6-18) simply says $p(D)f = 0$. Therefore, V is invariant under D. Let us now regard D as a linear operator on the subspace V. Then $p(D) = 0$.

If we are discussing differentiable complex-valued functions, then C_n and V are complex vector spaces, and $a_0, \ldots, a_{n-1}$ may be any complex numbers. We now write

$$p = (x - c_1)^{r_1} \cdots (x - c_k)^{r_k}$$

where $c_1, \ldots, c_k$ are distinct complex numbers. If W_j is the null space of $(D - c_jI)^{r_j}$, then Theorem 12 says that

$$V = W_1 \oplus \cdots \oplus W_k.$$

In other words, if f satisfies the differential equation (6-18), then f is uniquely expressible in the form

$$f = f_1 + \cdots + f_k$$

where f_j satisfies the differential equation $(D - c_jI)^{r_j}f_j = 0$. Thus, the study of the solutions to the equation (6-18) is reduced to the study of the space of solutions of a differential equation of the form

$$(D - cI)^r f = 0. \tag{6-19}$$

This reduction has been accomplished by the general methods of linear algebra, i.e., by the primary decomposition theorem.

To describe the space of solutions to (6-19), one must know something about differential equations, that is, one must know something about D other than the fact that it is a linear operator. However, one does not need to know very much. It is very easy to establish by induction on r that if f is in C_r then

$$(D - cI)^r f = e^{ct}D^r(e^{-ct}f)$$

that is,

$$\frac{df}{dt} - cf(t) = e^{ct}\frac{d}{dt}(e^{-ct}f), \quad \text{etc.}$$

Thus $(D - cI)^r f = 0$ if and only if $D^r(e^{-ct}f) = 0$. A function g such that $D^r g = 0$, i.e., $d^r g/dt^r = 0$, must be a polynomial function of degree $(r - 1)$ or less:

$$g(t) = b_0 + b_1 t + \cdots + b_{r-1}t^{r-1}.$$

Thus f satisfies (6-19) if and only if f has the form

$$f(t) = e^{ct}(b_0 + b_1 t + \cdots + b_{r-1}t^{r-1}).$$

Accordingly, the 'functions' $e^{ct}, te^{ct}, \ldots, t^{r-1}e^{ct}$ span the space of solutions of (6-19). Since $1, t, \ldots, t^{r-1}$ are linearly independent functions and the exponential function has no zeros, these r functions $t^j e^{ct}$, $0 \leq j \leq r - 1$, form a basis for the space of solutions.

Returning to the differential equation (6-18), which is

$$p(D)f = 0$$
$$p = (x - c_1)^{r_1} \cdots (x - c_k)^{r_k}$$

we see that the n functions $t^m e^{c_j t}$, $0 \leq m \leq r_j - 1$, $1 \leq j \leq k$, form a basis for the space of solutions to (6-18). In particular, the space of solutions is finite-dimensional and has dimension equal to the degree of the polynomial p.

Exercises

1. Let T be a linear operator on R^3 which is represented in the standard ordered basis by the matrix

$$\begin{bmatrix} 6 & -3 & -2 \\ 4 & -1 & -2 \\ 10 & -5 & -3 \end{bmatrix}.$$

Express the minimal polynomial p for T in the form $p = p_1 p_2$, where p_1 and p_2 are monic and irreducible over the field of real numbers. Let W_i be the null space of $p_i(T)$. Find bases $\mathcal{B}_i$ for the spaces W_1 and W_2. If T_i is the operator induced on W_i by T, find the matrix of T_i in the basis $\mathcal{B}_i$ (above).

2. Let T be the linear operator on R^3 which is represented by the matrix

$$\begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}$$

in the standard ordered basis. Show that there is a diagonalizable operator D on R^3 and a nilpotent operator N on R^3 such that $T = D + N$ and $DN = ND$. Find the matrices of D and N in the standard basis. (Just repeat the proof of Theorem 12 for this special case.)

3. If V is the space of all polynomials of degree less than or equal to n over a field F, prove that the differentiation operator on V is nilpotent.

4. Let T be a linear operator on the finite-dimensional space V with characteristic polynomial

$$f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}$$

and minimal polynomial

$$p = (x - c_1)^{r_1} \cdots (x - c_k)^{r_k}.$$

Let W_i be the null space of $(T - c_i I)^{r_i}$.

(a) Prove that W_i is the set of all vectors α in V such that $(T - c_iI)^m\alpha = 0$ for some positive integer m (which may depend upon α).

(b) Prove that the dimension of W_i is d_i. (*Hint:* If T_i is the operator induced on W_i by T, then $T_i - c_iI$ is nilpotent; thus the characteristic polynomial for $T_i - c_iI$ must be x^{e_i} where e_i is the dimension of W_i (proof?); thus the characteristic polynomial of T_i is $(x - c_i)^{e_i}$; now use the fact that the characteristic polynomial for T is the product of the characteristic polynomials of the T_i to show that $e_i = d_i$.)

5. Let V be a finite-dimensional vector space over the field of complex numbers. Let T be a linear operator on V and let D be the diagonalizable part of T. Prove that if g is any polynomial with complex coefficients, then the diagonalizable part of $g(T)$ is $g(D)$.

6. Let V be a finite-dimensional vector space over the field F, and let T be a linear operator on V such that rank $(T) = 1$. Prove that either T is diagonalizable or T is nilpotent, not both.

7. Let V be a finite-dimensional vector space over F, and let T be a linear operator on V. Suppose that T commutes with every diagonalizable linear operator on V. Prove that T is a scalar multiple of the identity operator.

8. Let V be the space of $n \times n$ matrices over a field F, and let A be a fixed $n \times n$ matrix over F. Define a linear operator T on V by $T(B) = AB - BA$. Prove that if A is a nilpotent matrix, then T is a nilpotent operator.

9. Give an example of two 4×4 nilpotent matrices which have the same minimal polynomial (they necessarily have the same characteristic polynomial) but which are not similar.

10. Let T be a linear operator on the finite-dimensional space V, let $p = p_1^{r_1} \cdots p_k^{r_k}$ be the minimal polynomial for T, and let $V = W_1 \oplus \cdots \oplus W_k$ be the primary decomposition for T, i.e., W_j is the null space of $p_j(T)^{r_j}$. Let W be any subspace of V which is invariant under T. Prove that

$$W = (W \cap W_1) \oplus (W \cap W_2) \oplus \cdots \oplus (W \cap W_k).$$

11. What's wrong with the following proof of Theorem 13? Suppose that the minimal polynomial for T is a product of linear factors. Then, by Theorem 5, T is triangulable. Let $\mathcal{B}$ be an ordered basis such that $A = [T]_{\mathcal{B}}$ is upper-triangular. Let D be the diagonal matrix with diagonal entries $a_{11}, \ldots, a_{nn}$. Then $A = D + N$, where N is strictly upper-triangular. Evidently N is nilpotent.

12. If you thought about Exercise 11, think about it again, after you observe what Theorem 7 tells you about the diagonalizable and nilpotent parts of T.

13. Let T be a linear operator on V with minimal polynomial of the form p^n, where p is irreducible over the scalar field. Show that there is a vector α in V such that the T-annihilator of α is p^n.

14. Use the primary decomposition theorem and the result of Exercise 13 to prove the following. If T is any linear operator on a finite-dimensional vector space V, then there is a vector α in V with T-annihilator equal to the minimal polynomial for T.

15. If N is a nilpotent linear operator on an n-dimensional vector space V, then the characteristic polynomial for N is x^n.