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

4. Polynomials

4.1. Algebras

The purpose of this chapter is to establish a few of the basic properties of the algebra of polynomials over a field. The discussion will be facilitated if we first introduce the concept of a linear algebra over a field.

Definition. *Let* F *be a field. A* **linear algebra over the field** F *is a vector space* $\mathcal{A}$ *over* F *with an additional operation called* **multiplication of vectors** *which associates with each pair of vectors* α, β *in* $\mathcal{A}$ *a vector* $\alpha\beta$ *in* $\mathcal{A}$ *called the* **product** *of* α *and* β *in such a way that*

(a) *multiplication is associative,*

$$\alpha(\beta\gamma) = (\alpha\beta)\gamma$$

(b) *multiplication is distributive with respect to addition,*

$$\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma \quad \textit{and} \quad (\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma$$

(c) *for each scalar* c *in* F,

$$c(\alpha\beta) = (c\alpha)\beta = \alpha(c\beta).$$

If there is an element 1 *in* $\mathcal{A}$ *such that* $1\alpha = \alpha 1 = \alpha$ *for each* α *in* $\mathcal{A}$, *we call* $\mathcal{A}$ *a* **linear algebra with identity over** F, *and call* 1 *the* **identity** *of* $\mathcal{A}$. *The algebra* $\mathcal{A}$ *is called* **commutative** *if* $\alpha\beta = \beta\alpha$ *for all* α *and* β *in* $\mathcal{A}$.

EXAMPLE 1. The set of $n \times n$ matrices over a field, with the usual operations, is a linear algebra with identity; in particular the field itself is an algebra with identity. This algebra is not commutative if $n \geq 2$. The field itself is (of course) commutative.

EXAMPLE 2. The space of all linear operators on a vector space, with composition as the product, is a linear algebra with identity. It is commutative if and only if the space is one-dimensional.

The reader may have had some experience with the dot product and cross product of vectors in R^3. If so, he should observe that neither of these products is of the type described in the definition of a linear algebra. The dot product is a 'scalar product,' that is, it associates with a pair of vectors a scalar, and thus it is certainly not the type of product we are presently discussing. The cross product does associate a vector with each pair of vectors in R^3; however, this is not an associative multiplication.

The rest of this section will be devoted to the construction of an algebra which is significantly different from the algebras in either of the preceding examples. Let F be a field and S the set of non-negative integers. By Example 3 of Chapter 2, the set of all functions from S into F is a vector space over F. We shall denote this vector space by F^∞. The vectors in F^∞ are therefore infinite sequences $f = (f_0, f_1, f_2, \ldots)$ of scalars f_i in F. If $g = (g_0, g_1, g_2, \ldots)$, g_i in F, and a, b are scalars in F, $af + bg$ is the infinite sequence given by

$$(4\text{-}1) \qquad af + bg = (af_0 + bg_0, af_1 + bg_1, af_2 + bg_2, \ldots).$$

We define a product in F^∞ by associating with each pair of vectors f and g in F^∞ the vector fg which is given by

$$(4\text{-}2) \qquad (fg)_n = \sum_{i=0}^{n} f_i g_{n-i}, \qquad n = 0, 1, 2, \ldots .$$

Thus

$$fg = (f_0g_0, f_0g_1 + f_1g_0, f_0g_2 + f_1g_1 + f_2g_0, \ldots)$$

and as

$$(gf)_n = \sum_{i=0}^{n} g_i f_{n-i} = \sum_{i=0}^{n} f_i g_{n-i} = (fg)_n$$

for $n = 0, 1, 2, \ldots$, it follows that multiplication is commutative, $fg = gf$. If h also belongs to F^∞, then

$$\begin{aligned}
[(fg)h]_n &= \sum_{i=0}^{n} (fg)_i h_{n-i} \\
&= \sum_{i=0}^{n} \left(\sum_{j=0}^{i} f_j g_{i-j} \right) h_{n-i} \\
&= \sum_{i=0}^{n} \sum_{j=0}^{i} f_i g_{i-j} h_{n-i} \\
&= \sum_{j=0}^{n} f_j \sum_{i=0}^{n-j} g_i h_{n-i-j} \\
&= \sum_{j=0}^{n} f_j (gh)_{n-j} = [f(gh)]_n
\end{aligned}$$

for $n = 0, 1, 2, \ldots$, so that

$$(fg)h = f(gh). \tag{4-3}$$

We leave it to the reader to verify that the multiplication defined by (4-2) satisfies (b) and (c) in the definition of a linear algebra, and that the vector $1 = (1, 0, 0, \ldots)$ serves as an identity for F^∞. Then F^∞, with the operations defined above, is a commutative linear algebra with identity over the field F.

The vector $(0, 1, 0, \ldots, 0, \ldots)$ plays a distinguished role in what follows and we shall consistently denote it by x. Throughout this chapter x will never be used to denote an element of the field F. The product of x with itself n times will be denoted by x^n and we shall put $x^0 = 1$. Then

$$x^2 = (0, 0, 1, 0, \ldots), \qquad x^3 = (0, 0, 0, 1, 0, \ldots)$$

and in general for each integer $k \geq 0$, $(x^k)_k = 1$ and $(x^k)_n = 0$ for all non-negative integers $n \neq k$. In concluding this section we observe that the set consisting of $1, x, x^2, \ldots$ is both independent and infinite. Thus the algebra F^∞ is not finite-dimensional.

The algebra F^∞ is sometimes called the **algebra of formal power series** over F. The element $f = (f_0, f_1, f_2, \ldots)$ is frequently written

$$f = \sum_{n=0}^{\infty} f_n x^n. \tag{4-4}$$

This notation is very convenient for dealing with the algebraic operations. When used, it must be remembered that it is purely formal. There are no 'infinite sums' in algebra, and the power series notation (4-4) is not intended to suggest anything about convergence, if the reader knows what that is. By using sequences, we were able to define carefully an algebra in which the operations behave like addition and multiplication of formal power series, without running the risk of confusion over such things as infinite sums.

4.2. The Algebra of Polynomials

We are now in a position to define a polynomial over the field F.

Definition. *Let* F[x] *be the subspace of* F^∞ *spanned by the vectors* 1, x, x^2, ... *An element of* F[x] *is called a* **polynomial over** F.

Since $F[x]$ consists of all (finite) linear combinations of x and its powers, a non-zero vector f in F^∞ is a polynomial if and only if there is an integer $n \geq 0$ such that $f_n \neq 0$ and such that $f_k = 0$ for all integers $k > n$; this integer (when it exists) is obviously unique and is called the **degree** of f. We denote the degree of a polynomial f by deg f, and do not assign a degree to the 0-polynomial. If f is a non-zero polynomial of degree n it follows that

$$f = f_0x^0 + f_1x + f_2x^2 + \cdots + f_nx^n, \qquad f_n \neq 0. \tag{4-5}$$

The scalars $f_0, f_1, \ldots, f_n$ are sometimes called the **coefficients** of f, and we may say that f is a polynomial with coefficients in F. We shall call polynomials of the form cx^0 **scalar polynomials,** and frequently write c for cx^0. A non-zero polynomial f of degree n such that $f_n = 1$ is said to be a **monic** polynomial.

The reader should note that polynomials are not the same sort of objects as the polynomial functions on F which we have discussed on several occasions. If F contains an infinite number of elements, there is a natural isomorphism between $F[x]$ and the algebra of polynomial functions on F. We shall discuss that in the next section. Let us verify that $F[x]$ is an algebra.

Theorem 1. *Let* f *and* g *be non-zero polynomials over* F. *Then*

(i) fg *is a non-zero polynomial;*

(ii) *deg* (fg) = *deg* f + *deg* g;

(iii) fg *is a monic polynomial if both* f *and* g *are monic polynomials;*

(iv) fg *is a scalar polynomial if and only if both* f *and* g *are scalar polynomials;*

(v) *if* f + g ≠ 0,

$$\text{deg}\,(f + g) \leq \max\,(\text{deg}\,f, \text{deg}\,g).$$

Proof. Suppose f has degree m and that g has degree n. If k is a non-negative integer,

$$(fg)_{m+n+k} = \sum_{i=0}^{m+n+k} f_ig_{m+n+k-i}.$$

In order that $f_ig_{m+n+k-i} \neq 0$, it is necessary that $i \leq m$ and $m + n + k - i \leq n$. Hence it is necessary that $m + k \leq i \leq m$, which implies $k = 0$ and $i = m$. Thus

$$(fg)_{m+n} = f_mg_n \tag{4-6}$$

and

$$(fg)_{m+n+k} = 0, \qquad k > 0. \tag{4-7}$$

The statements (i), (ii), (iii) follow immediately from (4-6) and (4-7), while (iv) is a consequence of (i) and (ii). We leave the verification of (v) to the reader. ▌

Corollary 1. *The set of all polynomials over a given field* F *equipped with the operations (4-1) and (4-2) is a commutative linear algebra with identity over* F.

Proof. Since the operations (4-1) and (4-2) are those defined in the algebra F^∞ and since $F[x]$ is a subspace of F^∞, it suffices to prove that the product of two polynomials is again a polynomial. This is trivial when one of the factors is 0 and otherwise follows from (i). ▌

Corollary 2. *Suppose* f, g, *and* h *are polynomials over the field* F *such that* f $\neq$ 0 *and* fg = fh. *Then* g = h.

Proof. Since $fg = fh$, $f(g - h) = 0$, and as $f \neq 0$ it follows at once from (i) that $g - h = 0$. ▌

Certain additional facts follow rather easily from the proof of Theorem 1, and we shall mention some of these.

Suppose

$$f = \sum_{i=0}^{m} f_i x^i \quad \text{and} \quad g = \sum_{j=0}^{n} g_j x^j.$$

Then from (4-7) we obtain,

$$fg = \sum_{s=0}^{m+n} \left(\sum_{r=0}^{s} f_r g_{s-r} \right) x^s. \tag{4-8}$$

The reader should verify, in the special case $f = cx^m$, $g = dx^n$ with c, d in F, that (4-8) reduces to

$$(cx^m)(dx^n) = cdx^{m+n}. \tag{4-9}$$

Now from (4-9) and the distributive laws in $F[x]$, it follows that the product in (4-8) is also given by

$$\sum_{i,j} f_i g_j x^{i+j} \tag{4-10}$$

where the sum is extended over all integer pairs i, j such that $0 \leq i \leq m$, and $0 \leq j \leq n$.

Definition. *Let* $\mathcal{A}$ *be a linear algebra with identity over the field* F. *We shall denote the identity of* $\mathcal{A}$ *by* 1 *and make the convention that* $\alpha^0 = 1$ *for each* α *in* $\mathcal{A}$. *Then to each polynomial* $\mathrm{f} = \sum_{\mathrm{i}=0}^{\mathrm{n}} \mathrm{f_i x^i}$ *over* F *and* α *in* $\mathcal{A}$ *we associate an element* $\mathrm{f}(\alpha)$ *in* $\mathcal{A}$ *by the rule*

$$f(\alpha) = \sum_{\mathrm{i}=0}^{\mathrm{n}} f_{\mathrm{i}} \alpha^{\mathrm{i}}.$$

EXAMPLE 3. Let C be the field of complex numbers and let $f = x^2 + 2$.

(a) If $\mathcal{A} = C$ and z belongs to C, $f(z) = z^2 + 2$, in particular $f(2) = 6$ and

$$f\left(\frac{1+i}{1-i}\right) = 1.$$

(b) If $\mathcal{A}$ is the algebra of all 2 × 2 matrices over C and if

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

then

$$f(B) = 2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}^2 = \begin{bmatrix} 3 & 0 \\ -3 & 6 \end{bmatrix}.$$

(c) If $\mathcal{A}$ is the algebra of all linear operators on C^3 and T is the element of $\mathcal{A}$ given by

$$T(c_1, c_2, c_3) = (i\sqrt{2}\, c_1, c_2, i\sqrt{2}\, c_3)$$

then $f(T)$ is the linear operator on C^3 defined by

$$f(T)(c_1, c_2, c_3) = (0, 3c_2, 0).$$

(d) If $\mathcal{A}$ is the algebra of all polynomials over C and $g = x^4 + 3i$, then $f(g)$ is the polynomial in $\mathcal{A}$ given by

$$f(g) = -7 + 6ix^4 + x^8.$$

The observant reader may notice in connection with this last example that if f is a polynomial over any field and x is the polynomial $(0, 1, 0, \ldots)$ then $f = f(x)$, but he is advised to forget this fact.

Theorem 2. *Let* F *be a field and* $\mathcal{A}$ *be a linear algebra with identity over* F. *Suppose* f *and* g *are polynomials over* F, *that* α *is an element of* $\mathcal{A}$, *and that* c *belongs to* F. *Then*

(i) $(\mathrm{cf} + \mathrm{g})(\alpha) = \mathrm{cf}(\alpha) + \mathrm{g}(\alpha)$;
(ii) $(\mathrm{fg})(\alpha) = \mathrm{f}(\alpha)\mathrm{g}(\alpha)$.

Proof. As (i) is quite easy to establish, we shall only prove (ii). Suppose

$$f = \sum_{i=0}^{m} f_i x^i \quad \text{and} \quad g = \sum_{j=0}^{n} g_j x^j.$$

By (4-10),

$$fg = \sum_{i,j} f_i g_j x^{i+j}$$

and hence by (i),

$$\begin{aligned} (fg)(\alpha) &= \sum_{i,j} f_i g_j \alpha^{i+j} \\ &= \left(\sum_{i=0}^{m} f_i \alpha^i\right)\left(\sum_{j=0}^{n} g_j \alpha^j\right) \\ &= f(\alpha)g(\alpha). \quad \blacksquare \end{aligned}$$

Exercises

1. Let F be a subfield of the complex numbers and let A be the following 2 × 2 matrix over F

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

For each of the following polynomials f over F, compute $f(A)$.

(a) $f = x^2 - x + 2$;

(b) $f = x^3 - 1$;

(c) $f = x^2 - 5x + 7$.

2. Let T be the linear operator on R^3 defined by

$$T(x_1, x_2, x_3) = (x_1, x_3, -2x_2 - x_3).$$

Let f be the polynomial over R defined by $f = -x^3 + 2$. Find $f(T)$.

3. Let A be an $n \times n$ diagonal matrix over the field F, i.e., a matrix satisfying $A_{ij} = 0$ for $i \neq j$. Let f be the polynomial over F defined by

$$f = (x - A_{11}) \cdots (x - A_{nn}).$$

What is the matrix $f(A)$?

4. If f and g are independent polynomials over a field F and h is a non-zero polynomial over F, show that fh and gh are independent.

5. If F is a field, show that the product of two non-zero elements of F^∞ is non-zero.

6. Let S be a set of non-zero polynomials over a field F. If no two elements of S have the same degree, show that S is an independent set in $F[x]$.

7. If a and b are elements of a field F and $a \neq 0$, show that the polynomials 1, $ax + b$, $(ax + b)^2$, $(ax + b)^3$, . . . form a basis of $F[x]$.

8. If F is a field and h is a polynomial over F of degree ≥ 1, show that the mapping $f \rightarrow f(h)$ is a one-one linear transformation of $F[x]$ into $F[x]$. Show that this transformation is an isomorphism of $F[x]$ onto $F[x]$ if and only if deg $h = 1$.

9. Let F be a subfield of the complex numbers and let T, D be the transformations on $F[x]$ defined by

$$T\left(\sum_{i=0}^{n} c_i x^i\right) = \sum_{i=0}^{n} \frac{c_i}{1 + i} x^{i+1}$$

and

$$D\left(\sum_{i=0}^{n} c_i x^i\right) = \sum_{i=1}^{n} i c_i x^{i-1}.$$

(a) Show that T is a non-singular linear operator on $F[x]$. Show also that T is not invertible.

(b) Show that D is a linear operator on $F[x]$ and find its null space.

(c) Show that $DT = I$, and $TD \neq I$.

(d) Show that $T[(Tf)g] = (Tf)(Tg) - T[f(Tg)]$ for all f, g in $F[x]$.

(e) State and prove a rule for D similar to the one given for T in (d).

(f) Suppose V is a non-zero subspace of $F[x]$ such that Tf belongs to V for each f in V. Show that V is not finite-dimensional.

(g) Suppose V is a finite-dimensional subspace of $F[x]$. Prove there is an integer $m \geq 0$ such that $D^m f = 0$ for each f in V.

4.3. Lagrange Interpolation

Throughout this section we shall assume F is a fixed field and that $t_0, t_1, \dots, t_n$ are $n + 1$ *distinct* elements of F. Let V be the subspace of $F[x]$ consisting of all polynomials of degree less than or equal to n (together with the 0-polynomial), and let L_i be the function from V into F defined for f in V by

$$L_i(f) = f(t_i), \qquad 0 \leq i \leq n.$$

By part (i) of Theorem 2, each L_i is a linear functional on V, and one of the things we intend to show is that the set consisting of $L_0, L_1, \dots, L_n$ is a basis for V^*, the dual space of V.

Of course in order that this be so, it is sufficient (cf. Theorem 15 of Chapter 3) that $\{L_0, L_1, \dots, L_n\}$ be the dual of a basis $\{P_0, P_1, \dots, P_n\}$ of V. There is at most one such basis, and if it exists it is characterized by

$$\text{(4-11)} \qquad L_j(P_i) = P_i(t_j) = \delta_{ij}.$$

The polynomials

$$\text{(4-12)} \qquad \begin{aligned} P_i &= \frac{(x - t_0) \cdots (x - t_{i-1})(x - t_{i+1}) \cdots (x - t_n)}{(t_i - t_0) \cdots (t_i - t_{i-1})(t_i - t_{i+1}) \cdots (t_i - t_n)} \\ &= \prod_{j \neq i} \left(\frac{x - t_j}{t_i - t_j}\right) \end{aligned}$$

are of degree n, hence belong to V, and by Theorem 2, they satisfy (4-11).

If $f = \sum_i c_i P_i$, then for each j

$$\text{(4-13)} \qquad f(t_j) = \sum_i c_i P_i(t_j) = c_j.$$

Since the 0-polynomial has the property that $0(t) = 0$ for each t in F, it follows from (4-13) that the polynomials $P_0, P_1, \dots, P_n$ are linearly independent. The polynomials $1, x, \dots, x^n$ form a basis of V and hence the dimension of V is $(n + 1)$. So, the independent set $\{P_0, P_1, \dots, P_n\}$ must also be a basis for V. Thus for each f in V

$$\text{(4-14)} \qquad f = \sum_{i=0}^{n} f(t_i) P_i.$$

The expression (4-14) is called **Lagrange's interpolation formula.** Setting $f = x^j$ in (4-14) we obtain

$$x^j = \sum_{i=0}^{n} (t_i)^j P_i.$$

Now from Theorem 7 of Chapter 2 it follows that the matrix

$$\text{(4-15)} \qquad \begin{bmatrix} 1 & t_0 & t_0^2 & \cdots & t_0^n \\ 1 & t_1 & t_1^2 & \cdots & t_1^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^n \end{bmatrix}$$

is invertible. The matrix in (4-15) is called a **Vandermonde matrix;** it is an interesting exercise to show directly that such a matrix is invertible, when $t_0, t_1, \ldots, t_n$ are $n + 1$ distinct elements of F.

If f is any polynomial over F we shall, in our present discussion, denote by $f^\sim$ the polynomial function from F into F taking each t in F into $f(t)$. By definition (cf. Example 4, Chapter 2) every polynomial function arises in this way; however, it may happen that $f^\sim = g^\sim$ for two polynomials f and g such that $f \neq g$. Fortunately, as we shall see, this unpleasant situation only occurs in the case where F is a field having only a finite number of distinct elements. In order to describe in a precise way the relation between polynomials and polynomial functions, we need to define the product of two polynomial functions. If f, g are polynomials over F, the product of $f^\sim$ and $g^\sim$ is the function $f^\sim g^\sim$ from F into F given by

$$(f^\sim g^\sim)(t) = f^\sim(t)g^\sim(t), \qquad t \text{ in } F. \tag{4-16}$$

By part (ii) of Theorem 2, $(fg)(t) = f(t)g(t)$, and hence

$$(fg)^\sim(t) = f^\sim(t)g^\sim(t)$$

for each t in F. Thus $f^\sim g^\sim = (fg)^\sim$, and is a polynomial function. At this point it is a straightforward matter, which we leave to the reader, to verify that the vector space of polynomial functions over F becomes a linear algebra with identity over F if multiplication is defined by (4-16).

Definition. *Let* F *be a field and let* $\mathfrak{A}$ *and* $\mathfrak{A}^\sim$ *be linear algebras over* F. *The algebras* $\mathfrak{A}$ *and* $\mathfrak{A}^\sim$ *are said to be* **isomorphic** *if there is a one-to-one mapping* $\alpha \to \alpha^\sim$ *of* $\mathfrak{A}$ *onto* $\mathfrak{A}^\sim$ *such that*

$$\text{(a)} \qquad (c\alpha + d\beta)^\sim = c\alpha^\sim + d\beta^\sim$$

$$\text{(b)} \qquad (\alpha\beta)^\sim = \alpha^\sim\beta^\sim$$

for all α, β *in* $\mathfrak{A}$ *and all scalars* c, d *in* F. *The mapping* $\alpha \to \alpha^\sim$ *is called an* **isomorphism** *of* $\mathfrak{A}$ *onto* $\mathfrak{A}^\sim$. *An isomorphism of* $\mathfrak{A}$ *onto* $\mathfrak{A}^\sim$ *is thus a vector-space isomorphism of* $\mathfrak{A}$ *onto* $\mathfrak{A}^\sim$ *which has the additional property* (b) *of 'preserving' products.*

EXAMPLE 4. Let V be an n-dimensional vector space over the field F. By Theorem 13 of Chapter 3 and subsequent remarks, each ordered basis $\mathfrak{B}$ of V determines an isomorphism $T \to [T]_{\mathfrak{B}}$ of the algebra of linear operators on V onto the algebra of $n \times n$ matrices over F. Suppose now that U is a fixed linear operator on V and that we are given a polynomial

$$f = \sum_{i=0}^{n} c_i x^i$$

with coefficients c_i in F. Then

$$f(U) = \sum_{i=0}^{n} c_i U^i$$

and since $T \to [T]_{\mathcal{B}}$ is a linear mapping

$$[f(U)]_{\mathcal{B}} = \sum_{i=0}^{n} c_i[U^i]_{\mathcal{B}}.$$

Now from the additional fact that

$$[T_1T_2]_{\mathcal{B}} = [T_1]_{\mathcal{B}}[T_2]_{\mathcal{B}}$$

for all T_1, T_2 in $L(V, V)$ it follows that

$$[U^i]_{\mathcal{B}} = ([U]_{\mathcal{B}})^i, \qquad 2 \leq i \leq n.$$

As this relation is also valid for $i = 0, 1$ we obtain the result that

$$(4\text{-}17) \qquad [f(U)]_{\mathcal{B}} = f([U]_{\mathcal{B}}).$$

In words, if U is a linear operator on V, the matrix of a polynomial in U, in a given basis, is the same polynomial in the matrix of U.

Theorem 3. *If* F *is a field containing an infinite number of distinct elements, the mapping* f $\to$ f$^\sim$ *is an isomorphism of the algebra of polynomials over* F *onto the algebra of polynomial functions over* F.

Proof. By definition, the mapping is onto, and if f, g belong to $F[x]$ it is evident that

$$(cf + dg)^\sim = df^\sim + dg^\sim$$

for all scalars c and d. Since we have already shown that $(fg)^\sim = f^\sim g^\sim$, we need only show that the mapping is one-to-one. To do this it suffices by linearity to show that $f^\sim = 0$ implies $f = 0$. Suppose then that f is a polynomial of degree n or less such that $f' = 0$. Let $t_0, t_1, \ldots, t_n$ be any $n + 1$ distinct elements of F. Since $f^\sim = 0$, $f(t_i) = 0$ for $i = 0, 1, \ldots, n$, and it is an immediate consequence of (4-14) that $f = 0$. ▌

From the results of the next section we shall obtain an altogether different proof of this theorem.

Exercises

1. Use the Lagrange interpolation formula to find a polynomial f with real coefficients such that f has degree ≤ 3 and $f(-1) = -6$, $f(0) = 2$, $f(1) = -2$, $f(2) = 6$.

2. Let $\alpha, \beta, \gamma, \delta$ be real numbers. We ask when it is possible to find a polynomial f over R, of *degree not more than* 2, such that $f(-1) = \alpha$, $f(1) = \beta$, $f(3) = \gamma$ and $f(0) = \delta$. Prove that this is possible if and only if

$$3\alpha + 6\beta - \gamma - 8\delta = 0.$$

3. Let F be the field of real numbers,

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

$$p = (x-2)(x-3)(x-1).$$

(a) Show that $p(A) = 0$.

(b) Let P_1, P_2, P_3 be the Lagrange polynomials for $t_1 = 2$, $t_2 = 3$, $t_3 = 1$. Compute $E_i = P_i(A)$, $i = 1, 2, 3$.

(c) Show that $E_1 + E_2 + E_3 = I$, $E_iE_j = 0$ if $i \neq j$, $E_i^2 = E_i$.

(d) Show that $A = 2E_1 + 3E_2 + E_3$.

4. Let $p = (x-2)(x-3)(x-1)$ and let T be any linear operator on R^4 such that $p(T) = 0$. Let P_1, P_2, P_3 be the Lagrange polynomials of Exercise 3, and let $E_i = P_i(T)$, $i = 1, 2, 3$. Prove that

$$E_1 + E_2 + E_3 = I, \qquad E_iE_j = 0 \quad \text{if} \quad i \neq j,$$
$$E_i^2 = E_i, \quad \text{and} \quad T = 2E_1 + 3E_2 + E_3.$$

5. Let n be a positive integer and F a field. Suppose A is an $n \times n$ matrix over F and P is an invertible $n \times n$ matrix over F. If f is any polynomial over F, prove that

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

6. Let F be a field. We have considered certain special linear functionals on $F[x]$ obtained via 'evaluation at t':

$$L(f) = f(t).$$

Such functionals are not only linear but also have the property that $L(fg) = L(f)L(g)$. Prove that if L is any linear functional on $F[x]$ such that

$$L(fg) = L(f)L(g)$$

for all f and g, then either $L = 0$ or there is a t in F such that $L(f) = f(t)$ for all f.

4.4. Polynomial Ideals

In this section we are concerned with results which depend primarily on the multiplicative structure of the algebra of polynomials over a field.

Lemma. *Suppose* f *and* d *are non-zero polynomials over a field* F *such that deg* d $\leq$ *deg* f. *Then there exists a polynomial* g *in* F[x] *such that either*

$$\mathrm{f} - \mathrm{dg} = 0 \quad \textit{or} \quad \textit{deg}\ (\mathrm{f} - \mathrm{dg}) < \textit{deg}\ \mathrm{f}.$$

Proof. Suppose

$$f = a_m x^m + \sum_{i=0}^{m-1} a_i x^i, \qquad a_m \neq 0$$

and that

$$d = b_n x^n + \sum_{i=0}^{n-1} b_i x^i, \qquad b_n \neq 0.$$

Then $m \geq n$, and

$$f - \left(\frac{a_m}{b_n}\right)x^{m-n}d = 0 \quad \text{or} \quad \deg\left[f - \left(\frac{a_m}{b_n}\right)x^{m-n}d\right] < \deg f.$$

Thus we may take $g = \left(\frac{a_m}{b_n}\right) x^{m-n}$. ▌

Using this lemma we can show that the familiar process of 'long division' of polynomials with real or complex coefficients is possible over any field.

Theorem 4. *If* f, d *are polynomials over a field* F *and* d *is different from* 0 *then there exist polynomials* q, r *in* F[x] *such that*

(i) f = dq + r.
(ii) *either* r = 0 *or deg* r < *deg* d.

The polynomials q, r *satisfying* (i) *and* (ii) *are unique.*

Proof. If f is 0 or $\deg f < \deg d$ we may take $q = 0$ and $r = f$. In case $f \neq 0$ and $\deg f \geq \deg d$, the preceding lemma shows we may choose a polynomial g such that $f - dg = 0$ or $\deg\,(f - dg) < \deg f$. If $f - dg \neq 0$ and $\deg\,(f - dg) \geq \deg d$ we choose a polynomial h such that $(f - dg) - dh = 0$ or

$$\deg\,[f - d(g + h)] < \deg\,(f - dg).$$

Continuing this process as long as necessary, we ultimately obtain polynomials q, r such that $r = 0$ or $\deg r < \deg d$, and $f = dq + r$. Now suppose we also have $f = dq_1 + r_1$ where $r_1 = 0$ or $\deg r_1 < \deg d$. Then $dq + r = dq_1 + r_1$, and $d(q - q_1) = r_1 - r$. If $q - q_1 \neq 0$ then $d(q - q_1) \neq 0$ and

$$\deg d + \deg\,(q - q_1) = \deg\,(r_1 - r).$$

But as the degree of $r_1 - r$ is less than the degree of d, this is impossible and $q - q_1 = 0$. Hence also $r_1 - r = 0$. ▌

Definition. *Let* d *be a non-zero polynomial over the field* F. *If* f *is in* F[x], *the preceding theorem shows there is at most one polynomial* q *in* F[x] *such that* f = dq. *If such a* q *exists we say that* d **divides** f, *that* f *is* **divisible** *by* d, *that* f *is a* **multiple** *of* d, *and call* q *the* **quotient** *of* f *and* d. *We also write* q = f/d.

Corollary 1. *Let* f *be a polynomial over the field* F, *and let* c *be an element of* F. *Then* f *is divisible by* x − c *if and only if* f(c) = 0.

Proof. By the theorem, $f = (x - c)q + r$ where r is a scalar polynomial. By Theorem 2,

$$f(c) = 0q(c) + r(c) = r(c).$$

Hence $r = 0$ if and only if $f(c) = 0$. ▌

Definition. *Let* F *be a field. An element* c *in* F *is said to be a* **root** *or a* **zero** *of a given polynomial* f *over* F *if* f(c) = 0.

Corollary 2. *A polynomial* f *of degree* n *over a field* F *has at most* n *roots in* F.

Proof. The result is obviously true for polynomials of degree 0 and degree 1. We assume it to be true for polynomials of degree $n - 1$. If a is a root of f, $f = (x - a)q$ where q has degree $n - 1$. Since $f(b) = 0$ if and only if $a = b$ or $q(b) = 0$, it follows by our inductive assumption that f has at most n roots. ▌

The reader should observe that the main step in the proof of Theorem 3 follows immediately from this corollary.

The formal derivatives of a polynomial are useful in discussing multiple roots. The **derivative** of the polynomial

$$f = c_0 + c_1x + \cdots + c_nx^n$$

is the polynomial

$$f' = c_1 + 2c_2x + \cdots + nc_nx^{n-1}.$$

We also use the notation $Df = f'$. Differentiation is linear, that is, D is a linear operator on $F[x]$. We have the higher order formal derivatives $f'' = D^2f$, $f^{(3)} = D^3f$, and so on.

Theorem 5 (Taylor's Formula). *Let* F *be a field of characteristic zero,* c *an element of* F, *and* n *a positive integer. If* f *is a polynomial over* f *with* deg f ≤ n, *then*

$$\mathrm{f} = \sum_{\mathrm{k}=0}^{\mathrm{n}} \frac{(\mathrm{D^kf})}{\mathrm{k}!}(\mathrm{c})(\mathrm{x} - \mathrm{c})^{\mathrm{k}}.$$

Proof. Taylor's formula is a consequence of the binomial theorem and the linearity of the operators $D, D^2, \ldots, D^n$. The binomial theorem is easily proved by induction and asserts that

$$(a + b)^m = \sum_{k=0}^{m} \binom{m}{k} a^{m-k}\, b^k$$

where

$$\binom{m}{k} = \frac{m!}{k!(m-k)!} = \frac{m(m-1)\cdots(m-k+1)}{1 \cdot 2 \cdots k}$$

is the familiar binomial coefficient giving the number of combinations of m objects taken k at a time. By the binomial theorem

$$\begin{aligned} x^m &= [c + (x - c)]^m \\ &= \sum_{k=0}^{m} \binom{m}{k} c^{m-k}(x - c)^k \\ &= c^m + mc^{m-1}(x - c) + \cdots + (x - c)^m \end{aligned}$$

and this is the statement of Taylor's formula for the case $f = x^m$. If

$$f = \sum_{m=0}^{n} a_m x^m$$

then

$$D^k f(c) = \sum_m a_m (D^k x^m)(c)$$

and

$$\begin{aligned}\sum_{k=0}^{n} \frac{D^k f(c)}{k!} (x - c)^k &= \sum_k \sum_m a_m \frac{(D^k x^m)}{k!} (c)(x - c)^k \\ &= \sum_m a_m \sum_k \frac{(D^k x^m)}{k!} (c)(x - c)^k \\ &= \sum_m a_m x^m \\ &= f. \quad \blacksquare\end{aligned}$$

It should be noted that because the polynomials $1, (x - c), \ldots, (x - c)^n$ are linearly independent (cf. Exercise 6, Section 4.2) Taylor's formula provides the unique method for writing f as a linear combination of the polynomials $(x - c)^k$ $(0 \leq k \leq n)$.

Although we shall not give any details, it is perhaps worth mentioning at this point that with the proper interpretation Taylor's formula is also valid for polynomials over fields of finite characteristic. If the field F has finite characteristic (the sum of some finite number of 1's in F is 0) then we may have $k! = 0$ in F, in which case the division of $(D^k f)$ (c) by $k!$ is meaningless. Nevertheless, sense can be made out of the division of $D^k f$ by $k!$, because every coefficient of $D^k f$ is an element of F multiplied by an integer divisible by $k!$ If all of this seems confusing, we advise the reader to restrict his attention to fields of characteristic 0 or to subfields of the complex numbers.

If c is a root of the polynomial f, the **multiplicity** of c as a root of f is the largest positive integer r such that $(x - c)^r$ divides f.

The multiplicity of a root is clearly less than or equal to the degree of f. For polynomials over fields of characteristic zero, the multiplicity of c as a root of f is related to the number of derivatives of f that are 0 at c.

Theorem 6. *Let* F *be a field of characteristic zero and* f *a polynomial over* F *with deg* f $\leq$ n. *Then the scalar* c *is a root of* f *of multiplicity* r *if and only if*

$$(D^k f)(c) = 0, \qquad 0 \leq k \leq r - 1$$

$$(D^r f)(c) \neq 0.$$

Proof. Suppose that r is the multiplicity of c as a root of f. Then there is a polynomial g such that $f = (x - c)^r g$ and $g(c) \neq 0$. For otherwise f would be divisible by $(x - c)^{r+1}$, by Corollary 1 of Theorem 4. By Taylor's formula applied to g

$$f = (x - c)^r \left[\sum_{m=0}^{n-r} \frac{(D^m g)}{m!} (c)\,(x - c)^m \right]$$
$$= \sum_{m=0}^{n-r} \frac{(D^m g)}{m!} (x - c)^{r+m}$$

Since there is only one way to write f as a linear combination of the powers $(x - c)^k$ $(0 \leq k \leq n)$ it follows that

$$\frac{(D^k f)(c)}{k!} = \begin{cases} 0 \text{ if } 0 \leq k \leq r - 1 \\ \dfrac{D^{k-r} g(c)}{(k - r)!} \text{ if } r \leq k \leq n. \end{cases}$$

Therefore, $D^k f(c) = 0$ for $0 \leq k \leq r - 1$, and $D^r f(c) = g(c) \neq 0$. Conversely, if these conditions are satisfied, it follows at once from Taylor's formula that there is a polynomial g such that $f = (x - c)^r g$ and $g(c) \neq 0$. Now suppose that r is not the largest positive integer such that $(x - c)^r$ divides f. Then there is a polynomial h such that $f = (x - c)^{r+1} h$. But this implies $g = (x - c)h$, by Corollary 2 of Theorem 1; hence $g(c) = 0$, a contradiction. ▌

Definition. *Let* F *be a field. An* **ideal** *in* F[x] *is a subspace* M *of* F[x] *such that* fg *belongs to* M *whenever* f *is in* F[x] *and* g *is in* M.

EXAMPLE 5. If F is a field and d is a polynomial over F, the set $M = dF[x]$, of all multiples df of d by arbitrary f in $F[x]$, is an ideal. For M is non-empty, M in fact contains d. If f, g belong to $F[x]$ and c is a scalar, then

$$c(df) - dg = d(cf - g)$$

belongs to M, so that M is a subspace. Finally M contains $(df)g = d(fg)$ as well. The ideal M is called the **principal ideal generated by** d.

EXAMPLE 6. Let $d_1, \ldots, d_n$ be a finite number of polynomials over F. Then the sum M of the subspaces $d_i F[x]$ is a subspace and is also an ideal. For suppose p belongs to M. Then there exist polynomials $f_1, \ldots, f_n$ in $F[x]$ such that $p = d_1 f_1 + \cdots + d_n f_n$. If g is an arbitrary polynomial over F, then

$$pg = d_1(f_1 g) + \cdots + d_n(f_n g)$$

so that pg also belongs to M. Thus M is an ideal, and we say that M is the ideal **generated** by the polynomials, $d_1, \ldots, d_n$.

EXAMPLE 7. Let F be a subfield of the complex numbers, and consider the ideal

$$M = (x + 2)F[x] + (x^2 + 8x + 16)F[x].$$

We assert that $M = F[x]$. For M contains

$$x^2 + 8x + 16 - x(x + 2) = 6x + 16$$

and hence M contains $6x + 16 - 6(x + 2) = 4$. Thus the scalar polynomial 1 belongs to M as well as all its multiples.

Theorem 7. *If* F *is a field, and* M *is any non-zero ideal in* F[x], *there is a unique monic polynomial* d *in* F[x] *such that* M *is the principal ideal generated by* d.

Proof. By assumption, M contains a non-zero polynomial; among all non-zero polynomials in M there is a polynomial d of minimal degree. We may assume d is monic, for otherwise we can multiply d by a scalar to make it monic. Now if f belongs to M, Theorem 4 shows that $f = dq + r$ where $r = 0$ or $\deg r < \deg d$. Since d is in M, dq and $f - dq = r$ also belong to M. Because d is an element of M of minimal degree we cannot have $\deg r < \deg d$, so $r = 0$. Thus $M = dF[x]$. If g is another monic polynomial such that $M = gF[x]$, then there exist non-zero polynomials p, q such that $d = gp$ and $g = dq$. Thus $d = dpq$ and

$$\deg d = \deg d + \deg p + \deg q.$$

Hence $\deg p = \deg q = 0$, and as d, g are monic, $p = q = 1$. Thus $d = g$. ▮

It is worth observing that in the proof just given we have used a special case of a more general and rather useful fact; namely, if p is a non-zero polynomial in an ideal M and if f is a polynomial in M which is not divisible by p, then $f = pq + r$ where the 'remainder' r belongs to M, is different from 0, and has smaller degree than p. We have already made use of this fact in Example 7 to show that the scalar polynomial 1 is the monic generator of the ideal considered there. In principle it is always possible to find the monic polynomial generating a given non-zero ideal. For one can ultimately obtain a polynomial in the ideal of minimal degree by a finite number of successive divisions.

Corollary. *If* $p_1, \ldots, p_n$ *are polynomials over a field* F, *not all of which are* 0, *there is a unique monic polynomial* d *in* F[x] *such that*

(a) d *is in the ideal generated by* $p_1, \ldots, p_n$;

(b) d *divides each of the polynomials* p_i.

Any polynomial satisfying (a) *and* (b) *necessarily satisfies*

(c) d *is divisible by every polynomial which divides each of the polynomials* $p_1, \ldots, p_n$.

Proof. Let d be the monic generator of the ideal

$$p_1F[x] + \cdots + p_nF[x].$$

Every member of this ideal is divisible by d; thus each of the polynomials p_i is divisible by d. Now suppose f is a polynomial which divides each of the polynomials $p_1, \ldots, p_n$. Then there exist polynomials $g_1, \ldots, g_n$ such that $p_i = fg_i$, $1 \leq i \leq n$. Also, since d is in the ideal

$$p_1F[x] + \cdots + p_nF[x],$$

there exist polynomials $q_1, \ldots, q_n$ in $F[x]$ such that

$$d = p_1q_1 + \cdots + p_nq_n.$$

Thus

$$d = f[g_1q_1 + \cdots + g_nq_n].$$

We have shown that d is a monic polynomial satisfying (a), (b), and (c). If d' is any polynomial satisfying (a) and (b) it follows, from (a) and the definition of d, that d' is a scalar multiple of d and satisfies (c) as well. Finally, in case d' is a monic polynomial, we have $d' = d$. ▌

Definition. *If* $\mathrm{p}_1, \ldots, \mathrm{p}_n$ *are polynomials over a field* F, *not all of which are* 0, *the monic generator* d *of the ideal*

$$\mathrm{p}_1\mathrm{F}[\mathrm{x}] + \cdots + \mathrm{p}_n\mathrm{F}[\mathrm{x}]$$

is called the **greatest common divisor** *(g.c.d.) of* $\mathrm{p}_1, \ldots, \mathrm{p}_n$. *This terminology is justified by the preceding corollary. We say that the polynomials* $\mathrm{p}_1, \ldots, \mathrm{p}_n$ *are* **relatively prime** *if their greatest common divisor is* 1, *or equivalently if the ideal they generate is all of* F[x].

EXAMPLE 8. Let C be the field of complex numbers. Then

(a) g.c.d. $(x + 2, x^2 + 8x + 16) = 1$ (see Example 7);

(b) g.c.d. $((x - 2)^2(x + i), (x - 2)(x^2 + 1)) = (x - 2)(x + i)$. For, the ideal

$$(x - 2)^2(x + i)F[x] + (x - 2)(x^2 + 1)F[x]$$

contains

$$(x - 2)^2(x + i) - (x - 2)(x^2 + 1) = (x - 2)(x + i)(i - 2).$$

Hence it contains $(x - 2)(x + i)$, which is monic and divides both

$$(x - 2)^2(x + i) \quad \text{and} \quad (x - 2)(x^2 + 1).$$

EXAMPLE 9. Let F be the field of rational numbers and in $F[x]$ let M be the ideal generated by

$$(x - 1)(x + 2)^2, \qquad (x + 2)^2(x - 3), \qquad \text{and} \qquad (x - 3).$$

Then M contains

$$\tfrac{1}{2}(x + 2)^2[(x - 1) - (x - 3)] = (x + 2)^2$$

and since

$$(x + 2)^2 = (x - 3)(x + 7) - 17$$

M contains the scalar polynomial 1. Thus $M = F[x]$ and the polynomials

$$(x-1)(x+2)^2, \qquad (x+2)^2(x-3), \qquad \text{and} \qquad (x-3)$$

are relatively prime.

Exercises

1. Let Q be the field of rational numbers. Determine which of the following subsets of $Q[x]$ are ideals. When the set is an ideal, find its monic generator.

(a) all f of even degree;
(b) all f of degree ≥ 5;
(c) all f such that $f(0) = 0$;
(d) all f such that $f(2) = f(4) = 0$;
(e) all f in the range of the linear operator T defined by

$$T\left(\sum_{i=0}^{n} c_i x^i\right) = \sum_{i=0}^{n} \frac{c_i}{i+1} x^{i+1}.$$

2. Find the g.c.d. of each of the following pairs of polynomials

(a) $2x^5 - x^3 - 3x^2 - 6x + 4$, $x^4 + x^3 - x^2 - 2x - 2$;
(b) $3x^4 + 8x^2 - 3$, $x^3 + 2x^2 + 3x + 6$;
(c) $x^4 - 2x^3 - 2x^2 - 2x - 3$, $x^3 + 6x^2 + 7x + 1$.

3. Let A be an $n \times n$ matrix over a field F. Show that the set of all polynomials f in $F[x]$ such that $f(A) = 0$ is an ideal.

4. Let F be a subfield of the complex numbers, and let

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

Find the monic generator of the ideal of all polynomials f in $F[x]$ such that $f(A) = 0$.

5. Let F be a field. Show that the intersection of any number of ideals in $F[x]$ is an ideal.

6. Let F be a field. Show that the ideal generated by a finite number of polynomials $f_1, \ldots, f_n$ in $F[x]$ is the intersection of all ideals containing $f_1, \ldots, f_n$.

7. Let K be a subfield of a field F, and suppose f, g are polynomials in $K[x]$. Let M_K be the ideal generated by f and g in $K[x]$ and M_F be the ideal they generate in $F[x]$. Show that M_K and M_F have the same monic generator.

4.5. The Prime Factorization of a Polynomial

In this section we shall prove that each polynomial over the field F can be written as a product of 'prime' polynomials. This factorization provides us with an effective tool for finding the greatest common divisor of a finite number of polynomials, and in particular, provides an effective means for deciding when the polynomials are relatively prime.

Definition. *Let* F *be a field. A polynomial* f *in* F[x] *is said to be* **reducible over** F *if there exist polynomials* g, h *in* F[x] *of degree* ≥ 1 *such that* f = gh, *and if not,* f *is said to be* **irreducible over** F. *A non-scalar irreducible polynomial over* F *is called a* **prime polynomial over** F, *and we sometimes say it is a* **prime in** F[x].

EXAMPLE 10. The polynomial $x^2 + 1$ is reducible over the field C of complex numbers. For

$$x^2 + 1 = (x + i)(x - i)$$

and the polynomials $x + i$, $x - i$ belong to $C[x]$. On the other hand, $x^2 + 1$ is irreducible over the field R of real numbers. For if

$$x^2 + 1 = (ax + b)(a'x + b')$$

with a, a', b, b' in R, then

$$aa' = 1, \qquad ab' + ba' = 0, \qquad bb' = 1.$$

These relations imply $a^2 + b^2 = 0$, which is impossible with real numbers a and b, unless $a = b = 0$.

Theorem 8. *Let* p, f, *and* g *be polynomials over the field* F. *Suppose that* p *is a prime polynomial and that* p *divides the product* fg. *Then either* p *divides* f *or* p *divides* g.

Proof. It is no loss of generality to assume that p is a monic prime polynomial. The fact that p is prime then simply says that the only monic divisors of p are 1 and p. Let d be the g.c.d. of f and p. Then either $d = 1$ or $d = p$, since d is a monic polynomial which divides p. If $d = p$, then p divides f and we are done. So suppose $d = 1$, i.e., suppose f and p are relatively prime. We shall prove that p divides g. Since $(f, p) = 1$, there are polynomials f_0 and p_0 such that $1 = f_0 f + p_0 p$. Multiplying by g, we obtain

$$\begin{aligned} g &= f_0 fg + p_0 pg \\ &= (fg)f_0 + p(p_0 g). \end{aligned}$$

Since p divides fg it divides $(fg)f_0$, and certainly p divides $p(p_0g)$. Thus p divides g. ▌

Corollary. *If* p *is a prime and divides a product* $f_1 \cdots f_n$, *then* p *divides one of the polynomials* $f_1, \ldots, f_n$.

Proof. The proof is by induction. When $n = 2$, the result is simply the statement of Theorem 6. Suppose we have proved the corollary for $n = k$, and that p divides the product $f_1 \cdots f_{k+1}$ of some $(k + 1)$ polynomials. Since p divides $(f_1 \cdots f_k)f_{k+1}$, either p divides f_{k+1} or p divides $f_1 \cdots f_k$. By the induction hypothesis, if p divides $f_1 \cdots f_k$, then p divides f_j for some j, $1 \leq j \leq k$. So we see that in any case p must divide some f_j, $1 \leq j \leq k + 1$. ▌

Theorem 9. *If* F *is a field, a non-scalar monic polynomial in* F[x] *can be factored as a product of monic primes in* F[x] *in one and, except for order, only one way.*

Proof. Suppose f is a non-scalar monic polynomial over F. As polynomials of degree one are irreducible, there is nothing to prove if $\deg f = 1$. Suppose f has degree $n > 1$. By induction we may assume the theorem is true for all non-scalar monic polynomials of degree less than n. If f is irreducible, it is already factored as a product of monic primes, and otherwise $f = gh$ where g and h are non-scalar monic polynomials of degree less than n. Thus g and h can be factored as products of monic primes in $F[x]$ and hence so can f. Now suppose

$$f = p_1 \cdots p_m = q_1 \cdots q_n$$

where $p_1, \ldots, p_m$ and $q_1, \ldots, q_n$ are monic primes in $F[x]$. Then p_m divides the product $q_1 \cdots q_n$. By the above corollary, p_m must divide some q_i. Since q_i and p_m are both monic primes, this means that

$$q_i = p_m. \tag{4-16}$$

From (4-16) we see that $m = n = 1$ if either $m = 1$ or $n = 1$. For

$$\deg f = \sum_{i=1}^{m} \deg p_i = \sum_{j=1}^{n} \deg q_j.$$

In this case there is nothing more to prove, so we may assume $m > 1$ and $n > 1$. By rearranging the q's we can then assume $p_m = q_n$, and that

$$p_1 \cdots p_{m-1}p_m = q_1 \cdots q_{n-1}p_m.$$

Now by Corollary 2 of Theorem 1 it follows that

$$p_1 \cdots p_{m-1} = q_1 \cdots q_{n-1}.$$

As the polynomial $p_1 \cdots p_{m-1}$ has degree less than n, our inductive assumption applies and shows that the sequence $q_1, \ldots, q_{n-1}$ is at most a rearrangement of the sequence $p_1, \ldots, p_{m-1}$. This together with (4-16) shows that the factorization of f as a product of monic primes is unique up to the order of the factors. ▌

In the above factorization of a given non-scalar monic polynomial f, some of the monic prime factors may be repeated. If $p_1, p_2, \ldots, p_r$ are the distinct monic primes occurring in this factorization of f, then

$$f = p_1^{n_1}p_2^{n_2} \cdots p_r^{n_r}, \tag{4-17}$$

the exponent n_i being the number of times the prime p_i occurs in the factorization. This decomposition is also clearly unique, and is called the **primary decomposition** of f. It is easily verified that every monic divisor of f has the form

$$p_1^{m_1}p_2^{m_2}\cdots p_r^{m_r}, \qquad 0 \leq m_i \leq n_i. \tag{4-18}$$

From (4-18) it follows that the g.c.d. of a finite number of non-scalar monic polynomials $f_1, \ldots, f_s$ is obtained by combining all those monic' primes which occur simultaneously in the factorizations of $f_1, \ldots, f_s$. The exponent to which each prime is to be taken is the largest for which the corresponding prime power is a factor of each f_i. If no (non-trivial) prime power is a factor of each f_i, the polynomials are relatively prime.

EXAMPLE 11. Suppose F is a field, and let a, b, c be distinct elements of F. Then the polynomials $x - a$, $x - b$, $x - c$ are distinct monic primes in $F[x]$. If m, n, and s are positive integers, $(x - c)^s$ is the g.c.d. of the polynomials.

$$(x - b)^n(x - c)^s \quad \text{and} \quad (x - a)^m(x - c)^s$$

whereas the three polynomials

$$(x - b)^n(x - c)^s, \qquad (x - a)^m(x - c)^s, \qquad (x - a)^m(x - b)^n$$

are relatively prime.

Theorem 10. *Let* f *be a non-scalar monic polynomial over the field* F *and let*

$$\mathrm{f} = \mathrm{p}_1^{\mathrm{n}_1} \cdots \mathrm{p}_k^{\mathrm{n}_k}$$

be the prime factorization of f. *For each* j, $1 \leq \mathrm{j} \leq \mathrm{k}$, *let*

$$\mathrm{f}_j = \mathrm{f}/\mathrm{p}_j^{\mathrm{n}_j} = \prod_{i \neq j} \mathrm{p}_i^{\mathrm{n}_i}.$$

Then $\mathrm{f}_1, \ldots, \mathrm{f}_k$ *are relatively prime.*

Proof. We leave the (easy) proof of this to the reader. We have stated this theorem largely because we wish to refer to it later. ∎

Theorem 11. *Let* f *be a polynomial over the field* F *with derivative* f′. *Then* f *is a product of distinct irreducible polynomials over* F *if and only if* f *and* f′ *are relatively prime.*

Proof. Suppose in the prime factorization of f over the field F that some (non-scalar) prime polynomial p is repeated. Then $f = p^2h$ for some h in $F[x]$. Then

$$f' = p^2h' + 2pp'h$$

and p is also a divisor of f'. Hence f and f' are not relatively prime.

Now suppose $f = p_1 \cdots p_k$, where $p_1, \ldots, p_k$ are distinct non-scalar irreducible polynomials over F. Let $f_j = f/p_j$. Then

$$f' = p_1'f_1 + p_2'f_2 + \cdots + p_k'f_k.$$

Let p be a prime polynomial which divides both f and f'. Then $p = p_i$ for some i. Now p_i divides f_j for $j \neq i$, and since p_i also divides

$$f' = \sum_{j=1}^{k} p_j' f_j$$

we see that p_i must divide $p_i' f_i$. Therefore p_i divides either f_i or p_i'. But p_i does not divide f_i since $p_1, \ldots, p_k$ are distinct. So p_i divides p_i'. This is not possible, since p_i' has degree one less than the degree of p_i. We conclude that no prime divides both f and f', or that, f and f' are relatively prime. ▌

Definition. *The field* F *is called* **algebraically closed** *if every prime polynomial over* F *has degree* 1.

To say that F is algebraically closed means every non-scalar irreducible monic polynomial over F is of the form $(x - c)$. We have already observed that each such polynomial is irreducible for any F. Accordingly, an equivalent definition of an algebraically closed field is a field F such that each non-scalar polynomial f in $F[x]$ can be expressed in the form

$$f = c(x - c_1)^{n_1} \cdots (x - c_k)^{n_k}$$

where c is a scalar, $c_1, \ldots, c_k$ are distinct elements of F, and $n_1, \ldots, n_k$ are positive integers. Still another formulation is that if f is a non-scalar polynomial over F, then there is an element c in F such that $f(c) = 0$.

The field R of real numbers is not algebraically closed, since the polynomial $(x^2 + 1)$ is irreducible over R but not of degree 1, or, because there is no real number c such that $c^2 + 1 = 0$. The so-called Fundamental Theorem of Algebra states that the field C of complex numbers is algebraically closed. We shall not prove this theorem, although we shall use it somewhat later in this book. The proof is omitted partly because of the limitations of time and partly because the proof depends upon a 'non-algebraic' property of the system of real numbers. For one possible proof the interested reader may consult the book by Schreier and Sperner in the Bibliography.

The Fundamental Theorem of Algebra also makes it clear what the possibilities are for the prime factorization of a polynomial with real coefficients. If f is a polynomial with real coefficients and c is a complex root of f, then the complex conjugate $\bar{c}$ is also a root of f. Therefore, those complex roots which are not real must occur in conjugate pairs, and the entire set of roots has the form $\{t_1, \ldots, t_k, c_1, \bar{c}_1, \ldots, c_r, \bar{c}_r\}$ where $t_1, \ldots, t_k$ are real and $c_1, \ldots, c_r$ are non-real complex numbers. Thus f factors

$$f = c(x - t_1) \cdots (x - t_k)p_1 \cdots p_r$$

where p_i is the quadratic polynomial

$$p_i = (x - c_i)(x - \bar{c}_i).$$

These polynomials p_i have real coefficients. We conclude that every irreducible polynomial over the real number field has degree 1 or 2. Each polynomial over R is the product of certain linear factors, obtained from the real roots of f, and certain irreducible quadratic polynomials.

Exercises

1. Let p be a monic polynomial over the field F, and let f and g be relatively prime polynomials over F. Prove that the g.c.d. of pf and pg is p.

2. Assuming the Fundamental Theorem of Algebra, prove the following. If f and g are polynomials over the field of complex numbers, then g.c.d. $(f, g) = 1$ if and only if f and g have no common root.

3. Let D be the differentiation operator on the space of polynomials over the field of complex numbers. Let f be a monic polynomial over the field of complex numbers. Prove that

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

where $c_1, \ldots, c_k$ are *distinct* complex numbers if and only if f and Df are relatively prime. In other words, f has no repeated root if and only if f and Df have no common root. (Assume the Fundamental Theorem of Algebra.)

4. Prove the following generalization of Taylor's formula. Let f, g, and h be polynomials over a subfield of the complex numbers, with $\deg f \leq n$. Then

$$f(g) = \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(h)(g - h)^k.$$

(Here $f(g)$ denotes 'f of g.')

For the remaining exercises, we shall need the following definition. If f, g, and p are polynomials over the field F with $p \neq 0$, we say that f is **congruent to** g **modulo** p if $(f - g)$ is divisible by p. If f is congruent to g modulo p, we write

$$f \equiv g \bmod p.$$

5. Prove, for any non-zero polynomial p, that congruence modulo p is an equivalence relation.

(a) It is reflexive: $f \equiv f \bmod p$.
(b) It is symmetric: if $f \equiv g \bmod p$, then $g \equiv f \bmod p$.
(c) It is transitive: if $f \equiv g \bmod p$ and $g \equiv h \bmod p$, then $f \equiv h \bmod p$.

6. Suppose $f \equiv g \bmod p$ and $f_1 \equiv g_1 \bmod p$.
(a) Prove that $f + f_1 \equiv g + g_1 \bmod p$.
(b) Prove that $ff_1 \equiv gg_1 \bmod p$.

7. Use Exercise 7 to prove the following. If f, g, h, and p are polynomials over the field F and $p \neq 0$, and if $f \equiv g \bmod p$, then $h(f) \equiv h(g) \bmod p$.

8. If p is an irreducible polynomial and $fg \equiv 0 \bmod p$, prove that either $f \equiv 0 \bmod p$ or $g \equiv 0 \bmod p$. Give an example which shows that this is false if p is not irreducible.

5. Determinants

5.1. Commutative Rings

In this chapter we shall prove the essential facts about determinants of square matrices. We shall do this not only for matrices over a field, but also for matrices with entries which are 'scalars' of a more general type. There are two reasons for this generality. First, at certain points in the next chapter, we shall find it necessary to deal with determinants of matrices with polynomial entries. Second, in the treatment of determinants which we present, one of the axioms for a field plays no role, namely, the axiom which guarantees a multiplicative inverse for each non-zero element. For these reasons, it is appropriate to develop the theory of determinants for matrices, the entries of which are elements from a commutative ring with identity.

Definition. *A* **ring** *is a set* K, *together with two operations* $(\mathrm{x}, \mathrm{y}) \rightarrow \mathrm{x} + \mathrm{y}$ *and* $(\mathrm{x}, \mathrm{y}) \rightarrow \mathrm{xy}$ *satisfying*

(a) K *is a commutative group under the operation* $(\mathrm{x}, \mathrm{y}) \rightarrow \mathrm{x} + \mathrm{y}$ (K *is a commutative group under addition*);

(b) $(\mathrm{xy})\mathrm{z} = \mathrm{x}(\mathrm{yz})$ (*multiplication is associative*);

(c) $\mathrm{x}(\mathrm{y} + \mathrm{z}) = \mathrm{xy} + \mathrm{xz}$; $(\mathrm{y} + \mathrm{z})\mathrm{x} = \mathrm{yx} + \mathrm{zx}$ (*the two distributive laws hold*).

If $\mathrm{xy} = \mathrm{yx}$ *for all* x *and* y *in* K, *we say that the ring* K *is* **commutative.** *If there is an element* 1 *in* K *such that* $1\mathrm{x} = \mathrm{x}1 = \mathrm{x}$ *for each* x, K *is said to be a* **ring with identity,** *and* 1 *is called the* **identity** *for* K.

We are interested here in commutative rings with identity. Such a ring can be described briefly as a set K, together with two operations which satisfy all the axioms for a field given in Chapter 1, except possibly for axiom (8) and the condition $1 \neq 0$. Thus, a field is a commutative ring with non-zero identity such that to each non-zero x there corresponds an element x^{-1} with $xx^{-1} = 1$. The set of integers, with the usual operations, is a commutative ring with identity which is not a field. Another commutative ring with identity is the set of all polynomials over a field, together with the addition and multiplication which we have defined for polynomials.

If K is a commutative ring with identity, we define an $m \times n$ matrix over K to be a function A from the set of pairs (i, j) of integers, $1 \leq i \leq m$, $1 \leq j \leq n$, into K. As usual we represent such a matrix by a rectangular array having m rows and n columns. The sum and product of matrices over K are defined as for matrices over a field

$$\begin{aligned}(A + B)_{ij} &= A_{ij} + B_{ij}\\ (AB)_{ij} &= \sum_k A_{ik}B_{kj}\end{aligned}$$

the sum being defined when A and B have the same number of rows and the same number of columns, the product being defined when the number of columns of A is equal to the number of rows of B. The basic algebraic properties of these operations are again valid. For example,

$$A(B + C) = AB + AC, \qquad (AB)C = A(BC), \qquad \text{etc.}$$

As in the case of fields, we shall refer to the elements of K as scalars. We may then define linear combinations of the rows or columns of a matrix as we did earlier. Roughly speaking, all that we previously did for matrices over a field is valid for matrices over K, excluding those results which depended upon the ability to 'divide' in K.

5.2. Determinant Functions

Let K be a commutative ring with identity. We wish to assign to each $n \times n$ (square) matrix over K a scalar (element of K) to be known as the determinant of the matrix. It is possible to define the determinant of a square matrix A by simply writing down a formula for this determinant in terms of the entries of A. One can then deduce the various properties of determinants from this formula. However, such a formula is rather complicated, and to gain some technical advantage we shall proceed as follows. We shall define a 'determinant function' on $K^{n \times n}$ as a function which assigns to each $n \times n$ matrix over K a scalar, the function having these special properties. It is linear as a function of each of the rows of the matrix; its value is 0 on any matrix having two equal rows; and its value on the $n \times n$ identity matrix is 1. We shall prove that such a function exists, and then that it is unique, i.e., that there is precisely one such function. As we prove the uniqueness, an explicit formula for the determinant will be obtained, along with many of its useful properties.

This section will be devoted to the definition of 'determinant function' and to the proof that at least one such function exists.

Definition. *Let* K *be a commutative ring with identity,* n *a positive integer, and let* D *be a function which assigns to each* n × n *matrix* A *over* K *a scalar* D(A) *in* K. *We say that* D *is* **n-linear** *if for each* i, $1 \leq \text{i} \leq \text{n}$, D *is a linear function of the* ith *row when the other* (n − 1) *rows are held fixed.*

This definition requires some clarification. If D is a function from $K^{n \times n}$ into K, and if $\alpha_1, \ldots, \alpha_n$ are the rows of the matrix A, let us also write

$$D(A) = D(\alpha_1, \ldots, \alpha_n)$$

that is, let us also think of D as the function of the rows of A. The statement that D is n-linear then means

$$\text{(5-1)} \qquad D(\alpha_1, \ldots, c\alpha_i + \alpha'_i, \ldots, \alpha_n) = cD(\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n) \\ + D(\alpha_1, \ldots, \alpha'_i, \ldots, \alpha_n).$$

If we fix all rows except row i and regard D as a function of the ith row, it is often convenient to write $D(\alpha_i)$ for $D(A)$. Thus, we may abbreviate (5-1) to

$$D(c\alpha_i + \alpha'_i) = cD(\alpha_i) + D(\alpha'_i)$$

so long as it is clear what the meaning is.

Example 1. Let $k_1, \ldots, k_n$ be positive integers, $1 \leq k_i \leq n$, and let a be an element of K. For each $n \times n$ matrix A over K, define

$$\text{(5-2)} \qquad D(A) = aA(1, k_1) \cdots A(n, k_n).$$

Then the function D defined by (5-2) is n-linear. For, if we regard D as a function of the ith row of A, the others being fixed, we may write

$$D(\alpha_i) = A(i, k_i)b$$

where b is some fixed element of K. Let $\alpha'_i = (A'_{i1}, \ldots, A'_{in})$. Then we have

$$\begin{aligned} D(c\alpha_i + \alpha'_i) &= [cA(i, k_i) + A'(i, k_i)]b \\ &= cD(\alpha_i) + D(\alpha'_i). \end{aligned}$$

Thus D is a linear function of each of the rows of A.

A particular n-linear function of this type is

$$D(A) = A_{11}A_{22} \cdots A_{nn}.$$

In other words, the 'product of the diagonal entries' is an n-linear function on $K^{n\times n}$.

Example 2. Let us find all 2-linear functions on 2×2 matrices over K. Let D be such a function. If we denote the rows of the 2×2 identity matrix by ϵ_1, ϵ_2, we have

$$D(A) = D(A_{11}\epsilon_1 + A_{12}\epsilon_2, A_{21}\epsilon_1 + A_{22}\epsilon_2).$$

Using the fact that D is 2-linear, (5-1), we have

$$\begin{aligned} D(A) &= A_{11}D(\epsilon_1, A_{21}\epsilon_1 + A_{22}\epsilon_2) + A_{12}D(\epsilon_2, A_{21}\epsilon_1 + A_{22}\epsilon_2) \\ &= A_{11}A_{21}D(\epsilon_1, \epsilon_1) + A_{11}A_{22}D(\epsilon_1, \epsilon_2) \\ &\qquad + A_{12}A_{21}D(\epsilon_2, \epsilon_1) + A_{12}A_{22}D(\epsilon_2, \epsilon_2). \end{aligned}$$

Thus D is completely determined by the four scalars

$$D(\epsilon_1, \epsilon_1), \quad D(\epsilon_1, \epsilon_2), \quad D(\epsilon_2, \epsilon_1), \quad \text{and} \quad D(\epsilon_2, \epsilon_2).$$

The reader should find it easy to verify the following. If a, b, c, d are any four scalars in K and if we define

$$D(A) = A_{11}A_{21}a + A_{11}A_{22}b + A_{12}A_{21}c + A_{12}A_{22}d$$

then D is a 2-linear function on 2×2 matrices over K and

$$\begin{aligned} D(\epsilon_1, \epsilon_1) &= a, \qquad D(\epsilon_1, \epsilon_2) = b \\ D(\epsilon_2, \epsilon_1) &= c, \qquad D(\epsilon_2, \epsilon_2) = d. \end{aligned}$$

Lemma. *A linear combination of* n*-linear functions is* n*-linear.*

Proof. It suffices to prove that a linear combination of two n-linear functions is n-linear. Let D and E be n-linear functions. If a and b belong to K, the linear combination $aD + bE$ is of course defined by

$$(aD + bE)(A) = aD(A) + bE(A).$$

Hence, if we fix all rows except row i

$$\begin{aligned} (aD + bE)(c\alpha_i + \alpha_i') &= aD(c\alpha_i + \alpha_i') + bE(c\alpha_i + \alpha_i') \\ &= acD(\alpha_i) + aD(\alpha_i') + bcE(\alpha_i) + bE(\alpha_i') \\ &= c(aD + bE)(\alpha_i) + (aD + bE)(\alpha_i'). \quad \blacksquare \end{aligned}$$

If K is a field and V is the set of $n \times n$ matrices over K, the above lemma says the following. The set of n-linear functions on V is a subspace of the space of all functions from V into K.

Example 3. Let D be the function defined on 2×2 matrices over K by

$$D(A) = A_{11}A_{22} - A_{12}A_{21}. \tag{5-3}$$

Now D is the sum of two functions of the type described in Example 1:

$$\begin{aligned} D &= D_1 + D_2 \\ D_1(A) &= A_{11}A_{22} \\ D_2(A) &= -A_{12}A_{21}. \end{aligned}$$

By the above lemma, D is a 2-linear function. The reader who has had any experience with determinants will not find this surprising, since he will recognize (5-3) as the usual definition of the determinant of a 2×2 matrix. Of course the function D we have just defined is not a typical 2-linear function. It has many special properties. Let us note some of these properties. First, if I is the 2×2 identity matrix, then $D(I) = 1$, i.e., $D(\epsilon_1, \epsilon_2) = 1$. Second, if the two rows of A are equal, then

$$D(A) = A_{11}A_{12} - A_{12}A_{11} = 0.$$

Third, if A' is the matrix obtained from a 2×2 matrix A by interchanging its rows, then $D(A') = -D(A)$; for

$$\begin{aligned} D(A') &= A'_{11}A'_{22} - A'_{12}A'_{21} \\ &= A_{21}A_{12} - A_{22}A_{11} \\ &= -D(A). \end{aligned}$$

Definition. *Let* D *be an* n*-linear function. We say* D *is* **alternating** (*or* **alternate**) *if the following two conditions are satisfied:*

(a) D(A) = 0 *whenever two rows of* A *are equal.*

(b) *If* A′ *is a matrix obtained from* A *by interchanging two rows of* A, *then* D(A′) = −D(A).

We shall prove below that any n-linear function D which satisfies (a) automatically satisfies (b). We have put both properties in the definition of alternating n-linear function as a matter of convenience. The reader will probably also note that if D satisfies (b) and A is a matrix with two equal rows, then $D(A) = -D(A)$. It is tempting to conclude that D satisfies condition (a) as well. This is true, for example, if K is a field in which $1 + 1 \neq 0$, but in general (a) is not a consequence of (b).

Definition. *Let* K *be a commutative ring with identity, and let* n *be a positive integer. Suppose* D *is a function from* n × n *matrices over* K *into* K. *We say that* D *is a* **determinant function** *if* D *is* n*-linear, alternating, and* D(I) = 1.

As we stated earlier, we shall ultimately show that there is exactly one determinant function on $n \times n$ matrices over K. This is easily seen for 1×1 matrices $A = [a]$ over K. The function D given by $D(A) = a$ is a determinant function, and clearly this is the only determinant function on 1×1 matrices. We are also in a position to dispose of the case $n = 2$. The function

$$D(A) = A_{11}A_{22} - A_{12}A_{21}$$

was shown in Example 3 to be a determinant function. Furthermore, the formula exhibited in Example 2 shows that D is the only determinant function on 2×2 matrices. For we showed that for any 2-linear function D

$$D(A) = A_{11}A_{21}D(\epsilon_1, \epsilon_1) + A_{11}A_{22}D(\epsilon_1, \epsilon_2) + A_{12}A_{21}D(\epsilon_2, \epsilon_1) + A_{12}A_{22}D(\epsilon_2, \epsilon_2).$$

If D is alternating, then

$$D(\epsilon_1, \epsilon_1) = D(\epsilon_2, \epsilon_2) = 0$$

and

$$D(\epsilon_2, \epsilon_1) = -D(\epsilon_1, \epsilon_2) = -D(I).$$

If D also satisfies $D(I) = 1$, then

$$D(A) = A_{11}A_{22} - A_{12}A_{21}.$$

Example 4. Let F be a field and let D be any alternating 3-linear function on 3×3 matrices over the polynomial ring $F[x]$.

Let

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

If we denote the rows of the 3×3 identity matrix by $\epsilon_1, \epsilon_2, \epsilon_3$, then

$$D(A) = D(x\epsilon_1 - x^2\epsilon_3, \epsilon_2, \epsilon_1 + x^3\epsilon_3).$$

Since D is linear as a function of each row,

$$\begin{aligned} D(A) &= xD(\epsilon_1, \epsilon_2, \epsilon_1 + x^3\epsilon_3) - x^2D(\epsilon_3, \epsilon_2, \epsilon_1 + x^3\epsilon_3) \\ &= xD(\epsilon_1, \epsilon_2, \epsilon_1) + x^4D(\epsilon_1, \epsilon_2, \epsilon_3) - x^2D(\epsilon_3, \epsilon_2, \epsilon_1) - x^5D(\epsilon_3, \epsilon_2, \epsilon_3). \end{aligned}$$

Because D is alternating it follows that

$$D(A) = (x^4 + x^2)D(\epsilon_1, \epsilon_2, \epsilon_3).$$

Lemma. *Let D be a 2-linear function with the property that D(A) = 0 for all 2 × 2 matrices A over K having equal rows. Then D is alternating.*

Proof. What we must show is that if A is a 2×2 matrix and A' is obtained by interchanging the rows of A, then $D(A') = -D(A)$. If the rows of A are α and β, this means we must show that $D(\beta, \alpha) = -D(\alpha, \beta)$. Since D is 2-linear,

$$D(\alpha + \beta, \alpha + \beta) = D(\alpha, \alpha) + D(\alpha, \beta) + D(\beta, \alpha) + D(\beta, \beta).$$

By our hypothesis $D(\alpha + \beta, \alpha + \beta) = D(\alpha, \alpha) = D(\beta, \beta) = 0$. So

$$0 = D(\alpha, \beta) + D(\beta, \alpha). \quad \blacksquare$$

Lemma. *Let D be an n-linear function on n × n matrices over K. Suppose D has the property that D(A) = 0 whenever two adjacent rows of A are equal. Then D is alternating.*

Proof. We must show that $D(A) = 0$ when any two rows of A are equal, and that $D(A') = -D(A)$ if A' is obtained by interchanging some two rows of A. First, let us suppose that A' is obtained by interchanging two adjacent rows of A. The reader should see that the argument used in the proof of the preceding lemma extends to the present case and gives us $D(A') = -D(A)$.

Now let B be obtained by interchanging rows i and j of A, where $i < j$. We can obtain B from A by a succession of interchanges of pairs of adjacent rows. We begin by interchanging row i with row $(i + 1)$ and continue until the rows are in the order

$$\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_j, \alpha_i, \alpha_{j+1}, \ldots, \alpha_n.$$

This requires $k = j - i$ interchanges of adjacent rows. We now move α_j to the ith position using $(k - 1)$ interchanges of adjacent rows. We have thus obtained B from A by $k + (k - 1) = 2k - 1$ interchanges of adjacent rows. Thus

$$D(B) = (-1)^{2k-1}D(A) = -D(A).$$

Suppose A is any $n \times n$ matrix with two equal rows, say $\alpha_i = \alpha_j$ with $i < j$. If $j = i + 1$, then A has two equal and adjacent rows and $D(A) = 0$. If $j > i + 1$, we interchange α_{i+1} and α_j and the resulting matrix B has two equal and adjacent rows, so $D(B) = 0$. On the other hand, $D(B) = -D(A)$, hence $D(A) = 0$. ▌

Definition. *If* $\mathrm{n} > 1$ *and* A *is an* $\mathrm{n} \times \mathrm{n}$ *matrix over* K, *we let* $\mathrm{A(i|j)}$ *denote the* $(\mathrm{n} - 1) \times (\mathrm{n} - 1)$ *matrix obtained by deleting the* i*th row and* j*th column of* A. *If* D *is an* $(\mathrm{n} - 1)$*-linear function and* A *is an* $\mathrm{n} \times \mathrm{n}$ *matrix, we put* $\mathrm{D_{ij}(A) = D[A(i|j)]}$.

Theorem 1. *Let* $\mathrm{n} > 1$ *and let* D *be an alternating* $(\mathrm{n} - 1)$*-linear function on* $(\mathrm{n} - 1) \times (\mathrm{n} - 1)$ *matrices over* K. *For each* j, $1 \leq \mathrm{j} \leq \mathrm{n}$, *the function* $\mathrm{E_j}$ *defined by*

$$\mathrm{E_j(A)} = \sum_{\mathrm{i}=1}^{\mathrm{n}} (-1)^{\mathrm{i+j}}\mathrm{A_{ij}D_{ij}(A)} \tag{5-4}$$

is an alternating n*-linear function on* $\mathrm{n} \times \mathrm{n}$ *matrices* A. *If* D *is a determinant function, so is each* $\mathrm{E_j}$.

Proof. If A is an $n \times n$ matrix, $D_{ij}(A)$ is independent of the ith row of A. Since D is $(n - 1)$-linear, it is clear that D_{ij} is linear as a function of any row except row i. Therefore $A_{ij}D_{ij}(A)$ is an n-linear function of A. A linear combination of n-linear functions is n-linear; hence, E_j is n-linear. To prove that E_j is alternating, it will suffice to show that $E_j(A) = 0$ whenever A has two equal and adjacent rows. Suppose $\alpha_k = \alpha_{k+1}$. If $i \neq k$ and $i \neq k + 1$, the matrix $A(i|j)$ has two equal rows, and thus $D_{ij}(A) = 0$. Therefore

$$E_j(A) = (-1)^{k+j}A_{kj}D_{kj}(A) + (-1)^{k+1+j}A_{(k+1)j}D_{(k+1)j}(A).$$

Since $\alpha_k = \alpha_{k+1}$,

$$A_{kj} = A_{(k+1)j} \quad \text{and} \quad A(k|j) = A(k+1|j).$$

Clearly then $E_j(A) = 0$.

Now suppose D is a determinant function. If $I^{(n)}$ is the $n \times n$ identity matrix, then $I^{(n)}(j|j)$ is the $(n-1) \times (n-1)$ identity matrix $I^{(n-1)}$. Since $I_{ij}^{(n)} = \delta_{ij}$, it follows from (5-4) that

$$E_j(I^{(n)}) = D(I^{(n-1)}). \tag{5-5}$$

Now $D(I^{(n-1)}) = 1$, so that $E_j(I^{(n)}) = 1$ and E_j is a determinant function. ▌

Corollary. *Let* K *be a commutative ring with identity and let* n *be a positive integer. There exists at least one determinant function on* $K^{n \times n}$.

Proof. We have shown the existence of a determinant function on 1×1 matrices over K, and even on 2×2 matrices over K. Theorem 1 tells us explicitly how to construct a determinant function on $n \times n$ matrices, given such a function on $(n-1) \times (n-1)$ matrices. The corollary follows by induction. ▌

EXAMPLE 5. If B is a 2×2 matrix over K, we let

$$|B| = B_{11}B_{22} - B_{12}B_{21}.$$

Then $|B| = D(B)$, where D is the determinant function on 2×2 matrices. We showed that this function on $K^{2\times 2}$ is unique. Let

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}$$

be a 3×3 matrix over K. If we define E_1, E_2, E_3 as in (5-4), then

$$E_1(A) = A_{11}\begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix} - A_{21}\begin{vmatrix} A_{12} & A_{13} \\ A_{32} & A_{33} \end{vmatrix} + A_{31}\begin{vmatrix} A_{12} & A_{13} \\ A_{22} & A_{23} \end{vmatrix} \tag{5-6}$$

$$E_2(A) = -A_{12}\begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{vmatrix} + A_{22}\begin{vmatrix} A_{11} & A_{13} \\ A_{31} & A_{33} \end{vmatrix} - A_{32}\begin{vmatrix} A_{11} & A_{13} \\ A_{21} & A_{23} \end{vmatrix} \tag{5-7}$$

$$E_3(A) = A_{13}\begin{vmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{vmatrix} - A_{23}\begin{vmatrix} A_{11} & A_{12} \\ A_{31} & A_{32} \end{vmatrix} + A_{33}\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix}. \tag{5-8}$$

It follows from Theorem 1 that E_1, E_2, and E_3 are determinant functions. Actually, as we shall show later, $E_1 = E_2 = E_3$, but this is not yet apparent even in this simple case. It could, however, be verified directly, by expanding each of the above expressions. Instead of doing this we give some specific examples.

(a) Let $K = R[x]$ and

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

Then

$$E_1(A) = (x-1)\begin{vmatrix} x-2 & 1 \\ 0 & x-3 \end{vmatrix} = (x-1)(x-2)(x-3)$$

$$\begin{aligned} E_2(A) &= -x^2\begin{vmatrix} 0 & 1 \\ 0 & x-3 \end{vmatrix} + (x-2)\begin{vmatrix} x-1 & x^3 \\ 0 & x-3 \end{vmatrix} \\ &= (x-1)(x-2)(x-3) \end{aligned}$$

and

$$\begin{aligned} E_3(A) &= x^3\begin{vmatrix} 0 & x-2 \\ 0 & 0 \end{vmatrix} - \begin{vmatrix} x-1 & x^2 \\ 0 & 0 \end{vmatrix} + (x-3)\begin{vmatrix} x-1 & x^2 \\ 0 & x-2 \end{vmatrix} \\ &= (x-1)(x-2)(x-3). \end{aligned}$$

(b) Let $K = R$ and

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

Then

$$\begin{aligned} E_1(A) &= \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 \\ E_2(A) &= -\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = 1 \\ E_3(A) &= -\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = 1. \end{aligned}$$

Exercises

1. Each of the following expressions defines a function D on the set of 3×3 matrices over the field of real numbers. In which of these cases is D a 3-linear function?

(a) $D(A) = A_{11} + A_{22} + A_{33}$;
(b) $D(A) = (A_{11})^2 + 3A_{11}A_{22}$;
(c) $D(A) = A_{11}A_{12}A_{33}$;
(d) $D(A) = A_{13}A_{22}A_{32} + 5A_{12}A_{22}A_{32}$;
(e) $D(A) = 0$;
(f) $D(A) = 1$.

2. Verify directly that the three functions E_1, E_2, E_3 defined by (5-6), (5-7), and (5-8) are identical.

3. Let K be a commutative ring with identity. If A is a 2×2 matrix over K, the **classical adjoint** of A is the 2×2 matrix adj A defined by

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

If det denotes the unique determinant function on 2×2 matrices over K, show that

(a) $(\operatorname{adj} A)A = A(\operatorname{adj} A) = (\det A)I$;

(b) $\det(\operatorname{adj} A) = \det(A)$;

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

(A^t denotes the transpose of A.)

4. Let A be a 2×2 matrix over a field F. Show that A is invertible if and only if $\det A \neq 0$. When A is invertible, give a formula for A^{-1}.

5. Let A be a 2×2 matrix over a field F, and suppose that $A^2 = 0$. Show for each scalar c that $\det(cI - A) = c^2$.

6. Let K be a subfield of the complex numbers and n a positive integer. Let $j_1, \ldots, j_n$ and $k_1, \ldots, k_n$ be positive integers not exceeding n. For an $n \times n$ matrix A over K define

$$D(A) = A(j_1, k_1)A(j_2, k_2) \cdots A(j_n, k_n).$$

Prove that D is n-linear if and only if the integers $j_1, \ldots, j_n$ are distinct.

7. Let K be a commutative ring with identity. Show that the determinant function on 2×2 matrices A over K is alternating and 2-linear as a function of the columns of A.

8. Let K be a commutative ring with identity. Define a function D on 3×3 matrices over K by the rule

$$D(A) = A_{11} \det \begin{bmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{bmatrix} - A_{12} \det \begin{bmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{bmatrix} + A_{13} \det \begin{bmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix}.$$

Show that D is alternating and 3-linear as a function of the columns of A.

9. Let K be a commutative ring with identity and D an alternating n-linear function on $n \times n$ matrices over K. Show that

(a) $D(A) = 0$, if one of the rows of A is 0.

(b) $D(B) = D(A)$, if B is obtained from A by adding a scalar multiple of one row of A to another.

10. Let F be a field, A a 2×3 matrix over F, and (c_1, c_2, c_3) the vector in F^3 defined by

$$c_1 = \begin{vmatrix} A_{12} & A_{13} \\ A_{22} & A_{23} \end{vmatrix}, \qquad c_2 = \begin{vmatrix} A_{13} & A_{11} \\ A_{23} & A_{21} \end{vmatrix}, \qquad c_3 = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix}.$$

Show that

(a) $\operatorname{rank}(A) = 2$ if and only if $(c_1, c_2, c_3) \neq 0$;

(b) if A has rank 2, then (c_1, c_2, c_3) is a basis for the solution space of the system of equations $AX = 0$.

11. Let K be a commutative ring with identity, and let D be an alternating 2-linear function on 2×2 matrices over K. Show that $D(A) = (\det A)D(I)$ for all A. Now use this result (no computations with the entries allowed) to show that $\det(AB) = (\det A)(\det B)$ for any 2×2 matrices A and B over K.

12. Let F be a field and D a function on $n \times n$ matrices over F (with values in F). Suppose $D(AB) = D(A)D(B)$ for all A, B. Show that either $D(A) = 0$ for all A, or $D(I) = 1$. In the latter case show that $D(A) \neq 0$ whenever A is invertible.

13. Let R be the field of real numbers, and let D be a function on 2×2 matrices 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

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

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

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