ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS 20

FINITE FIELDS

RUDOLF LIDL & HARALD NIEDERREITER

CAMBRIDGE

more information – www.cambridge.org/9780521392310

This page intentionally left blank

ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS

EDITED BY G.-C. ROTA

VOLUME 20

Finite fields

ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS

4 W. Miller, Jr. *Symmetry and separation of variables*
6 H. Minc *Permanents*
11 W. B. Jones and W. J. Thron *Continured fractions*
12 N. F. G. Martin and J. W. England *Mathematical theory of entropy*
18 H. O. Fattorini *The Cauchy problem*
19 G. G. Lorentz, K. Jetter, and S. D. Riemenschneider *Birkhoff interpolation*
21 W. T. Tutte *Graph theory*
22 J. R. Bastida *Field extensions and Galois theory*
23 J. R. Cannon *The one-dimensional heat equation*
25 A. Salomaa *Computation and automata*
26 N. White (ed.) *Theory of matroids*
27 N. H. Bingham, C. M. Goldie, and J. L. Teugels *Regular variation*
28 P. P. Petrushev and V. A. Popov *Rational aproximation of real functions*
29 N. White (ed.) *Combinatorial geometries*
30 M. Pohst and H. Zassenhaus *Algorithmic algebraic number theory*
31 J. Aczel and J. Dhombres *Functional equations containing several variables*
32 M. Kuczma, B. Chozewski, and R. Ger *Iterative functional equations*
33 R. V. Ambartzumian *Factorization calculus and geometric probability*
34 G. Gripenberg, S.-O. London, and O. Staffans *Volerra integral and functional equations*
35 G. Gasper and M. Rahman *Basic hypergeometric series*
36 E. Torgersen *Comparison of statistical experiments*
37 A. Neumaier *Interval methods for systems of equations*
38 N. Korneichuk *Exact constants in appoximation theory*
39 R. A. Brualdi and H. J. Ryser *Combinatorial matrix theory*
40 N. White (ed.) *Matroid applications*
41 S. Sakai *Operator algebras in dynamical systems*
42 W. Hodges *Model theory*
43 H. Stahl and V. Totik *General orthogonal polynomials*
44 R. Schneider *Convex bodies*
45 G. Da Prato and J. Zabczyk *Stochastic equations in infinite dimensions*
46 A. Bjorner, M. Las Vergnas, B. Sturmfels, N. White, and G. Ziegler *Oriented matroids*
47 G. A. Edgar and L. Sucheston *Stopping times and directed processes*
48 C. Sims *Computation with finitely presented groups*
49 T. Palmer *Banach algebras and the general theory of*-algebras*
50 F. Borceux *Handbook of Categorical Algebra I*
51 F. Borceux *Handbook of Categorical Algebra II*
52 F. Borceux *Handbook of Categorical Algebra III*
54 A. Katok and B. Hasselblatt *Introduction to the modern theory of dynamical systems*
55 V. N. Sachkov *Combinatorial methods in discrete mathematics*
57 P. M. Cohn *Skew fields*
58 R. Gardner *Geometric tomography*
60 J. Krajiček *Bounded arithmetic, propositional logic and complexity theory*

ENCYCLOPEDIA OF MATHEMATICS AND ITS APPLICATIONS

Finite fields

Rudolf Lidl
University of Tasmania
Hobart, Australia

Harald Niederreiter
Austrian Academy of Sciences
Vienna, Austria

Foreword by
P. M. Cohn
University of London
London, England

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011–4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published by Addison-Wesley Publishing Inc. 1983
Second edition published by Cambridge University Press 1997
Reprinted 2000

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data available

ISBN 0 521 39231 4 hardback

Transferred to digital printing 2003

To Pamela and Gerlinde

Contents

Foreword

Most modern algebra texts devote a few pages (but no more) to finite fields. So at first it may come as a surprise to see an entire book on the subject, and even more for it to appear in the *Encyclopedia of Mathematics and Its Applications*. But the reader of this book will find that the authors performed the very timely task of drawing together the different threads of development that have emanated from the subject. Foremost among these developments is the rapid growth of coding theory which already has been treated in R. J. McEliece's volume in this series. The present volume deals with coding theory in the wider context of polynomial theory over finite fields, and also establishes the connection with linear recurring series and shift registers.

On the pure side there is a good deal of number theory that is most naturally expressed in terms of finite fields. Much of this—for example, equations over finite fields and exponential sums—can serve as a paradigm for the more general case; and the authors have gone as far in their treatment as is reasonable, using elementary algebraic methods only. As a result the book can also serve as an introduction to these topics.

But finite fields also have properties that are not shared with other types of algebra; thus they (like finite Boolean algebras) are functionally complete. This means that every mapping of a finite field can be expressed as a polynomial. While the proof is not hard (it is an immediate consequence of the Lagrange interpolation formula), practical questions arise when we try to find polynomials effecting permutations. Such permutation polynomials

are useful in several contexts, and methods of obtaining them are discussed here. True to its nature as a handbook of applications, this volume also gives various algorithms for factorizing polynomials (over both large and small finite fields).

The lengthy notes at the end of each chapter contain interesting historical perspectives, and the comprehensive bibliography helps to make this volume truly the handbook of finite fields.

P. M. COHN

Preface

The theory of finite fields is a branch of modern algebra that has come to the fore in the last 50 years because of its diverse applications in combinatorics, coding theory, and the mathematical study of switching circuits, among others. The origins of the subject reach back into the 17th and 18th century, with such eminent mathematicians as Pierre de Fermat (1601–1665), Leonhard Euler (1707–1783), Joseph-Louis Lagrange (1736–1813), and Adrien-Marie Legendre (1752–1833) contributing to the structure theory of special finite fields—namely, the so-called finite prime fields. The general theory of finite fields may be said to begin with the work of Carl Friedrich Gauss (1777–1855) and Evariste Galois (1811–1832), but it only became of interest for applied mathematicians in recent decades with the emergence of discrete mathematics as a serious discipline.

In this book, which is the first one devoted entirely to finite fields, we have aimed at presenting both the classical and the applications-oriented aspect of the subject. Thus, in addition to what has to be considered the essential core of the theory, the reader will find results and techniques that are of importance mainly because of their use in applications. Because of the vastness of the subject, limitations had to be imposed on the choice of material. In trying to make the book as self-contained as possible, we have refrained from discussing results or methods that belong properly to algebraic geometry or to the theory of algebraic function fields. Applications are described to the extent to which this can be done without too much

digression. The only noteworthy prerequisite for the book is a background in linear algebra, on the level of a first course on this topic. A rudimentary knowledge of analysis is needed in a few passages. Prior exposure to abstract algebra is certainly helpful, although all the necessary information is summarized in Chapter 1.

Chapter 2 is basic for the rest of the book as it contains the general structure theory of finite fields as well as the discussion of concepts that are used throughout the book. Chapter 3 on the theory of polynomials and Chapter 4 on factorization algorithms for polynomials are closely linked and should best be studied together. A similar unit is formed by Chapters 5 and 6. Chapters 7 and 8 can be read independently of each other and depend mostly on Chapters 2 and 3. The applications presented in Chapter 9 draw on various material in the previous chapters. Chapter 10 supplements parts of Chapters 2 and 3.

Each chapter starts with a brief description of its contents, hence it should not be necessary to give a synopsis of the book here. As this volume is part of an encyclopedic series, we have attempted to provide as much information as possible in a limited space, which meant, in particular, the omission of a few cumbersome proofs. Bibliographical references have been relegated to the notes at the end of each chapter so as not to clutter the main text. These notes also provide the researcher in the field with a survey of the literature and a summary of further results. The bibliography at the end of the volume collects all the references given in the notes.

In order to enhance the attractiveness of this monograph as a textbook, we have inserted worked-out examples at appropriate points in the text and included lists of exercises for Chapters 1–9. These exercises range from routine problems to alternative proofs of key theorems, but contain also material going beyond what is covered in the text.

With regard to cross-references, we have numbered all items in the main text consecutively by chapters, regardless of whether they are definitions, theorems, examples, and so on. Thus, "Definition 2.41" refers to item 41 in Chapter 2 (which happens to be a definition) and "Remark 6.28" refers to item 28 in Chapter 6 (which happens to be a remark). In the same vein, "Exercise 5.31" refers to the list of exercises in Chapter 5.

It is with great pleasure that we express our gratitude to Professor Gian-Carlo Rota for inviting us to write this book and for his patience in waiting for the result of our effort. We gratefully acknowledge the help of Mrs. Melanie Barton, who typed the manuscript with great care and efficiency. The staff of Addison-Wesley deserves our thanks for its professionalism in the production of the book.

R. LIDL
H. NIEDERREITER

Chapter 1

Algebraic Foundations

This introductory chapter contains a survey of some basic algebraic concepts that will be employed throughout the book. Elementary algebra uses the operations of arithmetic such as addition and multiplication, but replaces particular numbers by symbols and thereby obtains formulas that, by substitution, provide solutions to specific numerical problems. In modern algebra the level of abstraction is raised further: instead of dealing with the familiar operations on real numbers, one treats general operations —processes of combining two or more elements to yield another element—in general sets. The aim is to study the common properties of all systems consisting of sets on which are defined a fixed number of operations interrelated in some definite way—for instance, sets with two binary operations behaving like $+$ and $\cdot$ for the real numbers.

Only the most fundamental definitions and properties of algebraic systems—that is, of sets together with one or more operations on the set—will be introduced, and the theory will be discussed only to the extent needed for our special purposes in the study of finite fields later on. We state some standard results without proof. With regard to sets we adopt the naive standpoint. We use the following sets of numbers: the set $\mathbb{N}$ of natural numbers, the set $\mathbb{Z}$ of integers, the set $\mathbb{Q}$ of rational numbers, the set $\mathbb{R}$ of real numbers, and the set $\mathbb{C}$ of complex numbers.

1. GROUPS

In the set of all integers the two operations addition and multiplication are well known. We can generalize the concept of operation to arbitrary sets. Let S be a set and let $S \times S$ denote the set of all ordered pairs (s, t) with $s \in S$, $t \in S$. Then a mapping from $S \times S$ into S will be called a (*binary*) *operation* on S. Under this definition we require that the image of $(s, t) \in S \times S$ must be in S; this is the *closure property* of an operation. By an *algebraic structure* or *algebraic system* we mean a set S together with one or more operations on S.

In elementary arithmetic we are provided with two operations, addition and multiplication, that have associativity as one of their most important properties. Of the various possible algebraic systems having a single associative operation, the type known as a group has been by far the most extensively studied and developed. The theory of groups is one of the oldest parts of abstract algebra as well as one particularly rich in applications.

1.1. Definition. A *group* is a set G together with a binary operation $*$ on G such that the following three properties hold:

1. $*$ is *associative*; that is, for any $a, b, c \in G$,
$$a * (b * c) = (a * b) * c.$$
2. There is an *identity* (or *unity*) *element* e in G such that for all $a \in G$,
$$a * e = e * a = a.$$
3. For each $a \in G$, there exists an *inverse element* $a^{-1} \in G$ such that
$$a * a^{-1} = a^{-1} * a = e.$$

If the group also satisfies

4. For all $a, b \in G$,
$$a * b = b * a,$$

then the group is called *abelian* (or *commutative*).

It is easily shown that the identity element e and the inverse element a^{-1} of a given element $a \in G$ are uniquely determined by the properties above. Furthermore, $(a * b)^{-1} = b^{-1} * a^{-1}$ for all $a, b \in G$. For simplicity, we shall frequently use the notation of ordinary multiplication to designate the operation in the group, writing simply ab instead of $a * b$. But it must be emphasized that by doing so we do not assume that the operation actually is ordinary multiplication. Sometimes it is also convenient to write $a + b$ instead of $a * b$ and $-a$ instead of a^{-1}, but this additive notation is usually reserved for abelian groups.

The associative law guarantees that expressions such as $a_1 a_2 \cdots a_n$ with $a_j \in G$, $1 \leqslant j \leqslant n$, are unambiguous, since no matter how we insert parentheses, the expression will always represent the same element of G. To indicate the n-fold composite of an element $a \in G$ with itself, where $n \in \mathbb{N}$, we shall write

$$a^n = aa \cdots a \qquad (n \text{ factors } a)$$

if using multiplicative notation, and we call a^n the nth power of a. If using additive notation for the operation $*$ on G, we write

$$na = a + a + \cdots + a \qquad (n \text{ summands } a).$$

Following customary notation, we have the following rules:

Multiplicative Notation	*Additive Notation*
$a^{-n} = (a^{-1})^n$	$(-n)a = n(-a)$
$a^n a^m = a^{n+m}$	$na + ma = (n+m)a$
$(a^n)^m = a^{nm}$	$m(na) = (mn)a$

For $n = 0 \in \mathbf{Z}$, one adopts the convention $a^0 = e$ in the multiplicative notation and $0a = 0$ in the additive notation, where the last "zero" represents the identity element of G.

1.2. Examples

(i) Let G be the set of integers with the operation of addition. The ordinary sum of two integers is a unique integer and the associativity is a familiar fact. The identity element is 0 (zero), and the inverse of an integer a is the integer $-a$. We denote this group by $\mathbf{Z}$.

(ii) The set consisting of a single element e, with the operation $*$ defined by $e * e = e$, forms a group.

(iii) Let G be the set of remainders of all the integers on division by 6—that is, $G = \{0, 1, 2, 3, 4, 5\}$—and let $a * b$ be the remainder on division by 6 of the ordinary sum of a and b. The existence of an identity element and of inverses is again obvious. In this case, it requires some computation to establish the associativity of $*$. This group can be readily generalized by replacing the integer 6 by any positive integer n. □

These examples lead to an interesting class of groups in which every element is a power of some fixed element of the group. If the group operation is written as addition, we refer to "multiple" instead of "power" of an element.

1.3. Definition. A multiplicative group G is said to be *cyclic* if there is an element $a \in G$ such that for any $b \in G$ there is some integer j with $b = a^j$.

Such an element a is called a *generator* of the cyclic group, and we write $G = \langle a \rangle$.

It follows at once from the definition that every cyclic group is commutative. We also note that a cyclic group may very well have more than one element that is a generator of the group. For instance, in the additive group $\mathbf{Z}$ both 1 and -1 are generators.

With regard to the "additive" group of remainders of the integers on division by n, the generalization of Example 1.2(iii), we find that the type of operation used there leads to an equivalence relation on the set of integers. In general, a subset R of $S \times S$ is called an *equivalence relation* on a set S if it has the following three properties:

(a) $(s, s) \in R$ for all $s \in S$ (*reflexivity*).
(b) If $(s, t) \in R$, then $(t, s) \in R$ (*symmetry*).
(c) If $(s, t), (t, u) \in R$, then $(s, u) \in R$ (*transitivity*).

The most obvious example of an equivalence relation is that of equality. It is an important fact that an equivalence relation R on a set S induces a partition of S—that is, a representation of S as the union of nonempty, mutually disjoint subsets of S. If we collect all elements of S equivalent to a fixed $s \in S$, we obtain the *equivalence class* of s, denoted by

$$[s] = \{t \in S : (s, t) \in R\}.$$

The collection of all distinct equivalence classes forms then the desired partition of S. We note that $[s] = [t]$ precisely if $(s, t) \in R$. Example 1.2(iii) suggests the following concept.

1.4. Definition. For arbitrary integers a, b and a positive integer n, we say that a is *congruent* to b modulo n, and write $a \equiv b \bmod n$, if the difference $a - b$ is a multiple of n—that is, if $a = b + kn$ for some integer k.

It is easily verified that "congruence modulo n" is an equivalence relation on the set $\mathbf{Z}$ of integers. The relation is obviously reflexive and symmetric. The transitivity also follows easily: if $a = b + kn$ and $b = c + ln$ for some integers k and l, then $a = c + (k + l)n$, so that $a \equiv b \bmod n$ and $b \equiv c \bmod n$ together imply $a \equiv c \bmod n$.

Consider now the equivalence classes into which the relation of congruence modulo n partitions the set $\mathbf{Z}$. These will be the sets

$$\begin{aligned}
[0] &= \{\ldots, -2n, -n, 0, n, 2n, \ldots\}, \\
[1] &= \{\ldots, -2n+1, -n+1, 1, n+1, 2n+1, \ldots\}, \\
&\vdots \\
[n-1] &= \{\ldots, -n-1, -1, n-1, 2n-1, 3n-1, \ldots\}.
\end{aligned}$$

We may define on the set $\{[0], [1], \ldots, [n-1]\}$ of equivalence classes a binary

operation (which we shall again write as $+$, although it is certainly not ordinary addition) by

$$[a]+[b]=[a+b], \tag{1.1}$$

where a and b are any elements of the respective sets $[a]$ and $[b]$ and the sum $a+b$ on the right is the ordinary sum of a and b. In order to show that we have actually defined an operation—that is, that this operation is well defined—we must verify that the image element of the pair $([a],[b])$ is uniquely determined by $[a]$ and $[b]$ alone and does not depend in any way on the representatives a and b. We leave this proof as an exercise. Associativity of the operation in (1.1) follows from the associativity of ordinary addition. The identity element is $[0]$ and the inverse of $[a]$ is $[-a]$. Thus the elements of the set $\{[0],[1],\ldots,[n-1]\}$ form a group.

1.5. Definition. The group formed by the set $\{[0],[1],\ldots,[n-1]\}$ of equivalence classes modulo n with the operation (1.1) is called the *group of integers modulo n* and denoted by $\mathbf{Z}_n$.

$\mathbf{Z}_n$ is actually a cyclic group with the equivalence class $[1]$ as a generator, and it is a group of order n according to the following definition.

1.6. Definition. A group is called *finite* (resp. *infinite*) if it contains finitely (resp. infinitely) many elements. The number of elements in a finite group is called its *order*. We shall write $|G|$ for the order of the finite group G.

There is a convenient way of presenting a finite group. A table displaying the group operation, nowadays referred to as a *Cayley table*, is constructed by indexing the rows and the columns of the table by the group elements. The element appearing in the row indexed by a and the column indexed by b is then taken to be ab.

1.7. Example. The Cayley table for the group $\mathbf{Z}_6$ is:

$+$	$[0]$	$[1]$	$[2]$	$[3]$	$[4]$	$[5]$
$[0]$	$[0]$	$[1]$	$[2]$	$[3]$	$[4]$	$[5]$
$[1]$	$[1]$	$[2]$	$[3]$	$[4]$	$[5]$	$[0]$
$[2]$	$[2]$	$[3]$	$[4]$	$[5]$	$[0]$	$[1]$
$[3]$	$[3]$	$[4]$	$[5]$	$[0]$	$[1]$	$[2]$
$[4]$	$[4]$	$[5]$	$[0]$	$[1]$	$[2]$	$[3]$
$[5]$	$[5]$	$[0]$	$[1]$	$[2]$	$[3]$	$[4]$

□

A group G contains certain subsets that form groups in their own right under the operation of G. For instance, the subset $\{[0],[2],[4]\}$ of $\mathbf{Z}_6$ is easily seen to have this property.

1.8. Definition. A subset H of the group G is a *subgroup* of G if H is itself a group with respect to the operation of G. Subgroups of G other than the *trivial subgroups* $\langle e \rangle$ and G itself are called *nontrivial subgroups* of G.

One verifies at once that for any fixed a in a group G, the set of all powers of a is a subgroup of G.

1.9. Definition. The subgroup of G consisting of all powers of the element a of G is called the subgroup *generated by* a and is denoted by $\langle a \rangle$. This subgroup is necessarily cyclic. If $\langle a \rangle$ is finite, then its order is called the *order* of the element a. Otherwise, a is called an element of *infinite order*.

Thus, a is of finite order k if k is the least positive integer such that $a^k = e$. Any other integer m with $a^m = e$ is then a multiple of k. If S is a nonempty subset of a group G, then the subgroup H of G consisting of all finite products of powers of elements of S is called the subgroup *generated by* S, denoted by $H = \langle S \rangle$. If $\langle S \rangle = G$, we say that S *generates* G, or that G is *generated by* S.

For a positive element n of the additive group $\mathbf{Z}$ of integers, the subgroup $\langle n \rangle$ is closely associated with the notion of congruence modulo n, since $a \equiv b \bmod n$ if and only if $a - b \in \langle n \rangle$. Thus the subgroup $\langle n \rangle$ defines an equivalence relation on $\mathbf{Z}$. This situation can be generalized as follows.

1.10. Theorem. *If H is a subgroup of G, then the relation R_H on G defined by $(a, b) \in R_H$ if and only if $a = bh$ for some $h \in H$, is an equivalence relation.*

The proof is immediate. The equivalence relation R_H is called *left congruence* modulo H. Like any equivalence relation, it induces a partition of G into nonempty, mutually disjoint subsets. These subsets (= equivalence classes) are called the *left cosets* of G modulo H and they are denoted by

$$aH = \{ah : h \in H\}$$

(or $a + H = \{a + h : h \in H\}$ if G is written additively), where a is a fixed element of G. Similarly, there is a decomposition of G into *right cosets* modulo H, which have the form $Ha = \{ha : h \in H\}$. If G is abelian, then the distinction between left and right cosets modulo H is unnecessary.

1.11. Example. Let $G = \mathbf{Z}_{12}$ and let H be the subgroup $\{[0],[3],[6],[9]\}$. Then the distinct (left) cosets of G modulo H are given by:

$$[0] + H = \{[0],[3],[6],[9]\},$$
$$[1] + H = \{[1],[4],[7],[10]\},$$
$$[2] + H = \{[2],[5],[8],[11]\}. \qquad \square$$

1.12. Theorem. *If H is a finite subgroup of G, then every (left or right) coset of G modulo H has the same number of elements as H.*

1.13. Definition. If the subgroup H of G only yields finitely many distinct left cosets of G modulo H, then the number of such cosets is called the *index* of H in G.

Since the left cosets of G modulo H form a partition of G, Theorem 1.12 implies the following important result.

1.14. Theorem. *The order of a finite group G is equal to the product of the order of any subgroup H and the index of H in G. In particular, the order of H divides the order of G and the order of any element $a \in G$ divides the order of G.*

The subgroups and the orders of elements are easy to describe for cyclic groups. We summarize the relevant facts in the subsequent theorem.

1.15. Theorem

(i) *Every subgroup of a cyclic group is cyclic.*

(ii) *In a finite cyclic group $\langle a \rangle$ of order m, the element a^k generates a subgroup of order $m/\gcd(k, m)$, where $\gcd(k, m)$ denotes the greatest common divisor of k and m.*

(iii) *If d is a positive divisor of the order m of a finite cyclic group $\langle a \rangle$, then $\langle a \rangle$ contains one and only one subgroup of index d. For any positive divisor f of m, $\langle a \rangle$ contains precisely one subgroup of order f.*

(iv) *Let f be a positive divisor of the order of a finite cyclic group $\langle a \rangle$. Then $\langle a \rangle$ contains $\phi(f)$ elements of order f. Here $\phi(f)$ is* Euler's function *and indicates the number of integers n with $1 \leqslant n \leqslant f$ that are relatively prime to f.*

(v) *A finite cyclic group $\langle a \rangle$ of order m contains $\phi(m)$ generators — that is, elements a^r such that $\langle a^r \rangle = \langle a \rangle$. The generators are the powers a^r with $\gcd(r, m) = 1$.*

Proof. (i) Let H be a subgroup of the cyclic group $\langle a \rangle$ with $H \neq \{e\}$. If $a^n \in H$, then $a^{-n} \in H$; hence H contains at least one power of a with a positive exponent. Let d be the least positive exponent such that $a^d \in H$, and let $a^s \in H$. Dividing s by d gives $s = qd + r$, $0 \leqslant r < d$, and $q, r \in \mathbb{Z}$. Thus $a^s(a^{-d})^q = a^r \in H$, which contradicts the minimality of d, unless $r = 0$. Therefore the exponents of all powers of a that belong to H are divisible by d, and so $H = \langle a^d \rangle$.

(ii) Put $d = \gcd(k, m)$. The order of $\langle a^k \rangle$ is the least positive integer n such that $a^{kn} = e$. The latter identity holds if and only if m divides kn, or equivalently, if and only if m/d divides n. The least positive n with this property is $n = m/d$.

(iii) If d is given, then $\langle a^d \rangle$ is a subgroup of order m/d, and so of index d, because of (ii). If $\langle a^k \rangle$ is another subgroup of index d, then its

order is m/d, and so $d = \gcd(k, m)$ by (ii). In particular, d divides k, so that $a^k \in \langle a^d \rangle$ and $\langle a^k \rangle$ is a subgroup of $\langle a^d \rangle$. But since both groups have the same order, they are identical. The second part follows immediately because the subgroups of order f are precisely the subgroups of index m/f.

(iv) Let $|\langle a \rangle| = m$ and $m = df$. By (ii), an element a^k is of order f if and only if $\gcd(k, m) = d$. Hence, the number of elements of order f is equal to the number of integers k with $1 \leqslant k \leqslant m$ and $\gcd(k, m) = d$. We may write $k = dh$ with $1 \leqslant h \leqslant f$, the condition $\gcd(k, m) = d$ being now equivalent to $\gcd(h, f) = 1$. The number of these h is equal to $\phi(f)$.

(v) The generators of $\langle a \rangle$ are precisely the elements of order m, so that the first part is implied by (iv). The second part follows from (ii). □

When comparing the structures of two groups, mappings between the groups that preserve the operations play an important role.

1.16. Definition. A mapping $f: G \to H$ of the group G into the group H is called a *homomorphism* of G into H if f preserves the operation of G. That is, if $*$ and $\cdot$ are the operations of G and H, respectively, then f preserves the operation of G if for all $a, b \in G$ we have $f(a * b) = f(a) \cdot f(b)$. If, in addition, f is onto H, then f is called an *epimorphism* (or *homomorphism "onto"*) and H is a *homomorphic image* of G. A homomorphism of G into G is called an *endomorphism*. If f is a one-to-one homomorphism of G onto H, then f is called an *isomorphism* and we say that G and H are *isomorphic*. An isomorphism of G onto G is called an *automorphism*.

Consider, for instance, the mapping f of the additive group $\mathbf{Z}$ of the integers onto the group $\mathbf{Z}_n$ of the integers modulo n, defined by $f(a) = [a]$. Then

$$f(a + b) = [a + b] = [a] + [b] = f(a) + f(b) \quad \text{for } a, b \in \mathbf{Z},$$

and f is a homomorphism.

If $f: G \to H$ is a homomorphism and e is the identity element in G, then $ee = e$ implies $f(e)f(e) = f(e)$, so that $f(e) = e'$, the identity element in H. From $aa^{-1} = e$ we get $f(a^{-1}) = (f(a))^{-1}$ for all $a \in G$.

The automorphisms of a group G are often of particular interest, partly because they themselves form a group with respect to the usual composition of mappings, as can be easily verified. Important examples of automorphisms are the *inner automorphisms*. For fixed $a \in G$, define f_a by $f_a(b) = aba^{-1}$ for $b \in G$. Then f_a is an automorphism of G of the indicated type, and we get all inner automorphisms of G by letting a run through all elements of G. The elements b and aba^{-1} are said to be *conjugate*, and for a nonempty subset S of G the set $aSa^{-1} = \{asa^{-1} : s \in S\}$ is called a *conjugate of* S. Thus, the conjugates of S are just the images of S under the various inner automorphisms of G.

1.17. Definition. The *kernel* of the homomorphism $f: G \to H$ of the group G into the group H is the set

$$\ker f = \{a \in G : f(a) = e'\},$$

where e' is the identity element in H.

1.18. Example. For the homomorphism $f: \mathbb{Z} \to \mathbb{Z}_n$ given by $f(a) = [a]$, $\ker f$ consists of all $a \in \mathbb{Z}$ with $[a] = [0]$. Since this condition holds exactly for all multiples a of n, we have $\ker f = \langle n \rangle$, the subgroup of $\mathbb{Z}$ generated by n. □

It is easily checked that $\ker f$ is always a subgroup of G. Moreover, $\ker f$ has a special property: whenever $a \in G$ and $b \in \ker f$, then $aba^{-1} \in \ker f$. This leads to the following concept.

1.19. Definition. The subgroup H of the group G is called a *normal* subgroup of G if $aha^{-1} \in H$ for all $a \in G$ and all $h \in H$.

Every subgroup of an abelian group is normal since we then have $aha^{-1} = aa^{-1}h = eh = h$. We shall state some alternative characterizations of the property of normality of a subgroup.

1.20. Theorem

(i) *The subgroup H of G is normal if and only if H is equal to its conjugates, or equivalently, if and only if H is invariant under all the inner automorphisms of G.*

(ii) *The subgroup H of G is normal if and only if the left coset aH is equal to the right coset Ha for every $a \in G$.*

One important feature of a normal subgroup is the fact that the set of its (left) cosets can be endowed with a group structure.

1.21. Theorem. *If H is a normal subgroup of G, then the set of (left) cosets of G modulo H forms a group with respect to the operation $(aH)(bH) = (ab)H$.*

1.22. Definition. For a normal subgroup H of G, the group formed by the (left) cosets of G modulo H under the operation in Theorem 1.21 is called the *factor group* (or *quotient group*) of G modulo H and denoted by G/H.

If G/H is finite, then its order is equal to the index of H in G. Thus, by Theorem 1.14, we get for a finite group G,

$$|G/H| = \frac{|G|}{|H|}.$$

Each normal subgroup of a group G determines in a natural way a homomorphism of G and vice versa.

1.23. ***Theorem*** (Homomorphism Theorem). *Let* $f: G \to f(G) = G_1$ *be a homomorphism of a group* G *onto a group* G_1. *Then* $\ker f$ *is a normal subgroup of* G, *and the group* G_1 *is isomorphic to the factor group* $G/\ker f$. *Conversely, if* H *is any normal subgroup of* G, *then the mapping* $\psi: G \to G/H$ *defined by* $\psi(a) = aH$ *for* $a \in G$ *is a homomorphism of* G *onto* G/H *with* $\ker \psi = H$.

We shall now derive a relation known as the *class equation* for a finite group, which will be needed in Chapter 2, Section 6.

1.24. Definition. Let S be a nonempty subset of a group G. The *normalizer* of S in G is the set $N(S) = \{a \in G: aSa^{-1} = S\}$.

1.25. ***Theorem.*** *For any nonempty subset* S *of the group* G, $N(S)$ *is a subgroup of* G *and there is a one-to-one correspondence between the left cosets of* G *modulo* $N(S)$ *and the distinct conjugates* aSa^{-1} *of* S.

Proof. We have $e \in N(S)$, and if $a, b \in N(S)$, then a^{-1} and ab are also in $N(S)$, so that $N(S)$ is a subgroup of G. Now

$$aSa^{-1} = bSb^{-1} \Leftrightarrow S = a^{-1}bSb^{-1}a = (a^{-1}b)S(a^{-1}b)^{-1}$$
$$\Leftrightarrow a^{-1}b \in N(S) \Leftrightarrow b \in aN(S).$$

Thus, conjugates of S are equal if and only if they are defined by elements in the same left coset of G modulo $N(S)$, and so the second part of the theorem is shown. □

If we collect all elements conjugate to a fixed element a, we obtain a set called the *conjugacy class* of a. For certain elements the corresponding conjugacy class has only one member, and this will happen precisely for the elements of the center of the group.

1.26. Definition. For any group G, the *center* of G is defined as the set $C = \{c \in G: ac = ca \text{ for all } a \in G\}$.

It is straightforward to check that the center C is a normal subgroup of G. Clearly, G is abelian if and only if $C = G$. A counting argument leads to the following result.

1.27. ***Theorem*** (Class Equation). *Let* G *be a finite group with center* C. *Then*

$$|G| = |C| + \sum_{i=1}^{k} n_i,$$

where each n_i *is* $\geqslant 2$ *and a divisor of* $|G|$. *In fact,* $n_1, n_2, \ldots, n_k$ *are the numbers of elements of the distinct conjugacy classes in* G *containing more than one member.*

Proof. Since the relation "a is conjugate to b" is an equivalence relation on G, the distinct conjugacy classes in G form a partition of G. Thus, $|G|$ is equal to the sum of the numbers of elements of the distinct conjugacy classes. There are $|C|$ conjugacy classes (corresponding to the elements of C) containing only one member, whereas $n_1, n_2, \ldots, n_k$ are the numbers of elements of the remaining conjugacy classes. This yields the class equation. To show that each n_i divides $|G|$, it suffices to note that n_i is the number of conjugates of some $a \in G$ and so equal to the number of left cosets of G modulo $N(\langle a \rangle)$ by Theorem 1.25. □

2. RINGS AND FIELDS

In most of the number systems used in elementary arithmetic there are two distinct binary operations: addition and multiplication. Examples are provided by the integers, the rational numbers, and the real numbers. We now define a type of algebraic structure known as a ring that shares some of the basic properties of these number systems.

1.28. Definition. A *ring* $(R, +, \cdot)$ is a set R, together with two binary operations, denoted by $+$ and $\cdot$, such that:

1. R is an abelian group with respect to $+$.
2. $\cdot$ is associative—that is, $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in R$.
3. The *distributive laws* hold; that is, for all $a, b, c \in R$ we have $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(b + c) \cdot a = b \cdot a + c \cdot a$.

We shall use R as a designation for the ring $(R, +, \cdot)$ and stress that the operations $+$ and $\cdot$ are not necessarily the ordinary operations with numbers. In following convention, we use 0 (called the *zero element*) to denote the identity element of the abelian group R with respect to addition, and the additive inverse of a is denoted by $-a$; also, $a + (-b)$ is abbreviated by $a - b$. Instead of $a \cdot b$ we will usually write ab. As a consequence of the definition of a ring one obtains the general property $a0 = 0a = 0$ for all $a \in R$. This, in turn, implies $(-a)b = a(-b) = -ab$ for all $a, b \in R$.

The most natural example of a ring is perhaps the ring of ordinary integers. If we examine the properties of this ring, we realize that it has properties not enjoyed by rings in general. Thus, rings can be further classified according to the following definitions.

1.29. Definition

(i) A ring is called a *ring with identity* if the ring has a multiplicative identity—that is, if there is an element e such that $ae = ea = a$ for all $a \in R$.

(ii) A ring is called *commutative* if $\cdot$ is commutative.

(iii) A ring is called an *integral domain* if it is a commutative ring with identity $e \neq 0$ in which $ab = 0$ implies $a = 0$ or $b = 0$.
(iv) A ring is called a *division ring* (or *skew field*) if the nonzero elements of R form a group under $\cdot$.
(v) A commutative division ring is called a *field*.

Since our study is devoted to fields, we emphasize again the definition of this concept. In the first place, a *field* is a set F on which two binary operations, called addition and multiplication, are defined and which contains two distinguished elements 0 and e with $0 \neq e$. Furthermore, F is an abelian group with respect to addition having 0 as the identity element, and the elements of F that are $\neq 0$ form an abelian group with respect to multiplication having e as the identity element. The two operations of addition and multiplication are linked by the distributive law $a(b + c) = ab + ac$. The second distributive law $(b + c)a = ba + ca$ follows automatically from the commutativity of multiplication. The element 0 is called the *zero element* and e is called the *multiplicative identity element* or simply the *identity*. Later on, the identity will usually be denoted by 1.

The property appearing in Definition 1.29(iii)—namely, that $ab = 0$ implies $a = 0$ or $b = 0$—is expressed by saying that there are *no zero divisors*. In particular, a field has no zero divisors, for if $ab = 0$ and $a \neq 0$, then multiplication by a^{-1} yields $b = a^{-1}0 = 0$.

In order to give an indication of the generality of the concept of ring, we present some examples.

1.30. Examples

(i) Let R be any abelian group with group operation $+$. Define $ab = 0$ for all $a, b \in R$; then R is a ring.
(ii) The integers form an integral domain, but not a field.
(iii) The even integers form a commutative ring without identity.
(iv) The functions from the real numbers into the real numbers form a commutative ring with identity under the definitions for $f + g$ and fg given by $(f + g)(x) = f(x) + g(x)$ and $(fg)(x) = f(x)g(x)$ for $x \in \mathbb{R}$.
(v) The set of all 2×2 matrices with real numbers as entries forms a noncommutative ring with identity with respect to matrix addition and multiplication. □

We have seen above that a field is, in particular, an integral domain. The converse is not true in general (see Example 1.30(ii)), but it will hold if the structures contain only finitely many elements.

1.31. Theorem. *Every finite integral domain is a field.*

Proof. Let the elements of the finite integral domain R be $a_1, a_2, \ldots, a_n$. For a fixed nonzero element $a \in R$, consider the products $aa_1, aa_2, \ldots, aa_n$. These are distinct, for if $aa_i = aa_j$, then $a(a_i - a_j) = 0$, and

since $a \neq 0$ we must have $a_i - a_j = 0$, or $a_i = a_j$. Thus each element of R is of the form aa_i, in particular, $e = aa_i$ for some i with $1 \leqslant i \leqslant n$, where e is the identity of R. Since R is commutative, we have also $a_i a = e$, and so a_i is the multiplicative inverse of a. Thus the nonzero elements of R form a commutative group, and R is a field. □

1.32. Definition. A subset S of a ring R is called a *subring* of R provided S is closed under $+$ and $\cdot$ and forms a ring under these operations.

1.33. Definition. A subset J of a ring R is called an *ideal* provided J is a subring of R and for all $a \in J$ and $r \in R$ we have $ar \in J$ and $ra \in J$.

1.34. Examples

(i) Let R be the field $\mathbb{Q}$ of rational numbers. Then the set $\mathbb{Z}$ of integers is a subring of $\mathbb{Q}$, but not an ideal since, for example, $1 \in \mathbb{Z}$, $\frac{1}{2} \in \mathbb{Q}$, but $\frac{1}{2} \cdot 1 = \frac{1}{2} \notin \mathbb{Z}$.

(ii) Let R be a commutative ring, $a \in R$, and let $J = \{ra : r \in R\}$, then J is an ideal.

(iii) Let R be a commutative ring. Then the smallest ideal containing a given element $a \in R$ is the ideal $(a) = \{ra + na : r \in R, n \in \mathbb{Z}\}$. If R contains an identity, then $(a) = \{ra : r \in R\}$. □

1.35. Definition. Let R be a commutative ring. An ideal J of R is said to be *principal* if there is an $a \in R$ such that $J = (a)$. In this case, J is also called the principal ideal *generated by a*.

Since ideals are normal subgroups of the additive group of a ring, it follows immediately that an ideal J of the ring R defines a partition of R into disjoint cosets, called *residue classes* modulo J. The residue class of the element a of R modulo J will be denoted by $[a] = a + J$, since it consists of all elements of R that are of the form $a + c$ for some $c \in J$. Elements $a, b \in R$ are called *congruent* modulo J, written $a \equiv b \bmod J$, if they are in the same residue class modulo J, or equivalently, if $a - b \in J$ (compare with Definition 1.4). One can verify that $a \equiv b \bmod J$ implies $a + r \equiv b + r \bmod J$, $ar \equiv br \bmod J$, and $ra \equiv rb \bmod J$ for any $r \in R$ and $na \equiv nb \bmod J$ for any $n \in \mathbb{Z}$. If, in addition, $r \equiv s \bmod J$, then $a + r \equiv b + s \bmod J$ and $ar \equiv bs \bmod J$.

It is shown by a straightforward argument that the set of residue classes of a ring R modulo an ideal J forms a ring with respect to the operations

$$(a + J) + (b + J) = (a + b) + J, \tag{1.2}$$

$$(a + J)(b + J) = ab + J. \tag{1.3}$$

1.36. Definition. The ring of residue classes of the ring R modulo the ideal J under the operations (1.2) and (1.3) is called the *residue class ring* (or *factor ring*) of R modulo J and is denoted by R/J.

1.37. Example (The residue class ring $\mathbb{Z}/(n)$). As in the case of groups (compare with Definition 1.5), we denote the coset or residue class of the integer a modulo the positive integer n by $[a]$, as well as by $a+(n)$, where (n) is the principal ideal generated by n. The elements of $\mathbb{Z}/(n)$ are

$$[0]=0+(n),\ [1]=1+(n),\ldots,[n-1]=n-1+(n). \qquad \square$$

1.38. Theorem. *$\mathbb{Z}/(p)$, the ring of residue classes of the integers modulo the principal ideal generated by a prime p, is a field.*

Proof. By Theorem 1.31 it suffices to show that $\mathbb{Z}/(p)$ is an integral domain. Now $[1]$ is an identity of $\mathbb{Z}/(p)$, and $[a][b]=[ab]=[0]$ if and only if $ab=kp$ for some integer k. But since p is prime, p divides ab if and only if p divides at least one of the factors. Therefore, either $[a]=[0]$ or $[b]=[0]$, so that $\mathbb{Z}/(p)$ contains no zero divisors. $\square$

1.39. Example. Let $p=3$. Then $\mathbb{Z}/(p)$ consists of the elements [0], [1], and [2]. The operations in this field can be described by operation tables that are similar to Cayley tables for finite groups (see Example 1.7):

+	[0]	[1]	[2]
[0]	[0]	[1]	[2]
[1]	[1]	[2]	[0]
[2]	[2]	[0]	[1]

·	[0]	[1]	[2]
[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]
[2]	[0]	[2]	[1]

$\square$

The residue class fields $\mathbb{Z}/(p)$ are our first examples of *finite fields* —that is, of fields that contain only finitely many elements. The general theory of such fields will be developed later on.

The reader is cautioned not to assume that in the formation of residue class rings all the properties of the original ring will be preserved in all cases. For example, the lack of zero divisors is not always preserved, as may be seen by considering the ring $\mathbb{Z}/(n)$, where n is a composite integer.

There is an obvious extension from groups to rings of the definition of a homomorphism. A mapping $\varphi: R \to S$ from a ring R into a ring S is called a *homomorphism* if for any $a, b \in R$ we have

$$\varphi(a+b)=\varphi(a)+\varphi(b) \quad \text{and} \quad \varphi(ab)=\varphi(a)\varphi(b).$$

Thus a homomorphism $\varphi: R \to S$ preserves both operations + and · of R and induces a homomorphism of the additive group of R into the additive group of S. The set

$$\ker\varphi=\{a \in R: \varphi(a)=0 \in S\}$$

is called the *kernel* of φ. Other concepts, such as that of an *isomorphism*, are analogous to those in Definition 1.16. The homomorphism theorem for rings, similar to Theorem 1.23 for groups, runs as follows.

1.40. Theorem (Homomorphism Theorem for Rings). *If φ is a homomorphism of a ring R onto a ring S, then* $\ker\varphi$ *is an ideal of R and S is*

isomorphic to the factor ring $R/\ker\varphi$. Conversely, if J is an ideal of the ring R, then the mapping $\psi: R \to R/J$ defined by $\psi(a) = a + J$ for $a \in R$ is a homomorphism of R onto R/J with kernel J.

Mappings can be used to transfer a structure from an algebraic system to a set without structure. For instance, let R be a ring and let φ be a one-to-one and onto mapping from R to a set S; then by means of φ one can define a ring structure on S that converts φ into an isomorphism. In detail, let s_1 and s_2 be two elements of S and let r_1 and r_2 be the elements of R uniquely determined by $\varphi(r_1) = s_1$ and $\varphi(r_2) = s_2$. Then one defines $s_1 + s_2$ to be $\varphi(r_1 + r_2)$ and $s_1 s_2$ to be $\varphi(r_1 r_2)$, and all the desired properties are satisfied. This structure on S may be called the ring structure *induced by* φ. In case R has additional properties, such as being an integral domain or a field, then these properties are inherited by S. We use this principle in order to arrive at a more convenient representation for the finite fields $\mathbb{Z}/(p)$.

1.41. Definition. For a prime p, let $\mathbb{F}_p$ be the set $\{0, 1, \ldots, p-1\}$ of integers and let $\varphi: \mathbb{Z}/(p) \to \mathbb{F}_p$ be the mapping defined by $\varphi([a]) = a$ for $a = 0, 1, \ldots, p-1$. Then $\mathbb{F}_p$, endowed with the field structure induced by φ, is a finite field, called the *Galois field of order p*.

By what we have said before, the mapping $\varphi: \mathbb{Z}/(p) \to \mathbb{F}_p$ is then an isomorphism, so that $\varphi([a] + [b]) = \varphi([a]) + \varphi([b])$ and $\varphi([a][b]) = \varphi([a])\varphi([b])$. The finite field $\mathbb{F}_p$ has zero element 0, identity 1, and its structure is exactly the structure of $\mathbb{Z}/(p)$. Computing with elements of $\mathbb{F}_p$ therefore means ordinary arithmetic of integers with reduction modulo p.

1.42. Examples

(i) Consider $\mathbb{Z}/(5)$, isomorphic to $\mathbb{F}_5 = \{0, 1, 2, 3, 4\}$, with the isomorphism given by: $[0] \to 0$, $[1] \to 1$, $[2] \to 2$, $[3] \to 3$, $[4] \to 4$. The tables for the two operations $+$ and $\cdot$ for elements in $\mathbb{F}_5$ are as follows:

+	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

·	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

(ii) An even simpler and more important example is the finite field $\mathbb{F}_2$. The elements of this field of order two are 0 and 1, and the operation tables have the following form:

+	0	1
0	0	1
1	1	0

·	0	1
0	0	0
1	0	1

In this context, the elements 0 and 1 are called *binary elements*. □

If b is any nonzero element of the ring $\mathbb{Z}$ of integers, then the additive order of b is infinite; that is, $nb = 0$ implies $n = 0$. However, in the ring $\mathbb{Z}/(p)$, p prime, the additive order of every nonzero element b is p; that is, $pb = 0$, and p is the least positive integer for which this holds. It is of interest to formalize this property.

1.43. Definition. If R is an arbitrary ring and there exists a positive integer n such that $nr = 0$ for every $r \in R$, then the least such positive integer n is called the *characteristic* of R and R is said to have (positive) characteristic n. If no such positive integer n exists, R is said to have characteristic 0.

1.44. Theorem. *A ring $R \neq \{0\}$ of positive characteristic having an identity and no zero divisors must have prime characteristic.*

Proof. Since R contains nonzero elements, R has characteristic $n \geqslant 2$. If n were not prime, we could write $n = km$ with $k, m \in \mathbb{Z}$, $1 < k, m < n$. Then $0 = ne = (km)e = (ke)(me)$, and this implies that either $ke = 0$ or $me = 0$ since R has no zero divisors. It follows that either $kr = (ke)r = 0$ for all $r \in R$ or $mr = (me)r = 0$ for all $r \in R$, in contradiction to the definition of the characteristic n. □

1.45. Corollary. *A finite field has prime characteristic.*

Proof. By Theorem 1.44 it suffices to show that a finite field F has a positive characteristic. Consider the multiples $e, 2e, 3e, \ldots$ of the identity. Since F contains only finitely many distinct elements, there exist integers k and m with $1 \leqslant k < m$ such that $ke = me$, or $(m-k)e = 0$, and so F has a positive characteristic. □

The finite field $\mathbb{Z}/(p)$ (or, equivalently, $\mathbb{F}_p$) obviously has characteristic p, whereas the ring $\mathbb{Z}$ of integers and the field $\mathbb{Q}$ of rational numbers have characteristic 0. We note that in a ring R of characteristic 2 we have $2a = a + a = 0$, hence $a = -a$ for all $a \in R$. A useful property of commutative rings of prime characteristic is the following.

1.46. Theorem. *Let R be a commutative ring of prime characteristic p. Then*

$$(a+b)^{p^n} = a^{p^n} + b^{p^n} \quad \text{and} \quad (a-b)^{p^n} = a^{p^n} - b^{p^n}$$

for $a, b \in R$ and $n \in \mathbb{N}$.

Proof. We use the fact that

$$\binom{p}{i} = \frac{p(p-1)\cdots(p-i+1)}{1 \cdot 2 \cdot \cdots \cdot i} \equiv 0 \bmod p$$

for all $i \in \mathbb{Z}$ with $0 < i < p$, which follows from $\binom{p}{i}$ being an integer and the observation that the factor p in the numerator cannot be cancelled. Then by

the binomial theorem (see Exercise 1.8),

$$(a+b)^p = a^p + \binom{p}{1} a^{p-1}b + \cdots + \binom{p}{p-1} ab^{p-1} + b^p = a^p + b^p,$$

and induction on n completes the proof of the first identity. By what we have shown, we get

$$a^{p^n} = ((a-b)+b)^{p^n} = (a-b)^{p^n} + b^{p^n},$$

and the second identity follows. □

Next we will show for the case of commutative rings with identity which ideals give rise to factor rings that are integral domains or fields. For this we need some definitions from ring theory.

Let R be a commutative ring with identity. An element $a \in R$ is called a *divisor* of $b \in R$ if there exists $c \in R$ such that $ac = b$. A *unit* of R is a divisor of the identity; two elements $a, b \in R$ are said to be *associates* if there is a unit ε of R such that $a = b\varepsilon$. An element $c \in R$ is called a *prime element* if it is no unit and if it has only the units of R and the associates of c as divisors. An ideal $P \neq R$ of the ring R is called a *prime ideal* if for $a, b \in R$ we have $ab \in P$ only if either $a \in P$ or $b \in P$. An ideal $M \neq R$ of R is called a *maximal ideal* of R if for any ideal J of R the property $M \subseteq J$ implies $J = R$ or $J = M$. Furthermore, R is said to be a *principal ideal domain* if R is an integral domain and if every ideal J of R is principal—that is, if there is a generating element a for J such that $J = (a) = \{ra : r \in R\}$.

1.47. Theorem. *Let R be a commutative ring with identity. Then:*

(i) *An ideal M of R is a maximal ideal if and only if R/M is a field.*
(ii) *An ideal P of R is a prime ideal if and only if R/P is an integral domain.*
(iii) *Every maximal ideal of R is a prime ideal.*
(iv) *If R is a principal ideal domain, then $R/(c)$ is a field if and only if c is a prime element of R.*

Proof.

(i) Let M be a maximal ideal of R. Then for $a \notin M$, $a \in R$, the set $J = \{ar + m : r \in R, m \in M\}$ is an ideal of R properly containing M, and therefore $J = R$. In particular, $ar + m = 1$ for some suitable $r \in R$, $m \in M$, where 1 denotes the multiplicative identity element of R. In other words, if $a + M \neq 0 + M$ is an element of R/M different from the zero element in R/M, then it possesses a multiplicative inverse, because $(a + M)(r + M) = ar + M = (1 - m) + M = 1 + M$. Therefore, R/M is a field. Conversely, let R/M be a field and let $J \supseteq M$, $J \neq M$, be an ideal of R. Then for $a \in J$, $a \notin M$, the residue class $a + M$ has a multi-

plicative inverse, so that $(a + M)(r + M) = 1 + M$ for some $r \in R$. This implies $ar + m = 1$ for some $m \in M$. Since J is an ideal, we have $1 \in J$ and therefore $(1) = R \subseteq J$, hence $J = R$. Thus M is a maximal ideal of R.

(ii) Let P be a prime ideal of R; then R/P is a commutative ring with identity $1 + P \neq 0 + P$. Let $(a + P)(b + P) = 0 + P$, hence $ab \in P$. Since P is a prime ideal, either $a \in P$ or $b \in P$; that is, either $a + P = 0 + P$ or $b + P = 0 + P$. Thus, R/P has no zero divisors and is therefore an integral domain. The converse follows immediately by reversing the steps of this proof.

(iii) This follows from (i) and (ii) since every field is an integral domain.

(iv) Let $c \in R$. If c is a unit, then $(c) = R$ and the ring $R/(c)$ consists only of one element and is no field. If c is neither a unit nor a prime element, then c has a divisor $a \in R$ that is neither a unit nor an associate of c. We note that $a \neq 0$, for if $a = 0$, then $c = 0$ and a would be an associate of c. We can write $c = ab$ with $b \in R$. Next we claim that $a \notin (c)$. For otherwise $a = cd = abd$ for some $d \in R$, or $a(1 - bd) = 0$. Since $a \neq 0$, this would imply $bd = 1$, so that d would be a unit, which contradicts the fact that a is not an associate of c. It follows that $(c) \subseteq (a) \subseteq R$, where all containments are proper, and so $R/(c)$ cannot be a field because of (i). Finally, we are left with the case where c is a prime element. Then $(c) \neq R$ since c is no unit. Furthermore, if $J \supseteq (c)$ is an ideal of R, then $J = (a)$ for some $a \in R$ since R is a principal ideal domain. It follows that $c \in (a)$, and so a is a divisor of c. Consequently, a is either a unit or an associate of c, so that either $J = R$ or $J = (c)$. This shows that (c) is a maximal ideal of R. Hence $R/(c)$ is a field by (i). □

As an application of this theorem, let us consider the case $R = \mathbb{Z}$. We note that $\mathbb{Z}$ is a principal ideal domain since the additive subgroups of $\mathbb{Z}$ are already generated by a single element because of Theorem 1.15(i). A prime number p fits the definition of a prime element, and so Theorem 1.47(iv) yields another proof of the known result that $\mathbb{Z}/(p)$ is a field. Consequently, (p) is a maximal ideal and a prime ideal of $\mathbb{Z}$. For a composite integer n, the ideal (n) is not a prime ideal of $\mathbb{Z}$, and so $\mathbb{Z}/(n)$ is not even an integral domain. Other applications will follow in the next section when we consider residue class rings of polynomial rings over fields.

3. POLYNOMIALS

In elementary algebra one regards a polynomial as an expression of the form $a_0 + a_1x + \cdots + a_nx^n$. The a_i's are called coefficients and are usually

real or complex numbers; x is viewed as a variable: that is, substituting an arbitrary number α for x, a well-defined number $a_0 + a_1\alpha + \cdots + a_n\alpha^n$ is obtained. The arithmetic of polynomials is governed by familiar rules. The concept of polynomial and the associated operations can be generalized to a formal algebraic setting in a straightforward manner.

Let R be an arbitrary ring. A *polynomial* over R is an expression of the form

$$f(x) = \sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + \cdots + a_n x^n,$$

where n is a nonnegative integer, the *coefficients* a_i, $0 \leqslant i \leqslant n$, are elements of R, and x is a symbol not belonging to R, called an *indeterminate* over R. Whenever it is clear which indeterminate is meant, we can use f as a designation for the polynomial $f(x)$. We adopt the convention that a term $a_i x^i$ with $a_i = 0$ need not be written down. In particular, the polynomial $f(x)$ above may then also be given in the equivalent form $f(x) = a_0 + a_1 x + \cdots + a_n x^n + 0x^{n+1} + \cdots + 0x^{n+h}$, where h is any positive integer. When comparing two polynomials $f(x)$ and $g(x)$ over R, it is therefore possible to assume that they both involve the same powers of x. The polynomials

$$f(x) = \sum_{i=0}^{n} a_i x^i \quad \text{and} \quad g(x) = \sum_{i=0}^{n} b_i x^i$$

over R are considered equal if and only if $a_i = b_i$ for $0 \leqslant i \leqslant n$. We define the *sum* of $f(x)$ and $g(x)$ by

$$f(x) + g(x) = \sum_{i=0}^{n} (a_i + b_i) x^i.$$

To define the *product* of two polynomials over R, let

$$f(x) = \sum_{i=0}^{n} a_i x^i \quad \text{and} \quad g(x) = \sum_{j=0}^{m} b_j x^j$$

and set

$$f(x)g(x) = \sum_{k=0}^{n+m} c_k x^k, \quad \text{where } c_k = \sum_{\substack{i+j=k \\ 0 \leqslant i \leqslant n,\, 0 \leqslant j \leqslant m}} a_i b_j.$$

It is easily seen that with these operations the set of polynomials over R forms a ring.

1.48. Definition. The ring formed by the polynomials over R with the above operations is called the *polynomial ring* over R and denoted by $R[x]$.

The zero element of $R[x]$ is the polynomial all of whose coefficients are 0. This polynomial is called the *zero polynomial* and denoted by 0. It should always be clear from the context whether 0 stands for the zero element of R or the zero polynomial.

1.49. Definition. Let $f(x) = \sum_{i=0}^{n} a_i x^i$ be a polynomial over R that is not the zero polynomial, so that we can suppose $a_n \neq 0$. Then a_n is called the *leading coefficient* of $f(x)$ and a_0 the *constant term*, while n is called the *degree* of $f(x)$, in symbols $n = \deg(f(x)) = \deg(f)$. By convention, we set $\deg(0) = -\infty$. Polynomials of degree $\leqslant 0$ are called *constant polynomials*. If R has the identity 1 and if the leading coefficient of $f(x)$ is 1, then $f(x)$ is called a *monic polynomial*.

By computing the leading coefficient of the sum and the product of two polynomials, one finds the following result.

1.50. Theorem. *Let $f, g \in R[x]$. Then*

$$\deg(f+g) \leqslant \max(\deg(f), \deg(g)),$$
$$\deg(fg) \leqslant \deg(f) + \deg(g).$$

If R is an integral domain, we have

$$\deg(fg) = \deg(f) + \deg(g). \tag{1.4}$$

If one identifies constant polynomials with elements of R, then R can be viewed as a subring of $R[x]$. Certain properties of R are inherited by $R[x]$. The essential step in the proof of part (iii) of the subsequent theorem depends on (1.4).

1.51. Theorem. *Let R be a ring. Then*:

(i) *$R[x]$ is commutative if and only if R is commutative.*
(ii) *$R[x]$ is a ring with identity if and only if R has an identity.*
(iii) *$R[x]$ is an integral domain if and only if R is an integral domain.*

In the following chapters we will deal almost exclusively with polynomials over fields. Let F denote a field (not necessarily finite). The concept of divisibility, when specialized to the ring $F[x]$, leads to the following. The polynomial $g \in F[x]$ *divides* the polynomial $f \in F[x]$ if there exists a polynomial $h \in F[x]$ such that $f = gh$. We also say that g is a *divisor* of f, or that f is a *multiple* of g, or that f is *divisible* by g. The units of $F[x]$ are the divisors of the constant polynomial 1, which are precisely all nonzero constant polynomials.

As for the ring of integers, there is a division with remainder in polynomial rings over fields.

1.52. Theorem (Division Algorithm). *Let $g \neq 0$ be a polynomial in $F[x]$. Then for any $f \in F[x]$ there exist polynomials $q, r \in F[x]$ such that*

$$f = qg + r, \quad \textit{where } \deg(r) < \deg(g).$$

1.53. Example. Consider $f(x) = 2x^5 + x^4 + 4x + 3 \in \mathbb{F}_5[x]$, $g(x) = 3x^2 + 1 \in \mathbb{F}_5[x]$. We compute the polynomials $q, r \in \mathbb{F}_5[x]$ with $f = qg + r$ by using

long division:

$$
\begin{array}{r}
4x^3+2x^2+2x+1 \\
3x^2+1\,\overline{\big|\;2x^5+x^4\qquad\qquad\qquad+4x+3} \\
\underline{-2x^5\quad -4x^3\qquad\qquad\qquad} \\
x^4+x^3 \\
\underline{-x^4\qquad -2x^2\qquad\qquad} \\
x^3+3x^2+4x \\
\underline{-x^3\qquad -2x\qquad} \\
3x^2+2x+3 \\
\underline{-3x^2\qquad -1} \\
2x+2
\end{array}
$$

Thus $q(x)=4x^3+2x^2+2x+1$, $r(x)=2x+2$, and obviously $\deg(r)<\deg(g)$. □

The fact that $F[x]$ permits a division algorithm implies by a standard argument that every ideal of $F[x]$ is principal.

1.54. Theorem. *$F[x]$ is a principal ideal domain. In fact, for every ideal $J\neq(0)$ of $F[x]$ there exists a uniquely determined monic polynomial $g\in F[x]$ with $J=(g)$.*

Proof. $F[x]$ is an integral domain by Theorem 1.51(iii). Suppose $J\neq(0)$ is an ideal of $F[x]$. Let $h(x)$ be a nonzero polynomial of least degree contained in J, let b be the leading coefficient of $h(x)$, and set $g(x)=b^{-1}h(x)$. Then $g\in J$ and g is monic. If $f\in J$ is arbitrary, the division algorithm yields $q,r\in F[x]$ with $f=qg+r$ and $\deg(r)<\deg(g)=\deg(h)$. Since J is an ideal, we get $f-qg=r\in J$, and by the definition of h we must have $r=0$. Therefore, f is a multiple of g, and so $J=(g)$. If $g_1\in F[x]$ is another monic polynomial with $J=(g_1)$, then $g=c_1g_1$ and $g_1=c_2g$ with $c_1,c_2\in F[x]$. This implies $g=c_1c_2g$, hence $c_1c_2=1$, and c_1 and c_2 are constant polynomials. Since both g and g_1 are monic, it follows that $g=g_1$, and the uniqueness of g is established. □

1.55. Theorem. *Let $f_1,\ldots,f_n$ be polynomials in $F[x]$ not all of which are 0. Then there exists a uniquely determined monic polynomial $d\in F[x]$ with the following properties: (i) d divides each f_j, $1\leqslant j\leqslant n$; (ii) any polynomial $c\in F[x]$ dividing each f_j, $1\leqslant j\leqslant n$, divides d. Moreover, d can be expressed in the form*

$$d=b_1f_1+\cdots+b_nf_n \quad \text{with } b_1,\ldots,b_n\in F[x]. \tag{1.5}$$

Proof. The set J consisting of all polynomials of the form $c_1f_1+\cdots+c_nf_n$ with $c_1,\ldots,c_n\in F[x]$ is easily seen to be an ideal of $F[x]$. Since not all f_j are 0, we have $J\neq(0)$, and Theorem 1.54 implies that $J=(d)$

for some monic polynomial $d \in F[x]$. Property (i) and the representation (1.5) follow immediately from the construction of d. Property (ii) follows from (1.5). If d_1 is another monic polynomial in $F[x]$ satisfying (i) and (ii), then these properties imply that d and d_1 are divisible by each other, and so $(d) = (d_1)$. An application of the uniqueness part of Theorem 1.54 yields $d = d_1$. □

The monic polynomial d appearing in the theorem above is called the *greatest common divisor* of $f_1, \ldots, f_n$, in symbols $d = \gcd(f_1, \ldots, f_n)$. If $\gcd(f_1, \ldots, f_n) = 1$, then the polynomials $f_1, \ldots, f_n$ are said to be *relatively prime*. They are called *pairwise relatively prime* if $\gcd(f_i, f_j) = 1$ for $1 \leqslant i < j \leqslant n$.

The greatest common divisor of two polynomials $f, g \in F[x]$ can be computed by the *Euclidean algorithm*. Suppose, without loss of generality, that $g \neq 0$ and that g does not divide f. Then we repeatedly use the division algorithm in the following manner:

$$\begin{aligned}
f &= q_1 g + r_1 && 0 \leqslant \deg(r_1) < \deg(g) \\
g &= q_2 r_1 + r_2 && 0 \leqslant \deg(r_2) < \deg(r_1) \\
r_1 &= q_3 r_2 + r_3 && 0 \leqslant \deg(r_3) < \deg(r_2) \\
&\vdots && \vdots \\
r_{s-2} &= q_s r_{s-1} + r_s && 0 \leqslant \deg(r_s) < \deg(r_{s-1}) \\
r_{s-1} &= q_{s+1} r_s.
\end{aligned}$$

Here $q_1, \ldots, q_{s+1}$ and $r_1, \ldots, r_s$ are polynomials in $F[x]$. Since $\deg(g)$ is finite, the procedure must stop after finitely many steps. If the last nonzero remainder r_s has leading coefficient b, then $\gcd(f, g) = b^{-1} r_s$. In order to find $\gcd(f_1, \ldots, f_n)$ for $n > 2$ and nonzero polynomials f_i, one first computes $\gcd(f_1, f_2)$, then $\gcd(\gcd(f_1, f_2), f_3)$, and so on, by the Euclidean algorithm.

1.56. Example. The Euclidean algorithm applied to

$$f(x) = 2x^6 + x^3 + x^2 + 2 \in \mathbb{F}_3[x], \qquad g(x) = x^4 + x^2 + 2x \in \mathbb{F}_3[x]$$

yields:

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

Therefore $\gcd(f, g) = 1$ and f and g are relatively prime. □

A counterpart to the notion of greatest common divisor is that of least common multiple. Let $f_1, \ldots, f_n$ be nonzero polynomials in $F[x]$. Then one shows (see Exercise 1.25) that there exists a uniquely determined monic

polynomial $m \in F[x]$ with the following properties: (i) m is a multiple of each f_j, $1 \leqslant j \leqslant n$; (ii) any polynomial $b \in F[x]$ that is a multiple of each f_j, $1 \leqslant j \leqslant n$, is a multiple of m. The polynomial m is called the *least common multiple* of $f_1,\ldots,f_n$ and denoted by $m = \operatorname{lcm}(f_1,\ldots,f_n)$. For two nonzero polynomials $f, g \in F[x]$ we have

$$a^{-1}fg = \operatorname{lcm}(f, g)\gcd(f, g), \tag{1.6}$$

where a is the leading coefficient of fg. This relation conveniently reduces the calculation of $\operatorname{lcm}(f, g)$ to that of $\gcd(f, g)$. There is no direct analog of (1.6) for three or more polynomials. In this case, one uses the identity $\operatorname{lcm}(f_1,\ldots,f_n) = \operatorname{lcm}(\operatorname{lcm}(f_1,\ldots,f_{n-1}), f_n)$ to compute the least common multiple.

The prime elements of the ring $F[x]$ are usually called irreducible polynomials. To emphasize this important concept, we give the definition again for the present context.

1.57. Definition. A polynomial $p \in F[x]$ is said to be *irreducible over* F (or *irreducible in* $F[x]$, or *prime in* $F[x]$) if p has positive degree and $p = bc$ with $b, c \in F[x]$ implies that either b or c is a constant polynomial.

Briefly stated, a polynomial of positive degree is irreducible over F if it allows only trivial factorizations. A polynomial in $F[x]$ of positive degree that is not irreducible over F is called *reducible over* F. The reducibility or irreducibility of a given polynomial depends heavily on the field under consideration. For instance, the polynomial $x^2 - 2 \in \mathbf{Q}[x]$ is irreducible over the field $\mathbf{Q}$ of rational numbers, but $x^2 - 2 = (x + \sqrt{2})(x - \sqrt{2})$ is reducible over the field of real numbers.

Irreducible polynomials are of fundamental importance for the structure of the ring $F[x]$ since the polynomials in $F[x]$ can be written as products of irreducible polynomials in an essentially unique manner. For the proof we need the following result.

1.58. Lemma. *If an irreducible polynomial p in $F[x]$ divides a product $f_1 \cdots f_m$ of polynomials in $F[x]$, then at least one of the factors f_j is divisible by p.*

Proof. Since p divides $f_1 \cdots f_m$, we get the identity $(f_1 + (p)) \cdots (f_m + (p)) = 0 + (p)$ in the factor ring $F[x]/(p)$. Now $F[x]/(p)$ is a field by Theorem 1.47(iv), and so $f_j + (p) = 0 + (p)$ for some j; that is, p divides f_j. □

1.59. Theorem (Unique Factorization in $F[x]$). *Any polynomial $f \in F[x]$ of positive degree can be written in the form*

$$f = ap_1^{e_1} \cdots p_k^{e_k}, \tag{1.7}$$

where $a \in F$, $p_1,\ldots,p_k$ are distinct monic irreducible polynomials in $F[x]$, and $e_1,\ldots,e_k$ are positive integers. Moreover, this factorization is unique apart from the order in which the factors occur.

Proof. The fact that any nonconstant $f \in F[x]$ can be represented in the form (1.7) is shown by induction on the degree of f. The case $\deg(f) = 1$ is trivial since any polynomial in $F[x]$ of degree 1 is irreducible over F. Now suppose the desired factorization is established for all nonconstant polynomials in $F[x]$ of degree $< n$. If $\deg(f) = n$ and f is irreducible over F, then we are done since we can write $f = a(a^{-1}f)$, where a is the leading coefficient of f and $a^{-1}f$ is a monic irreducible polynomial in $F[x]$. Otherwise, f allows a factorization $f = gh$ with $1 \leqslant \deg(g) < n$, $1 \leqslant \deg(h) < n$, and $g, h \in F[x]$. By the induction hypothesis, g and h can be factored in the form (1.7), and so f can be factored in this form.

To prove uniqueness, suppose f has two factorizations of the form (1.7), say

$$f = ap_1^{e_1} \cdots p_k^{e_k} = bq_1^{d_1} \cdots q_r^{d_r}. \tag{1.8}$$

By comparing leading coefficients, we get $a = b$. Furthermore, the irreducible polynomial p_1 in $F[x]$ divides the right-hand side of (1.8), and so Lemma 1.58 shows that p_1 divides q_j for some $j, 1 \leqslant j \leqslant r$. But q_j is also irreducible in $F[x]$, so that we must have $q_j = cp_1$ with a constant polynomial c. Since q_j and p_1 are both monic, it follows that $q_j = p_1$. Thus we can cancel p_1 against q_j in (1.8) and continue in the same manner with the remaining identity. After finitely many steps of this type, we obtain that the two factorizations are identical apart from the order of the factors. □

We shall refer to (1.7) as the *canonical factorization* of the polynomial f in $F[x]$. If $F = \mathbb{Q}$, there is a method due to Kronecker for finding the canonical factorization of a polynomial in finitely many steps. This method is briefly described in Exercise 1.30. For polynomials over finite fields, factorization algorithms will be discussed in Chapter 4.

A central question about polynomials in $F[x]$ is to decide whether a given polynomial is irreducible or reducible over F. For our purposes, irreducible polynomials over $\mathbb{F}_p$ are of particular interest. To determine all monic irreducible polynomials over $\mathbb{F}_p$ of fixed degree n, one may first compute all monic reducible polynomials over $\mathbb{F}_p$ of degree n and then eliminate them from the set of monic polynomials in $\mathbb{F}_p[x]$ of degree n. If p or n is large, this method is not feasible, and we will develop more powerful methods in Chapter 3, Sections 2 and 3.

1.60. Example. Find all irreducible polynomials over $\mathbb{F}_2$ of degree 4 (note that a nonzero polynomial in $\mathbb{F}_2[x]$ is automatically monic). There are $2^4 = 16$ polynomials in $\mathbb{F}_2[x]$ of degree 4. Such a polynomial is reducible over $\mathbb{F}_2$ if and only if it has a divisor of degree 1 or 2. Therefore, we compute all products $(a_0 + a_1x + a_2x^2 + x^3)(b_0 + x)$ and $(a_0 + a_1x + x^2)(b_0 + b_1x + x^2)$ with $a_i, b_j \in \mathbb{F}_2$ and obtain all reducible polynomials over $\mathbb{F}_2$ of degree 4. Comparison with the 16 polynomials of degree 4 leaves

us with the irreducible polynomials $f_1(x) = x^4 + x + 1$, $f_2(x) = x^4 + x^3 + 1$, $f_3(x) = x^4 + x^3 + x^2 + x + 1$ in $\mathbb{F}_2[x]$. □

Since the irreducible polynomials over a field F are exactly the prime elements of $F[x]$, the following result, one part of which was already used in Lemma 1.58, is an immediate consequence of Theorems 1.47(iv) and 1.54.

1.61. Theorem. *For $f \in F[x]$, the residue class ring $F[x]/(f)$ is a field if and only if f is irreducible over F.*

As a preparation for the next section, we shall take a closer look at the structure of the residue class ring $F[x]/(f)$, where f is an arbitrary nonzero polynomial in $F[x]$. We recall that as a residue class ring $F[x]/(f)$ consists of residue classes $g + (f)$ (also denoted by $[g]$) with $g \in F[x]$, where the operations are defined as in (1.2) and (1.3). Two residue classes $g + (f)$ and $h + (f)$ are identical precisely if $g \equiv h \bmod f$—that is, precisely if $g - h$ is divisible by f. This is equivalent to the requirement that g and h leave the same remainder after division by f. Each residue class $g + (f)$ contains a unique representative $r \in F[x]$ with $\deg(r) < \deg(f)$, which is simply the remainder in the division of g by f. The process of passing from g to r is called *reduction* mod f. The uniqueness of r follows from the observation that if $r_1 \in g + (f)$ with $\deg(r_1) < \deg(f)$, then $r - r_1$ is divisible by f and $\deg(r - r_1) < \deg(f)$, which is only possible if $r = r_1$. The distinct residue classes comprising $F[x]/(f)$ can now be described explicitly; namely, they are exactly the residue classes $r + (f)$, where r runs through all polynomials in $F[x]$ with $\deg(r) < \deg(f)$. Thus, if $F = \mathbb{F}_p$ and $\deg(f) = n \geqslant 0$, then the number of elements of $\mathbb{F}_p[x]/(f)$ is equal to the number of polynomials in $\mathbb{F}_p[x]$ of degree $< n$, which is p^n.

1.62. Examples

(i) Let $f(x) = x \in \mathbb{F}_2[x]$. The $p^n = 2^1$ polynomials in $\mathbb{F}_2[x]$ of degree < 1 determine all residue classes comprising $\mathbb{F}_2[x]/(x)$. Thus, $\mathbb{F}_2[x]/(x)$ consists of the residue classes [0] and [1] and is isomorphic to $\mathbb{F}_2$.

(ii) Let $f(x) = x^2 + x + 1 \in \mathbb{F}_2[x]$. Then $\mathbb{F}_2[x]/(f)$ has the $p^n = 2^2$ elements [0], [1], $[x]$, $[x + 1]$. The operation tables for this residue class ring are obtained by performing the required operations with the polynomials determining the residue classes and by carrying out reduction mod f if necessary:

+	[0]	[1]	$[x]$	$[x+1]$
[0]	[0]	[1]	$[x]$	$[x+1]$
[1]	[1]	[0]	$[x+1]$	$[x]$
$[x]$	$[x]$	$[x+1]$	[0]	[1]
$[x+1]$	$[x+1]$	$[x]$	[1]	[0]

·	[0]	[1]	[x]	[x+1]
[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[x]	[x+1]
[x]	[0]	[x]	[x+1]	[1]
[x+1]	[0]	[x+1]	[1]	[x]

By inspecting these tables, or from the irreducibility of f over $\mathbb{F}_2$ and Theorem 1.61, it follows that $\mathbb{F}_2[x]/(f)$ is a field. This is our first example of a finite field for which the number of elements is not a prime.

(iii) Let $f(x) = x^2 + 2 \in \mathbb{F}_3[x]$. Then $\mathbb{F}_3[x]/(f)$ consists of the $p^n = 3^2$ residue classes [0], [1], [2], [x], [x + 1], [x + 2], [2x], [2x + 1], [2x + 2]. The operation tables for $\mathbb{F}_3[x]/(f)$ are again produced by performing polynomial operations and using reduction mod f whenever necessary. Since $\mathbb{F}_3[x]/(f)$ is a commutative ring, we only have to compute the entries on and above the main diagonal.

+	[0]	[1]	[2]	[x]	[x+1]	[x+2]	[2x]	[2x+1]	[2x+2]
[0]	[0]	[1]	[2]	[x]	[x+1]	[x+2]	[2x]	[2x+1]	[2x+2]
[1]		[2]	[0]	[x+1]	[x+2]	[x]	[2x+1]	[2x+2]	[2x]
[2]			[1]	[x+2]	[x]	[x+1]	[2x+2]	[2x]	[2x+1]
[x]				[2x]	[2x+1]	[2x+2]	[0]	[1]	[2]
[x+1]					[2x+2]	[2x]	[1]	[2]	[0]
[x+2]						[2x+1]	[2]	[0]	[1]
[2x]							[x]	[x+1]	[x+2]
[2x+1]								[x+2]	[x]
[2x+2]									[x+1]

·	[0]	[1]	[2]	[x]	[x+1]	[x+2]	[2x]	[2x+1]	[2x+2]
[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]
[1]		[1]	[2]	[x]	[x+1]	[x+2]	[2x]	[2x+1]	[2x+2]
[2]			[1]	[2x]	[2x+2]	[2x+1]	[x]	[x+2]	[x+1]
[x]				[1]	[x+1]	[2x+1]	[2]	[x+2]	[2x+2]
[x+1]					[2x+2]	[0]	[2x+2]	[0]	[x+1]
[x+2]						[x+2]	[x+2]	[2x+1]	[0]
[2x]							[1]	[2x+1]	[x+1]
[2x+1]								[x+2]	[0]
[2x+2]									[2x+2]

Note that $\mathbb{F}_3[x]/(f)$ is not a field (and not even an integral domain). This is in accordance with Theorem 1.61 since $x^2 + 2 = (x + 1)(x + 2)$ is reducible over $\mathbb{F}_3$. □

If F is again an arbitrary field and $f(x) \in F[x]$, then replacement of the indeterminate x in $f(x)$ by a fixed element of F yields a well-defined

element of F. In detail, if $f(x) = a_0 + a_1x + \cdots + a_nx^n \in F[x]$ and $b \in F$, then replacing x by b we get $f(b) = a_0 + a_1b + \cdots + a_nb^n \in F$. In any polynomial identity in $F[x]$ we can substitute a fixed $b \in F$ for x and obtain a valid identity in F (*principle of substitution*).

1.63. Definition. An element $b \in F$ is called a *root* (or a *zero*) of the polynomial $f \in F[x]$ if $f(b) = 0$.

An important connection between roots and divisibility is given by the following theorem.

1.64. Theorem. *An element $b \in F$ is a root of the polynomial $f \in F[x]$ if and only if $x - b$ divides $f(x)$.*

Proof. We use the division algorithm (see Theorem 1.52) to write $f(x) = q(x)(x - b) + c$ with $q \in F[x]$ and $c \in F$. Substituting b for x, we get $f(b) = c$, hence $f(x) = q(x)(x - b) + f(b)$. The theorem follows now from this identity. □

1.65. Definition. Let $b \in F$ be a root of the polynomial $f \in F[x]$. If k is a positive integer such that $f(x)$ is divisible by $(x - b)^k$, but not by $(x - b)^{k+1}$, then k is called the *multiplicity* of b. If $k = 1$, then b is called a *simple root* (or a *simple zero*) of f, and if $k \geqslant 2$, then b is called a *multiple root* (or a *multiple zero*) of f.

1.66. Theorem. *Let $f \in F[x]$ with $\deg f = n \geqslant 0$. If $b_1, \ldots, b_m \in F$ are distinct roots of f with multiplicities $k_1, \ldots, k_m$, respectively, then $(x - b_1)^{k_1} \cdots (x - b_m)^{k_m}$ divides $f(x)$. Consequently, $k_1 + \cdots + k_m \leqslant n$, and f can have at most n distinct roots in F.*

Proof. We note that each polynomial $x - b_j$, $1 \leqslant j \leqslant m$, is irreducible over F, and so $(x - b_j)^{k_j}$ occurs as a factor in the canonical factorization of f. Altogether, the factor $(x - b_1)^{k_1} \cdots (x - b_m)^{k_m}$ appears in the canonical factorization of f and is thus a divisor of f. By comparing degrees, we get $k_1 + \cdots + k_m \leqslant n$, and $m \leqslant k_1 + \cdots + k_m \leqslant n$ shows the last statement. □

1.67. Definition. If $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \in F[x]$, then the *derivative* f' of f is defined by $f' = f'(x) = a_1 + 2a_2x + \cdots + na_nx^{n-1} \in F[x]$.

1.68. Theorem. *The element $b \in F$ is a multiple root of $f \in F[x]$ if and only if it is a root of both f and f'.*

There is a relation between the nonexistence of roots and irreducibility. If f is an irreducible polynomial in $F[x]$ of degree $\geqslant 2$, then Theorem 1.64 shows that f has no root in F. The converse holds for polynomials of degree 2 or 3, but not necessarily for polynomials of higher degree.

1.69. ***Theorem.*** *The polynomial $f \in F[x]$ of degree 2 or 3 is irreducible in $F[x]$ if and only if f has no root in F.*

Proof. The necessity of the condition was already noted. Conversely, if f has no root in F and were reducible in $F[x]$, we could write $f = gh$ with $g, h \in F[x]$ and $1 \leqslant \deg(g) \leqslant \deg(h)$. But $\deg(g)+\deg(h) = \deg(f) \leqslant 3$, hence $\deg(g) = 1$; that is, $g(x) = ax + b$ with $a, b \in F$, $a \neq 0$. Then $-ba^{-1}$ is a root of g, and so a root of f in F, a contradiction. □

1.70. Example. Because of Theorem 1.69, the irreducible polynomials in $\mathbb{F}_2[x]$ of degree 2 or 3 can be obtained by eliminating the polynomials with roots in $\mathbb{F}_2$ from the set of all polynomials in $\mathbb{F}_2[x]$ of degree 2 or 3. The only irreducible polynomial in $\mathbb{F}_2[x]$ of degree 2 is $f(x) = x^2 + x + 1$, and the irreducible polynomials in $\mathbb{F}_2[x]$ of degree 3 are $f_1(x) = x^3 + x + 1$ and $f_2(x) = x^3 + x^2 + 1$. □

In elementary analysis there is a well-known method for constructing a polynomial with real coefficients which assumes certain assigned values for given values of the indeterminate. The same method carries over to any field.

1.71. ***Theorem*** (Lagrange Interpolation Formula). *For $n \geqslant 0$, let $a_0, \ldots, a_n$ be $n+1$ distinct elements of F, and let $b_0, \ldots, b_n$ be $n+1$ arbitrary elements of F. Then there exists exactly one polynomial $f \in F[x]$ of degree $\leqslant n$ such that $f(a_i) = b_i$ for $i = 0, \ldots, n$. This polynomial is given by*

$$f(x) = \sum_{i=0}^{n} b_i \prod_{\substack{k=0 \\ k \neq i}}^{n} (a_i - a_k)^{-1}(x - a_k).$$

One can also consider polynomials in several indeterminates. Let R denote a commutative ring with identity and let $x_1, \ldots, x_n$ be symbols that will serve as indeterminates. We form the polynomial ring $R[x_1]$, then the polynomial ring $R[x_1, x_2] = R[x_1][x_2]$, and so on, until we arrive at $R[x_1, \ldots, x_n] = R[x_1, \ldots, x_{n-1}][x_n]$. The elements of $R[x_1, \ldots, x_n]$ are then expressions of the form

$$f = f(x_1, \ldots, x_n) = \sum a_{i_1 \cdots i_n} x_1^{i_1} \cdots x_n^{i_n}$$

with coefficients $a_{i_1 \cdots i_n} \in R$, where the summation is extended over finitely many n-tuples $(i_1, \ldots, i_n)$ of nonnegative integers and the convention $x_j^0 = 1$ $(1 \leqslant j \leqslant n)$ is observed. Such an expression is called a *polynomial in $x_1, \ldots, x_n$* over R. Two polynomials $f, g \in R[x_1, \ldots, x_n]$ are equal if and only if all corresponding coefficients are equal. It is tacitly assumed that the indeterminates $x_1, \ldots, x_n$ commute with each other, so that, for instance, the expressions $x_1x_2x_3x_4$ and $x_4x_1x_3x_2$ are identified.

1.72. Definition. Let $f \in R[x_1, \ldots, x_n]$ be given by

$$f(x_1, \ldots, x_n) = \sum a_{i_1 \cdots i_n} x_1^{i_1} \cdots x_n^{i_n}.$$

If $a_{i_1\cdots i_n} \neq 0$, then $a_{i_1\cdots i_n} x_1^{i_1} \cdots x_n^{i_n}$ is called a *term* of f and $i_1 + \cdots + i_n$ is the degree of the term. For $f \neq 0$ one defines the *degree* of f, denoted by $\deg(f)$, to be the maximum of the degrees of the terms of f. For $f = 0$ one sets $\deg(f) = -\infty$. If $f = 0$ or if all terms of f have the same degree, then f is called *homogeneous*.

Any $f \in R[x_1,\ldots,x_n]$ can be written as a finite sum of homogeneous polynomials. The degrees of polynomials in $R[x_1,\ldots,x_n]$ satisfy again the inequalities in Theorem 1.50, and if R is an integral domain, then (1.4) is valid and $R[x_1,\ldots,x_n]$ is an integral domain. If F is a field, then the polynomials in $F[x_1,\ldots,x_n]$ of positive degree can again be factored uniquely into a constant factor and a product of "monic" prime elements (using a suitable definition of "monic"), but for $n \geqslant 2$ there is no analog of the division algorithm (in the case of commuting indeterminates) and $F[x_1,\ldots,x_n]$ is not a principal ideal domain.

An important special class of polynomials in n indeterminates is that of symmetric polynomials.

1.73. Definition. A polynomial $f \in R[x_1,\ldots,x_n]$ is called *symmetric* if $f(x_{i_1},\ldots,x_{i_n}) = f(x_1,\ldots,x_n)$ for any permutation $i_1,\ldots,i_n$ of the integers $1,\ldots,n$.

1.74. Example. Let z be an indeterminate over $R[x_1,\ldots,x_n]$, and let $g(z) = (z - x_1)(z - x_2)\cdots(z - x_n)$. Then

$$g(z) = z^n - \sigma_1 z^{n-1} + \sigma_2 z^{n-2} + \cdots + (-1)^n \sigma_n$$

with

$$\sigma_k = \sigma_k(x_1,\ldots,x_n) = \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} x_{i_1} \cdots x_{i_k} \quad (k = 1,2,\ldots,n).$$

Thus:

$$\sigma_1 = x_1 + x_2 + \cdots + x_n,$$
$$\sigma_2 = x_1x_2 + x_1x_3 + \cdots + x_1x_n + x_2x_3 + \cdots + x_2x_n + \cdots + x_{n-1}x_n,$$
$$\vdots$$
$$\sigma_n = x_1x_2\cdots x_n.$$

As g remains unaltered under any permutation of the x_i, all the σ_k are symmetric polynomials; they are also homogeneous. The polynomial $\sigma_k = \sigma_k(x_1,\ldots,x_n) \in R[x_1,\ldots,x_n]$ is called the kth *elementary symmetric polynomial* in the indeterminates $x_1,\ldots,x_n$ over R. The adjective "elementary" is used because of the so-called "fundamental theorem on symmetric polynomials," which states that for any symmetric polynomial $f \in R[x_1,\ldots,x_n]$ there exists a uniquely determined polynomial $h \in R[x_1,\ldots,x_n]$ such that $f(x_1,\ldots,x_n) = h(\sigma_1,\ldots,\sigma_n)$. □

1.75. Theorem (Newton's Formula). *Let $\sigma_1,\ldots,\sigma_n$ be the elementary symmetric polynomials in $x_1,\ldots,x_n$ over R, and let $s_0 = n \in \mathbb{Z}$ and*

$s_k = s_k(x_1,\ldots,x_n) = x_1^k + \cdots + x_n^k \in R[x_1,\ldots,x_n]$ *for* $k \geqslant 1$. *Then the formula*

$$s_k - s_{k-1}\sigma_1 + s_{k-2}\sigma_2 + \cdots + (-1)^{m-1}s_{k-m+1}\sigma_{m-1} + (-1)^m \frac{m}{n} s_{k-m}\sigma_m = 0$$

holds for $k \geqslant 1$, *where* $m = \min(k, n)$.

1.76. Theorem (Waring's Formula). *With the same notation as in Theorem* 1.75, *we have*

$$s_k = \sum (-1)^{i_2+i_4+i_6+\cdots} \frac{(i_1+i_2+\cdots+i_n-1)!k}{i_1!i_2!\cdots i_n!} \sigma_1^{i_1}\sigma_2^{i_2}\cdots\sigma_n^{i_n}$$

for $k \geqslant 1$, *where the summation is extended over all n-tuples* $(i_1,\ldots,i_n)$ *of nonnegative integers with* $i_1 + 2i_2 + \cdots + ni_n = k$. *The coefficient of* $\sigma_1^{i_1}\sigma_2^{i_2}\cdots\sigma_n^{i_n}$ *is always an integer.*

4. FIELD EXTENSIONS

Let F be a field. A subset K of F that is itself a field under the operations of F will be called a *subfield* of F. In this context, F is called an *extension* (*field*) of K. If $K \neq F$, we say that K is a *proper subfield* of F.

If K is a subfield of the finite field $\mathbb{F}_p$, p prime, then K must contain the elements 0 and 1, and so all other elements of $\mathbb{F}_p$ by the closure of K under addition. It follows that $\mathbb{F}_p$ contains no proper subfields. We are thus led to the following concept.

1.77. Definition. A field containing no proper subfields is called a *prime field*.

By the above argument, any finite field of order p, p prime, is a prime field. Another example of a prime field is the field $\mathbb{Q}$ of rational numbers.

The intersection of any nonempty collection of subfields of a given field F is again a subfield of F. If we form the intersection of *all* subfields of F, we obtain the *prime subfield* of F. It is obviously a prime field.

1.78. Theorem. *The prime subfield of a field F is isomorphic to either* $\mathbb{F}_p$ *or* $\mathbb{Q}$, *according as the characteristic of F is a prime p or* 0.

1.79. Definition. Let K be a subfield of the field F and M any subset of F. Then the field $K(M)$ is defined as the intersection of all subfields of F containing both K and M and is called the extension (field) of K obtained by *adjoining* the elements in M. For finite $M = \{\theta_1,\ldots,\theta_n\}$ we write $K(M) = K(\theta_1,\ldots,\theta_n)$. If M consists of a single element $\theta \in F$, then $L = K(\theta)$ is said to be a *simple extension* of K and θ is called a *defining element* of L over K.

Obviously, $K(M)$ is the smallest subfield of F containing both K and M. We define now an important type of extension.

1.80. Definition. Let K be a subfield of F and $\theta \in F$. If θ satisfies a nontrivial polynomial equation with coefficients in K, that is, if $a_n\theta^n + \cdots + a_1\theta + a_0 = 0$ with $a_i \in K$ not all being 0, then θ is said to be *algebraic* over K. An extension L of K is called *algebraic* over K (or an *algebraic extension* of K) if every element of L is algebraic over K.

Suppose $\theta \in F$ is algebraic over K, and consider the set $J = \{f \in K[x]: f(\theta) = 0\}$. It is easily checked that J is an ideal of $K[x]$, and we have $J \neq (0)$ since θ is algebraic over K. It follows then from Theorem 1.54 that there exists a uniquely determined monic polynomial $g \in K[x]$ such that J is equal to the principal ideal (g). It is important to note that g is irreducible in $K[x]$. For, in the first place, g is of positive degree since it has the root θ; and if $g = h_1h_2$ in $K[x]$ with $1 \leqslant \deg(h_i) < \deg(g)$ $(i = 1, 2)$, then $0 = g(\theta) = h_1(\theta)h_2(\theta)$ implies that either h_1 or h_2 is in J and so divisible by g, which is impossible.

1.81. Definition. If $\theta \in F$ is algebraic over K, then the uniquely determined monic polynomial $g \in K[x]$ generating the ideal $J = \{f \in K[x]: f(\theta) = 0\}$ of $K[x]$ is called the *minimal polynomial* (or *defining polynomial*, or *irreducible polynomial*) of θ over K. By the *degree* of θ over K we mean the degree of g.

1.82. Theorem. *If $\theta \in F$ is algebraic over K, then its minimal polynomial g over K has the following properties*:

(i) *g is irreducible in $K[x]$.*
(ii) *For $f \in K[x]$ we have $f(\theta) = 0$ if and only if g divides f.*
(iii) *g is the monic polynomial in $K[x]$ of least degree having θ as a root.*

Proof. Property (i) was already noted and (ii) follows from the definition of g. As to (iii), it suffices to note that any monic polynomial in $K[x]$ having θ as a root must be a multiple of g, and so it is either equal to g or its degree is larger than that of g. □

We note that both the minimal polynomial and the degree of an algebraic element θ depend on the field K over which it is considered, so that one must be careful not to speak of the minimal polynomial or the degree of θ without specifying K, unless the latter is amply clear from the context.

If L is an extension field of K, then L may be viewed as a vector space over K. For the elements of L (= "vectors") form, first of all, an abelian group under addition. Moreover, each "vector" $\alpha \in L$ can be multiplied by a "scalar" $r \in K$ so that $r\alpha$ is again in L (here $r\alpha$ is simply the

product of the field elements r and α of L) and the laws for multiplication by scalars are satisfied: $r(\alpha+\beta)=r\alpha+r\beta$, $(r+s)\alpha=r\alpha+s\alpha$, $(rs)\alpha=r(s\alpha)$, and $1\alpha=\alpha$, where $r, s \in K$ and $\alpha, \beta \in L$.

1.83. Definition. Let L be an extension field of K. If L, considered as a vector space over K, is finite-dimensional, then L is called a *finite extension* of K. The dimension of the vector space L over K is then called the *degree* of L over K, in symbols $[L: K]$.

1.84. Theorem. *If L is a finite extension of K and M is a finite extension of L, then M is a finite extension of K with*

$$[M: K]=[M: L][L: K].$$

Proof. Put $[M: L]=m$, $[L: K]=n$, and let $\{\alpha_1,\ldots,\alpha_m\}$ be a basis of M over L and $\{\beta_1,\ldots,\beta_n\}$ a basis of L over K. Then every $\alpha \in M$ is a linear combination $\alpha=\gamma_1\alpha_1+\cdots+\gamma_m\alpha_m$ with $\gamma_i \in L$ for $1 \leqslant i \leqslant m$, and writing each γ_i in terms of the basis elements β_j we get

$$\alpha=\sum_{i=1}^{m}\gamma_i\alpha_i=\sum_{i=1}^{m}\left(\sum_{j=1}^{n}r_{ij}\beta_j\right)\alpha_i=\sum_{i=1}^{m}\sum_{j=1}^{n}r_{ij}\beta_j\alpha_i$$

with coefficients $r_{ij} \in K$. If we can show that the mn elements $\beta_j\alpha_i$, $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$, are linearly independent over K, then we are done. So suppose we have

$$\sum_{i=1}^{m}\sum_{j=1}^{n}s_{ij}\beta_j\alpha_i=0$$

with coefficients $s_{ij} \in K$. Then

$$\sum_{i=1}^{m}\left(\sum_{j=1}^{n}s_{ij}\beta_j\right)\alpha_i=0,$$

and from the linear independence of the α_i over L we infer

$$\sum_{j=1}^{n}s_{ij}\beta_j=0 \quad \text{for } 1 \leqslant i \leqslant m.$$

But since the β_j are linearly independent over K, we conclude that all s_{ij} are 0. □

1.85. Theorem. *Every finite extension of K is algebraic over K.*

Proof. Let L be a finite extension of K and put $[L: K]=m$. For $\theta \in L$, the $m+1$ elements $1, \theta, \ldots, \theta^m$ must then be linearly dependent over K, and so we get a relation $a_0+a_1\theta+\cdots+a_m\theta^m=0$ with $a_i \in K$ not all being 0. This just says that θ is algebraic over K. □

For the study of the structure of a simple extension $K(\theta)$ of K obtained by adjoining an algebraic element, let F be an extension of K and let $\theta \in F$ be algebraic over K. It turns out that $K(\theta)$ is a finite (and therefore an algebraic) extension of K.

1.86. Theorem. *Let $\theta \in F$ be algebraic of degree n over K and let g be the minimal polynomial of θ over K. Then:*

(i) *$K(\theta)$ is isomorphic to $K[x]/(g)$.*
(ii) *$[K(\theta): K] = n$ and $\{1, \theta, \ldots, \theta^{n-1}\}$ is a basis of $K(\theta)$ over K.*
(iii) *Every $\alpha \in K(\theta)$ is algebraic over K and its degree over K is a divisor of n.*

Proof. (i) Consider the mapping $\tau: K[x] \to K(\theta)$, defined by $\tau(f) = f(\theta)$ for $f \in K[x]$, which is easily seen to be a ring homomorphism. We have $\ker \tau = \{f \in K[x]: f(\theta) = 0\} = (g)$ by the definition of the minimal polynomial. Let S be the image of τ; that is, S is the set of polynomial expressions in θ with coefficients in K. Then the homomorphism theorem for rings (see Theorem 1.40) yields that S is isomorphic to $K[x]/(g)$. But $K[x]/(g)$ is a field by Theorems 1.61 and 1.82(i), and so S is a field. Since $K \subseteq S \subseteq K(\theta)$ and $\theta \in S$, it follows from the definition of $K(\theta)$ that $S = K(\theta)$, and (i) is thus shown.

(ii) Since $S = K(\theta)$, any given $\alpha \in K(\theta)$ can be written in the form $\alpha = f(\theta)$ for some $f \in K[x]$. By the division algorithm, $f = qg + r$ with $q, r \in K[x]$ and $\deg(r) < \deg(g) = n$. Then $\alpha = f(\theta) = q(\theta)g(\theta) + r(\theta) = r(\theta)$, and so α is a linear combination of $1, \theta, \ldots, \theta^{n-1}$ with coefficients in K. On the other hand, if $a_0 + a_1\theta + \cdots + a_{n-1}\theta^{n-1} = 0$ for certain $a_i \in K$, then the polynomial $h(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} \in K[x]$ has θ as a root and is thus a multiple of g by Theorem 1.82(ii). Since $\deg(h) < n = \deg(g)$, this is only possible if $h = 0$—that is, if all $a_i = 0$. Therefore, the elements $1, \theta, \ldots, \theta^{n-1}$ are linearly independent over K and (ii) follows.

(iii) $K(\theta)$ is a finite extension of K by (ii), and so $\alpha \in K(\theta)$ is algebraic over K by Theorem 1.85. Furthermore, $K(\alpha)$ is a subfield of $K(\theta)$. If d is the degree of α over K, then (ii) and Theorem 1.84 imply that $n = [K(\theta): K] = [K(\theta): K(\alpha)][K(\alpha): K] = [K(\theta): K(\alpha)]d$, hence d divides n. □

The elements of the simple algebraic extension $K(\theta)$ of K are therefore polynomial expressions in θ. Any element of $K(\theta)$ can be uniquely represented in the form $a_0 + a_1\theta + \cdots + a_{n-1}\theta^{n-1}$ with $a_i \in K$ for $0 \leqslant i \leqslant n-1$.

It should be pointed out that Theorem 1.86 operates under the assumption that both K and θ are embedded in a larger field F. This is necessary in order that algebraic expressions involving θ make sense. We now want to construct a simple algebraic extension *ab ovo*—that is, without

reference to a previously given larger field. The clue to this is contained in part (i) of Theorem 1.86.

1.87. Theorem. *Let $f \in K[x]$ be irreducible over the field K. Then there exists a simple algebraic extension of K with a root of f as a defining element.*

Proof. Consider the residue class ring $L = K[x]/(f)$, which is a field by Theorem 1.61. The elements of L are the residue classes $[h] = h + (f)$ with $h \in K[x]$. For any $a \in K$ we can form the residue class $[a]$ determined by the constant polynomial a, and if $a, b \in K$ are distinct, then $[a] \neq [b]$ since f has positive degree. The mapping $a \mapsto [a]$ gives an isomorphism from K onto a subfield K' of L, so that K' may be identified with K. In other words, we can view L as an extension of K. For every $h(x) = a_0 + a_1x + \cdots + a_mx^m \in K[x]$ we have $[h] = [a_0 + a_1x + \cdots + a_mx^m] = [a_0] + [a_1][x] + \cdots + [a_m][x]^m = a_0 + a_1[x] + \cdots + a_m[x]^m$ by the rules for operating with residue classes and the identification $[a_i] = a_i$. Thus, every element of L can be written as a polynomial expression in $[x]$ with coefficients in K. Since any field containing both K and $[x]$ must contain these polynomial expressions, L is a simple extension of K obtained by adjoining $[x]$. If $f(x) = b_0 + b_1x + \cdots + b_nx^n$, then $f([x]) = b_0 + b_1[x] + \cdots + b_n[x]^n = [b_0 + b_1x + \cdots + b_nx^n] = [f] = [0]$, so that $[x]$ is a root of f and L is a simple algebraic extension of K. □

1.88. Example. As an example of the formal process of root adjunction in Theorem 1.87, consider the prime field $\mathbb{F}_3$ and the polynomial $f(x) = x^2 + x + 2 \in \mathbb{F}_3[x]$, which is irreducible over $\mathbb{F}_3$. Let θ be a "root" of f; that is, θ is the residue class $x + (f)$ in $L = \mathbb{F}_3[x]/(f)$. The other root of f in L is then $2\theta + 2$, since $f(2\theta + 2) = (2\theta + 2)^2 + (2\theta + 2) + 2 = \theta^2 + \theta + 2 = 0$. By Theorem 1.86(ii), or by the known structure of a residue class field, the simple algebraic extension $L = \mathbb{F}_3(\theta)$ consists of the nine elements $0, 1, 2, \theta, \theta + 1, \theta + 2, 2\theta, 2\theta + 1, 2\theta + 2$. The operation tables for L can be constructed as in Example 1.62. □

We observe that in the above example we may adjoin either the root θ or the root $2\theta + 2$ of f and we would still obtain the same field. This situation is covered by the following result, which is easily established.

1.89. Theorem. *Let α and β be two roots of the polynomial $f \in K[x]$ that is irreducible over K. Then $K(\alpha)$ and $K(\beta)$ are isomorphic under an isomorphism mapping α to β and keeping the elements of K fixed.*

We are now asking for an extension field to which all roots of a given polynomial belong.

1.90. Definition. Let $f \in K[x]$ be of positive degree and F an extension field of K. Then f is said to *split in F* if f can be written as a product of

linear factors in $F[x]$—that is, if there exist elements $\alpha_1, \alpha_2, \ldots, \alpha_n \in F$ such that

$$f(x) = a(x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_n),$$

where a is the leading coefficient of f. The field F is a *splitting field* of f over K if f splits in F and if, moreover, $F = K(\alpha_1, \alpha_2, \ldots, \alpha_n)$.

It is clear that a splitting field F of f over K is in the following sense the smallest field containing all the roots of f: no proper subfield of F that is an extension of K contains all the roots of f. By repeatedly applying the process used in Theorem 1.87, one obtains the first part of the subsequent result. The second part is an extension of Theorem 1.89.

1.91. Theorem (Existence and Uniqueness of Splitting Field). *If K is a field and f any polynomial of positive degree in $K[x]$, then there exists a splitting field of f over K. Any two splitting fields of f over K are isomorphic under an isomorphism which keeps the elements of K fixed and maps roots of f into each other.*

Since isomorphic fields may be identified, we can speak of *the* splitting field of f over K. It is obtained from K by adjoining finitely many algebraic elements over K, and therefore one can show on the basis of Theorems 1.84 and 1.86(ii) that the splitting field of f over K is a finite extension of K.

As an illustration of the usefulness of splitting fields, we consider the question of deciding whether a given polynomial has a multiple root (compare with Definition 1.65).

1.92. Definition. Let $f \in K[x]$ be a polynomial of degree $n \geqslant 2$ and suppose that $f(x) = a_0(x-\alpha_1)\cdots(x-\alpha_n)$ with $\alpha_1, \ldots, \alpha_n$ in the splitting field of f over K. Then the *discriminant* $D(f)$ of f is defined by

$$D(f) = a_0^{2n-2} \prod_{1 \leqslant i < j \leqslant n} (\alpha_i - \alpha_j)^2.$$

It is obvious from the definition of $D(f)$ that f has a multiple root if and only if $D(f) = 0$. Although $D(f)$ is defined in terms of elements of an extension of K, it is actually an element of K itself. For small n this can be seen by direct calculation. For instance, if $n = 2$ and $f(x) = ax^2 + bx + c = a(x-\alpha_1)(x-\alpha_2)$, then $D(f) = a^2(\alpha_1 - \alpha_2)^2 = a^2((\alpha_1+\alpha_2)^2 - 4\alpha_1\alpha_2) = a^2(b^2a^{-2} - 4ca^{-1})$, hence

$$D(ax^2 + bx + c) = b^2 - 4ac,$$

a well-known expression from the theory of quadratic equations. If $n = 3$ and $f(x) = ax^3 + bx^2 + cx + d = a(x-\alpha_1)(x-\alpha_2)(x-\alpha_3)$, then $D(f) = a^4(\alpha_1-\alpha_2)^2(\alpha_1-\alpha_3)^2(\alpha_2-\alpha_3)^2$, and a more involved computation yields

$$D(ax^3 + bx^2 + cx + d) = b^2c^2 - 4b^3d - 4ac^3 - 27a^2d^2 + 18abcd. \quad (1.9)$$

In the general case, consider first the polynomial $s \in K[x_1,\ldots,x_n]$ given by

$$s(x_1,\ldots,x_n) = a_0^{2n-2} \prod_{1 \leqslant i < j \leqslant n} (x_i - x_j)^2.$$

Then s is a symmetric polynomial, and by a result in Example 1.74 it can be written as a polynomial expression in the elementary symmetric polynomials $\sigma_1,\ldots,\sigma_n$—that is, $s = h(\sigma_1,\ldots,\sigma_n)$ for some $h \in K[x_1,\ldots,x_n]$. If $f(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n = a_0(x-\alpha_1)\cdots(x-\alpha_n)$, then the definition of the elementary symmetric polynomials (see again Example 1.74) implies that $\sigma_k(\alpha_1,\ldots,\alpha_n) = (-1)^k a_k a_0^{-1} \in K$ for $1 \leqslant k \leqslant n$. Thus,

$$\begin{aligned} D(f) = s(\alpha_1,\ldots,\alpha_n) &= h(\sigma_1(\alpha_1,\ldots,\alpha_n),\ldots,\sigma_n(\alpha_1,\ldots,\alpha_n)) \\ &= h\big(-a_1a_0^{-1},\ldots,(-1)^n a_n a_0^{-1}\big) \in K. \end{aligned}$$

Since $D(f) \in K$, it should be possible to calculate $D(f)$ without having to pass to an extension field of K. This can be done via the notion of resultant. We note first that if a polynomial $f \in K[x]$ is given in the form $f(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n$ and we accept the possibility that $a_0 = 0$, then n need not be the degree of f. We speak of n as the *formal degree* of f; it is always greater than or equal to $\deg(f)$.

1.93. Definition. Let $f(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n \in K[x]$ and $g(x) = b_0x^m + b_1x^{m-1} + \cdots + b_m \in K[x]$ be two polynomials of formal degree n resp. m with $n, m \in \mathbb{N}$. Then the *resultant* $R(f,g)$ of the two polynomials is defined by the determinant

$$R(f,g) = \begin{vmatrix} a_0 & a_1 & \cdots & a_n & 0 & & \cdots & 0 \\ 0 & a_0 & a_1 & \cdots & a_n & 0 & \cdots & 0 \\ \vdots & & & & & & & \vdots \\ 0 & \cdots & 0 & a_0 & a_1 & & \cdots & a_n \\ b_0 & b_1 & \cdots & & b_m & 0 & \cdots & 0 \\ 0 & b_0 & b_1 & \cdots & & b_m & \cdots & 0 \\ \vdots & & & & & & & \vdots \\ 0 & \cdots & 0 & b_0 & b_1 & & \cdots & b_m \end{vmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}}\right\} m \text{ rows} \\ \left.\vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}}\right\} n \text{ rows} \end{matrix}$$

of order $m+n$.

If $\deg(f) = n$ (i.e., if $a_0 \neq 0$) and $f(x) = a_0(x-\alpha_1)\cdots(x-\alpha_n)$ in the splitting field of f over K, then $R(f,g)$ is also given by the formula

$$R(f,g) = a_0^m \prod_{i=1}^{n} g(\alpha_i). \tag{1.10}$$

In this case, we obviously have $R(f,g) = 0$ if and only if f and g have a common root, which is the same as saying that f and g have a common divisor in $K[x]$ of positive degree.

Theorem 1.68 suggests a connection between the discriminant $D(f)$ and the resultant $R(f, f')$. Let $f \in K[x]$ with $\deg(f) = n \geqslant 2$ and leading coefficient a_0. Then we have, in fact, the identity

$$D(f) = (-1)^{n(n-1)/2} a_0^{-1} R(f, f'), \tag{1.11}$$

where f' is viewed as a polynomial of formal degree $n-1$. The last remark is needed since we may have $\deg(f') < n-1$ and even $f' = 0$ in case K has prime characteristic. At any rate, the identity (1.11) shows that we can obtain $D(f)$ by calculating a determinant of order $2n-1$ with entries in K.

NOTES

1. The definitions and theorems in this chapter can be found in nearly any of the introductory books on modern algebra. To mention a few: Birkhoff and MacLane [1], Fraleigh [1], Herstein [4], Kochendörffer [1], Lang [4], Rédei [10], van der Waerden [2].

There are various alternative definitions of a group; for example, a group may be defined as a nonempty set G together with an associative binary operation such that for all $a, b \in G$ the equations $ax = b$ and $ya = b$ have solutions in G. Apart from the examples already given, important illustrations of the group concept are furnished by matrix groups—that is, sets of matrices with entries in a field that form groups under matrix multiplication. Such groups will occur in Chapter 8. For many other examples of groups we refer to the textbooks mentioned above.

A square table such that in every row and in every column each element of a certain set occurs exactly once is called a *latin square*. Hence, the Cayley table for any finite group forms a latin square. However, not every latin square may be regarded as a Cayley table since the associative law need not hold. See Chapter 9, Section 4, and Dénes and Keedwell [1] for more information on latin squares.

In connection with cyclic groups one can prove easily that any infinite cyclic group is isomorphic to the additive group $\mathbf{Z}$ of the integers and any cyclic group of order n is isomorphic to $\mathbf{Z}_n$.

We mention the definitions of algebraic systems that are even simpler than groups, insofar as only a part of the group axioms is assumed. A set with a binary operation is called a *groupoid*; if, in addition, associativity is assumed, then we speak of a *semigroup*. A semigroup with an identity element is called a *monoid*.

2. There are various definitions of a ring. For instance, some authors drop the associativity of multiplication and call the structure introduced in Definition 1.28 an associative ring. The requirement of the

existence of a multiplicative identity in an integral domain is sometimes omitted.

The first abstract definition of a field was given by Weber [3]. The finite fields $\mathbb{F}_p$, p prime, were already studied extensively by Gauss [1] in the context of congruences in $\mathbb{Z}$ with respect to prime moduli.

The characteristic of a field is equal to the characteristic of its prime subfield. There are fields of prime characteristic that are not finite. To get examples, consider suitable extensions of $\mathbb{F}_p$, such as the field of rational functions over $\mathbb{F}_p$ or the algebraic closure of $\mathbb{F}_p$ (compare with the notes on Section 4).

Many properties of the integers can be translated into properties of the corresponding principal ideals in the ring $\mathbb{Z}$. This is based on the fact that the integer a divides the integer b if and only if the principal ideal (a) contains the principal ideal (b). Of particular interest are the prime numbers. According to the usual definition, a prime is an integer >1 that has no nontrivial divisors. Alternatively, one could define a prime as an integer >1 that divides a product of integers only if it divides at least one of the factors. Phrased in terms of ideals, these characterizations lead to the definition of maximal and prime ideals.

3. In the usual definition of a polynomial as an expression of the form $a_0 + a_1x + \cdots + a_nx^n$, the question of how the coefficients a_i and the indeterminate x are connected is glossed over or altogether avoided. There is, however, a way of giving a rigorous definition of a polynomial as an element of a polynomial ring.

For this definition of a polynomial ring, we consider the set S of all infinite sequences of the form

$$(a_0, a_1, \ldots, a_n, \ldots),$$

where the components a_i are elements of a commutative ring R with identity 1 and at most finitely many a_i are allowed to be different from 0. One can easily show that the set S forms a commutative ring with identity with respect to the following operations of addition and multiplication:

$$(a_0, a_1, \ldots) + (b_0, b_1, \ldots) = (a_0 + b_0, a_1 + b_1, \ldots),$$
$$(a_0, a_1, \ldots)(b_0, b_1, \ldots) = (a_0b_0, a_0b_1 + a_1b_0, \ldots),$$

the $(n+1)$st component in the product being $a_0b_n + a_1b_{n-1} + \cdots + a_{n-1}b_1 + a_nb_0$. The zero element of this ring S is obviously $(0,0,\ldots)$ and the identity is the sequence $(1,0,0,\ldots)$.

The set P of special sequences $(a_0,0,0,\ldots)$, where at most the first component is different from 0, forms a subring of S. This subring P and the given ring R are isomorphic via the mapping $(a_0,0,0,\ldots) \mapsto a_0$ from P onto R. Thus we identify these two rings and write $(a_0,0,0,\ldots) = a_0$. Hence R can be regarded as a subring of S, and S is an extension ring of R.

We introduce the notation $x = (0,1,0,\ldots)$ for this special sequence and verify that

$$x^n = (0,\ldots,0,1,0,\ldots) \text{ for } n \geqslant 1,$$

where 1 is the $(n+1)$st component. If we define $x^0 = (1,0,0,\ldots) = 1$, we have

$$\begin{aligned}(a_0, a_1, a_2,\ldots) &= (a_0,0,0,\ldots)+(0,a_1,0,\ldots)+(0,0,a_2,0,\ldots)+\cdots\\ &= (a_0,0,0,\ldots)(1,0,0,\ldots)+(a_1,0,0,\ldots)(0,1,0,\ldots)\\ &\quad +(a_2,0,0,\ldots)(0,0,1,0,\ldots)+\cdots\\ &= (a_0,0,0,\ldots)1+(a_1,0,0,\ldots)x+(a_2,0,0,\ldots)x^2+\cdots\\ &= a_0+a_1x+a_2x^2+\cdots+a_nx^n\\ &= f(x)\end{aligned}$$

for any sequence belonging to S. Thus the elements of the ring S are the polynomials $f(x) \in R[x]$, defined as infinite sequences with only finitely many components $a_i \neq 0$.

We emphasize again that the reason for this kind of definition of polynomials $f(x)$ over R is to clarify the relation between the elements of R and the new element x. The process of passing from R to the ring S of polynomials in x is called ring adjunction of x to R. The polynomial ring $R[x]$ can also be regarded as a subring of the ring of formal power series over R, which will be introduced in Chapter 8.

By considering the properties of the ring of integers and of polynomial rings over fields, one soon notices similarities. Actually, both types of rings belong to the same special class of Euclidean rings. A *Euclidean ring* is a commutative ring R with at least two elements, that has no zero divisors, and for which there exists a mapping ν from the set of nonzero elements of R to the set of nonnegative integers such that: (i) if $a, b \in R$ with $ab \neq 0$, then $\nu(ab) \geqslant \nu(a)$; (ii) for $a, b \in R$ with $b \neq 0$, there exist elements $q, r \in R$ with $a = qb + r$ and either $r = 0$ or $\nu(r) < \nu(b)$. The mapping ν is often called a (*Euclidean*) *valuation* on the ring R. We see at once that the integers form a Euclidean ring with "absolute value" as a valuation, and a polynomial ring over a field is a Euclidean ring with "degree" as a valuation. As a general result, one shows that any Euclidean ring is a principal ideal domain.

The property stated in Theorem 1.59 also holds in more general contexts and leads to the following definition. An integral domain in which a unique factorization theorem holds—that is, in which every nonunit $\neq 0$ can be expressed uniquely (up to units and the order of the factors) as a product of prime elements—is called a *unique factorization domain*. Thus, put succinctly, Theorem 1.59 says that $F[x]$ is a unique factorization domain. More generally, any principal ideal domain is also a unique

factorization domain. The Chinese remainder theorem for $F[x]$ (see Exercise 1.37) is a special case of a general result of this type shown in Lang [4, Ch. 2].

Good sources for facts about polynomials in one and several indeterminates are Rédei [10] and van der Waerden [2]. A more advanced monograph on polynomials is Lausch and Nöbauer [1].

4. In this section, Theorems 1.86 and 1.87 are the key theorems. In fact, one could say that Theorem 1.87 constitutes one of the most fundamental results in the theory of fields. For this result, due to Kronecker [8], assures us that given any nonconstant polynomial over any field, there must be an extension field in which the polynomial has a root. Moreover, the proof of the theorem does more than merely prove existence, as it also provides a method for constructing the required field.

One can classify the elements in an extension F of a field K according to their relation to K. If $\theta \in F$, then either $K(\theta)$ is isomorphic to $K(x)$, the field of rational functions over K (also called the quotient field of $K[x]$), or θ is a root of an irreducible polynomial g in $K[x]$ and $K(\theta)$ is isomorphic to $K[x]/(g)$, as stated in Theorem 1.86. In the first case, θ is called *transcendental* over K, in the second case θ is algebraic over K, as we already know. Extensions F of K that are not algebraic extensions are called *transcendental extensions* of K. Examples of transcendental elements exist in abundance. For instance, most real numbers (such as e, π, $2^{\sqrt{2}}, \ldots$) are transcendental over the field $\mathbf{Q}$ of rationals.

Splitting fields not only exist for a single nonconstant polynomial in $K[x]$, but for any collection of nonconstant polynomials over K. The splitting field over K of the collection of all nonconstant polynomials in $K[x]$ is called the *algebraic closure* $\overline{K}$ of K. It is an algebraic extension of K with the additional property that any nonconstant polynomial in $\overline{K}[x]$ splits in $\overline{K}$. For $K = \mathbf{Q}$ and $K = \mathbf{F}_p$, the algebraic closure $\overline{K}$ is an example of an algebraic extension that is not a finite extension of K.

The abstract theory of field extensions was developed in the fundamental paper of Steinitz [1]. Earlier investigations in this direction were carried out by Kneser [1], Kronecker [5], [8], and Weber [3].

EXERCISES

1.1. Prove that the identity element of a group is uniquely determined.

1.2. For a multiplicative group G, prove that a nonempty subset H of G is a subgroup of G if and only if $a, b \in H$ implies $ab^{-1} \in H$. If H is finite, then the condition can be replaced by: $a, b \in H$ implies $ab \in H$.

1.3. Let a be an element of finite order k in the multiplicative group G. Show that for $m \in \mathbf{Z}$ we have $a^m = e$ if and only if k divides m.

1.4. For $m \in \mathbb{N}$, Euler's function $\phi(m)$ is defined to be the number of integers k with $1 \leqslant k \leqslant m$ and $\gcd(k, m) = 1$. Show the following properties for $m, n, s \in \mathbb{N}$ and a prime p:

(a) $\phi(p^s) = p^s\left(1 - \frac{1}{p}\right)$;

(b) $\phi(mn) = \phi(m)\phi(n)$ if $\gcd(m, n) = 1$;

(c) $\phi(m) = m\left(1 - \frac{1}{p_1}\right) \cdots \left(1 - \frac{1}{p_r}\right)$, where $m = p_1^{e_1} \cdots p_r^{e_r}$ is the prime factor decomposition of m.

1.5. Calculate $\phi(490)$ and $\phi(768)$.

1.6. Use the class equation to show the following: if the order of a finite group is a prime power p^s, p prime, $s \geqslant 1$, then the order of its center is divisible by p.

1.7. Prove that in a ring R we have $(-a)(-b) = ab$ for all $a, b \in R$.

1.8. Prove that in a commutative ring R the formula

$$(a+b)^n = a^n + \binom{n}{1} a^{n-1}b + \cdots + \binom{n}{n-1} ab^{n-1} + b^n$$

holds for all $a, b \in R$ and $n \in \mathbb{N}$. (Binomial Theorem)

1.9. Let p be a prime number in $\mathbb{Z}$. For all integers a not divisible by p, show that p divides $a^{p-1} - 1$. (Fermat's Little Theorem)

1.10. Prove that for any prime p we have $(p-1)! \equiv -1 \bmod p$. (Wilson's Theorem)

1.11. Prove: if p is a prime, we have $\binom{p-1}{j} \equiv (-1)^j \bmod p$ for $0 \leqslant j \leqslant p-1, j \in \mathbb{Z}$.

1.12. A conjecture of Fermat stated that for all $n \geqslant 0$ the integer $2^{2^n} + 1$ is a prime. Euler found to the contrary that 641 divides $2^{32} + 1$. Confirm this by using congruences.

1.13. Prove: if $m_1, \ldots, m_k$ are positive integers that are pairwise relatively prime—that is, $\gcd(m_i, m_j) = 1$ for $1 \leqslant i < j \leqslant k$—then for any integers $a_1, \ldots, a_k$ the system of congruences $y \equiv a_i \bmod m_i$, $i = 1, 2, \ldots, k$, has a simultaneous solution y that is uniquely determined modulo $m = m_1 \cdots m_k$. (Chinese Remainder Theorem)

1.14. Solve the system of congruences $5x \equiv 20 \bmod 6$, $6x \equiv 6 \bmod 5$, $4x \equiv 5 \bmod 77$.

1.15. For a commutative ring R of prime characteristic p, show that

$$(a_1 + \cdots + a_s)^{p^n} = a_1^{p^n} + \cdots + a_s^{p^n}$$

for all $a_1, \ldots, a_s \in R$ and $n \in \mathbb{N}$.

1.16. Deduce from Exercise 1.11 that in a commutative ring R of prime characteristic p we have

$$(a-b)^{p-1} = \sum_{j=0}^{p-1} a^j b^{p-1-j} \quad \text{for all } a, b \in R.$$

1.17. Let F be a field and $f \in F[x]$. Prove that $\{g(f(x)) : g \in F[x]\}$ is equal to $F[x]$ if and only if $\deg(f) = 1$.

1.18. Show that $p^2(x) - xq^2(x) = xr^2(x)$ for $p, q, r \in \mathbb{R}[x]$ implies $p = q = r = 0$.

1.19. Show that if $f, g \in F[x]$, then the principal ideal (f) is contained in the principal ideal (g) if and only if g divides f.

1.20. Prove: if $f, g \in F[x]$ are relatively prime and not both constant, then there exist $a, b \in F[x]$ such that $af + bg = 1$ and $\deg(a) < \deg(g)$, $\deg(b) < \deg(f)$.

1.21. Let $f_1, \ldots, f_n \in F[x]$ with $\gcd(f_1, \ldots, f_n) = d$, so that $f_i = dg_i$ with $g_i \in F[x]$ for $1 \leqslant i \leqslant n$. Prove that $g_1, \ldots, g_n$ are relatively prime.

1.22. Prove that $\gcd(f_1, \ldots, f_n) = \gcd(\gcd(f_1, \ldots, f_{n-1}), f_n)$ for $n \geqslant 3$.

1.23. Prove: if $f, g, h \in F[x]$, f divides gh, and $\gcd(f, g) = 1$, then f divides h.

1.24. Use the Euclidean algorithm to compute $\gcd(f, g)$ for the polynomials f and g with coefficients in the indicated field F:

(a) $F = \mathbb{Q}$, $f(x) = x^7 + 2x^5 + 2x^2 - x + 2$, $g(x) = x^6 - 2x^5 - x^4 + x^2 + 2x + 3$

(b) $F = \mathbb{F}_2$, $f(x) = x^7 + 1$, $g(x) = x^5 + x^3 + x + 1$

(c) $F = \mathbb{F}_2$, $f(x) = x^5 + x + 1$, $g(x) = x^6 + x^5 + x^4 + 1$

(d) $F = \mathbb{F}_3$, $f(x) = x^8 + 2x^5 + x^3 + x^2 + 1$, $g(x) = 2x^6 + x^5 + 2x^3 + 2x^2 + 2$

1.25. Let $f_1, \ldots, f_n$ be nonzero polynomials in $F[x]$. By considering the intersection $(f_1) \cap \cdots \cap (f_n)$ of principal ideals, prove the existence and uniqueness of the monic polynomial $m \in F[x]$ with the properties attributed to the least common multiple of $f_1, \ldots, f_n$.

1.26. Prove (1.6).

1.27. If $f_1, \ldots, f_n \in F[x]$ are nonzero polynomials that are pairwise relatively prime, show that $\operatorname{lcm}(f_1, \ldots, f_n) = a^{-1} f_1 \cdots f_n$, where a is the leading coefficient of $f_1 \cdots f_n$.

1.28. Prove that $\operatorname{lcm}(f_1, \ldots, f_n) = \operatorname{lcm}(\operatorname{lcm}(f_1, \ldots, f_{n-1}), f_n)$ for $n \geqslant 3$.

1.29. Let $f_1, \ldots, f_n \in F[x]$ be nonzero polynomials. Write the canonical factorization of each f_i, $1 \leqslant i \leqslant n$, in the form

$$f_i = a_i \prod p^{e_i(p)},$$

where $a_i \in F$, the product is extended over all monic irreducible polynomials p in $F[x]$, the $e_i(p)$ are nonnegative integers, and for each i we have $e_i(p) > 0$ for only finitely many p. For each p set $m(p) = \min(e_1(p), \ldots, e_n(p))$ and $M(p) = \max(e_1(p), \ldots, e_n(p))$. Prove that

$$\gcd(f_1, \ldots, f_n) = \prod p^{m(p)},$$

$$\operatorname{lcm}(f_1, \ldots, f_n) = \prod p^{M(p)}.$$

1.30. Kronecker's method for finding divisors of degree $\leqslant s$ of a nonconstant polynomial $f \in \mathbb{Q}[x]$ proceeds as follows:
(1) By multiplying f by a constant, we can assume $f \in \mathbb{Z}[x]$.
(2) Choose distinct elements $a_0, \ldots, a_s \in \mathbb{Z}$ that are not roots of f and determine all divisors of $f(a_i)$ for each $i, 0 \leqslant i \leqslant s$.
(3) For each $(s+1)$-tuple $(b_0, \ldots, b_s)$ with b_i dividing $f(a_i)$ for $0 \leqslant i \leqslant s$, determine the polynomial $g \in \mathbb{Q}[x]$ with $\deg(g) \leqslant s$ and $g(a_i) = b_i$ for $0 \leqslant i \leqslant s$ (for instance, by the Lagrange interpolation formula).
(4) Decide which of these polynomials g in (3) are divisors of f.
If $\deg(f) = n \geqslant 1$ and s is taken to be the greatest integer $\leqslant n/2$, then f is irreducible in $\mathbb{Q}[x]$ in case the method only yields constant polynomials as divisors. Otherwise, Kronecker's method yields a nontrivial factorization. By applying the method again to the factors and repeating the process, one eventually gets the canonical factorization of f. Use this procedure to find the canonical factorization of

$$f(x) = \tfrac{1}{3}x^6 - \tfrac{5}{3}x^5 + 2x^4 - x^3 + 5x^2 - \tfrac{17}{3}x - 1 \in \mathbb{Q}[x].$$

1.31. Construct the addition and multiplication table for $\mathbb{F}_2[x]/(x^3 + x^2 + x)$. Determine whether or not this ring is a field.

1.32. Let $[x+1]$ be the residue class of $x+1$ in $\mathbb{F}_2[x]/(x^4+1)$. Find the residue classes comprising the principal ideal $([x+1])$ in $\mathbb{F}_2[x]/(x^4+1)$.

1.33. Let F be a field and $a, b, g \in F[x]$ with $g \neq 0$. Prove that the congruence $af \equiv b \bmod g$ has a solution $f \in F[x]$ if and only if $\gcd(a, g)$ divides b.

1.34. Solve the congruence $(x^2+1)f(x) \equiv 1 \bmod (x^3+1)$ in $\mathbb{F}_3[x]$, if possible.

1.35. Solve $(x^4 + x^3 + x^2 + 1)f(x) \equiv (x^2+1) \bmod (x^3+1)$ in $\mathbb{F}_2[x]$, if possible.

1.36. Prove that $R[x]/(x^4 + x^3 + x + 1)$ cannot be a field, no matter what the commutative ring R with identity is.

1.37. Prove: given a field F, nonzero polynomials $f_1, \ldots, f_k \in F[x]$ that are pairwise relatively prime, and arbitrary polynomials $g_1, \ldots, g_k \in F[x]$, then the simultaneous congruences $h \equiv g_i \bmod f_i$, $i = 1, 2, \ldots, k$, have a unique solution $h \in F[x]$ modulo $f = f_1 \cdots f_k$. (Chinese Remainder Theorem for $F[x]$)

1.38. Evaluate $f(3)$ for $f(x) = x^{214} + 3x^{152} + 2x^{47} + 2 \in \mathbb{F}_5[x]$.

1.39. Let p be a prime and $a_0, \ldots, a_n$ integers with p not dividing a_n. Show that $a_0 + a_1 y + \cdots + a_n y^n \equiv 0 \bmod p$ has at most n different solutions y modulo p.

1.40. If $p > 2$ is a prime, show that there are exactly two elements $a \in \mathbb{F}_p$ such that $a^2 = 1$.

1.41. Show: if $f \in \mathbb{Z}[x]$ and $f(0) \equiv f(1) \equiv 1 \bmod 2$, then f has no roots in $\mathbb{Z}$.

1.42. Let p be a prime and $f \in \mathbb{Z}[x]$. Show: $f(a) \equiv 0 \bmod p$ holds for all $a \in \mathbb{Z}$ if and only if $f(x) = (x^p - x)g(x) + ph(x)$ with $g, h \in \mathbb{Z}[x]$.

1.43. Let p be a prime integer and c an element of the field F. Show that $x^p - c$ is irreducible over F if and only if $x^p - c$ has no root in F.

1.44. Show that for a polynomial $f \in F[x]$ of positive degree the following conditions are equivalent:
(a) f is irreducible over F;
(b) the principal ideal (f) of $F[x]$ is a maximal ideal;
(c) the principal ideal (f) of $F[x]$ is a prime ideal.

1.45. Show the following properties of the derivative for polynomials in $F[x]$:
(a) $(f_1 + \cdots + f_m)' = f_1' + \cdots + f_m'$;
(b) $(fg)' = f'g + fg'$;
(c) $(f_1 \cdots f_m)' = \sum_{i=1}^{m} f_1 \cdots f_{i-1} f_i' f_{i+1} \cdots f_m$.

1.46. For $f \in F[x]$ and F of characteristic 0, prove that $f' = 0$ if and only if f is a constant polynomial. If F has prime characteristic p, prove that $f' = 0$ if and only if $f(x) = g(x^p)$ for some $g \in F[x]$.

1.47. Prove Theorem 1.68.

1.48. Prove that the nonzero polynomial $f \in F[x]$ has a multiple root (in some extension field of F) if and only if f and f' are not relatively prime.

1.49. Use the criterion in the previous exercise to determine whether the following polynomials have a multiple root:
(a) $f(x) = x^4 - 5x^3 + 6x^2 + 4x - 8 \in \mathbb{Q}[x]$
(b) $f(x) = x^6 + x^5 + x^4 + x^3 + 1 \in \mathbb{F}_2[x]$

1.50. The nth derivative $f^{(n)}$ of $f \in F[x]$ is defined recursively as follows: $f^{(0)} = f$, $f^{(n)} = (f^{(n-1)})'$ for $n \geqslant 1$. Prove that for $f, g \in F[x]$ we have

$$(fg)^{(n)} = \sum_{i=0}^{n} \binom{n}{i} f^{(n-i)} g^{(i)}.$$

1.51. Let F be a field and k a positive integer such that $k < p$ in case F has prime characteristic p. Prove: $b \in F$ is a root of $f \in F[x]$ of multiplicity k if and only if $f^{(i)}(b) = 0$ for $0 \leqslant i \leqslant k-1$ and $f^{(k)}(b) \neq 0$.

1.52. Show that the Lagrange interpolation formula can also be written in the form

$$f(x) = \sum_{i=0}^{n} b_i \big(g'(a_i)\big)^{-1} \frac{g(x)}{x - a_i} \quad \text{with } g(x) = \prod_{k=0}^{n} (x - a_k).$$

1.53. Determine a polynomial $f \in \mathbb{F}_5[x]$ with $f(0) = f(1) = f(4) = 1$ and $f(2) = f(3) = 3$.

1.54. Determine a polynomial $f \in \mathbb{Q}[x]$ of degree $\leqslant 3$ such that $f(-1) = -1$, $f(0) = 3$, $f(1) = 3$, and $f(2) = 5$.

1.55. Express $s_5(x_1, x_2, x_3, x_4) = x_1^5 + x_2^5 + x_3^5 + x_4^5 \in \mathbb{F}_3[x_1, x_2, x_3, x_4]$ in terms of the elementary symmetric polynomials $\sigma_1, \sigma_2, \sigma_3, \sigma_4$.

1.56. Prove that a subset K of a field F is a subfield if and only if the following conditions are satisfied:
(a) K contains at least two elements;
(b) if $a, b \in K$, then $a - b \in K$;
(c) if $a, b \in K$ and $b \neq 0$, then $ab^{-1} \in K$.

1.57. Prove that an extension L of the field K is a finite extension if and only if L can be obtained from K by adjoining finitely many algebraic elements over K.

1.58. Prove: if θ is algebraic over L and L is an algebraic extension of K, then θ is algebraic over K. Thus show that if F is an algebraic extension of L, then F is an algebraic extension of K.

1.59. Prove: if the degree $[L: K]$ is a prime, then the only fields F with $K \subseteq F \subseteq L$ are $F = K$ and $F = L$.

1.60. Construct the operation tables for the field $L = \mathbb{F}_3(\theta)$ in Example 1.88.

1.61. Show that $f(x) = x^4 + x + 1 \in \mathbb{F}_2[x]$ is irreducible over $\mathbb{F}_2$. Then construct the operation tables for the simple extension $\mathbb{F}_2(\theta)$, where θ is a root of f.

1.62. Calculate the discriminant $D(f)$ and decide whether or not f has a multiple root:
(a) $f(x) = 2x^3 - 3x^2 + x + 1 \in \mathbb{Q}[x]$
(b) $f(x) = 2x^4 + x^3 + x^2 + 2x + 2 \in \mathbb{F}_3[x]$

1.63. Deduce (1.9) from (1.11).

1.64. Prove that $f, g \in K[x]$ have a common root (in some extension field of K) if and only if f and g have a common divisor in $K[x]$ of positive degree.

1.65. Determine the common roots of the polynomials $x^7 - 2x^4 - x^3 + 2$ and $x^5 - 3x^4 - x + 3$ in $\mathbb{Q}[x]$.

1.66. Prove: if f and g are as in Definition 1.93, then $R(f, g) = (-1)^{mn} R(g, f)$.

1.67. Let $f, g \in K[x]$ be of positive degree and suppose that $f(x) = a_0(x - \alpha_1) \cdots (x - \alpha_n)$, $a_0 \neq 0$, and $g(x) = b_0(x - \beta_1) \cdots (x - \beta_m)$, $b_0 \neq 0$, in the splitting field of fg over K. Prove that

$$R(f, g) = (-1)^{mn} b_0^n \prod_{j=1}^{m} f(\beta_j) = a_0^m b_0^n \prod_{i=1}^{n} \prod_{j=1}^{m} (\alpha_i - \beta_j),$$

where n and m are also taken as the formal degrees of f and g, respectively.

1.68. Calculate the resultant $R(f, g)$ of the two given polynomials f and g (with the formal degree equal to the degree) and decide whether or not f and g have a common root:
(a) $f(x) = x^3 + x + 1$, $g(x) = 2x^5 + x^2 + 2 \in \mathbb{F}_3[x]$
(b) $f(x) = x^4 + x^3 + 1$, $g(x) = x^4 + x^2 + x + 1 \in \mathbb{F}_2[x]$

1.69. For $f \in K[x_1,\ldots,x_n]$, $n \geqslant 2$, an n-tuple $(\alpha_1,\ldots,\alpha_n)$ of elements α_i belonging to some extension L of K may be called a *zero* of f if $f(\alpha_1,\ldots,\alpha_n) = 0$. Now let $f, g \in K[x_1,\ldots,x_n]$ with x_n actually appearing in f and g. Then f and g can be regarded as polynomials $\bar{f}(x_n)$ and $\bar{g}(x_n)$ in $K[x_1,\ldots,x_{n-1}][x_n]$ of positive degree. Their resultant with respect to x_n (with formal degree = degree) is $R(\bar{f}, \bar{g}) = R_{x_n}(f, g)$, which is a polynomial in $x_1,\ldots,x_{n-1}$. Show that f and g have a common zero $(\alpha_1,\ldots,\alpha_{n-1}, \alpha_n)$ if and only if $(\alpha_1,\ldots,\alpha_{n-1})$ is a zero of $R(\bar{f}, \bar{g})$.

1.70. Using the result of the previous exercise, determine the common zeros of the polynomials $f(x, y) = x(y^2 - x)^2 + y^5$ and $g(x, y) = y^4 + y^3 - x^2$ in $\mathbb{Q}[x, y]$.

Chapter 2

Structure of Finite Fields

This chapter is of central importance since it contains various fundamental properties of finite fields and a description of methods for constructing finite fields.

The field of integers modulo a prime number is, of course, the most familiar example of a finite field, but many of its properties extend to arbitrary finite fields. The characterization of finite fields (see Section 1) shows that every finite field is of prime-power order and that, conversely, for every prime power there exists a finite field whose number of elements is exactly that prime power. Furthermore, finite fields with the same number of elements are isomorphic and may therefore be identified. The next two sections provide information on roots of irreducible polynomials, leading to an interpretation of finite fields as splitting fields of irreducible polynomials, and on traces, norms, and bases relative to field extensions.

Section 4 treats roots of unity from the viewpoint of general field theory, which will be needed occasionally in Section 6 as well as in Chapter 5. Section 5 presents different ways of representing the elements of a finite field. In Section 6 we give two proofs of the famous theorem of Wedderburn according to which every finite division ring is a field.

Many discussions in this chapter will be followed up, continued, and partly generalized in later chapters.

1. CHARACTERIZATION OF FINITE FIELDS

In the previous chapter we have already encountered a basic class of finite fields—that is, of fields with finitely many elements. For every prime p the residue class ring $\mathbb{Z}/(p)$ forms a finite field with p elements (see Theorem 1.38), which may be identified with the Galois field $\mathbb{F}_p$ of order p (see Definition 1.41). The fields $\mathbb{F}_p$ play an important role in general field theory since every field of characteristic p must contain an isomorphic copy of $\mathbb{F}_p$ by Theorem 1.78 and can thus be thought of as an extension of $\mathbb{F}_p$. This observation, together with the fact that every finite field has prime characteristic (see Corollary 1.45), is fundamental for the classification of finite fields. We first establish a simple necessary condition on the number of elements of a finite field.

2.1. Lemma. *Let F be a finite field containing a subfield K with q elements. Then F has q^m elements, where $m = [F: K]$.*

Proof. F is a vector space over K, and since F is finite, it is finite-dimensional as a vector space over K. If $[F: K] = m$, then F has a basis over K consisting of m elements, say $b_1, b_2, \ldots, b_m$. Thus every element of F can be uniquely represented in the form $a_1b_1 + a_2b_2 + \cdots + a_mb_m$, where $a_1, a_2, \ldots, a_m \in K$. Since each a_i can have q values, F has exactly q^m elements. □

2.2. Theorem. *Let F be a finite field. Then F has p^n elements, where the prime p is the characteristic of F and n is the degree of F over its prime subfield.*

Proof. Since F is finite, its characteristic is a prime p according to Corollary 1.45. Therefore the prime subfield K of F is isomorphic to $\mathbb{F}_p$ by Theorem 1.78 and thus contains p elements. The rest follows from Lemma 2.1. □

Starting from the prime fields $\mathbb{F}_p$, we can construct other finite fields by the process of root adjunction described in Chapter 1, Section 4. If $f \in \mathbb{F}_p[x]$ is an irreducible polynomial over $\mathbb{F}_p$ of degree n, then by adjoining a root of f to $\mathbb{F}_p$ we get a finite field with p^n elements. However, at this stage it is not clear whether for every positive integer n there exists an irreducible polynomial in $\mathbb{F}_p[x]$ of degree n. In order to establish that for every prime p and every $n \in \mathbb{N}$ there is a finite field with p^n elements, we use an approach suggested by the following results.

2.3. Lemma. *If F is a finite field with q elements, then every $a \in F$ satisfies $a^q = a$.*

Proof. The identity $a^q = a$ is trivial for $a = 0$. On the other hand, the nonzero elements of F form a group of order $q - 1$ under multiplication.

Thus $a^{q-1} = 1$ for all $a \in F$ with $a \neq 0$, and multiplication by a yields the desired result. □

2.4. *Lemma.* *If F is a finite field with q elements and K is a subfield of F, then the polynomial $x^q - x$ in $K[x]$ factors in $F[x]$ as*

$$x^q - x = \prod_{a \in F} (x - a)$$

and F is a splitting field of $x^q - x$ over K.

Proof. The polynomial $x^q - x$ of degree q has at most q roots in F. By Lemma 2.3 we know q such roots—namely, all the elements of F. Thus the given polynomial splits in F in the indicated manner, and it cannot split in any smaller field. □

We are now able to prove the main characterization theorem for finite fields, the leading idea being contained in Lemma 2.4.

2.5. *Theorem* (Existence and Uniqueness of Finite Fields). *For every prime p and every positive integer n there exists a finite field with p^n elements. Any finite field with $q = p^n$ elements is isomorphic to the splitting field of $x^q - x$ over $\mathbb{F}_p$.*

Proof. (*Existence*) For $q = p^n$ consider $x^q - x$ in $\mathbb{F}_p[x]$, and let F be its splitting field over $\mathbb{F}_p$. This polynomial has q distinct roots in F since its derivative is $qx^{q-1} - 1 = -1$ in $\mathbb{F}_p[x]$ and so can have no common root with $x^q - x$ (compare with Theorem 1.68). Let $S = \{a \in F: a^q - a = 0\}$. Then S is a subfield of F since: (i) S contains 0 and 1; (ii) $a, b \in S$ implies by Theorem 1.46 that $(a - b)^q = a^q - b^q = a - b$, and so $a - b \in S$; (iii) for $a, b \in S$ and $b \neq 0$ we have $(ab^{-1})^q = a^q b^{-q} = ab^{-1}$, and so $ab^{-1} \in S$. But, on the other hand, $x^q - x$ must split in S since S contains all its roots. Thus $F = S$, and since S has q elements, F is a finite field with q elements.

(*Uniqueness*) Let F be a finite field with $q = p^n$ elements. Then F has characteristic p by Theorem 2.2 and so contains $\mathbb{F}_p$ as a subfield. It follows from Lemma 2.4 that F is a splitting field of $x^q - x$ over $\mathbb{F}_p$. Thus the desired result is a consequence of the uniqueness (up to isomorphisms) of splitting fields, which was noted in Theorem 1.91. □

The uniqueness part of Theorem 2.5 provides the justification for speaking of *the* finite field (or *the* Galois field) with q elements, or of *the* finite field (or *the* Galois field) of order q. We shall denote this field by $\mathbb{F}_q$, where it is of course understood that q is a power of the prime characteristic p of $\mathbb{F}_q$.

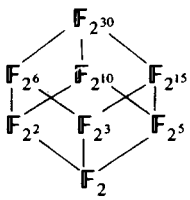
2.6. *Theorem* (Subfield Criterion). *Let $\mathbb{F}_q$ be the finite field with $q = p^n$ elements. Then every subfield of $\mathbb{F}_q$ has order p^m, where m is a positive divisor of n. Conversely, if m is a positive divisor of n, then there is exactly one subfield of $\mathbb{F}_q$ with p^m elements.*

Proof. It is clear that a subfield K of $\mathbb{F}_q$ has order p^m for some positive integer $m \leqslant n$. Lemma 2.1 shows that $q = p^n$ must be a power of p^m, and so m is necessarily a divisor of n.

Conversely, if m is a positive divisor of n, then $p^m - 1$ divides $p^n - 1$, and so $x^{p^m-1} - 1$ divides $x^{p^n-1} - 1$ in $\mathbb{F}_p[x]$. Consequently, $x^{p^m} - x$ divides $x^{p^n} - x = x^q - x$ in $\mathbb{F}_p[x]$. Thus, every root of $x^{p^m} - x$ is a root of $x^q - x$ and so belongs to $\mathbb{F}_q$. It follows that $\mathbb{F}_q$ must contain as a subfield a splitting field of $x^{p^m} - x$ over $\mathbb{F}_p$, and as we have seen in the proof of Theorem 2.5, such a splitting field has order p^m. If there were two distinct subfields of order p^m in $\mathbb{F}_q$, they would together contain more than p^m roots of $x^{p^m} - x$ in $\mathbb{F}_q$, an obvious contradiction. □

The proof of Theorem 2.6 shows that the unique subfield of $\mathbb{F}_{p^n}$ of order p^m, where m is a positive divisor of n, consists precisely of the roots of the polynomial $x^{p^m} - x \in \mathbb{F}_p[x]$ in $\mathbb{F}_{p^n}$.

2.7. Example. The subfields of the finite field $\mathbb{F}_{2^{30}}$ can be determined by listing all positive divisors of 30. The containment relations between these various subfields are displayed in the following diagram.



By Theorem 2.6, the containment relations are equivalent to divisibility relations among the positive divisors of 30. □

For a finite field $\mathbb{F}_q$ we denote by $\mathbb{F}_q^*$ the multiplicative group of nonzero elements of $\mathbb{F}_q$. The following result enunciates a useful property of this group.

2.8. Theorem. *For every finite field $\mathbb{F}_q$ the multiplicative group $\mathbb{F}_q^*$ of nonzero elements of $\mathbb{F}_q$ is cyclic.*

Proof. We may assume $q \geqslant 3$. Let $h = p_1^{r_1} p_2^{r_2} \cdots p_m^{r_m}$ be the prime factor decomposition of the order $h = q - 1$ of the group $\mathbb{F}_q^*$. For every i, $1 \leqslant i \leqslant m$, the polynomial $x^{h/p_i} - 1$ has at most h/p_i roots in $\mathbb{F}_q$. Since $h/p_i < h$, it follows that there are nonzero elements in $\mathbb{F}_q$ that are not roots of this polynomial. Let a_i be such an element and set $b_i = a_i^{h/p_i^{r_i}}$. We have $b_i^{p_i^{r_i}} = 1$, hence the order of b_i is a divisor of $p_i^{r_i}$ and is therefore of the form $p_i^{s_i}$ with $0 \leqslant s_i \leqslant r_i$. On the other hand,

$$b_i^{p_i^{r_i-1}} = a_i^{h/p_i} \neq 1,$$

and so the order of b_i is $p_i^{r_i}$. We claim that the element $b = b_1 b_2 \cdots b_m$ has order h. Suppose, on the contrary, that the order of b is a proper divisor of h

and is therefore a divisor of at least one of the m integers h/p_i, $1 \leqslant i \leqslant m$, say of h/p_1. Then we have

$$1 = b^{h/p_1} = b_1^{h/p_1} b_2^{h/p_1} \cdots b_m^{h/p_1}.$$

Now if $2 \leqslant i \leqslant m$, then $p_i^{r_i}$ divides h/p_1, and hence $b_i^{h/p_1} = 1$. Therefore $b_1^{h/p_1} = 1$. This implies that the order of b_1 must divide h/p_1, which is impossible since the order of b_1 is $p_1^{r_1}$. Thus, $\mathbb{F}_q^*$ is a cyclic group with generator b. □

2.9. Definition. A generator of the cyclic group $\mathbb{F}_q^*$ is called a *primitive element* of $\mathbb{F}_q$.

It follows from Theorem 1.15(v) that $\mathbb{F}_q$ contains $\phi(q-1)$ primitive elements, where ϕ is Euler's function. The existence of primitive elements can be used to show a result that implies, in particular, that every finite field can be thought of as a simple algebraic extension of its prime subfield.

2.10. Theorem. *Let $\mathbb{F}_q$ be a finite field and $\mathbb{F}_r$ a finite extension field. Then $\mathbb{F}_r$ is a simple algebraic extension of $\mathbb{F}_q$ and every primitive element of $\mathbb{F}_r$ can serve as a defining element of $\mathbb{F}_r$ over $\mathbb{F}_q$.*

Proof. Let ζ be a primitive element of $\mathbb{F}_r$. We clearly have $\mathbb{F}_q(\zeta) \subseteq \mathbb{F}_r$. On the other hand, $\mathbb{F}_q(\zeta)$ contains 0 and all powers of ζ, and so all elements of $\mathbb{F}_r$. Therefore $\mathbb{F}_r = \mathbb{F}_q(\zeta)$. □

2.11. Corollary. *For every finite field $\mathbb{F}_q$ and every positive integer n there exists an irreducible polynomial in $\mathbb{F}_q[x]$ of degree n.*

Proof. Let $\mathbb{F}_r$ be the extension field of $\mathbb{F}_q$ of order q^n, so that $[\mathbb{F}_r : \mathbb{F}_q] = n$. By Theorem 2.10 we have $\mathbb{F}_r = \mathbb{F}_q(\zeta)$ for some $\zeta \in \mathbb{F}_r$. Then the minimal polynomial of ζ over $\mathbb{F}_q$ is an irreducible polynomial in $\mathbb{F}_q[x]$ of degree n, according to Theorems 1.82(i) and 1.86(ii). □

2. ROOTS OF IRREDUCIBLE POLYNOMIALS

In this section we collect some information about the set of roots of an irreducible polynomial over a finite field.

2.12. Lemma. *Let $f \in \mathbb{F}_q[x]$ be an irreducible polynomial over a finite field $\mathbb{F}_q$ and let α be a root of f in an extension field of $\mathbb{F}_q$. Then for a polynomial $h \in \mathbb{F}_q[x]$ we have $h(\alpha) = 0$ if and only if f divides h.*

Proof. Let a be the leading coefficient of f and set $g(x) = a^{-1} f(x)$. Then g is a monic irreducible polynomial in $\mathbb{F}_q[x]$ with $g(\alpha) = 0$ and so it is the minimal polynomial of α over $\mathbb{F}_q$ in the sense of Definition 1.81. The rest follows from Theorem 1.82(ii). □

2.13. Lemma. *Let $f \in \mathbb{F}_q[x]$ be an irreducible polynomial over $\mathbb{F}_q$ of degree m. Then $f(x)$ divides $x^{q^n} - x$ if and only if m divides n.*

Proof. Suppose $f(x)$ divides $x^{q^n} - x$. Let α be a root of f in the splitting field of f over $\mathbb{F}_q$. Then $\alpha^{q^n} = \alpha$, so that $\alpha \in \mathbb{F}_{q^n}$. It follows that $\mathbb{F}_q(\alpha)$ is a subfield of $\mathbb{F}_{q^n}$. But since $[\mathbb{F}_q(\alpha):\mathbb{F}_q] = m$ and $[\mathbb{F}_{q^n}:\mathbb{F}_q] = n$, Theorem 1.84 shows that m divides n.

Conversely, if m divides n, then Theorem 2.6 implies that $\mathbb{F}_{q^n}$ contains $\mathbb{F}_{q^m}$ as a subfield. If α is a root of f in the splitting field of f over $\mathbb{F}_q$, then $[\mathbb{F}_q(\alpha):\mathbb{F}_q] = m$, and so $\mathbb{F}_q(\alpha) = \mathbb{F}_{q^m}$. Consequently, we have $\alpha \in \mathbb{F}_{q^n}$, hence $\alpha^{q^n} = \alpha$, and thus α is a root of $x^{q^n} - x \in \mathbb{F}_q[x]$. We infer then from Lemma 2.12 that $f(x)$ divides $x^{q^n} - x$. □

2.14. Theorem. *If f is an irreducible polynomial in $\mathbb{F}_q[x]$ of degree m, then f has a root α in $\mathbb{F}_{q^m}$. Furthermore, all the roots of f are simple and are given by the m distinct elements $\alpha, \alpha^q, \alpha^{q^2}, \ldots, \alpha^{q^{m-1}}$ of $\mathbb{F}_{q^m}$.*

Proof. Let α be a root of f in the splitting field of f over $\mathbb{F}_q$. Then $[\mathbb{F}_q(\alpha):\mathbb{F}_q] = m$, hence $\mathbb{F}_q(\alpha) = \mathbb{F}_{q^m}$, and in particular $\alpha \in \mathbb{F}_{q^m}$. Next we show that if $\beta \in \mathbb{F}_{q^m}$ is a root of f, then β^q is also a root of f. Write $f(x) = a_m x^m + \cdots + a_1 x + a_0$ with $a_i \in \mathbb{F}_q$ for $0 \leqslant i \leqslant m$. Then, using Lemma 2.3 and Theorem 1.46, we get

$$f(\beta^q) = a_m \beta^{qm} + \cdots + a_1 \beta^q + a_0 = a_m^q \beta^{qm} + \cdots + a_1^q \beta^q + a_0^q$$

$$= (a_m \beta^m + \cdots + a_1 \beta + a_0)^q = f(\beta)^q = 0.$$

Therefore, the elements $\alpha, \alpha^q, \alpha^{q^2}, \ldots, \alpha^{q^{m-1}}$ are roots of f. It remains to prove that these elements are distinct. Suppose, on the contrary, that $\alpha^{q^j} = \alpha^{q^k}$ for some integers j and k with $0 \leqslant j < k \leqslant m - 1$. By raising this identity to the power q^{m-k}, we get

$$\alpha^{q^{m-k+j}} = \alpha^{q^m} = \alpha.$$

It follows then from Lemma 2.12 that $f(x)$ divides $x^{q^{m-k+j}} - x$. By Lemma 2.13, this is only possible if m divides $m - k + j$. But we have $0 < m - k + j < m$, and so we arrive at a contradiction. □

2.15. Corollary. *Let f be an irreducible polynomial in $\mathbb{F}_q[x]$ of degree m. Then the splitting field of f over $\mathbb{F}_q$ is given by $\mathbb{F}_{q^m}$.*

Proof. Theorem 2.14 shows that f splits in $\mathbb{F}_{q^m}$. Furthermore, $\mathbb{F}_q(\alpha, \alpha^q, \alpha^{q^2}, \ldots, \alpha^{q^{m-1}}) = \mathbb{F}_q(\alpha) = \mathbb{F}_{q^m}$ for a root α of f in $\mathbb{F}_{q^m}$, where the second identity is taken from the proof of Theorem 2.14. □

2.16. Corollary. *Any two irreducible polynomials in $\mathbb{F}_q[x]$ of the same degree have isomorphic splitting fields.*

We introduce a convenient terminology for the elements appearing in Theorem 2.14, regardless of whether $\alpha \in \mathbb{F}_{q^m}$ is a root of an irreducible polynomial in $\mathbb{F}_q[x]$ of degree m or not.

2.17. Definition. Let $\mathbb{F}_{q^m}$ be an extension of $\mathbb{F}_q$ and let $\alpha \in \mathbb{F}_{q^m}$. Then the elements $\alpha, \alpha^q, \alpha^{q^2}, \ldots, \alpha^{q^{m-1}}$ are called the *conjugates* of α with respect to $\mathbb{F}_q$.

The conjugates of $\alpha \in \mathbb{F}_{q^m}$ with respect to $\mathbb{F}_q$ are distinct if and only if the minimal polynomial of α over $\mathbb{F}_q$ has degree m. Otherwise, the degree d of this minimal polynomial is a proper divisor of m, and then the conjugates of α with respect to $\mathbb{F}_q$ are the distinct elements $\alpha, \alpha^q, \ldots, \alpha^{q^{d-1}}$, each repeated m/d times.

2.18. Theorem. *The conjugates of $\alpha \in \mathbb{F}_q^*$ with respect to any subfield of $\mathbb{F}_q$ have the same order in the group $\mathbb{F}_q^*$.*

Proof. Since $\mathbb{F}_q^*$ is a cyclic group by Theorem 2.8, the result follows from Theorem 1.15(ii) and the fact that every power of the characteristic of $\mathbb{F}_q$ is relatively prime to the order $q-1$ of $\mathbb{F}_q^*$. □

2.19. Corollary. *If α is a primitive element of $\mathbb{F}_q$, then so are all its conjugates with respect to any subfield of $\mathbb{F}_q$.*

2.20. Example. Let $\alpha \in \mathbb{F}_{16}$ be a root of $f(x) = x^4 + x + 1 \in \mathbb{F}_2[x]$. Then the conjugates of α with respect to $\mathbb{F}_2$ are $\alpha, \alpha^2, \alpha^4 = \alpha + 1$, and $\alpha^8 = \alpha^2 + 1$, each of them being a primitive element of $\mathbb{F}_{16}$. The conjugates of α with respect to $\mathbb{F}_4$ are α and $\alpha^4 = \alpha + 1$. □

There is an intimate relationship between conjugate elements and certain automorphisms of a finite field. Let $\mathbb{F}_{q^m}$ be an extension of $\mathbb{F}_q$. By an *automorphism* σ of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$ we mean an automorphism of $\mathbb{F}_{q^m}$ that fixes the elements of $\mathbb{F}_q$. Thus, in detail, we require that σ be a one-to-one mapping from $\mathbb{F}_{q^m}$ onto itself with $\sigma(\alpha + \beta) = \sigma(\alpha) + \sigma(\beta)$ and $\sigma(\alpha\beta) = \sigma(\alpha)\sigma(\beta)$ for all $\alpha, \beta \in \mathbb{F}_{q^m}$ and $\sigma(a) = a$ for all $a \in \mathbb{F}_q$.

2.21. Theorem. *The distinct automorphisms of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$ are exactly the mappings $\sigma_0, \sigma_1, \ldots, \sigma_{m-1}$, defined by $\sigma_j(\alpha) = \alpha^{q^j}$ for $\alpha \in \mathbb{F}_{q^m}$ and $0 \leqslant j \leqslant m-1$.*

Proof. For each σ_j and all $\alpha, \beta \in \mathbb{F}_{q^m}$ we obviously have $\sigma_j(\alpha\beta) = \sigma_j(\alpha)\sigma_j(\beta)$, and also $\sigma_j(\alpha + \beta) = \sigma_j(\alpha) + \sigma_j(\beta)$ because of Theorem 1.46, so that σ_j is an endomorphism of $\mathbb{F}_{q^m}$. Furthermore, $\sigma_j(\alpha) = 0$ if and only if $\alpha = 0$, and so σ_j is one-to-one. Since $\mathbb{F}_{q^m}$ is a finite set, σ_j is an epimorphism and therefore an automorphism of $\mathbb{F}_{q^m}$. Moreover, we have $\sigma_j(a) = a$ for all $a \in \mathbb{F}_q$ by Lemma 2.3, and so each σ_j is an automorphism of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$.

The mappings $\sigma_0, \sigma_1, \ldots, \sigma_{m-1}$ are distinct since they attain distinct values for a primitive element of $\mathbb{F}_{q^m}$.

Now suppose that σ is an arbitrary automorphism of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$. Let β be a primitive element of $\mathbb{F}_{q^m}$ and let $f(x) = x^m + a_{m-1}x^{m-1} + \cdots + a_0 \in \mathbb{F}_q[x]$ be its minimal polynomial over $\mathbb{F}_q$. Then

$$\begin{aligned} 0 &= \sigma\left(\beta^m + a_{m-1}\beta^{m-1} + \cdots + a_0\right) \\ &= \sigma(\beta)^m + a_{m-1}\sigma(\beta)^{m-1} + \cdots + a_0, \end{aligned}$$

so that $\sigma(\beta)$ is a root of f in $\mathbb{F}_{q^m}$. It follows from Theorem 2.14 that $\sigma(\beta) = \beta^{q^j}$ for some j, $0 \leqslant j \leqslant m-1$. Since σ is a homomorphism, we get then $\sigma(\alpha) = \alpha^{q^j}$ for all $\alpha \in \mathbb{F}_{q^m}$. □

On the basis of Theorem 2.21 it is evident that the conjugates of $\alpha \in \mathbb{F}_{q^m}$ with respect to $\mathbb{F}_q$ are obtained by applying all automorphisms of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$ to the element α. The automorphisms of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$ form a group with the operation being the usual composition of mappings. The information provided in Theorem 2.21 shows that this group of automorphisms of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$ is a cyclic group of order m generated by σ_1.

3. TRACES, NORMS, AND BASES

In this section we adopt again the viewpoint of regarding a finite extension $F = \mathbb{F}_{q^m}$ of the finite field $K = \mathbb{F}_q$ as a vector space over K (compare with Chapter 1, Section 4). Then F has dimension m over K, and if $\{\alpha_1, \ldots, \alpha_m\}$ is a basis of F over K, each element $\alpha \in F$ can be uniquely represented in the form

$$\alpha = c_1\alpha_1 + \cdots + c_m\alpha_m \quad \text{with } c_j \in K \quad \text{for } 1 \leqslant j \leqslant m.$$

We introduce an important mapping from F to K which will turn out to be linear.

2.22. Definition. For $\alpha \in F = \mathbb{F}_{q^m}$ and $K = \mathbb{F}_q$, the *trace* $\mathrm{Tr}_{F/K}(\alpha)$ of α over K is defined by

$$\mathrm{Tr}_{F/K}(\alpha) = \alpha + \alpha^q + \cdots + \alpha^{q^{m-1}}.$$

If K is the prime subfield of F, then $\mathrm{Tr}_{F/K}(\alpha)$ is called the *absolute trace* of α and simply denoted by $\mathrm{Tr}_F(\alpha)$.

In other words, the trace of α over K is the sum of the conjugates of α with respect to K. Still another description of the trace may be obtained as follows. Let $f \in K[x]$ be the minimal polynomial of α over K; its degree d is a divisor of m. Then $g(x) = f(x)^{m/d} \in K[x]$ is called the *characteristic polynomial* of α over K. By Theorem 2.14, the roots of f in F are given by

$\alpha, \alpha^q, \ldots, \alpha^{q^{d-1}}$, and then a remark following Definition 2.17 implies that the roots of g in F are precisely the conjugates of α with respect to K. Hence

$$\begin{aligned} g(x) &= x^m + a_{m-1}x^{m-1} + \cdots + a_0 \\ &= (x-\alpha)(x-\alpha^q)\ldots(x-\alpha^{q^{m-1}}), \end{aligned} \tag{2.1}$$

and a comparison of coefficients shows that

$$\mathrm{Tr}_{F/K}(\alpha) = -a_{m-1}. \tag{2.2}$$

In particular, $\mathrm{Tr}_{F/K}(\alpha)$ is always an element of K.

2.23. *Theorem.* *Let $K = \mathbf{F}_q$ and $F = \mathbf{F}_{q^m}$. Then the trace function $\mathrm{Tr}_{F/K}$ satisfies the following properties*:

(i) $\mathrm{Tr}_{F/K}(\alpha+\beta) = \mathrm{Tr}_{F/K}(\alpha) + \mathrm{Tr}_{F/K}(\beta)$ *for all* $\alpha, \beta \in F$;
(ii) $\mathrm{Tr}_{F/K}(c\alpha) = c\,\mathrm{Tr}_{F/K}(\alpha)$ *for all* $c \in K$, $\alpha \in F$;
(iii) $\mathrm{Tr}_{F/K}$ *is a linear transformation from F onto K, where both F and K are viewed as vector spaces over K*;
(iv) $\mathrm{Tr}_{F/K}(a) = ma$ *for all* $a \in K$;
(v) $\mathrm{Tr}_{F/K}(\alpha^q) = \mathrm{Tr}_{F/K}(\alpha)$ *for all* $\alpha \in F$.

Proof.

(i) For $\alpha, \beta \in F$ we use Theorem 1.46 to get

$$\begin{aligned} \mathrm{Tr}_{F/K}(\alpha+\beta) &= \alpha+\beta+(\alpha+\beta)^q + \cdots + (\alpha+\beta)^{q^{m-1}} \\ &= \alpha+\beta+\alpha^q+\beta^q+\cdots+\alpha^{q^{m-1}}+\beta^{q^{m-1}} \\ &= \mathrm{Tr}_{F/K}(\alpha)+\mathrm{Tr}_{F/K}(\beta). \end{aligned}$$

(ii) For $c \in K$ we have $c^{q^j} = c$ for all $j \geqslant 0$ by Lemma 2.3. Therefore we obtain for $\alpha \in F$,

$$\begin{aligned} \mathrm{Tr}_{F/K}(c\alpha) &= c\alpha + c^q\alpha^q + \cdots + c^{q^{m-1}}\alpha^{q^{m-1}} \\ &= c\alpha + c\alpha^q + \cdots + c\alpha^{q^{m-1}} \\ &= c\,\mathrm{Tr}_{F/K}(\alpha). \end{aligned}$$

(iii) The properties (i) and (ii), together with the fact that $\mathrm{Tr}_{F/K}(\alpha) \in K$ for all $\alpha \in F$, show that $\mathrm{Tr}_{F/K}$ is a linear transformation from F into K. To prove that this mapping is onto, it suffices then to show the existence of an $\alpha \in F$ with $\mathrm{Tr}_{F/K}(\alpha) \neq 0$. Now $\mathrm{Tr}_{F/K}(\alpha) = 0$ if and only if α is a root of the polynomial $x^{q^{m-1}} + \cdots + x^q + x \in K[x]$ in F. But since this polynomial can have at most q^{m-1} roots in F and F has q^m elements, we are done.

(iv) This follows immediately from the definition of the trace function and Lemma 2.3.

(v) For $\alpha \in F$ we have $\alpha^{q^m} = \alpha$ by Lemma 2.3, and so $\mathrm{Tr}_{F/K}(\alpha^q) = \alpha^q + \alpha^{q^2} + \cdots + \alpha^{q^m} = \mathrm{Tr}_{F/K}(\alpha)$. □

The trace function $\mathrm{Tr}_{F/K}$ is not only in itself a linear transformation from F onto K, but serves for a description of all linear transformations from F into K (or, in an equivalent terminology, of all linear functionals on F) that has the advantage of being independent of a chosen basis.

2.24. Theorem. *Let F be a finite extension of the finite field K, both considered as vector spaces over K. Then the linear transformations from F into K are exactly the mappings L_β, $\beta \in F$, where $L_\beta(\alpha) = \mathrm{Tr}_{F/K}(\beta\alpha)$ for all $\alpha \in F$. Furthermore, we have $L_\beta \neq L_\gamma$ whenever β and γ are distinct elements of F.*

Proof. Each mapping L_β is a linear transformation from F into K by Theorem 2.23(iii). For $\beta, \gamma \in F$ with $\beta \neq \gamma$, we have $L_\beta(\alpha) - L_\gamma(\alpha) = \mathrm{Tr}_{F/K}(\beta\alpha) - \mathrm{Tr}_{F/K}(\gamma\alpha) = \mathrm{Tr}_{F/K}((\beta - \gamma)\alpha) \neq 0$ for suitable $\alpha \in F$ since $\mathrm{Tr}_{F/K}$ maps F onto K, and so the mappings L_β and L_γ are different. If $K = \mathbb{F}_q$ and $F = \mathbb{F}_{q^m}$, then the mappings L_β yield q^m different linear transformations from F into K. On the other hand, every linear transformation from F into K can be obtained by assigning arbitrary elements of K to the m elements of a given basis of F over K. Since this can be done in q^m different ways, the mappings L_β already exhaust all possible linear transformations from F into K. □

2.25. Theorem. *Let F be a finite extension of $K = \mathbb{F}_q$. Then for $\alpha \in F$ we have $\mathrm{Tr}_{F/K}(\alpha) = 0$ if and only if $\alpha = \beta^q - \beta$ for some $\beta \in F$.*

Proof. The sufficiency of the condition is obvious by Theorem 2.23(v). To prove the necessity, suppose $\alpha \in F = \mathbb{F}_{q^m}$ with $\mathrm{Tr}_{F/K}(\alpha) = 0$ and let β be a root of $x^q - x - \alpha$ in some extension field of F. Then $\beta^q - \beta = \alpha$ and

$$\begin{aligned} 0 = \mathrm{Tr}_{F/K}(\alpha) &= \alpha + \alpha^q + \cdots + \alpha^{q^{m-1}} \\ &= (\beta^q - \beta) + (\beta^q - \beta)^q + \cdots + (\beta^q - \beta)^{q^{m-1}} \\ &= (\beta^q - \beta) + \left(\beta^{q^2} - \beta^q\right) + \cdots + \left(\beta^{q^m} - \beta^{q^{m-1}}\right) \\ &= \beta^{q^m} - \beta, \end{aligned}$$

so that $\beta \in F$. □

In case a chain of extension fields is considered, the composition of trace functions proceeds according to a very simple rule.

2.26. Theorem (Transitivity of Trace). *Let K be a finite field, let F be a finite extension of K and E a finite extension of F. Then*

$$\mathrm{Tr}_{E/K}(\alpha) = \mathrm{Tr}_{F/K}\left(\mathrm{Tr}_{E/F}(\alpha)\right) \quad \text{for all } \alpha \in E.$$

Proof. Let $K = \mathbb{F}_q$, let $[F: K] = m$ and $[E: F] = n$, so that $[E: K] = mn$ by Theorem 1.84. Then for $\alpha \in E$ we have

$$\mathrm{Tr}_{F/K}(\mathrm{Tr}_{E/F}(\alpha)) = \sum_{i=0}^{m-1} \mathrm{Tr}_{E/F}(\alpha)^{q^i} = \sum_{i=0}^{m-1} \left(\sum_{j=0}^{n-1} \alpha^{q^{jm}} \right)^{q^i}$$

$$= \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \alpha^{q^{jm+i}} = \sum_{k=0}^{mn-1} \alpha^{q^k} = \mathrm{Tr}_{E/K}(\alpha). \qquad \square$$

Another interesting function from a finite field to a subfield is obtained by forming the product of the conjugates of an element of the field with respect to the subfield.

2.27. Definition. For $\alpha \in F = \mathbb{F}_{q^m}$ and $K = \mathbb{F}_q$, the *norm* $\mathrm{N}_{F/K}(\alpha)$ of α over K is defined by

$$\mathrm{N}_{F/K}(\alpha) = \alpha \cdot \alpha^q \cdot \cdots \cdot \alpha^{q^{m-1}} = \alpha^{(q^m-1)/(q-1)}.$$

By comparing the constant terms in (2.1), we see that $\mathrm{N}_{F/K}(\alpha)$ can be read off from the characteristic polynomial g of α over K—namely,

$$\mathrm{N}_{F/K}(\alpha) = (-1)^m a_0. \tag{2.3}$$

It follows, in particular, that $\mathrm{N}_{F/K}(\alpha)$ is always an element of K.

2.28. Theorem. *Let $K = \mathbb{F}_q$ and $F = \mathbb{F}_{q^m}$. Then the norm function $\mathrm{N}_{F/K}$ satisfies the following properties:*

(i) $\mathrm{N}_{F/K}(\alpha\beta) = N_{F/K}(\alpha)N_{F/K}(\beta)$ *for all* $\alpha, \beta \in F$;
(ii) $\mathrm{N}_{F/K}$ *maps F onto K and F^* onto K^**;
(iii) $\mathrm{N}_{F/K}(a) = a^m$ *for all* $a \in K$;
(iv) $\mathrm{N}_{F/K}(\alpha^q) = \mathrm{N}_{F/K}(\alpha)$ *for all* $\alpha \in F$.

Proof. (i) follows immediately from the definition of the norm. We have already noted that $\mathrm{N}_{F/K}$ maps F into K. Since $\mathrm{N}_{F/K}(\alpha) = 0$ if and only if $\alpha = 0$, $\mathrm{N}_{F/K}$ maps F^* into K^*. Property (i) shows that $\mathrm{N}_{F/K}$ is a group homomorphism between these multiplicative groups. Since the elements of the kernel of $\mathrm{N}_{F/K}$ are exactly the roots of the polynomial $x^{(q^m-1)/(q-1)} - 1 \in K[x]$ in F, the order d of the kernel satisfies $d \leqslant (q^m - 1)/(q - 1)$. By Theorem 1.23, the image of $\mathrm{N}_{F/K}$ has order $(q^m - 1)/d$, which is $\geqslant q - 1$. Therefore, $\mathrm{N}_{F/K}$ maps F^* onto K^* and so F onto K. Property (iii) follows from the definition of the norm and the fact that for $a \in K$ the conjugates of a with respect to K are all equal to a. Finally, we have $\mathrm{N}_{F/K}(\alpha^q) = \mathrm{N}_{F/K}(\alpha)^q = \mathrm{N}_{F/K}(\alpha)$ because of (i) and $\mathrm{N}_{F/K}(\alpha) \in K$, and so (iv) is shown. $\square$

2.29. *Theorem* (Transitivity of Norm). *Let K be a finite field, let F be a finite extension of K and E a finite extension of F. Then*

$$N_{E/K}(\alpha) = N_{F/K}\big(N_{E/F}(\alpha)\big) \quad \textit{for all } \alpha \in E.$$

Proof. With the same notation as in the proof of Theorem 2.26, we have for $\alpha \in E$,

$$\begin{aligned} N_{F/K}\big(N_{E/F}(\alpha)\big) &= N_{F/K}\big(\alpha^{(q^{mn}-1)/(q^m-1)}\big) \\ &= \big(\alpha^{(q^{mn}-1)/(q^m-1)}\big)^{(q^m-1)/(q-1)} \\ &= \alpha^{(q^{mn}-1)/(q-1)} = N_{E/K}(\alpha). \end{aligned}$$ □

If $\{\alpha_1,\dots,\alpha_m\}$ is a basis of the finite field F over a subfield K, the question arises as to the calculation of the coefficients $c_j(\alpha) \in K$, $1 \leqslant j \leqslant m$, in the unique representation

$$\alpha = c_1(\alpha)\alpha_1 + \cdots + c_m(\alpha)\alpha_m \tag{2.4}$$

of an element $\alpha \in F$. We note that $c_j\colon \alpha \mapsto c_j(\alpha)$ is a linear transformation from F into K, and thus, according to Theorem 2.24, there exists a $\beta_j \in F$ such that $c_j(\alpha) = \mathrm{Tr}_{F/K}(\beta_j\alpha)$ for all $\alpha \in F$. Putting $\alpha = \alpha_i$, $1 \leqslant i \leqslant m$, we see that $\mathrm{Tr}_{F/K}(\beta_j\alpha_i) = 0$ for $i \neq j$ and 1 for $i = j$. Furthermore, $\{\beta_1,\dots,\beta_m\}$ is again a basis of F over K, for if

$$d_1\beta_1 + \cdots + d_m\beta_m = 0 \quad \text{with } d_i \in K \quad \text{for } 1 \leqslant i \leqslant m,$$

then by multiplying by a fixed α_i and applying the trace function $\mathrm{Tr}_{F/K}$, one shows that $d_i = 0$.

2.30. Definition. Let K be a finite field and F a finite extension of K. Then two bases $\{\alpha_1,\dots,\alpha_m\}$ and $\{\beta_1,\dots,\beta_m\}$ of F over K are said to be *dual* (or *complementary*) bases if for $1 \leqslant i, j \leqslant m$ we have

$$\mathrm{Tr}_{F/K}(\alpha_i\beta_j) = \begin{cases} 0 & \text{for } i \neq j, \\ 1 & \text{for } i = j. \end{cases}$$

In the discussion above we have shown that for any basis $\{\alpha_1,\dots,\alpha_m\}$ of F over K there exists a dual basis $\{\beta_1,\dots,\beta_m\}$. The dual basis is, in fact, uniquely determined since its definition implies that the coefficients $c_j(\alpha)$, $1 \leqslant j \leqslant m$, in (2.4) are given by $c_j(\alpha) = \mathrm{Tr}_{F/K}(\beta_j\alpha)$ for all $\alpha \in F$, and by Theorem 2.24 the element $\beta_j \in F$ is uniquely determined by the linear transformation c_j.

2.31. Example. Let $\alpha \in \mathbb{F}_8$ be a root of the irreducible polynomial $x^3 + x^2 + 1$ in $\mathbb{F}_2[x]$. Then $\{\alpha, \alpha^2, 1 + \alpha + \alpha^2\}$ is a basis of $\mathbb{F}_8$ over $\mathbb{F}_2$. One checks easily that its uniquely determined dual basis is again $\{\alpha, \alpha^2, 1 + \alpha + \alpha^2\}$. Such a basis that is its own dual basis is called a *self-dual basis*. The element $\alpha^5 \in \mathbb{F}_8$ can be uniquely represented in the form $\alpha^5 = c_1\alpha + c_2\alpha^2 +$

$c_3(1+\alpha+\alpha^2)$ with $c_1, c_2, c_3 \in \mathbf{F}_2$, and the coefficients are given by

$$c_1 = \mathrm{Tr}_{\mathbf{F}_8}(\alpha \cdot \alpha^5) = 0,$$

$$c_2 = \mathrm{Tr}_{\mathbf{F}_8}(\alpha^2 \cdot \alpha^5) = 1,$$

$$c_3 = \mathrm{Tr}_{\mathbf{F}_8}((1+\alpha+\alpha^2)\alpha^5) = 1,$$

so that $\alpha^5 = \alpha^2 + (1+\alpha+\alpha^2)$. □

The number of distinct bases of F over K is rather large (see Exercise 2.37), but there are two special types of bases of particular importance. The first is a *polynomial basis* $\{1, \alpha, \alpha^2, \ldots, \alpha^{m-1}\}$, made up of the powers of a defining element α of F over K. The element α is often taken to be a primitive element of F (compare with Theorem 2.10). Another type of basis is a normal basis defined by a suitable element of F.

2.32. Definition. Let $K = \mathbf{F}_q$ and $F = \mathbf{F}_{q^m}$. Then a basis of F over K of the form $\{\alpha, \alpha^q, \ldots, \alpha^{q^{m-1}}\}$, consisting of a suitable element $\alpha \in F$ and its conjugates with respect to K, is called a *normal basis* of F over K.

The basis $\{\alpha, \alpha^2, 1+\alpha+\alpha^2\}$ of $\mathbf{F}_8$ over $\mathbf{F}_2$ discussed in Example 2.31 is a normal basis of $\mathbf{F}_8$ over $\mathbf{F}_2$ since $1+\alpha+\alpha^2 = \alpha^4$. We shall show that a normal basis exists in the general case as well. The proof depends on two lemmas, one on a kind of linear independence property of certain group homomorphisms and one on linear operators.

2.33. Lemma (Artin Lemma). *Let $\psi_1, \ldots, \psi_m$ be distinct homomorphisms from a group G into the multiplicative group F^* of an arbitrary field F, and let $a_1, \ldots, a_m$ be elements of F that are not all 0. Then for some $g \in G$ we have*

$$a_1\psi_1(g) + \cdots + a_m\psi_m(g) \neq 0.$$

Proof. We proceed by induction on m. The case $m = 1$ being trivial, we assume that $m > 1$ and that the statement is shown for any $m-1$ distinct homomorphisms. Now take $\psi_1, \ldots, \psi_m$ and $a_1, \ldots, a_m$ as in the lemma. If $a_1 = 0$, the induction hypothesis immediately yields the desired result. Thus let $a_1 \neq 0$. Suppose we had

$$a_1\psi_1(g) + \cdots + a_m\psi_m(g) = 0 \quad \text{for all } g \in G. \tag{2.5}$$

Since $\psi_1 \neq \psi_m$, there exists $h \in G$ with $\psi_1(h) \neq \psi_m(h)$. Then, replacing g by hg in (2.5), we get

$$a_1\psi_1(h)\psi_1(g) + \cdots + a_m\psi_m(h)\psi_m(g) = 0 \quad \text{for all } g \in G.$$

After multiplication by $\psi_m(h)^{-1}$ we obtain

$$b_1\psi_1(g) + \cdots + b_{m-1}\psi_{m-1}(g) + a_m\psi_m(g) = 0 \quad \text{for all } g \in G,$$

where $b_i = a_i\psi_i(h)\psi_m(h)^{-1}$ for $1 \leqslant i \leqslant m-1$. By subtracting this identity

from (2.5), we arrive at

$$c_1\psi_1(g)+\cdots+c_{m-1}\psi_{m-1}(g)=0 \quad \text{for all } g\in G,$$

where $c_i=a_i-b_i$ for $1\leqslant i\leqslant m-1$. But $c_1=a_1-a_1\psi_1(h)\psi_m(h)^{-1}\neq 0$, and we have a contradiction to the induction hypothesis. □

We recall a few concepts and facts from linear algebra. If T is a linear operator on the finite-dimensional vector space V over the (arbitrary) field K, then a polynomial $f(x)=a_nx^n+\cdots+a_1x+a_0\in K[x]$ is said to *annihilate* T if $a_nT^n+\cdots+a_1T+a_0I=0$, where I is the identity operator and 0 the zero operator on V. The uniquely determined monic polynomial of least positive degree with this property is called the *minimal polynomial* for T. It divides any other polynomial in $K[x]$ annihilating T. In particular, the minimal polynomial for T divides the *characteristic polynomial* $g(x)$ for T (Cayley-Hamilton theorem), which is given by $g(x)=\det(xI-T)$ and is a monic polynomial of degree equal to the dimension of V. A vector $\alpha\in V$ is called a *cyclic vector* for T if the vectors $T^k\alpha$, $k=0,1,\ldots$, span V. The following is a standard result from linear algebra.

2.34. *Lemma.* *Let T be a linear operator on the finite-dimensional vector space V. Then T has a cyclic vector if and only if the characteristic and minimal polynomials for T are identical.*

2.35. *Theorem* (Normal Basis Theorem). *For any finite field K and any finite extension F of K, there exists a normal basis of F over K.*

Proof. Let $K=\mathbf{F}_q$ and $F=\mathbf{F}_{q^m}$ with $m\geqslant 2$. From Theorem 2.21 and the remarks following it, we know that the distinct automorphisms of F over K are given by $\varepsilon,\sigma,\sigma^2,\ldots,\sigma^{m-1}$, where ε is the identity mapping on F, $\sigma(\alpha)=\alpha^q$ for $\alpha\in F$, and a power σ^j refers to the j-fold composition of σ with itself. Because of $\sigma(\alpha+\beta)=\sigma(\alpha)+\sigma(\beta)$ and $\sigma(c\alpha)=\sigma(c)\sigma(\alpha)=c\sigma(\alpha)$ for $\alpha,\beta\in F$ and $c\in K$, the mapping σ may also be considered as a linear operator on the vector space F over K. Since $\sigma^m=\varepsilon$, the polynomial $x^m-1\in K[x]$ annihilates σ. Lemma 2.33, applied to $\varepsilon,\sigma,\sigma^2,\ldots,\sigma^{m-1}$ viewed as endomorphisms of F^*, shows that no nonzero polynomial in $K[x]$ of degree less than m annihilates σ. Consequently, x^m-1 is the minimal polynomial for the linear operator σ. Since the characteristic polynomial for σ is a monic polynomial of degree m that is divisible by the minimal polynomial for σ, it follows that the characteristic polynomial for σ is also given by x^m-1. Lemma 2.34 implies then the existence of an element $\alpha\in F$ such that $\alpha,\sigma(\alpha),\sigma^2(\alpha),\ldots$ span F. By dropping repeated elements, we see that $\alpha,\sigma(\alpha),\sigma^2(\alpha),\ldots,\sigma^{m-1}(\alpha)$ span F and thus form a basis of F over K. Since this basis consists of α and its conjugates with respect to K, it is a normal basis of F over K. □

An alternative proof of the normal basis theorem will be provided in Chapter 3, Section 4, by using so-called linearized polynomials.

We introduce an expression that allows us to decide whether a given set of elements forms a basis of an extension field.

2.36. Definition. Let K be a finite field and F an extension of K of degree m over K. Then the *discriminant* $\Delta_{F/K}(\alpha_1,\ldots,\alpha_m)$ of the elements $\alpha_1,\ldots,\alpha_m \in F$ is defined by the determinant of order m given by

$$\Delta_{F/K}(\alpha_1,\ldots,\alpha_m) = \begin{vmatrix} \mathrm{Tr}_{F/K}(\alpha_1\alpha_1) & \mathrm{Tr}_{F/K}(\alpha_1\alpha_2) & \cdots & \mathrm{Tr}_{F/K}(\alpha_1\alpha_m) \\ \mathrm{Tr}_{F/K}(\alpha_2\alpha_1) & \mathrm{Tr}_{F/K}(\alpha_2\alpha_2) & \cdots & \mathrm{Tr}_{F/K}(\alpha_2\alpha_m) \\ \vdots & \vdots & & \vdots \\ \mathrm{Tr}_{F/K}(\alpha_m\alpha_1) & \mathrm{Tr}_{F/K}(\alpha_m\alpha_2) & \cdots & \mathrm{Tr}_{F/K}(\alpha_m\alpha_m) \end{vmatrix}.$$

It follows from the definition that $\Delta_{F/K}(\alpha_1,\ldots,\alpha_m)$ is always an element of K. The following simple characterization of bases can now be given.

2.37. Theorem. *Let K be a finite field, F an extension of K of degree m over K, and $\alpha_1,\ldots,\alpha_m \in F$. Then $\{\alpha_1,\ldots,\alpha_m\}$ is a basis of F over K if and only if $\Delta_{F/K}(\alpha_1,\ldots,\alpha_m) \neq 0$.*

Proof. Let $\{\alpha_1,\ldots,\alpha_m\}$ be a basis of F over K. We prove that $\Delta_{F/K}(\alpha_1,\ldots,\alpha_m) \neq 0$ by showing that the row vectors of the determinant defining $\Delta_{F/K}(\alpha_1,\ldots,\alpha_m)$ are linearly independent. For suppose that

$$c_1\mathrm{Tr}_{F/K}(\alpha_1\alpha_j) + \cdots + c_m\mathrm{Tr}_{F/K}(\alpha_m\alpha_j) = 0 \quad \text{for } 1 \leqslant j \leqslant m,$$

where $c_1,\ldots,c_m \in K$. Then with $\beta = c_1\alpha_1 + \cdots + c_m\alpha_m$ we get $\mathrm{Tr}_{F/K}(\beta\alpha_j) = 0$ for $1 \leqslant j \leqslant m$, and since $\alpha_1,\ldots,\alpha_m$ span F, it follows that $\mathrm{Tr}_{F/K}(\beta\alpha) = 0$ for all $\alpha \in F$. However, this is only possible if $\beta = 0$, and then $c_1\alpha_1 + \cdots + c_m\alpha_m = 0$ implies $c_1 = \cdots = c_m = 0$.

Conversely, suppose that $\Delta_{F/K}(\alpha_1,\ldots,\alpha_m) \neq 0$ and $c_1\alpha_1 + \cdots + c_m\alpha_m = 0$ for some $c_1,\ldots,c_m \in K$. Then

$$c_1\alpha_1\alpha_j + \cdots + c_m\alpha_m\alpha_j = 0 \quad \text{for } 1 \leqslant j \leqslant m,$$

and by applying the trace function we get

$$c_1\mathrm{Tr}_{F/K}(\alpha_1\alpha_j) + \cdots + c_m\mathrm{Tr}_{F/K}(\alpha_m\alpha_j) = 0 \quad \text{for } 1 \leqslant j \leqslant m.$$

But since the row vectors of the determinant defining $\Delta_{F/K}(\alpha_1,\ldots,\alpha_m)$ are linearly independent, it follows that $c_1 = \cdots = c_m = 0$. Therefore, $\alpha_1,\ldots,\alpha_m$ are linearly independent over K. □

There is another determinant of order m that serves the same purpose as the discriminant $\Delta_{F/K}(\alpha_1,\ldots,\alpha_m)$. The entries of this determinant are, however, elements of the extension field F. For $\alpha_1,\ldots,\alpha_m \in F$, let