Introduction to finite fields and their applications

RUDOLF LIDL
University of Tasmania, Launceston, Australia

HARALD NIEDERREITER
Austrian Academy of Sciences, Vienna, Austria

Revised edition

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

© Cambridge University Press 1986, 1994

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 1986
Revised edition 1994
Reprinted 1997, 2000

Printed in the United Kingdom at the University Press, Cambridge

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

Library of Congress Cataloguing in Publication data available

ISBN 0 521 46094 8 hardback

Contents

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 $\mathbb{Z}_n$.

$\mathbb{Z}_n$ 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 $\mathbb{Z}_6$ 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 $\mathbb{Z}_6$ 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 $\mathbb{Z}$ 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 $\mathbb{Z}$. 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 $G = \mathbb{Z}_{12}$ and let H be the subgroup $\langle [0], [3], [6], [9] \rangle$. 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 $\mathbb{Z}$ of the integers onto the group $\mathbb{Z}_n$ of the integers modulo n, defined by $f(a) = [a]$. Then

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

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):

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

$\cdot$	$[0]$	$[1]$	$[2]$
$[0]$	$[0]$	$[0]$	$[0]$
$[1]$	$[0]$	$[1]$	$[2]$
$[2]$	$[0]$	$[2]$	$[1]$

$\square$

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 $\cdot$ 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