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

© 1971, 1961 by
Prentice-Hall, Inc.
Englewood Cliffs, New Jersey

All rights reserved. No part of this book may be
reproduced in any form or by any means without
permission in writing from the publisher.

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

Chapter 1. ***Linear Equations*** ***1***

1.1. Fields 1
1.2. Systems of Linear Equations 3
1.3. Matrices and Elementary Row Operations 6
1.4. Row-Reduced Echelon Matrices 11
1.5. Matrix Multiplication 16
1.6. Invertible Matrices 21

Chapter 2. ***Vector Spaces*** ***28***

2.1. Vector Spaces 28
2.2. Subspaces 34
2.3. Bases and Dimension 40
2.4. Coordinates 49
2.5. Summary of Row-Equivalence 55
2.6. Computations Concerning Subspaces 58

Chapter 3. ***Linear Transformations*** ***67***

3.1. Linear Transformations 67
3.2. The Algebra of Linear Transformations 74
3.3. Isomorphism 84
3.4. Representation of Transformations by Matrices 86
3.5. Linear Functionals 97
3.6. The Double Dual 107
3.7. The Transpose of a Linear Transformation 111

Chapter 4. Polynomials ***117***

4.1. Algebras 117
4.2. The Algebra of Polynomials 119
4.3. Lagrange Interpolation 124
4.4. Polynomial Ideals 127
4.5. The Prime Factorization of a Polynomial 134

Chapter 5. Determinants ***140***

5.1. Commutative Rings 140
5.2. Determinant Functions 141
5.3. Permutations and the Uniqueness of Determinants 150
5.4. Additional Properties of Determinants 156
5.5. Modules 164
5.6. Multilinear Functions 166
5.7. The Grassman Ring 173

Chapter 6. Elementary Canonical Forms ***181***

6.1. Introduction 181
6.2. Characteristic Values 182
6.3. Annihilating Polynomials 190
6.4. Invariant Subspaces 198
6.5. Simultaneous Triangulation; Simultaneous Diagonalization 206
6.6. Direct-Sum Decompositions 209
6.7. Invariant Direct Sums 213
6.8. The Primary Decomposition Theorem 219

Chapter 7. The Rational and Jordan Forms ***227***

7.1. Cyclic Subspaces and Annihilators 227
7.2. Cyclic Decompositions and the Rational Form 231
7.3. The Jordan Form 244
7.4. Computation of Invariant Factors 251
7.5. Summary; Semi-Simple Operators 262

Chapter 8. Inner Product Spaces ***270***

8.1. Inner Products 270
8.2. Inner Product Spaces 277
8.3. Linear Functionals and Adjoints 290
8.4. Unitary Operators 299
8.5. Normal Operators 311

over R, with values in R, such that $D(AB) = D(A)D(B)$ for all A, B. Suppose also that

$$D\left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right) \neq D\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right).$$

Prove the following.

(a) $D(0) = 0$;

(b) $D(A) = 0$ if $A^2 = 0$;

(c) $D(B) = -D(A)$ if B is obtained by interchanging the rows (or columns) of A;

(d) $D(A) = 0$ if one row (or one column) of A is 0;

(e) $D(A) = 0$ whenever A is singular.

14. Let A be a 2×2 matrix over a field F. Then the set of all matrices of the form $f(A)$, where f is a polynomial over F, is a commutative ring K with identity. If B is a 2×2 matrix over K, the determinant of B is then a 2×2 matrix over F, of the form $f(A)$. Suppose I is the 2×2 identity matrix over F and that B is the 2×2 matrix over K

$$B = \begin{bmatrix} A - A_{11}I & -A_{12}I \\ -A_{21}I & A - A_{22}I \end{bmatrix}.$$

Show that $\det B = f(A)$, where $f = x^2 - (A_{11} + A_{22})x + \det A$, and also that $f(A) = 0$.

5.3. Permutations and the Uniqueness of Determinants

In this section we prove the uniqueness of the determinant function on $n \times n$ matrices over K. The proof will lead us quite naturally to consider permutations and some of their basic properties.

Suppose D is an alternating n-linear function on $n \times n$ matrices over K. Let A be an $n \times n$ matrix over K with rows $\alpha_1, \alpha_2, \cdots, \alpha_n$. If we denote the rows of the $n \times n$ identity matrix over K by $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$, then

$$\text{(5-9)} \qquad \alpha_i = \sum_{j=1}^{n} A(i, j)\epsilon_j, \qquad 1 \leq i \leq n.$$

Hence

$$\begin{aligned} D(A) &= D\left(\sum_j A(1, j)\epsilon_j, \alpha_2, \ldots, \alpha_n\right) \\ &= \sum_j A(1, j)D(\epsilon_j, \alpha_2, \ldots, \alpha_n). \end{aligned}$$

If we now replace α_2 by $\sum\limits_k A(2, k)\epsilon_k$, we see that

$$D(\epsilon_j, \alpha_2, \ldots, \alpha_n) = \sum_k A(2, k)D(\epsilon_j, \epsilon_k, \ldots, \alpha_n).$$

Thus

$$D(A) = \sum_{j,k} A(1, j)A(2, k)D(\epsilon_j, \epsilon_k, \ldots, \alpha_n).$$

In $D(\epsilon_j, \epsilon_k, \ldots, \alpha_n)$ we next replace α_3 by $\sum A(3, l)\epsilon_l$ and so on. We finally obtain a complicated but theoretically important expression for $D(A)$, namely

$$D(A) = \sum_{k_1, k_2, \ldots, k_n} A(1, k_1)A(2, k_2) \cdots A(n, k_n)D(\epsilon_{k_1}, \epsilon_{k_2}, \ldots, \epsilon_{k_n}). \tag{5-10}$$

In (5-10) the sum is extended over all sequences $(k_1, k_2, \ldots, k_n)$ of positive integers not exceeding n. This shows that D is a finite sum of functions of the type described by (5-2). It should be noted that (5-10) is a consequence just of assumption that D is n-linear, and that a special case of (5-10) was obtained in Example 2. Since D is alternating,

$$D(\epsilon_{k_1}, \epsilon_{k_2}, \ldots, \epsilon_{k_n}) = 0$$

whenever two of the indices k_i are equal. A sequence $(k_1, k_2, \ldots, k_n)$ of positive integers not exceeding n, with the property that no two of the k_i are equal, is called a **permutation of degree *n*.** In (5-10) we need therefore sum only over those sequences which are permutations of degree n.

Since a finite sequence, or n-tuple, is a function defined on the first n positive integers, a permutation of degree n may be defined as a one-one function from the set $\{1, 2, \ldots, n\}$ onto itself. Such a function σ corresponds to the n-tuple $(\sigma 1, \sigma 2, \ldots, \sigma n)$ and is thus simply a rule for ordering $1, 2, \ldots, n$ in some well-defined way.

If D is an alternating n-linear function and A is an $n \times n$ matrix over K, we then have

$$D(A) = \sum_{\sigma} A(1, \sigma 1) \cdots A(n, \sigma n)D(\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}) \tag{5-11}$$

where the sum is extended over the distinct permutations σ of degree n.

Next we shall show that

$$D(\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}) = \pm D(\epsilon_1, \ldots, \epsilon_n) \tag{5-12}$$

where the sign ± depends only on the permutation σ. The reason for this is as follows. The sequence $(\sigma 1, \sigma 2, \ldots, \sigma n)$ can be obtained from the sequence $(1, 2, \ldots, n)$ by a finite number of interchanges of pairs of elements. For example, if $\sigma 1 \neq 1$, we can transpose 1 and $\sigma 1$, obtaining $(\sigma 1, \ldots, 1, \ldots)$. Proceeding in this way we shall arrive at the sequence $(\sigma 1, \ldots, \sigma n)$ after n or less such interchanges of pairs. Since D is alternating, the sign of its value changes each time that we interchange two of the rows ϵ_i and ϵ_j. Thus, if we pass from $(1, 2, \ldots, n)$ to $(\sigma 1, \sigma 2, \ldots, \sigma n)$ by means of m interchanges of pairs (i, j), we shall have

$$D(\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}) = (-1)^m D(\epsilon_1, \ldots, \epsilon_n).$$

In particular, if D is a determinant function

$$D(\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}) = (-1)^m \tag{5-13}$$

where m depends only upon σ, not upon D. Thus all determinant functions assign the same value to the matrix with rows $\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}$, and this value is either 1 or -1.

Now a basic fact about permutations is the following. If σ is a permutation of degree n, one can pass from the sequence $(1, 2, \ldots, n)$ to the sequence $(\sigma 1, \sigma 2, \ldots, \sigma n)$ by a succession of interchanges of pairs, and this can be done in a variety of ways; however, no matter how it is done, the number of interchanges used is either always even or always odd. The permutation is then called **even** or **odd,** respectively. One defines the **sign** of a permutation by

$$\operatorname{sgn} \sigma = \begin{cases} 1, & \text{if } \sigma \text{ is even} \\ -1, & \text{if } \sigma \text{ is odd} \end{cases}$$

the symbol '1' denoting here the integer 1.

We shall show below that this basic property of permutations can be deduced from what we already know about determinant functions. Let us assume this for the time being. Then the integer m occurring in (5-13) is always even if σ is an even permutation, and is always odd if σ is an odd permutation. For any alternating n-linear function D we then have

$$D(\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}) = (\operatorname{sgn} \sigma) D(\epsilon_1, \ldots, \epsilon_n)$$

and using (5-11)

$$D(A) = \left[\sum_{\sigma} (\operatorname{sgn} \sigma) A(1, \sigma 1) \cdots A(n, \sigma n)\right] D(I). \tag{5-14}$$

Of course I denotes the $n \times n$ identity matrix.

From (5-14) we see that there is precisely one determinant function on $n \times n$ matrices over K. If we denote this function by det, it is given by

$$\det(A) = \sum_{\sigma} (\operatorname{sgn} \sigma) A(1, \sigma 1) \cdots A(n, \sigma n) \tag{5-15}$$

the sum being extended over the distinct permutations σ of degree n. We can formally summarize as follows.

Theorem 2. *Let* K *be a commutative ring with identity and let* n *be a positive integer. There is precisely one determinant function on the set of* n × n *matrices over* K, *and it is the function* det *defined by (5-15). If* D *is any alternating* n*-linear function on* $K^{n \times n}$, *then for each* n × n *matrix* A

$$D(A) = (\det A) D(I).$$

This is the theorem we have been seeking, but we have left a gap in the proof. That gap is the proof that for a given permutation σ, when we pass from $(1, 2, \ldots, n)$ to $(\sigma 1, \sigma 2, \ldots, \sigma n)$ by interchanging pairs, the number of interchanges is always even or always odd. This basic combinatorial fact can be proved without any reference to determinants;

however, we should like to point out how it follows from the *existence* of a determinant function on $n \times n$ matrices.

Let us take K to be the ring of integers. Let D be a determinant function on $n \times n$ matrices over K. Let σ be a permutation of degree n, and suppose we pass from $(1, 2, \ldots, n)$ to $(\sigma 1, \sigma 2, \ldots, \sigma n)$ by m interchanges of pairs (i, j), $i \neq j$. As we showed in (5-13)

$$(-1)^m = D(\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n})$$

that is, the number $(-1)^m$ must be the value of D on the matrix with rows $\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}$. If

$$D(\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}) = 1,$$

then m must be even. If

$$D(\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}) = -1,$$

then m must be odd.

Since we have an explicit formula for the determinant of an $n \times n$ matrix and this formula involves the permutations of degree n, let us conclude this section by making a few more observations about permutations. First, let us note that there are precisely $n! = 1 \cdot 2 \cdots n$ permutations of degree n. For, if σ is such a permutation, there are n possible choices for $\sigma 1$; when this choice has been made, there are $(n - 1)$ choices for $\sigma 2$, then $(n - 2)$ choices for $\sigma 3$, and so on. So there are

$$n(n-1)(n-2) \cdots 2 \cdot 1 = n!$$

permutations σ. The formula (5-15) for det (A) thus gives det (A) as a sum of $n!$ terms, one for each permutation of degree n. A given term is a product

$$A(1, \sigma 1) \cdots A(n, \sigma n)$$

of n entries of A, one entry from each row and one from each column, and is prefixed by a '+' or '−' sign according as σ is an even or odd permutation.

When permutations are regarded as one-one functions from the set $\{1, 2, \ldots, n\}$ onto itself, one can define a product of permutations. The product of σ and τ will simply be the composed function $\sigma\tau$ defined by

$$(\sigma\tau)(i) = \sigma(\tau(i)).$$

If ϵ denotes the identity permutation, $\epsilon(i) = i$, then each σ has an inverse σ^{-1} such that

$$\sigma\sigma^{-1} = \sigma^{-1}\sigma = \epsilon.$$

One can summarize these observations by saying that, under the operation of composition, the set of permutations of degree n is a group. This group is usually called the **symmetric group of degree n.**

From the point of view of products of permutations, the basic property of the sign of a permutation is that

$$\text{(5-16)} \qquad \operatorname{sgn}(\sigma\tau) = (\operatorname{sgn}\sigma)(\operatorname{sgn}\tau).$$

In other words, $\sigma\tau$ is an even permutation if σ and τ are either both even or both odd, while $\sigma\tau$ is odd if one of the two permutations is odd and the other is even. One can see this from the definition of the sign in terms of successive interchanges of pairs (i, j). It may also be instructive if we point out how $\text{sgn}\ (\sigma\tau) = (\text{sgn}\ \sigma)(\text{sgn}\ \tau)$ follows from a fundamental property of determinants.

Let K be the ring of integers and let σ and τ be permutations of degree n. Let $\epsilon_1, \ldots, \epsilon_n$ be the rows of the $n \times n$ identity matrix over K, let A be the matrix with rows $\epsilon_{\tau 1}, \ldots, \epsilon_{\tau n}$, and let B be the matrix with rows $\epsilon_{\sigma 1}, \ldots, \epsilon_{\sigma n}$. The ith row of A contains exactly one non-zero entry, namely the 1 in column τi. From this it is easy to see that $\epsilon_{\sigma\tau i}$ is the ith row of the product matrix AB. Now

$$\det\ (A) = \text{sgn}\ \tau, \qquad \det\ (B) = \text{sgn}\ \sigma, \qquad \text{and} \quad \det\ (AB) = \text{sgn}\ (\sigma\tau).$$

So we shall have $\text{sgn}\ (\sigma\tau) = (\text{sgn}\ \sigma)(\text{sgn}\ \tau)$ as soon as we prove the following.

Theorem 3. *Let* K *be a commutative ring with identity, and let* A *and* B *be* n $\times$ n *matrices over* K. *Then*

$$\textit{det}\ (\mathrm{AB}) = (\textit{det}\ \mathrm{A})(\textit{det}\ \mathrm{B}).$$

Proof. Let B be a fixed $n \times n$ matrix over K, and for each $n \times n$ matrix A define $D(A) = \det(AB)$. If we denote the rows of A by $\alpha_1, \ldots, \alpha_n$, then

$$D(\alpha_1, \ldots, \alpha_n) = \det\ (\alpha_1 B, \ldots, \alpha_n B).$$

Here $\alpha_j B$ denotes the $1 \times n$ matrix which is the product of the $1 \times n$ matrix α_j and the $n \times n$ matrix B. Since

$$(c\alpha_i + \alpha_i')B = c\alpha_i B + \alpha_i' B$$

and det is n-linear, it is easy to see that D is n-linear. If $\alpha_i = \alpha_j$, then $\alpha_i B = \alpha_j B$, and since det is alternating,

$$D(\alpha_1, \ldots, \alpha_n) = 0.$$

Hence, D is alternating. Now D is an alternating n-linear function, and by Theorem 2

$$D(A) = (\det A)D(I).$$

But $D(I) = \det\ (IB) = \det B$, so

$$\det\ (AB) = D(A) = (\det A)(\det B). \quad ▌$$

The fact that $\text{sgn}\ (\sigma\tau) = (\text{sgn}\ \sigma)(\text{sgn}\ \tau)$ is only one of many corollaries to Theorem 3. We shall consider some of these corollaries in the next section.

Exercises

1. If K is a commutative ring with identity and A is the matrix over K given by

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

show that $\det A = 0$.

2. Prove that the determinant of the Vandermonde matrix

$$\begin{bmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{bmatrix}$$

is $(b - a)(c - a)(c - b)$.

3. List explicitly the six permutations of degree 3, state which are odd and which are even, and use this to give the complete formula (5-15) for the determinant of a 3×3 matrix.

4. Let σ and τ be the permutations of degree 4 defined by $\sigma 1 = 2$, $\sigma 2 = 3$, $\sigma 3 = 4$, $\sigma 4 = 1$, $\tau 1 = 3$, $\tau 2 = 1$, $\tau 3 = 2$, $\tau 4 = 4$.

(a) Is σ odd or even? Is τ odd or even?

(b) Find $\sigma\tau$ and $\tau\sigma$.

5. If A is an invertible $n \times n$ matrix over a field, show that $\det A \neq 0$.

6. Let A be a 2×2 matrix over a field. Prove that $\det (I + A) = 1 + \det A$ if and only if $\operatorname{trace} (A) = 0$.

7. An $n \times n$ matrix A is called **triangular** if $A_{ij} = 0$ whenever $i > j$ or if $A_{ij} = 0$ whenever $i < j$. Prove that the determinant of a triangular matrix is the product $A_{11}A_{22} \cdots A_{nn}$ of its diagonal entries.

8. Let A be a 3×3 matrix over the field of complex numbers. We form the matrix $xI - A$ with polynomial entries, the i, j entry of this matrix being the polynomial $\delta_{ij}x - A_{ij}$. If $f = \det (xI - A)$, show that f is a monic polynomial of degree 3. If we write

$$f = (x - c_1)(x - c_2)(x - c_3)$$

with complex numbers c_1, c_2, and c_3, prove that

$$c_1 + c_2 + c_3 = \operatorname{trace} (A) \quad \text{and} \quad c_1c_2c_3 = \det A.$$

9. Let n be a positive integer and F a field. If σ is a permutation of degree n, prove that the function

$$T(x_1, \ldots, x_n) = (x_{\sigma 1}, \ldots, x_{\sigma n})$$

is an invertible linear operator on F^n.

10. Let F be a field, n a positive integer, and S the set of $n \times n$ matrices over F. Let V be the vector space of all functions from S into F. Let W be the set of alternating n-linear functions on S. Prove that W is a subspace of V. What is the dimension of W?

11. Let T be a linear operator on F^n. Define

$$D_T(\alpha_1, \dots, \alpha_n) = \det (T\alpha_1, \dots, T\alpha_n).$$

(a) Show that D_T is an alternating n-linear function.

(b) If

$$c = \det (T\epsilon_1, \dots, T\epsilon_n)$$

show that for any n vectors $\alpha_1, \dots, \alpha_n$ we have

$$\det (T\alpha_1, \dots, T\alpha_n) = c \det (\alpha_1, \dots, \alpha_n).$$

(c) If $\mathcal{B}$ is any ordered basis for F^n and A is the matrix of T in the ordered basis $\mathcal{B}$, show that $\det A = c$.

(d) What do you think is a reasonable name for the scalar c?

12. If σ is a permutation of degree n and A is an $n \times n$ matrix over the field F with row vectors $\alpha_1, \dots, \alpha_n$, let $\sigma(A)$ denote the $n \times n$ matrix with row vectors $\alpha_{\sigma 1}, \dots, \alpha_{\sigma n}$.

(a) Prove that $\sigma(AB) = \sigma(A)B$, and in particular that $\sigma(A) = \sigma(I)A$.

(b) If T is the linear operator of Exercise 9, prove that the matrix of T in the standard ordered basis is $\sigma(I)$.

(c) Is $\sigma^{-1}(I)$ the inverse matrix of $\sigma(I)$?

(d) Is it true that $\sigma(A)$ is similar to A?

13. Prove that the sign function on permutations is unique in the following sense. If f is any function which assigns to each permutation of degree n an integer, and if $f(\sigma\tau) = f(\sigma)f(\tau)$, then f is identically 0, or f is identically 1, or f is the sign function.

5.4. Additional Properties of Determinants

In this section we shall relate some of the useful properties of the determinant function on $n \times n$ matrices. Perhaps the first thing we should point out is the following. In our discussion of $\det A$, the rows of A have played a privileged role. Since there is no fundamental difference between rows and columns, one might very well expect that $\det A$ is an alternating n-linear function of the columns of A. This is the case, and to prove it, it suffices to show that

$$\det (A^t) = \det (A) \tag{5-17}$$

where A^t denotes the transpose of A.

If σ is a permutation of degree n,

$$A^t(i, \sigma i) = A(\sigma i, i).$$

From the expression (5-15) one then has

$$\det (A^t) = \sum_{\sigma} (\operatorname{sgn} \sigma) A(\sigma 1, 1) \cdots A(\sigma n, n).$$

When $i = \sigma^{-1}j$, $A(\sigma i, i) = A(j, \sigma^{-1}j)$. Thus

$$A(\sigma 1, 1) \cdots A(\sigma n, n) = A(1, \sigma^{-1}1) \cdots A(n, \sigma^{-1}n).$$

Since $\sigma\sigma^{-1}$ is the identity permutation,

$$(\operatorname{sgn} \sigma)(\operatorname{sgn} \sigma^{-1}) = 1 \quad \text{or} \quad \operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma).$$

Furthermore, as σ varies over all permutations of degree n, so does σ^{-1}. Therefore

$$\begin{aligned} \det(A^t) &= \sum_{\sigma} (\operatorname{sgn} \sigma^{-1}) A(1, \sigma^{-1}1) \cdots A(n, \sigma^{-1}n) \\ &= \det A \end{aligned}$$

proving (5-17).

On certain occasions one needs to compute specific determinants. When this is necessary, it is frequently useful to take advantage of the following fact. *If B is obtained from A by adding a multiple of one row of A to another* (*or a multiple of one column to another*), *then*

$$\det B = \det A. \tag{5-18}$$

We shall prove the statement about rows. Let B be obtained from A by adding $c\alpha_j$ to α_i, where $i < j$. Since det is linear as a function of the ith row

$$\begin{aligned} \det B &= \det A + c \det(\alpha_1, \ldots, \alpha_j, \ldots, \alpha_j, \ldots, \alpha_n) \\ &= \det A. \end{aligned}$$

Another useful fact is the following. Suppose we have an $n \times n$ matrix of the block form

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

where A is an $r \times r$ matrix, C is an $s \times s$ matrix, B is $r \times s$, and 0 denotes the $s \times r$ zero matrix. Then

$$\det \begin{bmatrix} A & B \\ 0 & C \end{bmatrix} = (\det A)(\det C). \tag{5-19}$$

To prove this, define

$$D(A, B, C) = \det \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}.$$

If we fix A and B, then D is alternating and s-linear as a function of the rows of C. Thus, by Theorem 2

$$D(A, B, C) = (\det C) D(A, B, I)$$

where I is the $s \times s$ identity matrix. By subtracting multiples of the rows of I from the rows of B and using the statement above (5-18), we obtain

$$D(A, B, I) = D(A, 0, I).$$

Now $D(A, 0, I)$ is clearly alternating and r-linear as a function of the rows of A. Thus

$$D(A, 0, I) = (\det A) D(I, 0, I).$$

But $D(I, 0, I) = 1$, so

$$\begin{aligned} D(A, B, C) &= (\det C)D(A, B, I) \\ &= (\det C)D(A, 0, I) \\ &= (\det C)(\det A). \end{aligned}$$

By the same sort of argument, or by taking transposes

$$\det \begin{bmatrix} A & 0 \\ B & C \end{bmatrix} = (\det A)(\det C). \tag{5-20}$$

Example 6. Suppose K is the field of rational numbers and we wish to compute the determinant of the 4×4 matrix

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

By subtracting suitable multiples of row 1 from rows 2, 3, and 4, we obtain the matrix

$$\begin{bmatrix} 1 & -1 & 2 & 3 \\ 0 & 4 & -4 & -4 \\ 0 & 5 & -9 & -13 \\ 0 & 3 & 1 & -3 \end{bmatrix}$$

which we know by (5-18) will have the same determinant as A. If we subtract $\frac{5}{4}$ of row 2 from row 3 and then subtract $\frac{3}{4}$ of row 2 from row 4, we obtain

$$B = \begin{bmatrix} 1 & -1 & 2 & 3 \\ 0 & 4 & -4 & -4 \\ 0 & 0 & -4 & -8 \\ 0 & 0 & 4 & 0 \end{bmatrix}$$

and again $\det B = \det A$. The block form of B tells us that

$$\det A = \det B = \begin{vmatrix} 1 & -1 \\ 0 & 4 \end{vmatrix} \begin{vmatrix} -4 & -8 \\ 4 & 0 \end{vmatrix} = 4(32) = 128.$$

Now let $n > 1$ and let A be an $n \times n$ matrix over K. In Theorem 1, we showed how to construct a determinant function on $n \times n$ matrices, given one on $(n-1) \times (n-1)$ matrices. Now that we have proved the uniqueness of the determinant function, the formula (5-4) tells us the following. If we fix any column index j,

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \det A(i|j).$$

The scalar $(-1)^{i+j} \det A(i|j)$ is usually called the i, j **cofactor** of A or the cofactor of the i, j entry of A. The above formula for $\det A$ is then called the expansion of det A by cofactors of the jth column (or sometimes the expansion by minors of the jth column). If we set

$$C_{ij} = (-1)^{i+j} \det A(i|j)$$

then the above formula says that for each j

$$\det A = \sum_{i=1}^{n} A_{ij}C_{ij}$$

where the cofactor C_{ij} is $(-1)^{i+j}$ times the determinant of the $(n-1) \times (n-1)$ matrix obtained by deleting the ith row and jth column of A.

If $j \neq k$, then

$$\sum_{i=1}^{n} A_{ik}C_{ij} = 0.$$

For, replace the jth column of A by its kth column, and call the resulting matrix B. Then B has two equal columns and so $\det B = 0$. Since $B(i|j) = A(i|j)$, we have

$$\begin{aligned} 0 &= \det B \\ &= \sum_{i=1}^{n} (-1)^{i+j}B_{ij} \det B(i|j) \\ &= \sum_{i=1}^{n} (-1)^{i+j}A_{ik} \det A(i|j) \\ &= \sum_{i=1}^{n} A_{ik}C_{ij}. \end{aligned}$$

These properties of the cofactors can be summarized by

$$\sum_{i=1}^{n} A_{ik}C_{ij} = \delta_{jk} \det A. \tag{5-21}$$

The $n \times n$ matrix adj A, which is the transpose of the matrix of cofactors of A, is called the **classical adjoint** of A. Thus

$$(\operatorname{adj} A)_{ij} = C_{ji} = (-1)^{i+j} \det A(j|i). \tag{5-22}$$

The formulas (5-21) can be summarized in the matrix equation

$$(\operatorname{adj} A)A = (\det A)I. \tag{5-23}$$

We wish to see that $A(\operatorname{adj} A) = (\det A)I$ also. Since $A^t(i|j) = A(j|i)^t$, we have

$$(-1)^{i+j} \det A^t(i|j) = (-1)^{j+i} \det A(j|i)$$

which simply says that the i, j cofactor of A^t is the j, i cofactor of A. Thus

$$\operatorname{adj} (A^t) = (\operatorname{adj} A)^t \tag{5-24}$$

By applying (5-23) to A^t, we obtain

$$(\operatorname{adj} A^t)A^t = (\det A^t)I = (\det A)I$$

and transposing

$$A(\operatorname{adj} A^t)^t = (\det A)I.$$

Using (5-24), we have what we want:

$$(5\text{-}25) \qquad A(\operatorname{adj} A) = (\det A)I.$$

As for matrices over a field, an $n \times n$ matrix A over K is called **invertible over** K if there is an $n \times n$ matrix A^{-1} with entries in K such that $AA^{-1} = A^{-1}A = I$. If such an inverse matrix exists it is unique; for the same argument used in Chapter 1 shows that when $BA = AC = I$ we have $B = C$. The formulas (5-23) and (5-25) tell us the following about invertibility of matrices over K. If the element $\det A$ has a multiplicative inverse in K, then A is invertible and $A^{-1} = (\det A)^{-1} \operatorname{adj} A$ is the unique inverse of A. Conversely, it is easy to see that if A is invertible over K, the element $\det A$ is invertible in K. For, if $BA = I$ we have

$$1 = \det I = \det (AB) = (\det A)(\det B).$$

What we have proved is the following.

Theorem 4. *Let* A *be an* n × n *matrix over* K. *Then* A *is invertible over* K *if and only if det* A *is invertible in* K. *When* A *is invertible, the unique inverse for* A *is*

$$A^{-1} = (\det A)^{-1} \operatorname{adj} A.$$

In particular, an n × n *matrix over a field is invertible if and only if its determinant is different from zero.*

We should point out that this determinant criterion for invertibility proves that an $n \times n$ matrix with either a left or right inverse is invertible. This proof is completely independent of the proof which we gave in Chapter 1 for matrices over a field. We should also like to point out what invertibility means for matrices with polynomial entries. If K is the polynomial ring $F[x]$, the only elements of K which are invertible are the non-zero scalar polynomials. For if f and g are polynomials and $fg = 1$, we have $\deg f + \deg g = 0$ so that $\deg f = \deg g = 0$, i.e., f and g are scalar polynomials. So an $n \times n$ matrix over the polynomial ring $F[x]$ is invertible o\er $F[x]$ if and only if its determinant is a non-zero scalar polynomial.

EXAMPLE 7. Let $K = R[x]$, the ring of polynomials over the field of real numbers. Let

$$A = \begin{bmatrix} x^2 + x & x + 1 \\ x - 1 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} x^2 - 1 & x + 2 \\ x^2 - 2x + 3 & x \end{bmatrix}.$$

Then, by a short computation, $\det A = x + 1$ and $\det B = -6$. Thus A is not invertible over K, whereas B is invertible over K. Note that

$$\operatorname{adj} A = \begin{bmatrix} 1 & -x - 1 \\ -x + 1 & x^2 + x \end{bmatrix}, \qquad \operatorname{adj} B = \begin{bmatrix} x & -x - 2 \\ -x^2 + 2x - 3 & x^2 - 1 \end{bmatrix}$$

and (adj $A)A = (x + 1)I$, (adj $B)B = -6I$. Of course,

$$B^{-1} = -\frac{1}{6}\begin{bmatrix} x & -x-2 \\ -x^2+2x-3 & 1-x^2 \end{bmatrix}.$$

EXAMPLE 8. Let K be the ring of integers and

$$\mathrm{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Then $\det A = -2$ and

$$\operatorname{adj} A = \begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix}.$$

Thus A is not invertible as a matrix over the ring of integers; however, we can also regard A as a matrix over the field of rational numbers. If we do, then A is invertible and

$$A^{-1} = -\frac{1}{2}\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & \frac{1}{2} \end{bmatrix}.$$

In connection with invertible matrices, we should like to mention one further elementary fact. Similar matrices have the same determinant, that is, if P is invertible over K and $B = P^{-1}AP$, then $\det B = \det A$. This is clear since

$$\det (P^{-1}AP) = (\det P^{-1})(\det A)(\det P) = \det A.$$

This simple observation makes it possible to define the determinant of a linear operator on a finite dimensional vector space. If T is a linear operator on V, we define the determinant of T to be the determinant of any $n \times n$ matrix which represents T in an ordered basis for V. Since all such matrices are similar, they have the same determinant and our definition makes sense. In this connection, see Exercise 11 of section 5.3.

We should like now to discuss **Cramer's rule** for solving systems of linear equations. Suppose A is an $n \times n$ matrix over the field F and we wish to solve the system of linear equations $AX = Y$ for some given n-tuple $(y_1, \ldots, y_n)$. If $AX = Y$, then

$$(\operatorname{adj} A)AX = (\operatorname{adj} A)Y$$

and so

$$(\det A)X = (\operatorname{adj} A)Y.$$

Thus

$$\begin{aligned}(\det A)x_j &= \sum_{i=1}^{n} (\operatorname{adj} A)_{ji}y_i \\ &= \sum_{i=1}^{n} (-1)^{i+j}y_i \det A(i|j).\end{aligned}$$

This last expression is the determinant of the $n \times n$ matrix obtained by replacing the jth column of A by Y. If $\det A = 0$, all this tells us nothing; however, if $\det A \neq 0$, we have what is known as Cramer's rule. Let A

be an $n \times n$ matrix over the field F such that $\det A \neq 0$. If $y_1, \ldots, y_n$ are any scalars in F, the unique solution $X = A^{-1}Y$ of the system of equations $AX = Y$ is given by

$$x_j = \frac{\det B_j}{\det A}, \qquad j = 1, \ldots, n$$

where B_j is the $n \times n$ matrix obtained from A by replacing the jth column of A by Y.

In concluding this chapter, we should like to make some comments which serve to place determinants in what we believe to be the proper perspective. From time to time it is necessary to compute specific determinants, and this section has been partially devoted to techniques which will facilitate such work. However, the principal role of determinants in this book is theoretical. There is no disputing the beauty of facts such as Cramer's rule. But Cramer's rule is an inefficient tool for solving systems of linear equations, chiefly because it involves too many computations. So one should concentrate on what Cramer's rule says, rather than on how to compute with it. Indeed, while reflecting on this entire chapter, we hope that the reader will place more emphasis on understanding what the determinant function is and how it behaves than on how to compute determinants of specific matrices.

Exercises

1. Use the classical adjoint formula to compute the inverses of each of the following 3×3 real matrices.

$$\begin{bmatrix} -2 & 3 & 2 \\ 6 & 0 & 3 \\ 4 & 1 & -1 \end{bmatrix}, \quad \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix}$$

2. Use Cramer's rule to solve each of the following systems of linear equations over the field of rational numbers.

(a)
$$\begin{aligned} x + y + z &= 11 \\ 2x - 6y - z &= 0 \\ 3x + 4y + 2z &= 0. \end{aligned}$$

(b)
$$\begin{aligned} 3x - 2y &= 7 \\ 3y - 2z &= 6 \\ 3z - 2x &= -1. \end{aligned}$$

3. An $n \times n$ matrix A over a field F is **skew-symmetric** if $A^t = -A$. If A is a skew-symmetric $n \times n$ matrix with complex entries and n is odd, prove that $\det A = 0$.

4. An $n \times n$ matrix A over a field F is called **orthogonal** if $AA^t = I$. If A is orthogonal, show that $\det A = \pm 1$. Give an example of an orthogonal matrix for which $\det A = -1$.

5. An $n \times n$ matrix A over the field of complex numbers is said to be **unitary** if $AA^* = I$ (A^* denotes the conjugate transpose of A). If A is unitary, show that $|\det A| = 1$.

6. Let T and U be linear operators on the finite dimensional vector space V. Prove
(a) $\det(TU) = (\det T)(\det U)$;
(b) T is invertible if and only if $\det T \neq 0$.

7. Let A be an $n \times n$ matrix over K, a commutative ring with identity. Suppose A has the block form

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & A_k \end{bmatrix}$$

where A_j is an $r_j \times r_j$ matrix. Prove

$$\det A = (\det A_1)(\det A_2) \cdots (\det A_k).$$

8. Let V be the vector space of $n \times n$ matrices over the field F. Let B be a fixed element of V and let T_B be the linear operator on V defined by $T_B(A) = AB - BA$. Show that $\det T_B = 0$.

9. Let A be an $n \times n$ matrix over a field, $A \neq 0$. If r is any positive integer between 1 and n, an $r \times r$ **submatrix** of A is any $r \times r$ matrix obtained by deleting $(n - r)$ rows and $(n - r)$ columns of A. The **determinant rank** of A is the largest positive integer r such that some $r \times r$ submatrix of A has a non-zero determinant. Prove that the determinant rank of A is equal to the row rank of A (= column rank A).

10. Let A be an $n \times n$ matrix over the field F. Prove that there are at most n distinct scalars c in F such that $\det(cI - A) = 0$.

11. Let A and B be $n \times n$ matrices over the field F. Show that if A is invertible there are at most n scalars c in F for which the matrix $cA + B$ is not invertible.

12. If V is the vector space of $n \times n$ matrices over F and B is a fixed $n \times n$ matrix over F, let L_B and R_B be the linear operators on V defined by $L_B(A) = BA$ and $R_B(A) = AB$. Show that
(a) $\det L_B = (\det B)^n$;
(b) $\det R_B = (\det B)^n$.

13. Let V be the vector space of all $n \times n$ matrices over the field of complex numbers, and let B be a fixed $n \times n$ matrix over C. Define a linear operator M_B on V by $M_B(A) = BAB^*$, where $B^* = \overline{B^t}$. Show that

$$\det M_B = |\det B|^{2n}.$$

Now let H be the set of all Hermitian matrices in V, A being Hermitian if $A = A^*$. Then H is a vector space over the field of *real* numbers. Show that the function T_B defined by $T_B(A) = BAB^*$ is a linear operator on the real vector space H, and then show that $\det T_B = |\det B|^{2n}$. (*Hint:* In computing $\det T_B$, show that V has a basis consisting of Hermitian matrices and then show that $\det T_B = \det M_B$.)

14. Let A, B, C, D be *commuting* $n \times n$ matrices over the field F. Show that the determinant of the $2n \times 2n$ matrix

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

is det $(AD - BC)$.

5.5. Modules

If K is a commutative ring with identity, a module over K is an algebraic system which behaves like a vector space, with K playing the role of the scalar field. To be precise, we say that V is a **module over** K (or a K-**module**) if

1. there is an addition $(\alpha, \beta) \rightarrow \alpha + \beta$ on V, under which V is a commutative group;

2. there is a multiplication $(c, \alpha) \rightarrow c\alpha$ of elements α in V and c in K such that

$$\begin{aligned} (c_1 + c_2)\alpha &= c_1\alpha + c_2\alpha \\ c(\alpha_1 + \alpha_2) &= c\alpha_1 + c\alpha_2 \\ (c_1c_2)\alpha &= c_1(c_2\alpha) \\ 1\alpha &= \alpha. \end{aligned}$$

For us, the most important K-modules will be the n-tuple modules K^n. The matrix modules $K^{m\times n}$ will also be important. If V is any module, we speak of linear combinations, linear dependence and linear independence, just as we do in a vector space. We must be careful not to apply to V any vector space results which depend upon division by non-zero scalars, the one field operation which may be lacking in the ring K. For example, if $\alpha_1, \ldots, \alpha_k$ are linearly dependent, we cannot conclude that some α_i is a linear combination of the others. This makes it more difficult to find bases in modules.

A **basis** for the module V is a linearly independent subset which spans (or generates) the module. This is the same definition which we gave for vector spaces; and, the important property of a basis $\mathcal{B}$ is that each element of V can be expressed uniquely as a linear combination of (some finite number of) elements of $\mathcal{B}$. If one admits into mathematics the Axiom of Choice (see Appendix), it can be shown that every vector space has a basis. The reader is well aware that a basis exists in any vector space which is spanned by a finite number of vectors. But this is not the case for modules. Therefore we need special names for modules which have bases and for modules which are spanned by finite numbers of elements.

Definition. *The* K-*module* V *is called a* **free module** *if it has a basis. If* V *has a finite basis containing* n *elements, then* V *is called a* **free K-module with n generators.**

6. Elementary Canonical Forms

6.1. Introduction

We have mentioned earlier that our principal aim is to study linear transformations on finite-dimensional vector spaces. By this time, we have seen many specific examples of linear transformations, and we have proved a few theorems about the general linear transformation. In the finite-dimensional case we have utilized ordered bases to represent such transformations by matrices, and this representation adds to our insight into their behavior. We have explored the vector space $L(V, W)$, consisting of the linear transformations from one space into another, and we have explored the linear algebra $L(V, V)$, consisting of the linear transformations of a space into itself.

In the next two chapters, we shall be preoccupied with linear operators. Our program is to select a single linear operator T on a finite-dimensional vector space V and to 'take it apart to see what makes it tick.' At this early stage, it is easiest to express our goal in matrix language: Given the linear operator T, find an ordered basis for V in which the matrix of T assumes an especially simple form.

Here is an illustration of what we have in mind. Perhaps the simplest matrices to work with, beyond the scalar multiples of the identity, are the diagonal matrices:

$$D = \begin{bmatrix} c_1 & 0 & 0 & \cdots & 0 \\ 0 & c_2 & 0 & \cdots & 0 \\ 0 & 0 & c_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & c_n \end{bmatrix}. \tag{6-1}$$

Let T be a linear operator on an n-dimensional space V. If we could find an ordered basis $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ for V in which T were represented by a diagonal matrix D (6-1), we would gain considerable information about T. For instance, simple numbers associated with T, such as the rank of T or the determinant of T, could be determined with little more than a glance at the matrix D. We could describe explicitly the range and the null space of T. Since $[T]_{\mathcal{B}} = D$ if and only if

$$(6\text{-}2) \qquad T\alpha_k = c_k\alpha_k, \qquad k = 1, \ldots, n$$

the range would be the subspace spanned by those α_k's for which $c_k \neq 0$ and the null space would be spanned by the remaining α_k's. Indeed, it seems fair to say that, if we knew a basis $\mathcal{B}$ and a diagonal matrix D such that $[T]_{\mathcal{B}} = D$, we could answer readily any question about T which might arise.

Can each linear operator T be represented by a diagonal matrix in some ordered basis? If not, for which operators T does such a basis exist? How can we find such a basis if there is one? If no such basis exists, what is the simplest type of matrix by which we can represent T? These are some of the questions which we shall attack in this (and the next) chapter. The form of our questions will become more sophisticated as we learn what some of the difficulties are.

6.2. Characteristic Values

The introductory remarks of the previous section provide us with a starting point for our attempt to analyze the general linear operator T. We take our cue from (6-2), which suggests that we should study vectors which are sent by T into scalar multiples of themselves.

Definition. *Let* V *be a vector space over the field* F *and let* T *be a linear operator on* V. *A* **characteristic value** *of* T *is a scalar* c *in* F *such that there is a non-zero vector* α *in* V *with* $T\alpha = c\alpha$. *If* c *is a characteristic value of* T, *then*

(a) *any* α *such that* $T\alpha = c\alpha$ *is called a* **characteristic vector** *of* T *associated with the characteristic value* c;

(b) *the collection of all* α *such that* $T\alpha = c\alpha$ *is called the* **characteristic space** *associated with* c.

Characteristic values are often called characteristic roots, latent roots, eigenvalues, proper values, or spectral values. In this book we shall use only the name 'characteristic values.'

If T is any linear operator and c is any scalar, the set of vectors α such that $T\alpha = c\alpha$ is a subspace of V. It is the null space of the linear transformation $(T - cI)$. We call c a characteristic value of T if this subspace is different from the zero subspace, i.e., if $(T - cI)$ fails to be 1:1. If the underlying space V is finite-dimensional, $(T - cI)$ fails to be 1:1 precisely when its determinant is different from 0. Let us summarize.

Theorem 1. *Let* T *be a linear operator on a finite-dimensional space* V *and let* c *be a scalar. The following are equivalent.*

(i) c *is a characteristic value of* T.
(ii) *The operator* (T − cI) *is singular (not invertible).*
(iii) *det* (T − cI) = 0.

The determinant criterion (iii) is very important because it tells us where to look for the characteristic values of T. Since $\det (T - cI)$ is a polynomial of degree n in the variable c, we will find the characteristic values as the roots of that polynomial. Let us explain carefully.

If $\mathcal{B}$ is any ordered basis for V and $A = [T]_{\mathcal{B}}$, then $(T - cI)$ is invertible if and only if the matrix $(A - cI)$ is invertible. Accordingly, we make the following definition.

Definition. *If* A *is an* n × n *matrix over the field* F, *a* **characteristic value of** A **in** F *is a scalar* c *in* F *such that the matrix* (A − cI) *is singular (not invertible).*

Since c is a characteristic value of A if and only if $\det (A - cI) = 0$, or equivalently if and only if $\det (cI - A) = 0$, we form the matrix $(xI - A)$ with polynomial entries, and consider the polynomial $f = \det (xI - A)$. Clearly the characteristic values of A in F are just the scalars c in F such that $f(c) = 0$. For this reason f is called the **characteristic polynomial** of A. It is important to note that f is a monic polynomial which has degree exactly n. This is easily seen from the formula for the determinant of a matrix in terms of its entries.

Lemma. *Similar matrices have the same characteristic polynomial.*

Proof. If $B = P^{-1}AP$, then

$$\begin{aligned}\det (xI - B) &= \det (xI - P^{-1}AP)\\ &= \det (P^{-1}(xI - A)P)\\ &= \det P^{-1} \cdot \det (xI - A) \cdot \det P\\ &= \det (xI - A). \quad \blacksquare\end{aligned}$$

This lemma enables us to define sensibly the characteristic polynomial of the operator T as the characteristic polynomial of any $n \times n$ matrix which represents T in some ordered basis for V. Just as for matrices, the characteristic values of T will be the roots of the characteristic polynomial for T. In particular, this shows us that T cannot have more than n distinct characteristic values. It is important to point out that T may not have any characteristic values.

EXAMPLE 1. Let T be the linear operator on R^2 which is represented in the standard ordered basis by the matrix

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

The characteristic polynomial for T (or for A) is

$$\det(xI - A) = \begin{vmatrix} x & 1 \\ -1 & x \end{vmatrix} = x^2 + 1.$$

Since this polynomial has no real roots, T has no characteristic values. If U is the linear operator on C^2 which is represented by A in the standard ordered basis, then U has two characteristic values, i and $-i$. Here we see a subtle point. In discussing the characteristic values of a matrix A, we must be careful to stipulate the field involved. The matrix A above has no characteristic values in R, but has the two characteristic values i and $-i$ in C.

EXAMPLE 2. Let A be the (real) 3×3 matrix

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

Then the characteristic polynomial for A is

$$\begin{vmatrix} x-3 & -1 & 1 \\ -2 & x-2 & 1 \\ -2 & -2 & x \end{vmatrix} = x^3 - 5x^2 + 8x - 4 = (x-1)(x-2)^2.$$

Thus the characteristic values of A are 1 and 2.

Suppose that T is the linear operator on R^3 which is represented by A in the standard basis. Let us find the characteristic vectors of T associated with the characteristic values, 1 and 2. Now

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

It is obvious at a glance that $A - I$ has rank equal to 2 (and hence $T - I$ has nullity equal to 1). So the space of characteristic vectors associated with the characteristic value 1 is one-dimensional. The vector $\alpha_1 = (1, 0, 2)$ spans the null space of $T - I$. Thus $T\alpha = \alpha$ if and only if α is a scalar multiple of α_1. Now consider

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

Evidently $A - 2I$ also has rank 2, so that the space of characteristic vectors associated with the characteristic value 2 has dimension 1. Evidently $T\alpha = 2\alpha$ if and only if α is a scalar multiple of $\alpha_2 = (1, 1, 2)$.

Definition. *Let* T *be a linear operator on the finite-dimensional space* . V. *We say that* T *is* **diagonalizable** *if there is a basis for* V *each vector of which is a characteristic vector of* T.

The reason for the name should be apparent; for, if there is an ordered basis $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ for V in which each α_i is a characteristic vector of T, then the matrix of T in the ordered basis $\mathcal{B}$ is diagonal. If $T\alpha_i = c_i\alpha_i$, then

$$[T]_{\mathcal{B}} = \begin{bmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & c_n \end{bmatrix}.$$

We certainly do not require that the scalars $c_1, \ldots, c_n$ be distinct; indeed, they may all be the same scalar (when T is a scalar multiple of the identity operator).

One could also define T to be diagonalizable when the characteristic vectors of T span V. This is only superficially different from our definition, since we can select a basis out of any spanning set of vectors.

For Examples 1 and 2 we purposely chose linear operators T on R^n which are not diagonalizable. In Example 1, we have a linear operator on R^2 which is not diagonalizable, because it has no characteristic values. In Example 2, the operator T has characteristic values; in fact, the characteristic polynomial for T factors completely over the real number field: $f = (x - 1)(x - 2)^2$. Nevertheless T fails to be diagonalizable. There is only a one-dimensional space of characteristic vectors associated with each of the two characteristic values of T. Hence, we cannot possibly form a basis for R^3 which consists of characteristic vectors of T.

Suppose that T is a diagonalizable linear operator. Let $c_1, \ldots, c_k$ be the *distinct* characteristic values of T. Then there is an ordered basis $\mathcal{B}$ in which T is represented by a diagonal matrix which has for its diagonal entries the scalars c_i, each repeated a certain number of times. If c_i is repeated d_i times, then (we may arrange that) the matrix has the block form

$$[T]_{\mathcal{B}} = \begin{bmatrix} c_1I_1 & 0 & \cdots & 0 \\ 0 & c_2I_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & c_kI_k \end{bmatrix} \tag{6-3}$$

where I_j is the $d_j \times d_j$ identity matrix. From that matrix we see two things. First, the characteristic polynomial for T is the product of (possibly repeated) linear factors:

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

If the scalar field F is algebraically closed, e.g., the field of complex numbers, every polynomial over F can be so factored (see Section 4.5); however, if F is not algebraically closed, we are citing a special property of T when we say that its characteristic polynomial has such a factorization. The second thing we see from (6-3) is that d_i, the number of times which c_i is repeated as root of f, is equal to the dimension of the space of characteristic vectors associated with the characteristic value c_i. That is because the nullity of a diagonal matrix is equal to the number of zeros which it has on its main diagonal, and the matrix $[T - c_iI]_{\mathcal{B}}$ has d_i zeros on its main diagonal. This relation between the dimension of the characteristic space and the multiplicity of the characteristic value as a root of f does not seem exciting at first; however, it will provide us with a simpler way of determining whether a given operator is diagonalizable.

Lemma. *Suppose that* $T\alpha = c\alpha$. *If* f *is any polynomial, then* $f(T)\alpha = f(c)\alpha$.

Proof. Exercise.

Lemma. *Let* T *be a linear operator on the finite-dimensional space* V. *Let* $c_1, \ldots, c_k$ *be the distinct characteristic values of* T *and let* W_i *be the space of characteristic vectors associated with the characteristic value* c_i. *If* $W = W_1 + \cdots + W_k$, *then*

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

In fact, if $\mathcal{B}_i$ *is an ordered basis for* W_i, *then* $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$ *is an ordered basis for* W.

Proof. The space $W = W_1 + \cdots + W_k$ is the subspace spanned by all of the characteristic vectors of T. Usually when one forms the sum W of subspaces W_i, one expects that $\dim W < \dim W_1 + \cdots + \dim W_k$ because of linear relations which may exist between vectors in the various spaces. This lemma states that the characteristic spaces associated with different characteristic values are independent of one another.

Suppose that (for each i) we have a vector β_i in W_i, and assume that $\beta_1 + \cdots + \beta_k = 0$. We shall show that $\beta_i = 0$ for each i. Let f be any polynomial. Since $T\beta_i = c_i\beta_i$, the preceding lemma tells us that

$$\begin{aligned} 0 = f(T)0 &= f(T)\beta_1 + \cdots + f(T)\beta_k \\ &= f(c_1)\beta_1 + \cdots + f(c_k)\beta_k. \end{aligned}$$

Choose polynomials $f_1, \ldots, f_k$ such that

$$f_i(c_j) = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases}$$

Then

$$0 = f_i(T)0 = \sum_j \delta_{ij}\beta_j$$
$$= \beta_i.$$

Now, let $\mathcal{B}_i$ be an ordered basis for W_i, and let $\mathcal{B}$ be the sequence $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$. Then $\mathcal{B}$ spans the subspace $W = W_1 + \cdots + W_k$. Also, $\mathcal{B}$ is a linearly independent sequence of vectors, for the following reason. 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 $\mathcal{B}_i$. From what we just did, we know that $\beta_i = 0$ for each i. Since each $\mathcal{B}_i$ is linearly independent, we see that we have only the trivial linear relation between the vectors in $\mathcal{B}$. ▌

Theorem 2. *Let* T *be a linear operator on a finite-dimensional space* V. *Let* $c_1, \ldots, c_k$ *be the distinct characteristic values of* T *and let* W_i *be the null space of* $(T - c_iI)$. *The following are equivalent.*

(i) T *is diagonalizable.*

(ii) *The characteristic polynomial for* T *is*

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

and dim $W_i = d_i$, $i = 1, \ldots, k$.

(iii) *dim* $W_1 + \cdots +$ *dim* $W_k =$ *dim* V.

Proof. We have observed that (i) implies (ii). If the characteristic polynomial f is the product of linear factors, as in (ii), then $d_1 + \cdots + d_k = \dim V$. For, the sum of the d_i's is the degree of the characteristic polynomial, and that degree is dim V. Therefore (ii) implies (iii). Suppose (iii) holds. By the lemma, we must have $V = W_1 + \cdots + W_k$, i.e., the characteristic vectors of T span V. ▌

The matrix analogue of Theorem 2 may be formulated as follows. Let A be an $n \times n$ matrix with entries in a field F, and let $c_1, \ldots, c_k$ be the distinct characteristic values of A in F. For each i, let W_i be the space of column matrices X (with entries in F) such that

$$(A - c_iI)X = 0,$$

and let $\mathcal{B}_i$ be an ordered basis for W_i. The bases $\mathcal{B}_1, \ldots, \mathcal{B}_k$ collectively string together to form the sequence of columns of a matrix P:

$$P = [P_1, P_2, \ldots] = (\mathcal{B}_1, \ldots, \mathcal{B}_k).$$

The matrix A is similar over F to a diagonal matrix if and only if P is a square matrix. When P is square, P is invertible and $P^{-1}AP$ is diagonal.

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

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

Let us indicate how one might compute the characteristic polynomial, using various row and column operations:

$$\begin{aligned}
\begin{vmatrix} x-5 & 6 & 6 \\ 1 & x-4 & -2 \\ -3 & 6 & x+4 \end{vmatrix} &= \begin{vmatrix} x-5 & 0 & 6 \\ 1 & x-2 & -2 \\ -3 & 2-x & x+4 \end{vmatrix} \\
&= (x-2)\begin{vmatrix} x-5 & 0 & 6 \\ 1 & 1 & -2 \\ -3 & -1 & x+4 \end{vmatrix} \\
&= (x-2)\begin{vmatrix} x-5 & 0 & 6 \\ 1 & 1 & -2 \\ -2 & 0 & x+2 \end{vmatrix} \\
&= (x-2)\begin{vmatrix} x-5 & 6 \\ -2 & x+2 \end{vmatrix} \\
&= (x-2)(x^2-3x+2) \\
&= (x-2)^2(x-1).
\end{aligned}$$

What are the dimensions of the spaces of characteristic vectors associated with the two characteristic values? We have

$$A - I = \begin{bmatrix} 4 & -6 & -6 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{bmatrix}$$

$$A - 2I = \begin{bmatrix} 3 & -6 & -6 \\ -1 & 2 & 2 \\ 3 & -6 & -6 \end{bmatrix}.$$

We know that $A - I$ is singular and obviously rank $(A - I) \geq 2$. Therefore, rank $(A - I) = 2$. It is evident that rank $(A - 2I) = 1$.

Let W_1, W_2 be the spaces of characteristic vectors associated with the characteristic values 1, 2. We know that dim $W_1 = 1$ and dim $W_2 = 2$. By Theorem 2, T is diagonalizable. It is easy to exhibit a basis for R^3 in which T is represented by a diagonal matrix. The null space of $(T - I)$ is spanned by the vector $\alpha_1 = (3, -1, 3)$ and so $\{\alpha_1\}$ is a basis for W_1. The null space of $T - 2I$ (i.e., the space W_2) consists of the vectors (x_1, x_2, x_3) with $x_1 = 2x_2 + 2x_3$. Thus, one example of a basis for W_2 is

$$\begin{aligned} \alpha_2 &= (2, 1, 0) \\ \alpha_3 &= (2, 0, 1). \end{aligned}$$

If $\mathcal{B} = \{\alpha_1, \alpha_2, \alpha_3\}$, then $[T]_{\mathcal{B}}$ is the diagonal matrix

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

The fact that T is diagonalizable means that the original matrix A is similar (over R) to the diagonal matrix D. The matrix P which enables us to change coordinates from the basis $\mathcal{B}$ to the standard basis is (of course) the matrix which has the transposes of α_1, α_2, α_3 as its column vectors:

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

Furthermore, $AP = PD$, so that

$$P^{-1}AP = D.$$

Exercises

1. In each of the following cases, let T be the linear operator on R^2 which is represented by the matrix A in the standard ordered basis for R^2, and let U be the linear operator on C^2 represented by A in the standard ordered basis. Find the characteristic polynomial for T and that for U, find the characteristic values of each operator, and for each such characteristic value c find a basis for the corresponding space of characteristic vectors.

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

2. Let V be an n-dimensional vector space over F. What is the characteristic polynomial of the identity operator on V? What is the characteristic polynomial for the zero operator?

3. Let A be an $n \times n$ triangular matrix over the field F. Prove that the characteristic values of A are the diagonal entries of A, i.e., the scalars A_{ii}.

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

$$\begin{bmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{bmatrix}.$$

Prove that T is diagonalizable by exhibiting a basis for R^3, each vector of which is a characteristic vector of T.

5. Let

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

Is A similar over the field R to a diagonal matrix? Is A similar over the field C to a diagonal matrix?

6. Let T be the linear operator on R^4 which is represented in the standard ordered basis by the matrix

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

Under what conditions on a, b, and c is T diagonalizable?

7. Let T be a linear operator on the n-dimensional vector space V, and suppose that T has n *distinct* characteristic values. Prove that T is diagonalizable.

8. Let A and B be $n \times n$ matrices over the field F. Prove that if $(I - AB)$ is invertible, then $I - BA$ is invertible and

$$(I - BA)^{-1} = I + B(I - AB)^{-1}A.$$

9. Use the result of Exercise 8 to prove that, if A and B are $n \times n$ matrices over the field F, then AB and BA have precisely the same characteristic values in F.

10. Suppose that A is a 2×2 matrix with real entries which is symmetric ($A^t = A$). Prove that A is similar over R to a diagonal matrix.

11. Let N be a 2×2 complex matrix such that $N^2 = 0$. Prove that either $N = 0$ or N is similar over C to

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

12. Use the result of Exercise 11 to prove the following: If A is a 2×2 matrix with complex entries, then A is similar over C to a matrix of one of the two types

$$\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \quad \begin{bmatrix} a & 0 \\ 1 & a \end{bmatrix}.$$

13. Let V be the vector space of all functions from R into R which are continuous, i.e., the space of continuous real-valued functions on the real line. Let T be the linear operator on V defined by

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

Prove that T has no characteristic values.

14. Let A be an $n \times n$ *diagonal* matrix with characteristic polynomial

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

where $c_1, \ldots, c_k$ are distinct. Let V be the space of $n \times n$ matrices B such that $AB = BA$. Prove that the dimension of V is $d_1^2 + \cdots + d_k^2$.

15. 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 be the linear operator 'left multiplication by A' on V. Is it true that A and T have the same characteristic values?

6.3. Annihilating Polynomials

In attempting to analyze a linear operator T, one of the most useful things to know is the class of polynomials which annihilate T. Specifically, suppose T is a linear operator on V, a vector space over the field F. If p is a polynomial over F, then $p(T)$ is again a linear operator on V. If q is another polynomial over F, then

$$\begin{aligned}(p + q)(T) &= p(T) + q(T)\\ (pq)(T) &= p(T)q(T).\end{aligned}$$

Therefore, the collection of polynomials p which annihilate T, in the sense that

$$p(T) = 0,$$

is an ideal in the polynomial algebra $F[x]$. It may be the zero ideal, i.e., it may be that T is not annihilated by any non-zero polynomial. But, that cannot happen if the space V is finite-dimensional.

Suppose T is a linear operator on the n-dimensional space V. Look at the first $(n^2 + 1)$ powers of T:

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

This is a sequence of $n^2 + 1$ operators in $L(V, V)$, the space of linear operators on V. The space $L(V, V)$ has dimension n^2. Therefore, that sequence of $n^2 + 1$ operators must be linearly dependent, i.e., we have

$$c_0I + c_1T + \cdots + c_{n^2}T^{n^2} = 0$$

for some scalars c_i, not all zero. So, the ideal of polynomials which annihilate T contains a non-zero polynomial of degree n^2 or less.

According to Theorem 5 of Chapter 4, every polynomial ideal consists of all multiples of some fixed monic polynomial, the generator of the ideal. Thus, there corresponds to the operator T a monic polynomial p with this property: If f is a polynomial over F, then $f(T) = 0$ if and only if $f = pg$, where g is some polynomial over F.

Definition. *Let* T *be a linear operator on a finite-dimensional vector space* V *over the field* F. *The* **minimal polynomial** *for* T *is the (unique) monic generator of the ideal of polynomials over* F *which annihilate* T.

The name 'minimal polynomial' stems from the fact that the generator of a polynomial ideal is characterized by being the monic polynomial of minimum degree in the ideal. That means that the minimal polynomial p for the linear operator T is uniquely determined by these three properties:

(1) p is a monic polynomial over the scalar field F.

(2) $p(T) = 0$.

(3) No polynomial over F which annihilates T has smaller degree than p has.

If A is an $n \times n$ matrix over F, we define the **minimal polynomial** for A in an analogous way, as the unique monic generator of the ideal of all polynomials over F which annihilate A. If the operator T is represented in some ordered basis by the matrix A, then T and A have the same minimal polynomial. That is because $f(T)$ is represented in the basis by the matrix $f(A)$, so that $f(T) = 0$ if and only if $f(A) = 0$.

From the last remark about operators and matrices it follows that similar matrices have the same minimal polynomial. That fact is also clear from the definitions because

$$f(P^{-1}AP) = P^{-1}f(A)P$$

for every polynomial f.

There is another basic remark which we should make about minimal polynomials of matrices. Suppose that A is an $n \times n$ matrix with entries in the field F. Suppose that F_1 is a field which contains F as a subfield. (For example, A might be a matrix with rational entries, while F_1 is the field of real numbers. Or, A might be a matrix with real entries, while F_1 is the field of complex numbers.) We may regard A either as an $n \times n$ matrix over F or as an $n \times n$ matrix over F_1. On the surface, it might appear that we obtain two different minimal polynomials for A. Fortunately that is not the case; and we must see why. What is the definition of the minimal polynomial for A, regarded as an $n \times n$ matrix over the field F? We consider all monic polynomials with coefficients in F which annihilate A, and we choose the one of least degree. If f is a monic polynomial over F:

$$f = x^k + \sum_{j=0}^{k-1} a_j x^j \tag{6-4}$$

then $f(A) = 0$ merely says that we have a linear relation between the powers of A:

$$A^k + a_{k-1}A^{k-1} + \cdots + a_1A + a_0I = 0. \tag{6-5}$$

The degree of the minimal polynomial is the least positive integer k such that there is a linear relation of the form (6-5) between the powers I, $A, \ldots, A^k$. Furthermore, by the uniqueness of the minimal polynomial, there is for that k one and only one relation of the form (6-5); i.e., once the minimal k is determined, there are unique scalars $a_0, \ldots, a_{k-1}$ in F such that (6-5) holds. They are the coefficients of the minimal polynomial.

Now (for each k) we have in (6-5) a system of n^2 linear equations for the 'unknowns' $a_0, \ldots, a_{k-1}$. Since the entries of A lie in F, the coefficients of the system of equations (6-5) are in F. Therefore, if the system has a solution with $a_0, \ldots, a_{k-1}$ in F_1 it has a solution with $a_0, \ldots, a_{k-1}$ in F. (See the end of Section 1.4.) It should now be clear that the two minimal polynomials are the same.

What do we know thus far about the minimal polynomial for a linear operator on an n-dimensional space? Only that its degree does not exceed n^2. That turns out to be a rather poor estimate, since the degree cannot exceed n. We shall prove shortly that the operator is annihilated by its characteristic polynomial. First, let us observe a more elementary fact.

Theorem 3. *Let* T *be a linear operator on an* n*-dimensional vector space* V *[or, let* A *be an* n × n *matrix]. The characteristic and minimal polynomials for* T *[for* A*] have the same roots, except for multiplicities.*

Proof. Let p be the minimal polynomial for T. Let c be a scalar. What we want to show is that $p(c) = 0$ if and only if c is a characteristic value of T.

First, suppose $p(c) = 0$. Then

$$p = (x - c)q$$

where q is a polynomial. Since $\deg q < \deg p$, the definition of the minimal polynomial p tells us that $q(T) \neq 0$. Choose a vector β such that $q(T)\beta \neq 0$. Let $\alpha = q(T)\beta$. Then

$$\begin{aligned} 0 &= p(T)\beta \\ &= (T - cI)q(T)\beta \\ &= (T - cI)\alpha \end{aligned}$$

and thus, c is a characteristic value of T.

Now, suppose that c is a characteristic value of T, say, $T\alpha = c\alpha$ with $\alpha \neq 0$. As we noted in a previous lemma,

$$p(T)\alpha = p(c)\alpha.$$

Since $p(T) = 0$ and $\alpha \neq 0$, we have $p(c) = 0$. ▌

Let T be a diagonalizable linear operator and let $c_1, \ldots, c_k$ be the distinct characteristic values of T. Then it is easy to see that the minimal polynomial for T is the polynomial

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

If α is a characteristic vector, then one of the operators $T - c_1I, \ldots, T - c_kI$ sends α into 0. Therefore

$$(T - c_1I) \cdots (T - c_kI)\alpha = 0$$

for every characteristic vector α. There is a basis for the underlying space which consists of characteristic vectors of T; hence

$$p(T) = (T - c_1I) \cdots (T - c_kI) = 0.$$

What we have concluded is this. If T is a diagonalizable linear operator, then the minimal polynomial for T is a product of distinct linear factors. As we shall soon see, that property characterizes diagonalizable operators.

Example 4. Let's try to find the minimal polynomials for the operators in Examples 1, 2, and 3. We shall discuss them in reverse order. The operator in Example 3 was found to be diagonalizable with characteristic polynomial

$$f = (x - 1)(x - 2)^2.$$

From the preceding paragraph, we know that the minimal polynomial for T is

$$p = (x - 1)(x - 2).$$

The reader might find it reassuring to verify directly that

$$(A - I)(A - 2I) = 0.$$

In Example 2, the operator T also had the characteristic polynomial $f = (x - 1)(x - 2)^2$. But, this T is not diagonalizable, so we don't know that the minimal polynomial is $(x - 1)\ (x - 2)$. What do we know about the minimal polynomial in this case? From Theorem 3 we know that its roots are 1 and 2, with some multiplicities allowed. Thus we search for p among polynomials of the form $(x - 1)^k(x - 2)^l$, $k \geq 1, l \geq 1$. Try $(x - 1)$ $(x - 2)$:

$$\begin{aligned}(A - I)(A - 2I) &= \begin{bmatrix} 2 & 1 & -1 \\ 2 & 1 & -1 \\ 2 & 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 2 & 2 & -2 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 & -1 \\ 2 & 0 & -1 \\ 4 & 0 & -2 \end{bmatrix}.\end{aligned}$$

Thus, the minimal polynomial has degree at least 3. So, next we should try either $(x - 1)^2(x - 2)$ or $(x - 1)(x - 2)^2$. The second, being the characteristic polynomial, would seem a less random choice. One can readily compute that $(A - I)(A - 2I)^2 = 0$. Thus the minimal polynomial for T is its characteristic polynomial.

In Example 1 we discussed the linear operator T on R^2 which is represented in the standard basis by the matrix

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

The characteristic polynomial is $x^2 + 1$, which has no real roots. To determine the minimal polynomial, forget about T and concentrate on A. As a complex 2×2 matrix, A has the characteristic values i and $-i$. Both roots must appear in the minimal polynomial. Thus the minimal polynomial is divisible by $x^2 + 1$. It is trivial to verify that $A^2 + I = 0$. So the minimal polynomial is $x^2 + 1$.

Theorem 4 (Cayley-Hamilton). *Let* T *be a linear operator on a finite dimensional vector space* V. *If* f *is the characteristic polynomial for* T, *then* f(T) = 0; *in other words, the minimal polynomial divides the characteristic polynomial for* T.

Proof. Later on we shall give two proofs of this result independent of the one to be given here. The present proof, although short, may be difficult to understand. Aside from brevity, it has the virtue of providing an illuminating and far from trivial application of the general theory of determinants developed in Chapter 5.

Let K be the commutative ring with identity consisting of all polynomials in T. Of course, K is actually a commutative algebra with identity over the scalar field. Choose an ordered basis $\{\alpha_1, \ldots, \alpha_n\}$ for V, and let A be the matrix which represents T in the given basis. Then

$$T\alpha_i = \sum_{j=1}^{n} A_{ji}\alpha_j, \qquad 1 \le i \le n.$$

These equations may be written in the equivalent form

$$\sum_{j=1}^{n} (\delta_{ij}T - A_{ji}I)\alpha_j = 0, \qquad 1 \le i \le n.$$

Let B denote the element of $K^{n\times n}$ with entries

$$B_{ij} = \delta_{ij}T - A_{ji}I.$$

When $n = 2$

$$B = \begin{bmatrix} T - A_{11}I & -A_{21}I \\ -A_{12}I & T - A_{22}I \end{bmatrix}$$

and

$$\begin{aligned} \det B &= (T - A_{11}I)(T - A_{22}I) - A_{12}A_{21}I \\ &= T^2 - (A_{11} + A_{22})T + (A_{11}A_{22} - A_{12}A_{21})I \\ &= f(T) \end{aligned}$$

where f is the characteristic polynomial:

$$f = x^2 - (\text{trace } A)x + \det A.$$

For the case $n > 2$, it is also clear that

$$\det B = f(T)$$

since f is the determinant of the matrix $xI - A$ whose entries are the polynomials

$$(xI - A)_{ij} = \delta_{ij}x - A_{ji}.$$

We wish to show that $f(T) = 0$. In order that $f(T)$ be the zero operator, it is necessary and sufficient that $(\det B)\alpha_k = 0$ for $k = 1, \ldots, n$. By the definition of B, the vectors $\alpha_1, \ldots, \alpha_n$ satisfy the equations

$$\sum_{j=1}^{n} B_{ij}\alpha_j = 0, \qquad 1 \le i \le n. \tag{6-6}$$

When $n = 2$, it is suggestive to write (6-6) in the form

$$\begin{bmatrix} T - A_{11}I & -A_{21}I \\ -A_{12}I & T - A_{22}I \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

In this case, the classical adjoint, adj B is the matrix

$$\tilde{B} = \begin{bmatrix} T - A_{22}I & A_{21}I \\ A_{12}I & T - A_{11}I \end{bmatrix}$$

and

$$\tilde{B}B = \begin{bmatrix} \det B & 0 \\ 0 & \det B \end{bmatrix}.$$

Hence, we have

$$\begin{aligned}(\det B)\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix} &= (\tilde{B}B)\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix}\\ &= \tilde{B}\left(B\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix}\right)\\ &= \begin{bmatrix}0\\ 0\end{bmatrix}.\end{aligned}$$

In the general case, let $\tilde{B} = \operatorname{adj} B$. Then by (6-6)

$$\sum_{j=1}^{n} \tilde{B}_{ki}B_{ij}\alpha_j = 0$$

for each pair k, i, and summing on i, we have

$$\begin{aligned}0 &= \sum_{i=1}^{n}\sum_{j=1}^{n} \tilde{B}_{ki}B_{ij}\alpha_j\\ &= \sum_{j=1}^{n}\left(\sum_{i=1}^{n} \tilde{B}_{ki}B_{ij}\right)\alpha_j.\end{aligned}$$

Now $\tilde{B}B = (\det B)I$, so that

$$\sum_{i=1}^{n} \tilde{B}_{ki}B_{ij} = \delta_{kj}\det B.$$

Therefore

$$\begin{aligned}0 &= \sum_{j=1}^{n} \delta_{kj}(\det B)\alpha_j\\ &= (\det B)\alpha_k, \qquad 1 \le k \le n. \quad \blacksquare\end{aligned}$$

The Cayley-Hamilton theorem is useful to us at this point primarily because it narrows down the search for the minimal polynomials of various operators. If we know the matrix A which represents T in some ordered basis, then we can compute the characteristic polynomial f. We know that the minimal polynomial p divides f and that the two polynomials have the same roots. There is no method for computing precisely the roots of a polynomial (unless its degree is small); however, if f factors

(6-7) $\quad f = (x - c_1)^{d_1} \cdots (x - c_k)^{d_k}, \qquad c_1, \ldots, c_k$ distinct, $d_i \ge 1$

then

(6-8) $\quad p = (x - c_1)^{r_1} \cdots (x - c_k)^{r_k}, \qquad 1 \le r_j \le d_j.$

That is all we can say in general. If f is the polynomial (6-7) and has degree n, then for every polynomial p as in (6-8) we can find an $n \times n$ matrix which has f as its characteristic polynomial and p as its minimal polynomial. We shall not prove this now. But, we want to emphasize the fact that the knowledge that the characteristic polynomial has the form (6-7) tells us that the minimal polynomial has the form (6-8), and it tells us nothing else about p.

Example 5. Let A be the 4×4 (rational) matrix

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

The powers of A are easy to compute:

$$A^2 = \begin{bmatrix} 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \\ 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \end{bmatrix}$$

$$A^3 = \begin{bmatrix} 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 \\ 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 \end{bmatrix}.$$

Thus $A^3 = 4A$, i.e., if $p = x^3 - 4x = x(x + 2)(x - 2)$, then $p(A) = 0$. The minimal polynomial for A must divide p. That minimal polynomial is obviously not of degree 1, since that would mean that A was a scalar multiple of the identity. Hence, the candidates for the minimal polynomial are: p, $x(x + 2)$, $x(x - 2)$, $x^2 - 4$. The three quadratic polynomials can be eliminated because it is obvious at a glance that $A^2 \neq -2A$, $A^2 \neq 2A$, $A^2 \neq 4I$. Therefore p is the minimal polynomial for A. In particular 0, 2, and -2 are the characteristic values of A. One of the factors x, $x - 2$, $x + 2$ must be repeated twice in the characteristic polynomial. Evidently, rank $(A) = 2$. Consequently there is a two-dimensional space of characteristic vectors associated with the characteristic value 0. From Theorem 2, it should now be clear that the characteristic polynomial is $x^2(x^2 - 4)$ and that A is similar over the field of rational numbers to the matrix

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

Exercises

1. Let V be a finite-dimensional vector space. What is the minimal polynomial for the identity operator on V? What is the minimal polynomial for the zero operator?

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

$$c_1 d_1 + \cdots + c_k d_k = \text{trace}\,(A).$$

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