Herstein

Topics in Algebra

Second Edition

WILEY

I. N. Herstein

Topics in Algebra

Second Edition

i. n. herstein
University of Chicago

TOPICS IN ALGEBRA

2nd edition

JOHN WILEY & SONS
New York • Chichester • Brisbane • Toronto • Singapore

To Marianne

Library of Congress Catalog Card Number: 74-82577
Printed in the United States of America.

20 19 18 17 16 15

Preface to the Second Edition

I approached revising *Topics in Algebra* with a certain amount of trepidation. On the whole, I was satisfied with the first edition and did not want to tamper with it. However, there were certain changes I felt should be made, changes which would not affect the general style or content, but which would make the book a little more complete. I hope that I have achieved this objective in the present version.

For the most part, the major changes take place in the chapter on group theory. When the first edition was written it was fairly uncommon for a student learning abstract algebra to have had any previous exposure to linear algebra. Nowadays quite the opposite is true; many students, perhaps even a majority, have learned something about 2×2 matrices at this stage. Thus I felt free here to draw on 2×2 matrices for examples and problems. These parts, which depend on some knowledge of linear algebra, are indicated with a #.

In the chapter on groups I have largely expanded one section, that on Sylow's theorem, and added two others, one on direct products and one on the structure of finite abelian groups.

In the previous treatment of Sylow's theorem, only the existence of a Sylow subgroup was shown. This was done following the proof of Wielandt. The conjugacy of the Sylow subgroups and their number were developed in a series of exercises, but not in the text proper. Now all the parts of Sylow's theorem are done in the text material.

In addition to the proof previously given for the existence, two other proofs of existence are carried out. One could accuse me of overkill at this point, probably rightfully so. The fact of the matter is that Sylow's theorem is important, that each proof illustrates a different aspect of group theory and, above all, that I love Sylow's theorem. The proof of the conjugacy and number of Sylow subgroups exploits double cosets. A by-product of this development is that a means is given for finding Sylow subgroups in a large set of symmetric groups.

For some mysterious reason known only to myself, I had omitted direct products in the first edition. Why is beyond me. The material is easy, straightforward, and important. This lacuna is now filled in the section treating direct products. With this in hand, I go on in the next section to prove the decomposition of a finite abelian group as a direct product of cyclic groups and also prove the uniqueness of the invariants associated with this decomposition. In point of fact, this decomposition was already in the first edition, at the end of the chapter on vector spaces, as a consequence of the structure of finitely generated modules over Euclidean rings. However, the case of a finite group is of great importance by itself; the section on finite abelian groups underlines this importance. Its presence in the chapter on groups, an early chapter, makes it more likely that it will be taught.

One other entire section has been added at the end of the chapter on field theory. I felt that the student should see an explicit polynomial over an explicit field whose Galois group was the symmetric group of degree 5, hence one whose roots could not be expressed by radicals. In order to do so, a theorem is first proved which gives a criterion that an irreducible polynomial of degree p, p a prime, over the rational field have S_p as its Galois group. As an application of this criterion, an irreducible polynomial of degree 5 is given, over the rational field, whose Galois group is the symmetric group of degree 5.

There are several other additions. More than 150 new problems are to be found here. They are of varying degrees of difficulty. Many are routine and computational, many are very difficult. Furthermore, some interpolatory remarks are made about problems that have given readers a great deal of difficulty. Some paragraphs have been inserted, others rewritten, at places where the writing had previously been obscure or too terse.

Above I have described what I have added. What gave me greater difficulty about the revision was, perhaps, that which I have not added. I debated for a long time with myself whether or not to add a chapter on category theory and some elementary functors, whether or not to enlarge the material on modules substantially. After a great deal of thought and soul-searching, I decided not to do so. The book, as stands, has a certain concreteness about it with which this new material would not blend. It could be made to blend, but this would require a complete reworking of the material

of the book and a complete change in its philosophy—something I did not want to do. A mere addition of this new material, as an adjunct with no applications and no discernible goals, would have violated my guiding principle that all matters discussed should lead to some clearly defined objectives, to some highlight, to some exciting theorems. Thus I decided to omit the additional topics.

Many people wrote me about the first edition pointing out typographical mistakes or making suggestions on how to improve the book. I should like to take this opportunity to thank them for their help and kindness.

Preface to the First Edition

The idea to write this book, and more important the desire to do so, is a direct outgrowth of a course I gave in the academic year 1959–1960 at Cornell University. The class taking this course consisted, in large part, of the most gifted sophomores in mathematics at Cornell. It was my desire to experiment by presenting to them material a little beyond that which is usually taught in algebra at the junior-senior level.

I have aimed this book to be, both in content and degree of sophistication, about halfway between two great classics, *A Survey of Modern Algebra*, by Birkhoff and MacLane, and *Modern Algebra*, by Van der Waerden.

The last few years have seen marked changes in the instruction given in mathematics at the American universities. This change is most notable at the upper undergraduate and beginning graduate levels. Topics that a few years ago were considered proper subject matter for semiadvanced graduate courses in algebra have filtered down to, and are being taught in, the very first course in abstract algebra. Convinced that this filtration will continue and will become intensified in the next few years, I have put into this book, which is designed to be used as the student's first introduction to algebra, material which hitherto has been considered a little advanced for that stage of the game.

There is always a great danger when treating abstract ideas to introduce them too suddenly and without a sufficient base of examples to render them credible or natural. In order to try to mitigate this, I have tried to motivate the concepts beforehand and to illustrate them in concrete situations. One of the most telling proofs of the worth of an abstract

concept is what it, and the results about it, tells us in familiar situations. In almost every chapter an attempt is made to bring out the significance of the general results by applying them to particular problems. For instance, in the chapter on rings, the two-square theorem of Fermat is exhibited as a direct consequence of the theory developed for Euclidean rings.

The subject matter chosen for discussion has been picked not only because it has become standard to present it at this level or because it is important in the whole general development but also with an eye to this "concreteness." For this reason I chose to omit the Jordan-Hölder theorem, which certainly could have easily been included in the results derived about groups. However, to appreciate this result for its own sake requires a great deal of hindsight and to see it used effectively would require too great a digression. True, one could develop the whole theory of dimension of a vector space as one of its corollaries, but, for the first time around, this seems like a much too fancy and unnatural approach to something so basic and down-to-earth. Likewise, there is no mention of tensor products or related constructions. There is so much time and opportunity to become abstract; why rush it at the beginning?

A word about the problems. There are a great number of them. It would be an extraordinary student indeed who could solve them all. Some are present merely to complete proofs in the text material, others to illustrate and to give practice in the results obtained. Many are introduced not so much to be solved as to be tackled. The value of a problem is not so much in coming up with the answer as in the ideas and attempted ideas it forces on the would-be solver. Others are included in anticipation of material to be developed later, the hope and rationale for this being both to lay the groundwork for the subsequent theory and also to make more natural ideas, definitions, and arguments as they are introduced. Several problems appear more than once. Problems that for some reason or other seem difficult to me are often starred (sometimes with two stars). However, even here there will be no agreement among mathematicians; many will feel that some unstarred problems should be starred and vice versa.

Naturally, I am indebted to many people for suggestions, comments and criticisms. To mention just a few of these: Charles Curtis, Marshall Hall, Nathan Jacobson, Arthur Mattuck, and Maxwell Rosenlicht. I owe a great deal to Daniel Gorenstein and Irving Kaplansky for the numerous conversations we have had about the book, its material and its approach. Above all, I thank George Seligman for the many incisive suggestions and remarks that he has made about the presentation both as to its style and to its content. I am also grateful to Francis McNary of the staff of Ginn and Company for his help and cooperation. Finally, I should like to express my thanks to the John Simon Guggenheim Memorial Foundation; this book was in part written with their support while the author was in Rome as a Guggenheim Fellow.

Contents

1

Preliminary Notions

One of the amazing features of twentieth century mathematics has been its recognition of the power of the abstract approach. This has given rise to a large body of new results and problems and has, in fact, led us to open up whole new areas of mathematics whose very existence had not even been suspected.

In the wake of these developments has come not only a new mathematics but a fresh outlook, and along with this, simple new proofs of difficult classical results. The isolation of a problem into its basic essentials has often revealed for us the proper setting, in the whole scheme of things, of results considered to have been special and apart and has shown us interrelations between areas previously thought to have been unconnected.

The algebra which has evolved as an outgrowth of all this is not only a subject with an independent life and vigor—it is one of the important current research areas in mathematics—but it also serves as the unifying thread which interlaces almost all of mathematics—geometry, number theory, analysis, topology, and even applied mathematics.

This book is intended as an introduction to that part of mathematics that today goes by the name of abstract algebra. The term "abstract" is a highly subjective one; what is abstract to one person is very often concrete and down-to-earth to another, and vice versa. In relation to the current research activity in algebra, it could be described as "not too abstract"; from the point of view of someone schooled in the

calculus and who is seeing the present material for the first time, it may very well be described as "quite abstract."

Be that as it may, we shall concern ourselves with the introduction and development of some of the important algebraic systems—groups, rings, vector spaces, fields. An algebraic system can be described as a set of objects together with some operations for combining them.

Prior to studying sets restricted in any way whatever—for instance, with operations—it will be necessary to consider sets in general and some notions about them. At the other end of the spectrum, we shall need some information about the particular set, the set of integers. It is the purpose of this chapter to discuss these and to derive some results about them which we can call upon, as the occasions arise, later in the book.

1.1 Set Theory

We shall not attempt a formal definition of a set nor shall we try to lay the groundwork for an axiomatic theory of sets. Instead we shall take the operational and intuitive approach that a set is some given collection of objects. In most of our applications we shall be dealing with rather specific things, and the nebulous notion of a set, in these, will emerge as something quite recognizable. For those whose tastes run more to the formal and abstract side, we can consider a set as a primitive notion which one does not define.

A few remarks about notation and terminology. Given a set S we shall use the notation throughout $a \in S$ *to read* "*a is an element of S.*" In the same vein, $a \notin S$ will read "a is *not* an element of S." The set A will be said to be a *subset* of the set S if every element in A is an element of S, that is, if $a \in A$ implies $a \in S$. We shall write this as $A \subset S$ (or, sometimes, as $S \supset A$), which may be read "A is contained in S" (or, S contains A). This notation is not meant to preclude the possibility that $A = S$. By the way, what is meant by the equality of two sets? For us this will always mean that they contain the same elements, that is, every element which is in one is in the other, and vice versa. In terms of the symbol for the containing relation, the two sets A and B are equal, written $A = B$, if both $A \subset B$ and $B \subset A$. The standard device for proving the equality of two sets, something we shall be required to do often, is to demonstrate that the two opposite containing relations hold for them. A subset A of S will be called a *proper* subset of S if $A \subset S$ but $A \neq S$ (A is not equal to S).

The *null set* is the set having no elements; it is a subset of every set. We shall often describe that a set S is the null set by saying it is *empty*.

One final, purely notational remark: Given a set S we shall constantly use the notation $A = \{a \in S \mid P(a)\}$ to read "A is the set of all elements in S for which the property P holds." For instance, if S is the set of integers

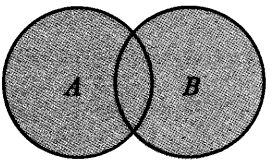
and if A is the subset of positive integers, then we can describe A as $A = \{a \in S \mid a > 0\}$. Another example of this: If S is the set consisting of the objects (1), (2), . . . , (10), then the subset A consisting of (1), (4), (7), (10) could be described by $A = \{(i) \in S \mid i = 3n + 1,\ n = 0, 1, 2, 3\}$.

Given two sets we can combine them to form new sets. There is nothing sacred or particular about this number two; we can carry out the same procedure for any number of sets, finite or infinite, and in fact we shall. We do so for two first because it illustrates the general construction but is not obscured by the additional notational difficulties.

DEFINITION The *union* of the two sets A and B, written as $A \cup B$, is the set $\{x \mid x \in A \text{ or } x \in B\}$.

A word about the use of "or." In ordinary English when we say that something is one or the other we imply that it is not both. The mathematical "or" is quite different, at least when we are speaking about set theory. *For when we say that x is in A or x is in B we mean x is in at least one of A or B, and may be in both.*

Let us consider a few examples of the union of two sets. For any set A, $A \cup A = A$; in fact, whenever B is a subset of A, $A \cup B = A$. If A is the set $\{x_1, x_2, x_3\}$ (i.e., the set whose elements are x_1, x_2, x_3) and if B is the set $\{y_1, y_2, x_1\}$, then $A \cup B = \{x_1, x_2, x_3, y_1, y_2\}$. If A is the set of all blonde-haired people and if B is the set of all people who smoke, then $A \cup B$ consists of all the people who either have blonde hair or smoke or both. Pictorially we can illustrate the union of the two sets A and B by

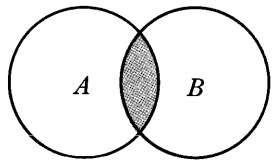


Here, A is the circle on the left, B that on the right, and $A \cup B$ is the shaded part.

DEFINITION The *intersection* of the two sets A and B, written as $A \cap B$, is the set $\{x \mid x \in A \text{ and } x \in B\}$.

The intersection of A and B is thus the set of all elements which are both in A and in B. In analogy with the examples used to illustrate the union of two sets, let us see what the intersections are in those very examples. For

any set A, $A \cap A = A$; in fact, if B is any subset of A, then $A \cap B = B$. If A is the set $\{x_1, x_2, x_3\}$ and B the set $\{y_1, y_2, x_1\}$, then $A \cap B = \{x_1\}$ (we are supposing no y is an x). If A is the set of all blonde-haired people and if B is the set of all people that smoke, then $A \cap B$ is the set of all blonde-haired people who smoke. Pictorially we can illustrate the intersection of the two sets A and B by



Here A is the circle on the left, B that on the right, while their intersection is the shaded part.

Two sets are said to be *disjoint* if their intersection is empty, that is, is the null set. For instance, if A is the set of positive integers and B the set of negative integers, then A and B are disjoint. Note however that if C is the set of nonnegative integers and if D is the set of nonpositive integers, then they are not disjoint, for their intersection consists of the integer 0, and so is not empty.

Before we generalize union and intersection from two sets to an arbitrary number of them, we should like to prove a little proposition interrelating union and intersection. This is the first of a whole host of such results that can be proved; some of these can be found in the problems at the end of this section.

PROPOSITION *For any three sets, A, B, C we have*

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

Proof. The proof will consist of showing, to begin with, the relation $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$ and then the converse relation $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$.

We first dispose of $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$. Because $B \subset B \cup C$, it is immediate that $A \cap B \subset A \cap (B \cup C)$. In a similar manner, $A \cap C \subset A \cap (B \cup C)$. Therefore

$$(A \cap B) \cup (A \cap C) \subset (A \cap (B \cup C)) \cup (A \cap (B \cup C)) = A \cap (B \cup C).$$

Now for the other direction. Given an element $x \in A \cap (B \cup C)$, first of all it must be an element of A. Secondly, as an element in $B \cup C$ it is either in B or in C. Suppose the former; then as an element both of A and of B, x must be in $A \cap B$. The second possibility, namely, $x \in C$, leads us

to $x \in A \cap C$. Thus in either eventuality $x \in (A \cap B) \cup (A \cap C)$, whence $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$.

The two opposite containing relations combine to give us the equality asserted in the proposition.

We continue the discussion of sets to extend the notion of union and of intersection to arbitrary collections of sets.

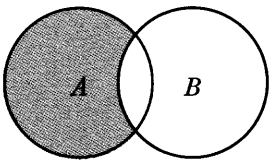
Given a set T we say that T serves as an *index set* for the family $\mathcal{F} = \{A_\alpha\}$ of sets if for every $\alpha \in T$ there exists a set of A_α in the family $\mathcal{F}$. The index set T can be any set, finite or infinite. Very often we use the set of non-negative integers as an index set, but, we repeat, T can be any (nonempty) set.

By the *union* of the sets A_α, where α is in T, we mean the set $\{x \mid x \in A_\alpha$ for at least one α in $T\}$. We shall denote it by $\bigcup_{\alpha \in T} A_\alpha$. By the *intersection* of the sets A_α, where α is in T, we mean the set $\{x \mid x \in A_\alpha$ for every $\alpha \in T\}$; we shall denote it by $\bigcap_{\alpha \in T} A_\alpha$. The sets A_α are *mutually disjoint* if for $\alpha \neq \beta$, $A_\alpha \cap A_\beta$ is the null set.

For instance, if S is the set of real numbers, and if T is the set of rational numbers, let, for $\alpha \in T$, $A_\alpha = \{x \in S \mid x \geq \alpha\}$. It is an easy exercise to see that $\bigcup_{\alpha \in T} A_\alpha = S$ whereas $\bigcap_{\alpha \in T} A_\alpha$ is the null set. The sets A_α are not mutually disjoint.

DEFINITION Given the two sets A, B then the *difference set*, $A - B$, is the set $\{x \in A \mid x \notin B\}$.

Returning to our little pictures, if A is the circle on the left, B that on the right, then $A - B$ is the shaded area.



Note that for any set B, the set A satisfies $A = (A \cap B) \cup (A - B)$. (Prove!) Note further that $B \cap (A - B)$ is the null set. A particular case of interest of the difference of two sets is when one of these is a subset of the other. In that case, when B is a subset of A, we call $A - B$ the *complement of B in A*.

We still want one more construct of two given sets A and B, their *Cartesian product* $A \times B$. This set $A \times B$ is defined as the set of all ordered pairs (a, b) where $a \in A$ and $b \in B$ and where we declare the pair (a_1, b_1) to be equal to (a_2, b_2) if and only if $a_1 = a_2$ and $b_1 = b_2$.

A few remarks about the Cartesian product. Given the two sets A and B we could construct the sets $A \times B$ and $B \times A$ from them. As sets these are distinct, yet we feel that they must be closely related. Given three sets A, B, C we can construct many Cartesian products from them: for instance, the set $A \times D$, where $D = B \times C$; the set $E \times C$, where $E = A \times B$; and also the set of all ordered triples (a, b, c) where $a \in A$, $b \in B$, and $c \in C$. These give us three distinct sets, yet here, also, we feel that these sets must be closely related. Of course, we can continue this process with more and more sets. To see the exact relation between them we shall have to wait until the next section, where we discuss one-to-one correspondences.

Given any index set T we could define the Cartesian product of the sets A_α as α varies over T; since we shall not need so general a product, we do not bother to define it.

Finally, we can consider the Cartesian product of a set A with itself, $A \times A$. Note that if the set A is a finite set having n elements, then the set $A \times A$ is also a finite set, but has n^2 elements. The set of elements (a, a) in $A \times A$ is called the *diagonal* of $A \times A$.

A subset R of $A \times A$ is said to define an *equivalence relation* on A if

1. $(a, a) \in R$ for all $a \in A$.
2. $(a, b) \in R$ implies $(b, a) \in R$.
3. $(a, b) \in R$ and $(b, c) \in R$ imply that $(a, c) \in R$.

Instead of speaking about subsets of $A \times A$ we can speak about a binary relation (one between two elements of A) on A itself, defining b to be related to a if $(a, b) \in R$. The properties 1, 2, 3 of the subset R immediately translate into the properties 1, 2, 3 of the definition below.

DEFINITION The binary relation $\sim$ on A is said to be an *equivalence relation* on A if for all a, b, c in A

1. $a \sim a$.
2. $a \sim b$ implies $b \sim a$.
3. $a \sim b$ and $b \sim c$ imply $a \sim c$.

The first of these properties is called *reflexivity*, the second, *symmetry*, and the third, *transitivity*.

The concept of an equivalence relation is an extremely important one and plays a central role in all of mathematics. We illustrate it with a few examples.

Example 1.1.1 Let S be any set and define $a \sim b$, for $a, b \in S$, if and only if $a = b$. This clearly defines an equivalence relation on S. In fact, an equivalence relation is a generalization of equality, measuring equality up to some property.

Example 1.1.2 Let S be the set of all integers. Given $a, b \in S$, define $a \sim b$ if $a - b$ is an even integer. We verify that this defines an equivalence relation of S.

1. Since $0 = a - a$ is even, $a \sim a$.
2. If $a \sim b$, that is, if $a - b$ is even, then $b - a = -(a - b)$ is also even, whence $b \sim a$.
3. If $a \sim b$ and $b \sim c$, then both $a - b$ and $b - c$ are even, whence $a - c = (a - b) + (b - c)$ is also even, proving that $a \sim c$.

Example 1.1.3 Let S be the set of all integers and let $n > 1$ be a fixed integer. Define for $a, b \in S$, $a \sim b$ if $a - b$ is a multiple of n. We leave it as an exercise to prove that this defines an equivalence relation on S.

Example 1.1.4 Let S be the set of all triangles in the plane. Two triangles are defined to be equivalent if they are similar (i.e., have corresponding angles equal). This defines an equivalence relation on S.

Example 1.1.5 Let S be the set of points in the plane. Two points a and b are defined to be equivalent if they are equidistant from the origin. A simple check verifies that this defines an equivalence relation on S.

There are many more equivalence relations; we shall encounter a few as we proceed in the book.

DEFINITION If A is a set and if $\sim$ is an equivalence relation on A, then the *equivalence class* of $a \in A$ is the set $\{x \in A \mid a \sim x\}$. We write it as $\mathrm{cl}(a)$.

In the examples just discussed, what are the equivalence classes? In Example 1.1.1, the equivalence class of a consists merely of a itself. In Example 1.1.2 the equivalence class of a consists of all the integers of the form $a + 2m$, where $m = 0, \pm 1, \pm 2, \ldots$; in this example there are only two distinct equivalence classes, namely, $\mathrm{cl}(0)$ and $\mathrm{cl}(1)$. In Example 1.1.3, the equivalence class of a consists of all integers of the form $a + kn$ where $k = 0, \pm 1, \pm 2, \ldots$; here there are n distinct equivalence classes, namely $\mathrm{cl}(0), \mathrm{cl}(1), \ldots, \mathrm{cl}(n - 1)$. In Example 1.1.5, the equivalence class of a consists of all the points in the plane which lie on the circle which has its center at the origin and passes through a.

Although we have made quite a few definitions, introduced some concepts, and have even established a simple little proposition, one could say in all fairness that up to this point we have not proved any result of real substance. We are now about to prove the first genuine result in the book. The proof of this theorem is not very difficult—actually it is quite easy—but nonetheless the result it embodies will be of great use to us.

THEOREM 1.1.1 *The distinct equivalence classes of an equivalence relation on A provide us with a decomposition of A as a union of mutually disjoint subsets. Conversely, given a decomposition of A as a union of mutually disjoint, nonempty subsets, we can define an equivalence relation on A for which these subsets are the distinct equivalence classes.*

Proof. Let the equivalence relation on A be denoted by $\sim$.

We first note that since for any $a \in A$, $a \sim a$, a must be in $\mathrm{cl}(a)$, whence the union of the $\mathrm{cl}(a)$'s is all of A. We now assert that given two equivalence classes they are either equal or disjoint. For, suppose that $\mathrm{cl}(a)$ and $\mathrm{cl}(b)$ are not disjoint; then there is an element $x \in \mathrm{cl}(a) \cap \mathrm{cl}(b)$. Since $x \in \mathrm{cl}(a)$, $a \sim x$; since $x \in \mathrm{cl}(b)$, $b \sim x$, whence by the symmetry of the relation, $x \sim b$. However, $a \sim x$ and $x \sim b$ by the transitivity of the relation forces $a \sim b$. Suppose, now that $y \in \mathrm{cl}(b)$; thus $b \sim y$. However, from $a \sim b$ and $b \sim y$, we deduce that $a \sim y$, that is, that $y \in \mathrm{cl}(a)$. Therefore, every element in $\mathrm{cl}(b)$ is in $\mathrm{cl}(a)$, which proves that $\mathrm{cl}(b) \subset \mathrm{cl}(a)$. The argument is clearly symmetric, whence we conclude that $\mathrm{cl}(a) \subset \mathrm{cl}(b)$. The two opposite containing relations imply that $\mathrm{cl}(a) = \mathrm{cl}(b)$.

We have thus shown that the distinct $\mathrm{cl}(a)$'s are mutually disjoint and that their union is A. This proves the first half of the theorem. Now for the other half!

Suppose that $A = \bigcup A_\alpha$ where the A_α are mutually disjoint, nonempty sets (α is in some index set T). How shall we use them to define an equivalence relation? The way is clear; given an element a in A it is in *exactly one* A_α. We define for $a, b \in A$, $a \sim b$ if a and b are in the same A_α. We leave it as an exercise to prove that this is an equivalence relation on A and that the distinct equivalence classes are the A_α's.

Problems

1. (a) If A is a subset of B and B is a subset of C, prove that A is a subset of C.
   (b) If $B \subset A$, prove that $A \cup B = A$, and conversely.
   (c) If $B \subset A$, prove that for any set C both $B \cup C \subset A \cup C$ and $B \cap C \subset A \cap C$.
2. (a) Prove that $A \cap B = B \cap A$ and $A \cup B = B \cup A$.
   (b) Prove that $(A \cap B) \cap C = A \cap (B \cap C)$.
3. Prove that $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
4. For a subset C of S let C' denote the complement of C in S. For any two subsets A, B of S prove the *De Morgan rules*:
   (a) $(A \cap B)' = A' \cup B'$.
   (b) $(A \cup B)' = A' \cap B'$.
5. For a finite set C let $o(C)$ indicate the number of elements in C. If A and B are finite sets prove $o(A \cup B) = o(A) + o(B) - o(A \cap B)$.

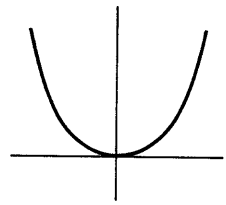
6. If A is a finite set having n elements, prove that A has exactly 2^n distinct subsets.
7. A survey shows that 63% of the American people like cheese whereas 76% like apples. What can you say about the percentage of the American people that like both cheese and apples? (The given statistics are not meant to be accurate.)
8. Given two sets A and B their *symmetric difference* is defined to be $(A - B) \cup (B - A)$. Prove that the symmetric difference of A and B equals $(A \cup B) - (A \cap B)$.
9. Let S be a set and let S^* be the set whose elements are the various subsets of S. In S^* we define an addition and multiplication as follows: If $A, B \in S^*$ (remember, this means that they are subsets of S):
   (1) $A + B = (A - B) \cup (B - A)$.
   (2) $A \cdot B = A \cap B$.
   Prove the following laws that govern these operations:
   (a) $(A + B) + C = A + (B + C)$.
   (b) $A \cdot (B + C) = A \cdot B + A \cdot C$.
   (c) $A \cdot A = A$.
   (d) $A + A =$ null set.
   (e) If $A + B = A + C$ then $B = C$.
   (The system just described is an example of a *Boolean algebra.*)
10. For the given set and relation below determine which define equivalence relations.
   (a) S is the set of all people in the world today, $a \sim b$ if a and b have an ancestor in common.
   (b) S is the set of all people in the world today, $a \sim b$ if a lives within 100 miles of b.
   (c) S is the set of all people in the world today, $a \sim b$ if a and b have the same father.
   (d) S is the set of real numbers, $a \sim b$ if $a = \pm b$.
   (e) S is the set of integers, $a \sim b$ if both $a > b$ and $b > a$.
   (f) S is the set of all straight lines in the plane, $a \sim b$ if a is parallel to b.
11. (a) Property 2 of an equivalence relation states that if $a \sim b$ then $b \sim a$; property 3 states that if $a \sim b$ and $b \sim c$ then $a \sim c$. What is wrong with the following proof that properties 2 and 3 imply property 1? Let $a \sim b$; then $b \sim a$, whence, by property 3 (using $a = c$), $a \sim a$.
   (b) Can you suggest an alternative of property 1 which will insure us that properties 2 and 3 do imply property 1?
12. In Example 1.1.3 of an equivalence relation given in the text, prove that the relation defined is an equivalence relation and that there are exactly n distinct equivalence classes, namely, $\text{cl}(0), \text{cl}(1), \ldots, \text{cl}(n-1)$.
13. Complete the proof of the second half of Theorem 1.1.1.

1.2 Mappings

We are about to introduce the concept of a mapping of one set into another. Without exaggeration this is probably the single most important and universal notion that runs through all of mathematics. It is hardly a new thing to any of us, for we have been considering mappings from the very earliest days of our mathematical training. When we were asked to plot the relation $y = x^2$ we were simply being asked to study the particular mapping which takes every real number onto its square.

Loosely speaking, a mapping from one set, S, into another, T, is a "rule" (whatever that may mean) that associates with each element in S a *unique* element t in T. We shall define a mapping somewhat more formally and precisely but the purpose of the definition is to allow us to think and speak in the above terms. We should think of them as rules or devices or mechanisms that transport us from one set to another.

Let us motivate a little the definition that we will make. The point of view we take is to consider the mapping to be defined by its "graph." We illustrate this with the familiar example $y = x^2$ defined on the real numbers S and taking its values also in S. For this set S, $S \times S$, the set of all pairs (a, b) can be viewed as the plane, the pair (a, b) corresponding to the point whose coordinates are a and b, respectively. In this plane we single out all those points whose coordinates are of the form (x, x^2) and call this set of points the graph of $y = x^2$. We even represent this set pictorially as



To find the "value" of the function or mapping at the point $x = a$, we look at the point in the graph whose first coordinate is a and read off the second coordinate as the value of the function at $x = a$.

This is, no more or less, the approach we take in the general setting to define a mapping from one set into another.

DEFINITION If S and T are nonempty sets, then a *mapping* from S to T is a subset, M, of $S \times T$ such that for *every* $s \in S$ there is a *unique* $t \in T$ such that the ordered pair (s, t) is in M.

This definition serves to make the concept of a mapping precise for us but we shall almost never use it in this form. Instead we do prefer to think of a

mapping as a rule which associates with any element s in S some element t in T, *the rule being, associate (or map)* $s \in S$ *with* $t \in T$ *if and only if* $(s, t) \in M$. We shall say that t is the *image* of s under the mapping.

Now for some notation for these things. Let σ be a mapping from S to T; we often denote this by writing $\sigma: S \to T$ or $S \xrightarrow{\sigma} T$. If t is the image of s under σ we shall sometimes write this as $\sigma: s \to t$; more often, we shall represent this fact by $t = s\sigma$. Note that we write the mapping σ on the *right*. There is no overall consistency in this usage; many people would write it as $t = \sigma(s)$. Algebraists often write mappings on the right; other mathematicians write them on the left. In fact, we shall not be absolutely consistent in this ourselves; when we shall want to emphasize the functional nature of σ we may very well write $t = \sigma(s)$.

Examples of Mappings

In all the examples the sets are assumed to be nonempty.

Example 1.2.1 Let S be any set; define $\iota: S \to S$ by $s = s\iota$ for any $s \in S$. This mapping ι is called the *identity mapping* of S.

Example 1.2.2 Let S and T be any sets and let t_0 be an element of T. Define $\tau: S \to T$ by $\tau: s \to t_0$ for every $s \in S$.

Example 1.2.3 Let S be the set of positive rational numbers and let $T = J \times J$ where J is the set of integers. Given a rational number s we can write it as $s = m/n$, where m and n have no common factor. Define $\tau: S \to T$ by $s\tau = (m, n)$.

Example 1.2.4 Let J be the set of integers and $S = \{(m, n) \in J \times J \mid n \neq 0\}$; let T be the set of rational numbers; define $\tau: S \to T$ by $(m, n)\tau = m/n$ for every (m, n) in S.

Example 1.2.5 Let J be the set of integers and $S = J \times J$. Define $\tau: S \to J$ by $(m, n)\tau = m + n$.

Note that in Example 1.2.5 the addition in J itself can be represented in terms of a mapping of $J \times J$ into J. Given an arbitrary set S we call a mapping of $S \times S$ into S a *binary operation* on S. Given such a mapping $\tau: S \times S \to S$ we could use it to define a "product" $*$ in S by declaring $a * b = c$ if $(a, b)\tau = c$.

Example 1.2.6 Let S and T be any sets; define $\tau: S \times T \to S$ by $(a, b)\tau = a$ for any $(a, b) \in S \times T$. This τ is called the *projection* of $S \times T$ on S. We could similarly define the projection of $S \times T$ on T.

Example 1.2.7 Let S be the set consisting of the elements x_1, x_2, x_3. Define $\tau: S \to S$ by $x_1\tau = x_2$, $x_2\tau = x_3$, $x_3\tau = x_1$.

Example 1.2.8 Let S be the set of integers and let T be the set consisting of the elements E and 0. Define $\tau: S \to T$ by declaring $n\tau = E$ if n is even and $n\tau = 0$ if n is odd.

If S is any set, let $\{x_1, \ldots, x_n\}$ be its subset consisting of the elements $x_1, x_2, \ldots, x_n$ of S. In particular, $\{x\}$ is the subset of S whose only element is x. Given S we can use it to construct a new set S^*, the set whose elements are the subsets of S. We call S^* the *set of subsets* of S. Thus for instance, if $S = \{x_1, x_2\}$ then S^* has exactly four elements, namely, a_1 = null set, a_2 = the subset, S, of S, $a_3 = \{x_1\}$, $a_4 = \{x_2\}$. The relation of S to S^*, in general, is a very interesting one; some of its properties are examined in the problems.

Example 1.2.9 Let S be a set, $T = S^*$; define $\tau: S \to T$ by $s\tau =$ complement of $\{s\}$ in $S = S - \{s\}$.

Example 1.2.10 Let S be a set with an equivalence relation, and let T be the set of equivalence classes in S (note that T is a subset of S^*). Define $\tau: S \to T$ by $s\tau = \text{cl}(s)$.

We leave the examples to continue the general discussion. Given a mapping $\tau: S \to T$ we define for $t \in T$, the *inverse image* of t with respect to τ to be the set $\{s \in S \mid t = s\tau\}$. In Example 1.2.8, the inverse image of E is the subset of S consisting of the even integers. It may happen that for some t in T that its inverse image with respect to τ is empty; that is, t is not the image under τ of any element in S. In Example 1.2.3, the element (4, 2) is not the image of any element in S under the τ used; in Example 1.2.9, S, as an element in S^*, is not the image under the τ used of any element in S.

DEFINITION The mapping τ of S into T is said to be *onto* T if given $t \in T$ there exists an element $s \in S$ such that $t = s\tau$.

If we call the subset $S\tau = \{x \in T \mid x = s\tau \text{ for some } s \in S\}$ the *image* of S under τ, then τ is onto if the image of S under τ is all of T. Note that in Examples 1.2.1, 1.2.4–1.2.8, and 1.2.10 the mappings used are all onto.

Another special type of mapping arises often and is important: the one-to-one mapping.

DEFINITION The mapping τ of S into T is said to be a *one-to-one mapping* if whenever $s_1 \neq s_2$, then $s_1\tau \neq s_2\tau$.

In terms of inverse images, the mapping τ is one-to-one if for any $t \in T$ the inverse image of t is either empty or is a set consisting of one element. In the examples discussed, the mappings in Examples 1.2.1, 1.2.3, 1.2.7, and 1.2.9 are all one-to-one.

When should we say that two mappings from S to T are equal? A natural definition for this is that they should have the same effect on every element of S; that is, the image of any element in S under each of these mappings should be the same. In a little more formal manner:

DEFINITION The two mappings σ and τ of S into T are said to be *equal* if $s\sigma = s\tau$ for every $s \in S$.

Consider the following situation: We have a mapping σ from S to T and another mapping τ from T to U. Can we compound these mappings to produce a mapping from S to U? The most natural and obvious way of doing this is to send a given element s, in S, in two stages into U, first by applying σ to s and then applying τ to the resulting element $s\sigma$ in T. This is the basis of the

DEFINITION If $\sigma: S \to T$ and $\tau: T \to U$ then the *composition* of σ and τ (also called their *product*) is the mapping $\sigma \circ \tau: S \to U$ defined by means of $s(\sigma \circ \tau) = (s\sigma)\tau$ for every $s \in S$.

Note that the order of events reads from left to right; $\sigma \circ \tau$ reads: first perform σ and then follow it up with τ. Here, too, the left-right business is not a uniform one. Mathematicians who write their mappings on the left would read $\sigma \circ \tau$ to mean first perform τ and then σ. Accordingly, in reading a given book in mathematics one must make absolutely sure as to what convention is being followed in writing the product of two mappings. We reiterate, *for us $\sigma \circ \tau$ will always mean: first apply σ and then τ.*

We illustrate the composition of σ and τ with a few examples.

Example 1.2.11 Let $S = \{x_1, x_2, x_3\}$ and let $T = S$. Let $\sigma: S \to S$ be defined by

$$x_1\sigma = x_2,$$
$$x_2\sigma = x_3,$$
$$x_3\sigma = x_1;$$

and $\tau: S \to S$ by

$$x_1\tau = x_1,$$
$$x_2\tau = x_3,$$
$$x_3\tau = x_2.$$

Thus

$$x_1(\sigma \circ \tau) = (x_1\sigma)\tau = x_2\tau = x_3,$$
$$x_2(\sigma \circ \tau) = (x_2\sigma)\tau = x_3\tau = x_2,$$
$$x_3(\sigma \circ \tau) = (x_3\sigma)\tau = x_1\tau = x_1.$$

At the same time we can compute $\tau \circ \sigma$, because in this case it also makes sense. Now

$$x_1(\tau \circ \sigma) = (x_1\tau)\sigma = (x_1\sigma) = x_2,$$
$$x_2(\tau \circ \sigma) = (x_2\tau)\sigma = x_3\sigma = x_1,$$
$$x_3(\tau \circ \sigma) = (x_3\tau)\sigma = x_2\sigma = x_3.$$

Note that $x_2 = x_1(\tau \circ \sigma)$, whereas $x_3 = x_1(\sigma \circ \tau)$ whence $\sigma \circ \tau \neq \tau \circ \sigma$.

Example 1.2.12 Let S be the set of integers, T the set $S \times S$, and suppose $\sigma: S \to T$ is defined by $m\sigma = (m - 1, 1)$. Let $U = S$ and suppose that $\tau: T \to U (= S)$ is defined by $(m, n)\tau = m + n$. Thus $\sigma \circ \tau: S \to S$ whereas $\tau \circ \sigma: T \to T$; even to speak about the equality of $\sigma \circ \tau$ and $\tau \circ \sigma$ would make no sense since they do not act on the same space. We now compute $\sigma \circ \tau$ as a mapping of S into itself and then $\tau \circ \sigma$ as one on T into itself.

Given $m \in S$, $m\sigma = (m - 1, 1)$ whence $m(\sigma \circ \tau) = (m\sigma)\tau = (m - 1, 1)\tau = (m - 1) + 1 = m$. Thus $\sigma \circ \tau$ is the identity mapping of S into itself. What about $\tau \circ \sigma$? Given $(m, n) \in T$, $(m, n)\tau = m + n$, whereby $(m, n)(\tau \circ \sigma) = ((m, n)\tau)\sigma = (m + n)\sigma = (m + n - 1, 1)$. Note that $\tau \circ \sigma$ is *not* the identity map of T into itself; it is not even an onto mapping of T.

Example 1.2.13 Let S be the set of real numbers, T the set of integers, and $U = \{E, 0\}$. Define $\sigma: S \to T$ by $s\sigma =$ largest integer less than or equal to s, and $\tau: T \to U$ defined by $n\tau = E$ if n is even, $n\tau = 0$ if n is odd. Note that in this case $\tau \circ \sigma$ cannot be defined. We compute $\sigma \circ \tau$ for two real numbers $s = \frac{8}{3}$ and $s = \pi$. Now since $\frac{8}{3} = 2 + \frac{2}{3}$, $(\frac{8}{3})\sigma = 2$, whence $(\frac{8}{3})(\sigma \circ \tau) = (\frac{8}{3}\sigma)\tau = (2)\tau = E$; $(\pi)\sigma = 3$, whence $\pi(\sigma \circ \tau) = (\pi\sigma)\tau = (3)\tau = 0$.

For mappings of sets, provided the requisite products make sense, a general *associative law* holds. This is the content of

LEMMA 1.2.1 (Associative Law) *If* $\sigma: S \to T$, $\tau: T \to U$, *and* $\mu: U \to V$, *then* $(\sigma \circ \tau) \circ \mu = \sigma \circ (\tau \circ \mu)$.

Proof. Note first that $\sigma \circ \tau$ makes sense and takes S into U, thus $(\sigma \circ \tau) \circ \mu$ also makes sense and takes S into V. Similarly $\sigma \circ (\tau \circ \mu)$ is meaningful and takes S into V. Thus we can speak about the equality, or lack of equality, of $(\sigma \circ \tau) \circ \mu$ and $\sigma \circ (\tau \circ \mu)$.

To prove the asserted equality we merely must show that for any $s \in S$, $s((\sigma \circ \tau) \circ \mu) = s(\sigma \circ (\tau \circ \mu))$. Now by the very definition of the composition

of maps, $s((\sigma \circ \tau) \circ \mu) = (s(\sigma \circ \tau))\mu = ((s\sigma)\tau)\mu$ whereas $s(\sigma \circ (\tau \circ \mu)) = (s\sigma)(\tau \circ \mu) = ((s\sigma)\tau)\mu$. Thus, the elements $s((\sigma \circ \tau) \circ \mu)$ and $s(\sigma \circ (\tau \circ \mu))$ are indeed equal. This proves the lemma.

We should like to show that if two mappings σ and τ are properly conditioned the very same conditions carry over to $\sigma \circ \tau$.

LEMMA 1.2.2 *Let $\sigma: S \to T$ and $\tau: T \to U$; then*

1. *$\sigma \circ \tau$ is onto if each of σ and τ is onto.*
2. *$\sigma \circ \tau$ is one-to-one if each of σ and τ is one-to-one.*

Proof. We prove only part 2, leaving the proof of part 1 as an exercise.

Suppose that $s_1, s_2 \in S$ and that $s_1 \neq s_2$. By the one-to-one nature of σ, $s_1\sigma \neq s_2\sigma$. Since τ is one-to-one and $s_1\sigma$ and $s_2\sigma$ are distinct elements of T, $(s_1\sigma)\tau \neq (s_2\sigma)\tau$ whence $s_1(\sigma \circ \tau) = (s_1\sigma)\tau \neq (s_2\sigma)\tau = s_2(\sigma \circ \tau)$, proving that $\sigma \circ \tau$ is indeed one-to-one, and establishing the lemma.

Suppose that σ is a one-to-one mapping of S *onto* T; we call σ a *one-to-one correspondence* between S and T. Given any $t \in T$, by the "onto-ness" of σ there exists an element $s \in S$ such that $t = s\sigma$; by the "one-to-oneness" of σ this s is unique. We define the mapping $\sigma^{-1}: T \to S$ by $s = t\sigma^{-1}$ if and only if $t = s\sigma$. The mapping σ^{-1} is called the *inverse* of σ. Let us compute $\sigma \circ \sigma^{-1}$ which maps S into itself. Given $s \in S$, let $t = s\sigma$, whence by definition $s = t\sigma^{-1}$; thus $s(\sigma \circ \sigma^{-1}) = (s\sigma)\sigma^{-1} = t\sigma^{-1} = s$. We have shown that $\sigma \circ \sigma^{-1}$ is the identity mapping of S onto itself. A similar computation reveals that $\sigma^{-1} \circ \sigma$ is the identity mapping of T onto itself.

Conversely, if $\sigma: S \to T$ is such that there exists a $\mu: T \to S$ with the property that $\sigma \circ \mu$ and $\mu \circ \sigma$ are the identity mappings on S and T, respectively, then we claim that σ is a one-to-one correspondence between S and T. First observe that σ is onto for, given $t \in T$, $t = t(\mu \circ \sigma) = (t\mu)\sigma$ (since $\mu \circ \sigma$ is the identity on T) and so t is the image under σ of the element $t\mu$ in S. Next observe that σ is one-to-one, for if $s_1\sigma = s_2\sigma$, using that $\sigma \circ \mu$ is the identity on S, we have $s_1 = s_1(\sigma \circ \mu) = (s_1\sigma)\mu = (s_2\sigma)\mu = s_2(\sigma \circ \mu) = s_2$. We have now proved

LEMMA 1.2.3 *The mapping $\sigma: S \to T$ is a one-to-one correspondence between S and T if and only if there exists a mapping $\mu: T \to S$ such that $\sigma \circ \mu$ and $\mu \circ \sigma$ are the identity mappings on S and T, respectively.*

DEFINITION If S is a nonempty set then $A(S)$ is the *set of all one-to-one mappings* of S onto itself.

Aside from its own intrinsic interest $A(S)$ plays a central and universal type of role in considering the mathematical system known as a group

(Chapter 2). For this reason we state the next theorem concerning its nature. All the constituent parts of the theorem have already been proved in the various lemmas, so we state the theorem without proof.

THEOREM 1.2.1 *If σ, τ, μ are elements of $A(S)$, then*

1. $\sigma \circ \tau$ *is in* $A(S)$.
2. $(\sigma \circ \tau) \circ \mu = \sigma \circ (\tau \circ \mu)$.
3. *There exists an element* ι *(the identity map) in* $A(S)$ *such that* $\sigma \circ \iota = \iota \circ \sigma = \sigma$.
4. *There exists an element* $\sigma^{-1} \in A(S)$ *such that* $\sigma \circ \sigma^{-1} = \sigma^{-1} \circ \sigma = \iota$.

We close the section with a remark about $A(S)$. Suppose that S has more than two elements; let x_1, x_2, x_3 be three distinct elements in S; define the mapping $\sigma: S \to S$ by $x_1\sigma = x_2$, $x_2\sigma = x_3$, $x_3\sigma = x_1$, $s\sigma = s$ for any $s \in S$ different from x_1, x_2, x_3. Define the mapping $\tau: S \to S$ by $x_2\tau = x_3$, $x_3\tau = x_2$, and $s\tau = s$ for any $s \in S$ different from x_2, x_3. Clearly both σ and τ are in $A(S)$. A simple computation shows that $x_1(\sigma \circ \tau) = x_3$ but that $x_1(\tau \circ \sigma) = x_2 \neq x_3$. Thus $\sigma \circ \tau \neq \tau \circ \sigma$. This is

LEMMA 1.2.4 *If S has more that two elements we can find two elements σ, τ in $A(S)$ such that $\sigma \circ \tau \neq \tau \circ \sigma$.*

Problems

1. In the following, where $\sigma: S \to T$, determine whether the σ is onto and/or one-to-one and determine the inverse image of any $t \in T$ under σ.
   (a) S = set of real numbers, T = set of nonnegative real numbers, $s\sigma = s^2$.
   (b) S = set of nonnegative real numbers, T = set of nonnegative real numbers, $s\sigma = s^2$.
   (c) S = set of integers, T = set of integers, $s\sigma = s^2$.
   (d) S = set of integers, T = set of integers, $s\sigma = 2s$.
2. If S and T are nonempty sets, prove that there exists a one-to-one correspondence between $S \times T$ and $T \times S$.
3. If S, T, U are nonempty sets, prove that there exists a one-to-one correspondence between
   (a) $(S \times T) \times U$ and $S \times (T \times U)$.
   (b) Either set in part (a) and the set of ordered triples (s, t, u) where $s \in S$, $t \in T$, $u \in U$.
4. (a) If there is a one-to-one correspondence between S and T, prove that there exists a one-to-one correspondence between T and S.

(b) If there is a one-to-one correspondence between S and T and between T and U, prove that there is a one-to-one correspondence between S and U.

5. If ι is the identity mapping on S, prove that for any $\sigma \in A(S)$, $\sigma \circ \iota = \iota \circ \sigma = \sigma$.

*6. If S is any set, prove that it is *impossible* to find a mapping of S *onto* S^*.

7. If the set S has n elements, prove that $A(S)$ has $n!$ (n factorial) elements.

8. If the set S has a finite number of elements, prove the following:
   (a) If σ maps S onto S, then σ is one-to-one.
   (b) If σ is a one-to-one mapping of S onto itself, then σ is onto.
   (c) Prove, by example, that both part (a) and part (b) are false if S does not have a finite number of elements.

9. Prove that the converse to both parts of Lemma 1.2.2 are false; namely,
   (a) If $\sigma \circ \tau$ is onto, it need not be that both σ and τ are onto.
   (b) If $\sigma \circ \tau$ is one-to-one, it need not be that both σ and τ are one-to-one.

10. Prove that there is a one-to-one correspondence between the set of integers and the set of rational numbers.

11. If $\sigma: S \to T$ and if A is a subset of S, the *restriction of* σ *to* A, σ_A, is defined by $a\sigma_A = a\sigma$ for any $a \in A$. Prove
   (a) σ_A defines a mapping of A into T.
   (b) σ_A is one-to-one if σ is.
   (c) σ_A may very well be one-to-one even if σ is not.

12. If $\sigma: S \to S$ and A is a subset of S such that $A\sigma \subset A$, prove that $(\sigma \circ \sigma)_A = \sigma_A \circ \sigma_A$.

13. A set S is said to be *infinite* if there is a one-to-one correspondence between S and a proper subset of S. Prove
   (a) The set of integers is infinite.
   (b) The set of real numbers is infinite.
   (c) If a set S has a subset A which is infinite, then S must be infinite.
   (*Note:* By the result of Problem 8, a set finite in the usual sense is not infinite.)

*14. If S is infinite and can be brought into one-to-one correspondence with the set of integers, prove that there is one-to-one correspondence between S and $S \times S$.

*15. Given two sets S and T we declare $S < T$ (S is smaller than T) if there is a mapping of T *onto* S but *no* mapping of S *onto* T. Prove that if $S < T$ and $T < U$ then $S < U$.

16. If S and T are finite sets having m and n elements, respectively, prove that if $m < n$ then $S < T$.

1.3 The Integers

We close this chapter with a brief discussion of the set of integers. We shall make no attempt to construct them axiomatically, assuming instead that we already have the set of integers and that we know many of the elementary facts about them. In this number we include the principle of mathematical induction (which will be used freely throughout the book) and the fact that a nonempty set of positive integers always contains a smallest element. As to notation, the familiar symbols: $a > b$, $a \leq b$, $|a|$, etc., will occur with their usual meaning. To avoid repeating that something is an integer, we make the assumption *that all symbols, in this section, written as lowercase Latin letters will be integers.*

Given a and b, with $b \neq 0$, we can divide a by b to get a nonnegative remainder r which is smaller in size than b; that is, we can find m and r such that $a = mb + r$ where $0 \leq r < |b|$. This fact is known as the *Euclidean algorithm* and we assume familiarity with it.

We say that $b \neq 0$ *divides* a if $a = mb$ for some m. We denote that b divides a by $b \mid a$, and that b does not divide a by $b \nmid a$. Note that if $a \mid 1$ then $a = \pm 1$, that when both $a \mid b$ and $b \mid a$, then $a = \pm b$, and that any b divides 0. If $b \mid a$, we call b a *divisor* of a. Note that if b is a divisor of g and of h, then it is a divisor of $mg + nh$ for arbitrary integers m and n. We leave the verification of these remarks as exercises.

DEFINITION The positive integer c is said to be the *greatest common divisor* of a and b if

1. c is a divisor of a and of b.
2. Any divisor of a and b is a divisor of c.

We shall use the notation (a, b) for the greatest common divisor of a and b. Since we insist that the greatest common divisor be positive, $(a, b) = (a, -b) = (-a, b) = (-a, -b)$. For instance, $(60, 24) = (60, -24) = 12$. Another comment: The mere fact that we have defined what is to be meant by the greatest common divisor does not guarantee that it exists. This will have to be proved. However, we can say that if it exists then it is unique, for, if we had c_1 and c_2 satisfying both conditions of the definition above, then $c_1 \mid c_2$ and $c_2 \mid c_1$, whence we would have $c_1 = \pm c_2$; the insistence on positivity would then force $c_1 = c_2$. Our first business at hand then is to dispose of the existence of (a, b). In doing so, in the next lemma, we actually prove a little more, namely that (a, b) must have a particular form.

LEMMA 1.3.1 *If a and b are integers, not both 0, then (a, b) exists; moreover, we can find integers m_0 and n_0 such that $(a, b) = m_0 a + n_0 b$.*

Proof. Let $\mathcal{M}$ be the set of all integers of the form $ma + nb$, where m and n range freely over the set of integers. Since one of a or b is not 0, there are nonzero integers in $\mathcal{M}$. Because $x = ma + nb$ is in $\mathcal{M}$, $-x = (-m)a + (-n)b$ is also in $\mathcal{M}$; therefore, $\mathcal{M}$ always has in it some positive integers. But then there is a smallest positive integer, c, in $\mathcal{M}$; being in $\mathcal{M}$, c has the form $c = m_0a + n_0b$. We claim that $c = (a, b)$.

Note first that if $d \mid a$ and $d \mid b$, the $d \mid (m_0a + n_0b)$, whence $d \mid c$. We now must show that $c \mid a$ and $c \mid b$. Given any element $x = ma + nb$ in $\mathcal{M}$, then by the Euclidean algorithm, $x = tc + r$ where $0 \leq r < c$. Writing this out explicitly, $ma + nb = t(m_0a + n_0b) + r$, whence $r = (m - tm_0)a + (n - tn_0)b$ and so must be in $\mathcal{M}$. Since $0 \leq r$ and $r < c$, by the choice of c, $r = 0$. Thus $x = tc$; we have proved that $c \mid x$ for any $x \in \mathcal{M}$. But $a = 1a + 0b \in \mathcal{M}$ and $b = 0a + 1b \in \mathcal{M}$, whence $c \mid a$ and $c \mid b$.

We have shown that c satisfies the requisite properties to be (a, b) and so we have proved the lemma.

DEFINITION The integers a and b are *relatively prime* if $(a, b) = 1$.

As an immediate consequence of Lemma 1.3.1, we have the

COROLLARY *If a and b are relatively prime, we can find integers m and n such that $ma + nb = 1$.*

We introduce another familiar notion, that of prime number. By this we shall mean an integer which has no nontrivial factorization. For technical reasons, we exclude 1 from the set of prime numbers. The sequence 2, 3, 5, 7, 11, ... are all prime numbers; equally, $-2, -3, -5, \ldots$ are prime numbers. Since, in factoring, the negative introduces no essential differences, for us prime numbers will always be positive.

DEFINITION The integer $p > 1$ is a *prime number* if its only divisors are $\pm 1, \pm p$.

Another way of putting this is to say that an integer p (larger than 1) is a prime number if and only if given any other integer n then either $(p, n) = 1$ or $p \mid n$. As we shall soon see, the prime numbers are the building blocks of the integers. But first we need the important observation,

LEMMA 1.3.2 *If a is relatively prime to b but $a \mid bc$, then $a \mid c$.*

Proof. Since a and b are relatively prime, by the corollary to Lemma 1.3.1, we can find integers m and n such that $ma + nb = 1$. Thus $mac + nbc = c$. Now $a \mid mac$ and, by assumption, $a \mid nbc$; consequently,

$a \mid (mac + nbc)$. Since $mac + nbc = c$, we conclude that $a \mid c$, which is precisely the assertion of the lemma.

Following immediately from the lemma and the definition of prime number is the important

COROLLARY *If a prime number divides the product of certain integers it must divide at least one of these integers.*

We leave the proof of the corollary to the reader.

We have asserted that the prime numbers serve as the building blocks for the set of integers. The precise statement of this is the *unique factorization theorem*:

THEOREM 1.3.1 *Any positive integer $a > 1$ can be factored in a unique way as $a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_t^{\alpha_t}$, where $p_1 > p_2 > \cdots > p_t$ are prime numbers and where each $\alpha_i > 0$.*

Proof. The theorem as stated actually consists of two distinct subtheorems; the first asserts the possibility of factoring the given integer as a product of prime powers; the second assures us that this decomposition is unique. We shall prove the theorem itself by proving each of these subtheorems separately.

An immediate question presents itself: How shall we go about proving the theorem? A natural method of attack is to use mathematical induction. A short word about this; we shall use the following version of mathematical induction: If the proposition $P(m_0)$ is true and if the truth of $P(r)$ for all r such that $m_0 \leq r < k$ implies the truth of $P(k)$, then $P(n)$ is true for all $n \geq m_0$. This variant of induction can be shown to be a consequence of the basic property of the integers which asserts that any nonempty set of positive integers has a minimal element (see Problem 10).

We first prove that every integer $a > 1$ can be factored as a product of prime powers; our approach is via mathematical induction.

Certainly $m_0 = 2$, being a prime number, has a representation as a product of prime powers.

Suppose that any integer r, $2 \leq r < k$ can be factored as a product of prime powers. If k itself is a prime number, then it is a product of prime powers. If k is not a prime number, then $k = uv$, where $1 < u < k$ and $1 < v < k$. By the induction hypothesis, since both u and v are less than k, each of these can be factored as a product of prime powers. Thus $k = uv$ is also such a product. We have shown that the truth of the proposition for all integers r, $2 \leq r < k$, implies its truth for k. Consequently, by the basic induction principle, the proposition is true for all integers $n \geq m_0 = 2$; that is, every integer $n \geq 2$ is a product of prime powers.

Now for the uniqueness. Here, too, we shall use mathematical induction, and in the form used above. Suppose that

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r} = q_1^{\beta_1} q_2^{\beta_2} \cdots q_s^{\beta_s},$$

where $p_1 > p_2 > \cdots p_r$, $q_1 > q_2 > \cdots > q_s$ are prime numbers, and where each $\alpha_i > 0$ and each $\beta_i > 0$. Our object is to prove

1. $r = s$.
2. $p_1 = q_1, p_2 = q_2, \ldots, p_r = q_r$.
3. $\alpha_1 = \beta_1, \alpha_2 = \beta_2, \ldots, \alpha_r = \beta_r$.

For $a = 2$ this is clearly true. Proceeding by induction we suppose it to be true for all integers u, $2 \leq u < a$. Now, since

$$a = p_1^{\alpha_1} \cdots p_r^{\alpha_r} = q_1^{\beta_1} \cdots q_s^{\beta_s}$$

and since $\alpha_1 > 0$, $p_1 \mid a$, hence $p_1 \mid q_1^{\beta_1} \cdots q_s^{\beta_s}$. However, since p_1 is a prime number, by the corollary to Lemma 1.3.2, it follows easily that $p_1 = q_i$ for some i. Thus $q_1 \geq q_i = p_1$. Similarly, since $q_1 \mid a$ we get $q_1 = p_j$ for some j, whence $p_1 \geq p_j = q_1$. In short, we have shown that $p_1 = q_1$. Therefore $a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r} = p_1^{\beta_1} q_2^{\beta_2} \cdots q_s^{\beta_s}$. We claim that this forces $\alpha_1 = \beta_1$. (Prove!) But then

$$b = \frac{a}{p^{\alpha_1}} = p_2^{\alpha_2} \cdots p_r^{\alpha_r} = q_2^{\beta_2} \cdots q_s^{\beta_s}.$$

If $b = 1$, then $\alpha_2 = \cdots = \alpha_r = 0$ and $\beta_2 = \cdots = \beta_s = 0$; that is, $r = s = 1$, and we are done. If $b > 1$, then since $b < a$ we can apply our induction hypothesis to b to get

1. The number of distinct prime power factors (in b) on both sides is equal, that is, $r - 1 = s - 1$, hence $r = s$.
2. $\alpha_2 = \beta_2, \ldots, \alpha_r = \beta_r$.
3. $p_2 = q_2, \ldots, p_r = q_r$.

Together with the information we already have obtained, namely, $p_1 = q_1$ and $\alpha_1 = \beta_1$, this is precisely what we were trying to prove. Thus we see that the assumption of the uniqueness of factorization for the integers less than a implied the uniqueness of factorization for a. In consequence, the induction is completed and the assertion of unique factorization is established.

We change direction a little to study the important notion of congruence modulo a given integer. As we shall see later, the relation that we now introduce is a special case of a much more general one that can be defined in a much broader context.

DEFINITION Let $n > 0$ be a fixed integer. We define $a \equiv b \bmod n$ if $n \mid (a - b)$.

The relation is referred to as *congruence modulo n*, n is called the *modulus* of the relation, and we read $a \equiv b \bmod n$ as "a is congruent to b modulo n." Note, for example, that $73 \equiv 4 \bmod 23$, $21 \equiv -9 \bmod 10$, etc.

This congruence relation enjoys the following basic properties:

LEMMA 1.3.3

1. *The relation congruence modulo n defines an equivalence relation on the set of integers.*
2. *This equivalence relation has n distinct equivalence classes.*
3. *If* $a \equiv b \bmod n$ *and* $c \equiv d \bmod n$, *then* $a + c \equiv b + d \bmod n$ *and* $ac \equiv bd \bmod n$.
4. *If* $ab \equiv ac \bmod n$ *and a is relatively prime to n, then* $b \equiv c \bmod n$.

Proof. We first verify that the relation congruence modulo n is an equivalence relation. Since $n \mid 0$, we indeed have that $n \mid (a - a)$ whence $a \equiv a \bmod n$ for every a. Further, if $a \equiv b \bmod n$ then $n \mid (a - b)$, and so $n \mid (b - a) = -(a - b)$; thus $b \equiv a \bmod n$. Finally, if $a \equiv b \bmod n$ and $b \equiv c \bmod n$, then $n \mid (a - b)$ and $n \mid (b - c)$ whence $n \mid \{(a - b) + (b - c)\}$, that is, $n \mid (a - c)$. This, of course, implies that $a \equiv c \bmod n$.

Let the equivalence class, under this relation, of a be denoted by $[a]$; we call it the *congruence class* (mod n) of a. Given any integer a, by the Euclidean algorithm, $a = kn + r$ where $0 \leq r < n$. But then, $a \in [r]$ and so $[a] = [r]$. Thus there are at most n distinct congruence classes; namely, $[0], [1], \ldots, [n - 1]$. However, these are distinct, for if $[i] = [j]$ with, say, $0 \leq i < j < n$, then $n \mid (j - i)$ where $j - i$ is a positive integer less than n, which is obviously impossible. Consequently, there are exactly the n distinct congruence classes $[0], [1], \ldots, [n - 1]$. We have now proved assertions 1 and 2 of the lemma.

We now prove part 3. Suppose that $a \equiv b \bmod n$ and $c \equiv d \bmod n$; therefore, $n \mid (a - b)$ and $n \mid (c - d)$ whence $n \mid \{(a - d) + (c - d)\}$, and so $n \mid \{(a + c) - (b + d)\}$. But then $a + c \equiv b + d \bmod n$. In addition, $n \mid \{(a - b)c + (c - d)b\} = ac - bd$, whence $ac \equiv bd \bmod n$.

Finally, notice that if $ab \equiv ac \bmod n$ and if a is relatively prime to n, then the fact that $n \mid a(b - c)$, by Lemma 1.3.2, implies that $n \mid (b - c)$ and so $b \equiv c \bmod n$.

If a is not relatively prime to n, the result of part 4 may be false; for instance, $2.3 \equiv 4.3 \bmod 6$, yet $2 \not\equiv 4 \bmod 6$.

Lemma 1.3.3 opens certain interesting possibilities for us. Let J_n be the

set of the congruence classes mod n; that is, $J_n = \{[0], [1], \ldots, [n-1]\}$. Given two elements, $[i]$ and $[j]$ in J_n, let us define

$$[i] + [j] = [i + j]; \tag{a}$$

$$[i][j] = [ij]. \tag{b}$$

We assert that the lemma assures us that this "addition" and "multiplication" are *well defined*; that is, if $[i] = [i']$ and $[j] = [j']$, then $[i] + [j] = [i + j] = [i' + j'] = [i'] + [j']$ and that $[i][j] = [i'][j']$. (Verify!) These operations in J_n have the following interesting properties (whose proofs we leave as exercises): for any $[i]$, $[j]$, $[k]$ in J_n,

1. $[i] + [j] = [j] + [i]$
2. $[i][j] = [j][i]$ } commutative laws.
3. $([i] + [j]) + [k] = [i] + ([j] + [k])$
4. $([i][j])[k] = [i]([j][k])$ } associative laws.
5. $[i]([j] + [k]) = [i][j] + [i][k]$ distributive law.
6. $[0] + [i] = [i]$.
7. $[1][i] = [i]$.

One more remark: if $n = p$ is a prime number and if $[a] \neq [0]$ is in J_p, then there is an element $[b]$ in J_p such that $[a][b] = [1]$.

The set J_n plays an important role in algebra and number theory. It is called the set of *integers* mod n; before we proceed much further we will have become well acquainted with it.

Problems

1. If $a \mid b$ and $b \mid a$, show that $a = \pm b$.
2. If b is a divisor of g and of h, show it is a divisor of $mg + nh$.
3. If a and b are integers, the *least common multiple* of a and b, written as $[a, b]$, is defined as that positive integer d such that
   (a) $a \mid d$ and $b \mid d$.
   (b) Whenever $a \mid x$ and $b \mid x$ then $d \mid x$.
   Prove that $[a, b]$ exists and that $[a, b] = ab/(a, b)$, if $a > 0$, $b > 0$.
4. If $a \mid x$ and $b \mid x$ and $(a, b) = 1$ prove that $(ab) \mid x$.
5. If $a = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$ and $b = p_1^{\beta_1} \cdots p_k^{\beta_k}$ where the p_i are distinct prime numbers and where each $\alpha_i \geq 0$, $\beta_i \geq 0$, prove
   (a) $(a, b) = p_1^{\delta_1} \cdots p_k^{\delta_k}$ where $\delta_i =$ minimum of α_i and β_i for each i.
   (b) $[a, b] = p_1^{\gamma_1} \cdots p_k^{\gamma_k}$ where $\gamma_i =$ maximum of α_i and β_i for each i.

6. Given a, b, on applying the Euclidean algorithm successively we have

$$\begin{aligned} a &= q_0 b + r_1, & 0 \leq r_1 < |b|, \\ b &= q_1 r_1 + r_2, & 0 \leq r_2 < r_1, \\ r_1 &= q_2 r_2 + r_3, & 0 \leq r_3 < r_2, \\ &\cdot \\ &\cdot \\ &\cdot \\ r_k &= q_{k+1} r_{k+1} + r_{k+2}, & 0 \leq r_{k+2} < r_{k+1}. \end{aligned}$$

Since the integers r_k are decreasing and are all nonnegative, there is a first integer n such that $r_{n+1} = 0$. Prove that $r_n = (a, b)$. (We consider, here, $r_0 = |b|$.)

7. Use the method in Problem 6 to calculate
(a) (1128, 33). (b) (6540, 1206).

8. To check that n is a prime number, prove that it is sufficient to show that it is not divisible by any prime number p, such that $p \leq \sqrt{n}$.

9. Show that $n > 1$ is a prime number if and only if for any a either $(a, n) = 1$ or $n \mid a$.

10. Assuming that any nonempty set of positive integers has a minimal element, prove
(a) If the proposition P is such that
(1) $P(m_0)$ is true,
(2) the truth of $P(m-1)$ implies the truth of $P(m)$,
then $P(n)$ is true for all $n \geq m_0$.
(b) If the proposition P is such that
(1) $P(m_0)$ is true,
(2) $P(m)$ is true whenever $P(a)$ is true for all a such that $m_0 \leq a < m$,
then $P(n)$ is true for all $n \geq m_0$.

11. Prove that the addition and multiplication used in J_n are well defined.

12. Prove the properties 1–7 for the addition and multiplication in J_n.

13. If $(a, n) = 1$, prove that one can find $[b] \in J_n$ such that $[a][b] = [1]$ in J_n.

*14. If p is a prime number, prove that for any integer a, $a^p \equiv a \bmod p$.

15. If $(m, n) = 1$, given a and b, prove that there exists an x such that $x \equiv a \bmod m$ and $x \equiv b \bmod n$.

16. Prove the corollary to Lemma 1.3.2.

17. Prove that n is a prime number if and only if in J_n, $[a][b] = [0]$ implies that $[a] = [b] = [0]$.

Supplementary Reading

For sets and cardinal numbers:

BIRKHOFF, G., and MACLANE, S., *A Brief Survey of Modern Algebra*, 2nd ed. New York: The Macmillan Company, 1965.

2

Group Theory

In this chapter we shall embark on the study of the algebraic object known as a group which serves as one of the fundamental building blocks for the subject today called abstract algebra. In later chapters we shall have a look at some of the others such as rings, fields, vector spaces, and linear algebras. Aside from the fact that it has become traditional to consider groups at the outset, there are natural, cogent reasons for this choice. To begin with, groups, being one-operational systems, lend themselves to the simplest formal description. Yet despite this simplicity of description the fundamental algebraic concepts such as homomorphism, quotient construction, and the like, which play such an important role in all algebraic structures—in fact, in all of mathematics—already enter here in a pure and revealing form.

At this point, before we become weighted down with details, let us take a quick look ahead. In abstract algebra we have certain basic systems which, in the history and development of mathematics, have achieved positions of paramount importance. These are usually sets on whose elements we can operate algebraically—by this we mean that we can combine two elements of the set, perhaps in several ways, to obtain a third element of the set—and, in addition, we assume that these algebraic operations are subject to certain rules, which are explicitly spelled out in what we call the axioms or postulates defining the system. In this abstract setting we then attempt to prove theorems about these very general structures, always hoping that when these results are applied to a particular, concrete realization of the abstract

system there will flow out facts and insights into the example at hand which would have been obscured from us by the mass of inessential information available to us in the particular, special case.

We should like to stress that these algebraic systems and the axioms which define them must have a certain naturality about them. They must come from the experience of looking at many examples; they should be rich in meaningful results. One does not just sit down, list a few axioms, and then proceed to study the system so described. This, admittedly, is done by some, but most mathematicians would dismiss these attempts as poor mathematics. The systems chosen for study are chosen because particular cases of these structures have appeared time and time again, because someone finally noted that these special cases were indeed special instances of a general phenomenon, because one notices analogies between two highly disparate mathematical objects and so is led to a search for the root of these analogies. To cite an example, case after case after case of the special object, which we know today as groups, was studied toward the end of the eighteenth, and at the beginning of the nineteenth, century, yet it was not until relatively late in the nineteenth century that the notion of an abstract group was introduced. The only algebraic structures, so far encountered, that have stood the test of time and have survived to become of importance, have been those based on a broad and tall pillar of special cases. Amongst mathematicians neither the beauty nor the significance of the first example which we have chosen to discuss—groups—is disputed.

2.1 Definition of a Group

At this juncture it is advisable to recall a situation discussed in the first chapter. For an arbitrary nonempty set S we defined $A(S)$ to be the set of all *one-to-one* mappings of the set S *onto* itself. For any two elements σ, $\tau \in A(S)$ we introduced a product, denoted by $\sigma \circ \tau$, and on further investigation it turned out that the following facts were true for the elements of $A(S)$ subject to this product:

1. Whenever $\sigma, \tau \in A(S)$, then it follows that $\sigma \circ \tau$ is also in $A(S)$. This is described by saying that $A(S)$ is *closed* under the product (or, sometimes, as closed under multiplication).
2. For any three elements $\sigma, \tau, \mu \in A(S)$, $\sigma \circ (\tau \circ \mu) = (\sigma \circ \tau) \circ \mu$. This relation is called the *associative law*.
3. There is a very special element $\iota \in A(S)$ which satisfies $\iota \circ \sigma = \sigma \circ \iota = \sigma$ for all $\sigma \in A(S)$. Such an element is called an *identity element* for $A(S)$.
4. For every $\sigma \in A(S)$ there is an element, written as σ^{-1}, also in $A(S)$, such that $\sigma \circ \sigma^{-1} = \sigma^{-1} \circ \sigma = \iota$. This is usually described by saying that every element in $A(S)$ has an *inverse* in $A(S)$.

One other fact about $A(S)$ stands out, namely, that whenever S has three or more elements we can find two elements $\alpha, \beta \in A(S)$ such that $\alpha \circ \beta \neq \beta \circ \alpha$. This possibility, which runs counter to our usual experience and intuition in mathematics so far, introduces a richness into $A(S)$ which would have not been present except for it.

With this example as a model, and with a great deal of hindsight, we abstract and make the

DEFINITION A nonempty set of elements G is said to form a *group* if in G there is defined a binary operation, called the product and denoted by $\cdot$, such that

1. $a, b \in G$ implies that $a \cdot b \in G$ (closed).
2. $a, b, c \in G$ implies that $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ (associative law).
3. There exists an element $e \in G$ such that $a \cdot e = e \cdot a = a$ for all $a \in G$ (the existence of an identity element in G).
4. For every $a \in G$ there exists an element $a^{-1} \in G$ such that $a \cdot a^{-1} = a^{-1} \cdot a = e$ (the existence of inverses in G).

Considering the source of this definition it is not surprising that for every nonempty set S the set $A(S)$ is a group. Thus we already have presented to us an infinite source of interesting, concrete groups. We shall see later (in a theorem due to Cayley) that these $A(S)$'s constitute, in some sense, a universal family of groups. If S has three or more elements, recall that we can find elements $\sigma, \tau \in A(S)$ such that $\sigma \circ \tau \neq \tau \circ \sigma$. This prompts us to single out a highly special, but very important, class of groups as in the next definition.

DEFINITION A group G is said to be *abelian* (or *commutative*) if for every $a, b \in G$, $a \cdot b = b \cdot a$.

A group which is not abelian is called, naturally enough, *non-abelian*; having seen a family of examples of such groups we know that non-abelian groups do indeed exist.

Another natural characteristic of a group G is the number of elements it contains. We call this the *order* of G and denote it by $o(G)$. This number is, of course, most interesting when it is finite. In that case we say that G is a *finite group*.

To see that finite groups which are not trivial do exist just note that if the set S contains n elements, then the group $A(S)$ has $n!$ elements. (Prove!) This highly important example will be denoted by S_n whenever it appears in this book, and will be called the *symmetric group* of degree n. In the next section we shall more or less dissect S_3, which is a non-abelian group of order 6.

2.2 Some Examples of Groups

Example 2.2.1 Let G consist of the integers $0, \pm 1, \pm 2, \ldots$ where we mean by $a \cdot b$ for $a, b \in G$ the usual sum of integers, that is, $a \cdot b = a + b$. Then the reader can quickly verify that G is an infinite abelian group in which 0 plays the role of e and $-a$ that of a^{-1}.

Example 2.2.2 Let G consist of the real numbers $1, -1$ under the multiplication of real numbers. G is then an abelian group of order 2.

Example 2.2.3 Let $G = S_3$, the group of all 1–1 mappings of the set $\{x_1, x_2, x_3\}$ onto itself, under the product which we defined in Chapter 1. G is a group of order 6. We digress a little before returning to S_3.

For a neater notation, not just in S_3, but in any group G, let us define for any $a \in G$, $a^0 = e$, $a^1 = a$, $a^2 = a \cdot a$, $a^3 = a \cdot a^2, \ldots, a^k = a \cdot a^{k-1}$, and $a^{-2} = (a^{-1})^2$, $a^{-3} = (a^{-1})^3$, etc. The reader may verify that the usual rules of exponents prevail; namely, for any two integers (positive, negative, or zero) m, n,

$$a^m \cdot a^n = a^{m+n}, \tag{1}$$

$$(a^m)^n = a^{mn}. \tag{2}$$

(It is worthwhile noting that, in this notation, if G is the group of Example 2.2.1, a^n means the integer na).

With this notation at our disposal let us examine S_3 more closely. Consider the mapping ϕ defined on the set x_1, x_2, x_3 by

$$\phi: \quad \begin{aligned} x_1 &\to x_2 \\ x_2 &\to x_1 \\ x_3 &\to x_3, \end{aligned}$$

and the mapping

$$\psi: \quad \begin{aligned} x_1 &\to x_2 \\ x_2 &\to x_3 \\ x_3 &\to x_1. \end{aligned}$$

Checking, we readily see that $\phi^2 = e$, $\psi^3 = e$, and that

$$\phi \cdot \psi: \quad \begin{aligned} x_1 &\to x_3 \\ x_2 &\to x_2 \\ x_3 &\to x_1, \end{aligned}$$

whereas

$$\psi \cdot \phi: \quad \begin{aligned} x_1 &\to x_1 \\ x_2 &\to x_3 \\ x_3 &\to x_2. \end{aligned}$$

It is clear that $\phi\cdot\psi \neq \psi\cdot\phi$ for they do not take x_1 into the same image. Since $\psi^3 = e$, it follows that $\psi^{-1} = \psi^2$. Let us now compute the action of $\psi^{-1}\cdot\phi$ on x_1, x_2, x_3. Since $\psi^{-1} = \psi^2$ and

$$\psi^2: \quad \begin{aligned} x_1 &\to x_3 \\ x_2 &\to x_1 \\ x_3 &\to x_2, \end{aligned}$$

we have that

$$\psi^{-1}\cdot\phi: \quad \begin{aligned} x_1 &\to x_3 \\ x_2 &\to x_2 \\ x_3 &\to x_1. \end{aligned}$$

In other words, $\phi\cdot\psi = \psi^{-1}\cdot\phi$. Consider the elements $e, \phi, \psi, \psi^2, \phi\cdot\psi, \psi\cdot\phi$; these are all distinct and are in G (since G is closed), which only has six elements. Thus this list enumerates all the elements of G. One might ask, for instance, What is the entry in the list for $\psi\cdot(\phi\cdot\psi)$? Using $\phi\cdot\psi = \psi^{-1}\cdot\phi$, we see that $\psi\cdot(\phi\cdot\psi) = \psi\cdot(\psi^{-1}\cdot\phi) = (\psi\cdot\psi^{-1})\cdot\phi = e\cdot\phi = \phi$. Of more interest is the form of $(\phi\cdot\psi)\cdot(\psi\cdot\phi) = \phi\cdot(\psi\cdot(\psi\cdot\phi)) = \phi\cdot(\psi^2\cdot\phi) = \phi\cdot(\psi^{-1}\cdot\phi) = \phi\cdot(\phi\cdot\psi) = \phi^2\cdot\psi = e\cdot\psi = \psi$. (The reader should not be frightened by the long, wearisome chain of equalities here. It is the last time we shall be so boringly conscientious.) Using the same techniques as we have used, the reader can compute to his heart's content others of the 25 products which do not involve e. Some of these will appear in the exercises.

Example 2.2.4 Let n be any integer. We construct a group of order n as follows: G will consist of all symbols a^i, $i = 0, 1, 2, \ldots, n-1$ where we insist that $a^0 = a^n = e$, $a^i\cdot a^j = a^{i+j}$ if $i + j \leq n$ and $a^i\cdot a^j = a^{i+j-n}$ if $i + j > n$. The reader may verify that this is a group. It is called a *cyclic group* of order n.

A geometric realization of the group in Example 2.2.4 may be achieved as follows: Let S be the circle, in the plane, of radius 1, and let ρ_n be a rotation through an angle of $2\pi/n$. Then $\rho_n \in A(S)$ and ρ_n in $A(S)$ generates a group of order n, namely, $\{e, \rho_n, \rho_n^2, \ldots, \rho_n^{n-1}\}$.

Example 2.2.5 Let S be the set of integers and, as usual, let $A(S)$ be the set of all one-to-one mappings of S onto itself. Let G be the set of all elements in $A(S)$ which move only a *finite* number of elements of S; that is, $\sigma \in G$ if and only if the number of x in S such that $x\sigma \neq x$ is finite. If $\sigma, \tau \in G$, let $\sigma\cdot\tau$ be the product of σ and τ as elements of $A(S)$. We claim that G is a group relative to this operation. We verify this now.

To begin with, if $\sigma, \tau \in G$, then σ and τ each moves only a finite number of elements of S. In consequence, $\sigma\cdot\tau$ can possibly move only those elements in S which are moved by at least one of σ or τ. Hence $\sigma\cdot\tau$ moves only a

finite number of elements in S; this puts $\sigma \cdot \tau$ in G. The identity element, ι, of $A(S)$ moves no element of S; thus ι certainly must be in G. Since the associative law holds universally in $A(S)$, it holds for elements of G. Finally, if $\sigma \in G$ and $x\sigma^{-1} \neq x$ for some $x \in S$, then $(x\sigma^{-1})\sigma \neq x\sigma$, which is to say, $x(\sigma^{-1} \cdot \sigma) \neq x\sigma$. This works out to say merely that $x \neq x\sigma$. In other words, σ^{-1} moves only those elements of S which are moved by σ. Because σ only moves a finite number of elements of S, this is also true for σ^{-1}. Therefore σ^{-1} must be in G.

We have verified that G satisfies the requisite four axioms which define a group, relative to the operation we specified. Thus G is a group. The reader should verify that G is an infinite, non-abelian group.

#Example 2.2.6 Let G be the set of all 2×2 matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ where a, b, c, d are real numbers, such that $ad - bc \neq 0$. For the operation in G we use the multiplication of matrices; that is,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} w & x \\ y & z \end{pmatrix} = \begin{pmatrix} aw + by & ax + bz \\ cw + dy & cx + dz \end{pmatrix}.$$

The entries of this 2×2 matrix are clearly real. To see that this matrix is in G we merely must show that

$$(aw + by)(cx + dz) - (ax + bz)(cw + dy) \neq 0$$

(this is the required relation on the entries of a matrix which puts it in G). A short computation reveals that

$$(aw + by)(cx + dz) - (ax + bz)(cw + dy) = (ad - bc)(wz - xy) \neq 0$$

since both

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} w & x \\ y & z \end{pmatrix}$$

are in G. The associative law of multiplication holds in matrices; therefore it holds in G. The element

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

is in G, since $1 \cdot 1 - 0 \cdot 0 = 1 \neq 0$; moreover, as the reader knows, or can verify, I acts as an identity element relative to the operation of G. Finally, if $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G$ then, since $ad - bc \neq 0$, the matrix

$$\begin{pmatrix} \dfrac{d}{ad - bc} & \dfrac{-b}{ad - bc} \\ \dfrac{-c}{ad - bc} & \dfrac{a}{ad - bc} \end{pmatrix}$$

makes sense. Moreover,

$$\left(\frac{d}{ad-bc}\right)\left(\frac{a}{ad-bc}\right)-\left(\frac{-b}{ad-bc}\right)\left(\frac{-c}{ad-bc}\right)=\frac{ad-bc}{(ad-bc)^2}=\frac{1}{ad-bc}\neq 0,$$

hence the matrix

$$\begin{pmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix}$$

is in G. An easy computation shows that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix}=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}=\begin{pmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix};$$

thus this element of G acts as the inverse of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. In short, G is a group. It is easy to see that G is an infinite, non-abelian group.

#Example 2.2.7 Let G be the set of all 2×2 matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, where a, b, c, d are real numbers such that $ad - bc = 1$. Define the operation $\cdot$ in G, as we did in Example 2.2.6, via the multiplication of matrices. We leave it to the reader to verify that G is a group. It is, in fact, an infinite, non-abelian group.

One should make a comment about the relationship of the group in Example 2.2.7 to that in Example 2.2.6. Clearly, the group of Example 2.2.7 is a subset of that in Example 2.2.6. However, more is true. Relative to the same operation, as an entity in its own right, it forms a group. One could describe the situation by declaring it to be a *subgroup* of the group of Example 2.2.6. We shall see much more about the concept of subgroup in a few pages.

#Example 2.2.8 Let G be the set of all 2×2 matrices $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, where a and b are real numbers, not both 0. (We can state this more succinctly by saying that $a^2 + b^2 \neq 0$.) Using the same operation as in the preceding two examples, we can easily show that G becomes a group. In fact, G is an infinite, abelian group.

Does the multiplication in G remind you of anything? Write $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ as $aI + bJ$ where $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and compute the product in these terms. Perhaps that will ring a bell with you.

#Example 2.2.9 Let G be the set of all 2×2 matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ where a, b, c, d are integers modulo p, p a prime number, such that $ad - bc \neq 0$. Define the multiplication in G as we did in Example 2.2.6, understanding the multiplication and addition of the entries to be those modulo p. We leave it to the reader to verify that G is a non-abelian *finite* group.

In fact, how many elements does G have? Perhaps it might be instructive for the reader to try the early cases $p = 2$ and $p = 3$. Here one can write down all the elements of G explicitly. (A word of warning! For $p = 3$, G already has 48 elements.) To get the case of a general prime, p will require an idea rather than a direct hacking-out of the answer. Try it!

2.3 Some Preliminary Lemmas

We have now been exposed to the theory of groups for several pages and as yet not a single, solitary fact has been proved about groups. It is high time to remedy this situation. Although the first few results we demonstrate are, admittedly, not very exciting (in fact, they are rather dull) they will be extremely useful. Learning the alphabet was probably not the most interesting part of our childhood education, yet, once this hurdle was cleared, fascinating vistas were opened before us.

We begin with

LEMMA 2.3.1 *If G is a group, then*

a. The identity element of G is unique.
b. Every $a \in G$ has a unique inverse in G.
c. For every $a \in G$, $(a^{-1})^{-1} = a$.
d. For all $a, b \in G$, $(a \cdot b)^{-1} = b^{-1} \cdot a^{-1}$.

Proof. Before we proceed with the proof itself it might be advisable to see what it is that we are going to prove. In part (a) we want to show that if two elements e and f in G enjoy the property that for every $a \in G$, $a = a \cdot e = e \cdot a = a \cdot f = f \cdot a$, then $e = f$. In part (b) our aim is to show that if $x \cdot a = a \cdot x = e$ and $y \cdot a = a \cdot y = e$, where all of a, x, y are in G, then $x = y$.

First let us consider part (a). Since $e \cdot a = a$ for every $a \in G$, then, in particular, $e \cdot f = f$. But, on the other hand, since $b \cdot f = b$ for every $b \in G$, we must have that $e \cdot f = e$. Piecing these two bits of information together we obtain $f = e \cdot f = e$, and so $e = f$.

Rather than proving part (b), we shall prove something stronger which immediately will imply part (b) as a consequence. Suppose that for a in G, $a \cdot x = e$ and $a \cdot y = e$; then, obviously, $a \cdot x = a \cdot y$. Let us make this our starting point, that is, assume that $a \cdot x = a \cdot y$ for a, x, y in G. There is an element $b \in G$ such that $b \cdot a = e$ (as far as we know yet there may be several such b's). Thus $b \cdot (a \cdot x) = b \cdot (a \cdot y)$; using the associative law this leads to

$$x = e \cdot x = (b \cdot a) \cdot x = b \cdot (a \cdot x) = b \cdot (a \cdot y) = (b \cdot a) \cdot y = e \cdot y = y.$$

We have, in fact, proved that $a \cdot x = a \cdot y$ in a group G forces $x = y$. Similarly we can prove that $x \cdot a = y \cdot a$ implies that $x = y$. This says that we can cancel, from the same side, in equations in groups. A note of caution, however, for we cannot conclude that $a \cdot x = y \cdot a$ implies $x = y$ for we have no way of knowing whether $a \cdot x = x \cdot a$. This is illustrated in S_3 with $a = \phi$, $x = \psi$, $y = \psi^{-1}$.

Part (c) follows from this by noting that $a^{-1} \cdot (a^{-1})^{-1} = e = a^{-1} \cdot a$; canceling off the a^{-1} on the left leaves us with $(a^{-1})^{-1} = a$. This is the analog in general groups of the familiar result $-(-5) = 5$, say, in the group of real numbers under addition.

Part (d) is the most trivial of these, for

$$(a \cdot b) \cdot (b^{-1} \cdot a^{-1}) = a \cdot ((b \cdot b^{-1}) \cdot a^{-1}) = a \cdot (e \cdot a^{-1}) = a \cdot a^{-1} = e,$$

and so by the very definition of the inverse, $(a \cdot b)^{-1} = b^{-1} \cdot a^{-1}$.

Certain results obtained in the proof just given are important enough to single out and we do so now in

LEMMA 2.3.2 *Given a, b in the group G, then the equations $a \cdot x = b$ and $y \cdot a = b$ have unique solutions for x and y in G. In particular, the two cancellation laws,*

$$a \cdot u = a \cdot w \text{ implies } u = w$$

and

$$u \cdot a = w \cdot a \text{ implies } u = w$$

hold in G.

The few details needed for the proof of this lemma are left to the reader.

Problems

1. In the following determine whether the systems described are groups. If they are not, point out which of the group axioms fail to hold.
   (a) G = set of all integers, $a \cdot b \equiv a - b$.
   (b) G = set of all positive integers, $a \cdot b = ab$, the usual product of integers.
   (c) $G = a_0, a_1, \ldots, a_6$ where

$$a_i \cdot a_j = a_{i+j} \quad \text{if} \quad i + j < 7,$$

$$a_i \cdot a_j = a_{i+j-7} \quad \text{if} \quad i + j \geq 7$$

   (for instance, $a_5 \cdot a_4 = a_{5+4-7} = a_2$ since $5 + 4 = 9 > 7$).
   (d) G = set of all rational numbers with odd denominators, $a \cdot b \equiv a + b$, the usual addition of rational numbers.
2. Prove that if G is an abelian group, then for all $a, b \in G$ and all integers n, $(a \cdot b)^n = a^n \cdot b^n$.
3. If G is a group such that $(a \cdot b)^2 = a^2 \cdot b^2$ for all $a, b \in G$, show that G must be abelian.
*4. If G is a group in which $(a \cdot b)^i = a^i \cdot b^i$ for three consecutive integers i for all $a, b \in G$, show that G is abelian.
5. Show that the conclusion of Problem 4 does not follow if we assume the relation $(a \cdot b)^i = a^i \cdot b^i$ for just two consecutive integers.
6. In S_3 give an example of two elements x, y such that $(x \cdot y)^2 \neq x^2 \cdot y^2$.
7. In S_3 show that there are four elements satisfying $x^2 = e$ and three elements satisfying $y^3 = e$.
8. If G is a finite group, show that there exists a positive integer N such that $a^N = e$ for all $a \in G$.
9. (a) If the group G has three elements, show it must be abelian.
   (b) Do part (a) if G has four elements.
   (c) Do part (a) if G has five elements.
10. Show that if every element of the group G is its own inverse, then G is abelian.
11. If G is a group of even order, prove it has an element $a \neq e$ satisfying $a^2 = e$.
12. Let G be a nonempty set closed under an associative product, which in addition satisfies:
   (a) There exists an $e \in G$ such that $a \cdot e = a$ for all $a \in G$.
   (b) Give $a \in G$, there exists an element $y(a) \in G$ such that $a \cdot y(a) = e$.
   Prove that G must be a group under this product.

13. Prove, by an example, that the conclusion of Problem 12 is false if we assume instead:
    (a′) There exists an $e \in G$ such that $a \cdot e = a$ for all $a \in G$.
    (b′) Given $a \in G$, there exists $y(a) \in G$ such that $y(a) \cdot a = e$.
14. Suppose a *finite* set G is closed under an associative product and that both cancellation laws hold in G. Prove that G must be a group.
15. (a) Using the result of Problem 14, prove that the nonzero integers modulo p, p a prime number, form a group under multiplication mod p.
    (b) Do part (a) for the nonzero integers relatively prime to n under multiplication mod n.
16. In Problem 14 show by an example that if one just assumed one of the cancellation laws, then the conclusion need not follow.
17. Prove that in Problem 14 infinite examples exist, satisfying the conditions, which are not groups.
18. For any $n > 2$ construct a non-abelian group of order $2n$. (*Hint:* imitate the relations in S_3.)
19. If S is a set closed under an associative operation, prove that no matter how you bracket $a_1 a_2 \cdots a_n$, retaining the order of the elements, you get the same element in S (e.g., $(a_1 \cdot a_2) \cdot (a_3 \cdot a_4) = a_1 \cdot (a_2 \cdot (a_3 \cdot a_4))$; use induction on n).

#20. Let G be the set of all real 2×2 matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, where $ad - bc \neq 0$ is a rational number. Prove that G forms a group under matrix multiplication.

#21. Let G be the set of all real 2×2 matrices $\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$ where $ad \neq 0$. Prove that G forms a group under matrix multiplication. Is G abelian?

#22. Let G be the set of all real 2×2 matrices $\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}$ where $a \neq 0$. Prove that G is an abelian group under matrix multiplication.

#23. Construct in the G of Problem 21 a subgroup of order 4.

#24. Let G be the set of all 2×2 matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ where a, b, c, d are integers modulo 2, such that $ad - bc \neq 0$. Using matrix multiplication as the operation in G, prove that G is a group of order 6.

#25. (a) Let G be the group of all 2×2 matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ where $ad - bc \neq 0$ and a, b, c, d are integers modulo 3, relative to matrix multiplication. Show that $o(G) = 48$.

(b) If we modify the example of G in part (a) by insisting that $ad - bc = 1$, then what is $o(G)$?

#*26. (a) Let G be the group of all 2×2 matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ where a, b, c, d are integers modulo p, p a prime number, such that $ad - bc \neq 0$. G forms a group relative to matrix multiplication. What is $o(G)$?

(b) Let H be the subgroup of the G of part (a) defined by

$$H = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G \mid ad - bc = 1 \right\}.$$

What is $o(H)$?

2.4 Subgroups

Before turning to the study of groups we should like to change our notation slightly. It is cumbersome to keep using the $\cdot$ for the group operation; henceforth we shall drop it and instead of writing $a \cdot b$ for $a, b \in G$ we shall simply denote this product as ab.

In general we shall not be interested in arbitrary subsets of a group G for they do not reflect the fact that G has an algebraic structure imposed on it. Whatever subsets we do consider will be those endowed with algebraic properties derived from those of G. The most natural such subsets are introduced in the

DEFINITION A nonempty subset H of a group G is said to be a *subgroup* of G if, under the product in G, H itself forms a group.

The following remark is clear: if H is a subgroup of G and K is a subgroup of H, then K is a subgroup of G.

It would be useful to have some criterion for deciding whether a given subset of a group is a subgroup. This is the purpose of the next two lemmas.

LEMMA 2.4.1 *A nonempty subset H of the group G is a subgroup of G if and only if*

1. *$a, b \in H$ implies that $ab \in H$.*
2. *$a \in H$ implies that $a^{-1} \in H$.*

Proof. If H is a subgroup of G, then it is obvious that (1) and (2) must hold.

Suppose conversely that H is a subset of G for which (1) and (2) hold. In order to establish that H is a subgroup, all that is needed is to verify that $e \in H$ and that the associative law holds for elements of H. Since the associative law does hold for G, it holds all the more so for H, which is a

subset of G. If $a \in H$, by part 2, $a^{-1} \in H$ and so by part 1, $e = aa^{-1} \in H$. This completes the proof.

In the special case of a finite group the situation becomes even nicer for there we can dispense with part 2.

LEMMA 2.4.2 *If H is a nonempty finite subset of a group G and H is closed under multiplication, then H is a subgroup of G.*

Proof. In light of Lemma 2.4.1 we need but show that whenever $a \in H$, then $a^{-1} \in H$. Suppose that $a \in H$; thus $a^2 = aa \in H$, $a^3 = a^2a \in H$, $\ldots$, $a^m \in H, \ldots$ since H is closed. Thus the infinite collection of elements $a, a^2, \ldots, a^m, \ldots$ must all fit into H, which is a finite subset of G. Thus there must be repetitions in this collection of elements; that is, for some integers r, s with $r > s > 0$, $a^r = a^s$. By the cancellation in G, $a^{r-s} = e$ (whence e is in H); since $r - s - 1 \geq 0$, $a^{r-s-1} \in H$ and $a^{-1} = a^{r-s-1}$ since $aa^{r-s-1} = a^{r-s} = e$. Thus $a^{-1} \in H$, completing the proof of the lemma.

The lemma tells us that to check whether a subset of a finite group is a subgroup we just see whether or not it is closed under multiplication.

We should, perhaps, now see some groups and some of their subgroups. G is always a subgroup of itself; likewise the set consisting of e is a subgroup of G. Neither is particularly interesting in the role of a subgroup, so we describe them as trivial subgroups. The subgroups between these two extremes we call nontrivial subgroups and it is in these we shall exhibit the most interest.

Example 2.4.1 Let G be the group of integers under addition, H the subset consisting of all the multiples of 5. The student should check that H is a subgroup.

In this example there is nothing extraordinary about 5; we could similarly define the subgroup H_n as the subset of G consisting of all the multiples of n. H_n is then a subgroup for every n. What can one say about $H_n \cap H_m$? It might be wise to try it for $H_6 \cap H_9$.

Example 2.4.2 Let S be any set, $A(S)$ the set of one-to-one mappings of S onto itself, made into a group under the composition of mappings. If $x_0 \in S$, let $H(x_0) = \{\phi \in A(S) \mid x_0\phi = x_0\}$. $H(x_0)$ is a subgroup of $A(S)$. If for $x_1 \neq x_0 \in S$ we similarly define $H(x_1)$, what is $H(x_0) \cap H(x_1)$?

Example 2.4.3 Let G be any group, $a \in G$. Let $(a) = \{a^i \mid i = 0, \pm 1, \pm 2, \ldots\}$. (a) is a subgroup of G (verify!); it is called the *cyclic subgroup generated by a*. This provides us with a ready means of producing subgroups

of G. If for some choice of a, $G = (a)$, then G is said to be a *cyclic group*. Such groups are very special but they play a very important role in the theory of groups, especially in that part which deals with abelian groups. Of course, cyclic groups are abelian, but the converse is false.

Example 2.4.4 Let G be a group, W a subset of G. Let (W) be the set of all elements of G representable as a product of elements of W raised to positive, zero, or negative integer exponents. (W) is the *subgroup of G generated by W* and is the smallest subgroup of G containing W. In fact, (W) is the intersection of all the subgroups of G which contain W (this intersection is not vacuous since G is a subgroup of G which contains W).

Example 2.4.5 Let G be the group of nonzero real numbers under multiplication, and let H be the subset of positive rational numbers. Then H is a subgroup of G.

Example 2.4.6 Let G be the group of all real numbers under addition, and let H be the set of all integers. Then H is a subgroup of G.

#Example 2.4.7 Let G be the group of all real 2×2 matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $ad - bc \neq 0$ under matrix multiplication. Let

$$H = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in G \mid ad \neq 0 \right\}.$$

Then, as is easily verified, H is a subgroup of G.

#Example 2.4.8 Let H be the group of Example 2.4.7, and let $K = \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \right\}$. Then K is a subgroup of H.

Example 2.4.9 Let G be the group of all nonzero complex numbers $a + bi$ (a, b real, not both 0) under multiplication, and let

$$H = \{a + bi \in G \mid a^2 + b^2 = 1\}.$$

Verify that H is a subgroup of G.

DEFINITION Let G be a group, H a subgroup of G; for $a, b \in G$ we say *a is congruent to b* mod H, written as $a \equiv b \bmod H$ if $ab^{-1} \in H$.

LEMMA 2.4.3 *The relation $a \equiv b \bmod H$ is an equivalence relation.*

Proof. If we look back in Chapter 1, we see that to prove Lemma 2.4.3 we must verify the following three conditions: For all $a, b, c \in G$,

1. $a \equiv a \bmod H$.
2. $a \equiv b \bmod H$ implies $b \equiv a \bmod H$.
3. $a \equiv b \bmod H$, $b \equiv c \bmod H$ implies $a \equiv c \bmod H$.

Let's go through each of these in turn.

1. To show that $a \equiv a \bmod H$ we must prove, using the very definition of congruence mod H, that $aa^{-1} \in H$. Since H is a subgroup of G, $e \in H$, and since $aa^{-1} = e$, $aa^{-1} \in H$, which is what we were required to demonstrate.

2. Suppose that $a \equiv b \bmod H$, that is, suppose $ab^{-1} \in H$; we want to get from this $b \equiv a \bmod H$, or, equivalently, $ba^{-1} \in H$. Since $ab^{-1} \in H$, which is a subgroup of G, $(ab^{-1})^{-1} \in H$; but, by Lemma 2.3.1, $(ab^{-1})^{-1} = (b^{-1})^{-1}a^{-1} = ba^{-1}$, and so $ba^{-1} \in H$ and $b \equiv a \bmod H$.

3. Finally we require that $a \equiv b \bmod H$ and $b \equiv c \bmod H$ forces $a \equiv c \bmod H$. The first congruence translates into $ab^{-1} \in H$, the second into $bc^{-1} \in H$; using that H is a subgroup of G, $(ab^{-1})(bc^{-1}) \in H$. However, $ac^{-1} = aec^{-1} = a(b^{-1}b)c^{-1} = (ab^{-1})(bc^{-1})$; hence $ac^{-1} \in H$, from which it follows that $a \equiv c \bmod H$.

This establishes that congruence mod H is a bona fide equivalence relation as defined in Chapter 1, and all results about equivalence relations have become available to us to be used in examining this particular relation.

A word about the notation we used. If G were the group of integers under addition, and $H = H_n$ were the subgroup consisting of all multiples of n, then in G, the relation $a \equiv b \bmod H$, that is, $ab^{-1} \in H$, under the additive notation, reads "$a - b$ is a multiple of n." This is the usual number theoretic congruence mod n. In other words, the relation we defined using an arbitrary group and subgroup is the natural generalization of a familiar relation in a familiar group.

DEFINITION If H is a subgroup of G, $a \in G$, then $Ha = \{ha \mid h \in H\}$. Ha is called a *right coset* of H in G.

LEMMA 2.4.4 *For all* $a \in G$,

$$Ha = \{x \in G \mid a \equiv x \bmod H\}.$$

Proof. Let $[a] = \{x \in G \mid a \equiv x \bmod H\}$. We first show that $Ha \subset [a]$. For, if $h \in H$, then $a(ha)^{-1} = a(a^{-1}h^{-1}) = h^{-1} \in H$ since H is a subgroup of G. By the definition of congruence mod H this implies that $ha \in [a]$ for every $h \in H$, and so $Ha \subset [a]$.

Suppose, now, that $x \in [a]$. Thus $ax^{-1} \in H$, so $(ax^{-1})^{-1} = xa^{-1}$ is

also in H. That is, $xa^{-1} = h$ for some $h \in H$. Multiplying both sides by a from the right we come up with $x = ha$, and so $x \in Ha$. Thus $[a] \subset Ha$. Having proved the two inclusions $[a] \subset Ha$ and $Ha \subset [a]$, we can conclude that $[a] = Ha$, which is the assertion of the lemma.

In the terminology of Chapter 1, $[a]$, and thus Ha, is the equivalence class of a in G. By Theorem 1.1.1 these equivalence classes yield a decomposition of G into disjoint subsets. *Thus any two right cosets of H in G either are identical or have no element in common.*

We now claim that between any two right cosets Ha and Hb of H in G there exists a one-to-one correspondence, namely, with any element $ha \in Ha$, where $h \in H$, associate the element $hb \in Hb$. Clearly this mapping is onto Hb. We aver that it is a one-to-one correspondence, for if $h_1 b = h_2 b$, with $h_1, h_2 \in H$, then by the cancellation law in G, $h_1 = h_2$ and so $h_1 a = h_2 a$. This proves

LEMMA 2.4.5 *There is a one-to-one correspondence between any two right cosets of H in G.*

Lemma 2.4.5 is of most interest when H is a finite group, for then it merely states that any two right cosets of H have the same number of elements. How many elements does a right coset of H have? Well, note that $H = He$ is itself a right coset of H, so any right coset of H in G has $o(H)$ elements. Suppose now that G is a finite group, and let k be the number of distinct right cosets of H in G. By Lemmas 2.4.4 and 2.4.5 any two distinct right cosets of H in G have no element in common, and each has $o(H)$ elements.

Since any $a \in G$ is in the unique right coset Ha, the right cosets fill out G. Thus if k represents the number of distinct right cosets of H in G we must have that $ko(H) = o(G)$. We have proved the famous theorem due to Lagrange, namely,

THEOREM 2.4.1 *If G is a finite group and H is a subgroup of G, then $o(H)$ is a divisor of $o(G)$.*

DEFINITION If H is a subgroup of G, the *index of H in G* is the number of distinct right cosets of H in G.

We shall denote it by $i_G(H)$. In case G is a finite group, $i_G(H) = o(G)/o(H)$, as became clear in the proof of Lagrange's theorem. It is quite possible for an infinite group G to have a subgroup $H \neq G$ which is of finite index in G.

It might be difficult, at this point, for the student to see the extreme importance of this result. As the subject is penetrated more deeply one will

become more and more aware of its basic character. Because the theorem is of such stature it merits a little closer scrutiny, a little more analysis, and so we give, below, a slightly different way of looking at its proof. In truth, the procedure outlined below is no different from the one already given. The introduction of the congruence mod H smooths out the listing of elements used below, and obviates the need for checking that the new elements introduced at each stage did not appear before.

So suppose again that G is a finite group and that H is a subgroup of G. Let $h_1, h_2, \ldots, h_r$ be a complete list of the elements of H, $r = o(H)$. If $H = G$, there is nothing to prove. Suppose, then, that $H \neq G$; thus there is an $a \in G$, $a \notin H$. List all the elements so far in two rows as

$$h_1, h_2, \ldots, h_r,$$
$$h_1a, h_2a, \ldots, h_ra.$$

We claim that all the entries in the second line are different from each other and are different from the entries in the first line. If any two in the second line were equal, then $h_ia = h_ja$ with $i \neq j$, but by the cancellation law this would lead to $h_i = h_j$, a contradiction. If an entry in the second line were equal to one in the first line, then $h_ia = h_j$, resulting in $a = h_i^{-1}h_j \in H$ since H is a subgroup of G; this violates $a \notin H$.

Thus we have, so far, listed $2o(H)$ elements; if these elements account for all the elements of G, we are done. If not, there is a $b \in G$ which did not occur in these two lines. Consider the new list

$$h_1, h_2, \ldots, h_r,$$
$$h_1a, h_2a, \ldots, h_ra,$$
$$h_1b, h_2b, \ldots, h_rb.$$

As before (we are now waving our hands) we could show that no two entries in the third line are equal to each other, and that no entry in the third line occurs in the first or second line. Thus we have listed $3o(H)$ elements. Continuing in this way, every new element introduced, in fact, produces $o(H)$ new elements. Since G is a finite group, we must eventually exhaust all the elements of G. But if we ended up using k lines to list all the elements of the group, we would have written down $ko(H)$ distinct elements, and so $ko(H) = o(G)$.

It is essential to point out that the converse to Lagrange's theorem is false—a group G need not have a subgroup of order m if m is a divisor of $o(G)$. For instance, a group of order 12 exists which has no subgroup of order 6. The reader might try to find an example of this phenomenon; the place to look is in S_4, the symmetric group of degree 4 which has a subgroup of order 12, which will fulfill our requirement.

Lagrange's theorem has some very important corollaries. Before we present these we make one definition.

DEFINITION If G is a group and $a \in G$, the *order* (or *period*) of a is the least positive integer m such that $a^m = e$.

If no such integer exists we say that a is of infinite order. We use the notation $o(a)$ for the order of a. Recall our other notation: for two integers u, v, $u \mid v$ reads "u is a divisor of v."

COROLLARY 1 *If G is a finite group and $a \in G$, then $o(a) \mid o(G)$.*

Proof. With Lagrange's theorem already in hand, it seems most natural to prove the corollary by exhibiting a subgroup of G whose order is $o(a)$. The element a itself furnishes us with this subgroup by considering the cyclic subgroup, (a), of G generated by a; (a) consists of $e, a, a^2, \ldots$. How many elements are there in (a)? We assert that this number is the order of a. Clearly, since $a^{o(a)} = e$, this subgroup has at most $o(a)$ elements. If it should actually have fewer than this number of elements, then $a^i = a^j$ for some integers $0 \leq i < j < o(a)$. Then $a^{j-i} = e$, yet $0 < j - i < o(a)$ which would contradict the very meaning of $o(a)$. Thus the cyclic subgroup generated by a has $o(a)$ elements, whence, by Lagrange's theorem, $o(a) \mid o(G)$.

COROLLARY 2 *If G is a finite group and $a \in G$, then $a^{o(G)} = e$.*

Proof. By Corollary 1, $o(a) \mid o(G)$; thus $o(G) = mo(a)$. Therefore, $a^{o(G)} = a^{mo(a)} = (a^{o(a)})^m = e^m = e$.

A particular case of Corollary 2 is of great interest in number theory. The Euler ϕ-function, $\phi(n)$, is defined for all integers n by the following: $\phi(1) = 1$; for $n > 1$, $\phi(n)$ = number of positive integers less than n and relatively prime to n. Thus, for instance, $\phi(8) = 4$ since only $1, 3, 5, 7$ are the numbers less than 8 which are relatively prime to 8. In Problem 15(b) at the end of Section 2.3 the reader was asked to prove that the numbers less than n and relatively prime to n formed a group under multiplication mod n. This group has order $\phi(n)$. If we apply Corollary 2 to this group we obtain

COROLLARY 3 (EULER) *If n is a positive integer and a is relatively prime to n, then $a^{\phi(n)} \equiv 1 \bmod n$.*

In order to apply Corollary 2 one should replace a by its remainder on division by n. If n should be a prime number p, then $\phi(p) = p - 1$. If a is an integer relatively prime to p, then by Corollary 3, $a^{p-1} \equiv 1 \bmod p$, whence $a^p \equiv a \bmod p$. If, on the other hand, a is not relatively prime to p,

since p is a prime number, we must have that $p \mid a$, so that $a \equiv 0 \bmod p$; hence $0 \equiv a^p \equiv a \bmod p$ here also. Thus

COROLLARY 4 (FERMAT) *If p is a prime number and a is any integer, then $a^p \equiv a \bmod p$.*

COROLLARY 5 *If G is a finite group whose order is a prime number p, then G is a cyclic group.*

Proof. First we claim that G has no nontrivial subgroups H; for $o(H)$ must divide $o(G) = p$ leaving only two possibilities, namely, $o(H) = 1$ or $o(H) = p$. The first of these implies $H = (e)$, whereas the second implies that $H = G$. Suppose now that $a \neq e \in G$, and let $H = (a)$. H is a subgroup of G, $H \neq (e)$ since $a \neq e \in H$. Thus $H = G$. This says that G is cyclic and that every element in G is a power of a.

This section is of great importance in all that comes later, not only for its results but also because the spirit of the proofs occurring here are genuinely group-theoretic. The student can expect to encounter other arguments having a similar flavor. It would be wise to assimilate the material and approach thoroughly, now, rather than a few theorems later when it will be too late.

2.5 A Counting Principle

As we have defined earlier, if H is a subgroup of G and $a \in G$, then Ha consists of all elements in G of the form ha where $h \in H$. Let us generalize this notion. If H, K are two subgroups of G, let

$$HK = \{x \in G \mid x = hk, h \in H, k \in K\}.$$

Let's pause and look at an example; in S_3 let $H = \{e, \phi\}$, $K = \{e, \phi\psi\}$. Since $\phi^2 = (\phi\psi)^2 = e$, both H and K are subgroups. What can we say about HK? Just using the definition of HK we can see that HK consists of the elements $e, \phi, \phi\psi, \phi^2\psi = \psi$. Since HK consists of four elements and 4 is not a divisor of 6, the order of S_3 by Lagrange's theorem HK could not be a subgroup of S_3. (Of course, we could verify this directly but it does not hurt to keep recalling Lagrange's theorem.) We might try to find out why HK is not a subgroup. Note that $KH = \{e, \phi, \phi\psi, \phi\psi\phi = \psi^{-1}\} \neq HK$. This is precisely why HK fails to be a subgroup, as we see in the next lemma.

LEMMA 2.5.1 *HK is a subgroup of G if and only if $HK = KH$.*

Proof. Suppose, first, that $HK = KH$; that is, if $h \in H$ and $k \in K$, then $hk = k_1h_1$ for some $k_1 \in K$, $h_1 \in H$ (it need not be that $k_1 = k$ or

$h_1 = h$!). To prove that HK is a subgroup we must verify that it is closed and every element in HK has its inverse in HK. Let's show the closure first; so suppose $x = hk \in HK$ and $y = h'k' \in HK$. Then $xy = hkh'k'$, but since $kh' \in KH = HK$, $kh' = h_2k_2$ with $h_2 \in H$, $k_2 \in K$. Hence $xy = h(h_2k_2)k' = (hh_2)(k_2k') \in HK$, and HK is closed. Also $x^{-1} = (hk)^{-1} = k^{-1}h^{-1} \in KH = HK$, so $x^{-1} \in HK$. Thus HK is a subgroup of G.

On the other hand, if HK is a subgroup of G, then for any $h \in H$, $k \in K$, $h^{-1}k^{-1} \in HK$ and so $kh = (h^{-1}k^{-1})^{-1} \in HK$. Thus $KH \subset HK$. Now if x is any element of HK, $x^{-1} = hk \in HK$ and so $x = (x^{-1})^{-1} = (hk)^{-1} = k^{-1}h^{-1} \in KH$, so $HK \subset KH$. Thus $HK = KH$.

An interesting special case is the situation when G is an abelian group for in that case trivially $HK = KH$. Thus as a consequence we have the

COROLLARY *If H, K are subgroups of the abelian group G, then HK is a subgroup of G.*

If H, K are subgroups of a group G, we have seen that the subset HK need not be a subgroup of G. Yet it is a perfect meaningful question to ask: How many distinct elements are there in the subset HK? If we denote this number by $o(HK)$, we prove

THEOREM 2.5.1 *If H and K are finite subgroups of G of orders $o(H)$ and $o(K)$, respectively, then*

$$o(HK) = \frac{o(H)o(K)}{o(H \cap K)}.$$

Proof. Although there is no need to pay special attention to the particular case in which $H \cap K = (e)$, looking at this case, which is devoid of some of the complexity of the general situation, is quite revealing. Here we should seek to show that $o(HK) = o(H)o(K)$. One should ask oneself: How could this fail to happen? The answer clearly must be that if we list all the elements hk, $h \in H$, $k \in K$ there should be some collapsing; that is, some element in the list must appear at least twice. Equivalently, for some $h \neq h_1 \in H$, $hk = h_1k_1$. But then $h_1^{-1}h = k_1k^{-1}$; now since $h_1 \in H$, h_1^{-1} must also be in H, thus $h_1^{-1}h \in H$. Similarly, $k_1k^{-1} \in K$. Since $h_1^{-1}h = k_1k^{-1}$, $h_1^{-1}h \in H \cap K = (e)$, so $h_1^{-1}h = e$, whence $h = h_1$, a contradiction. We have proved that no collapsing can occur, and so, here, $o(HK)$ is indeed $o(H)o(K)$.

With this experience behind us we are ready to attack the general case. As above we must ask: How often does a given element hk appear as a product in the list of HK? We assert it must appear $o(H \cap K)$ times! To see this we first remark that if $h_1 \in H \cap K$, then

$$hk = (hh_1)(h_1^{-1}k), \tag{1}$$

where $hh_1 \in H$, since $h \in H$, $h_1 \in H \cap K \subset H$ and $h_1^{-1}k \in K$ since $h_1^{-1} \in H \cap K \subset K$ and $k \in K$. Thus hk is duplicated in the product at least $o(H \cap K)$ times. However, if $hk = h'k'$, then $h^{-1}h' = k(k')^{-1} = u$, and $u \in H \cap K$, and so $h' = hu$, $k' = u^{-1}k$; thus all duplications were accounted for in (1). Consequently hk appears in the list of HK exactly $o(H \cap K)$ times. Thus the number of distinct elements in HK is the total number in the listing of HK, that is, $o(H)o(K)$ divided by the number of times a given element appears, namely, $o(H \cap K)$. This proves the theorem.

Suppose H, K are subgroups of the finite group G and $o(H) > \sqrt{o(G)}$, $o(K) > \sqrt{o(G)}$. Since $HK \subset G$, $o(HK) \leq o(G)$. However,

$$o(G) \geq o(HK) = \frac{o(H)o(K)}{o(H \cap K)} > \frac{\sqrt{o(G)}\sqrt{o(G)}}{o(H \cap K)} = \frac{o(G)}{o(H \cap K)},$$

thus $o(H \cap K) > 1$. Therefore, $H \cap K \neq (e)$. We have proved the

COROLLARY *If H and K are subgroups of G and $o(H) > \sqrt{o(G)}$, $o(K) > \sqrt{o(G)}$, then $H \cap K \neq (e)$.*

We apply this corollary to a very special group. Suppose G is a finite group of order pq where p and q are prime numbers with $p > q$. We claim that G can have at most one subgroup of order p. For suppose H, K are subgroups of order p. By the corollary, $H \cap K \neq (e)$, and being a subgroup of H, which having prime order has no nontrivial subgroups, we must conclude that $H \cap K = H$, and so $H \subset H \cap K \subset K$. Similarly $K \subset H$, whence $H = K$, proving that there is at most one subgroup of order p. Later on we shall see that there is at least one subgroup of order p, which, combined with the above, will tell us there is exactly one subgroup of order p in G. From this we shall be able to determine completely the structure of G.

Problems

1. If H and K are subgroups of G, show that $H \cap K$ is a subgroup of G. (Can you see that the same proof shows that the intersection of any number of subgroups of G, finite or infinite, is again a subgroup of G?)
2. Let G be a group such that the intersection of all its subgroups which are different from (e) is a subgroup different from (e). Prove that every element in G has finite order.
3. If G has no nontrivial subgroups, show that G must be finite of prime order.

4. (a) If H is a subgroup of G, and $a \in G$ let $aHa^{-1} = \{aha^{-1} \mid h \in H\}$. Show that aHa^{-1} is a subgroup of G.
   (b) If H is finite, what is $o(aHa^{-1})$?

5. For a subgroup H of G define the left coset aH of H in G as the set of all elements of the form ah, $h \in H$. Show that there is a one-to-one correspondence between the set of left cosets of H in G and the set of right cosets of H in G.

6. Write out all the right cosets of H in G where
   (a) $G = (a)$ is a cyclic group of order 10 and $H = (a^2)$ is the subgroup of G generated by a^2.
   (b) G as in part (a), $H = (a^5)$ is the subgroup of G generated by a^5.
   (c) $G = A(S)$, $S = \{x_1, x_2, x_3\}$, and $H = \{\sigma \in G \mid x_1\sigma = x_1\}$.

7. Write out all the left cosets of H in G for H and G as in parts (a), (b), (c) of Problem 6.

8. Is every right coset of H in G a left coset of H in G in the groups of Problem 6?

9. Suppose that H is a subgroup of G such that whenever $Ha \neq Hb$ then $aH \neq bH$. Prove that $gHg^{-1} \subset H$ for all $g \in G$.

10. Let G be the group of integers under addition, H_n the subgroup consisting of all multiples of a fixed integer n in G. Determine the index of H_n in G and write out all the right cosets of H_n in G.

11. In Problem 10, what is $H_n \cap H_m$?

12. If G is a group and H, K are two subgroups of finite index in G, prove that $H \cap K$ is of finite index in G. Can you find an upper bound for the index of $H \cap K$ in G?

13. If $a \in G$, define $N(a) = \{x \in G \mid xa = ax\}$. Show that $N(a)$ is a subgroup of G. $N(a)$ is usually called the *normalizer* or *centralizer* of a in G.

14. If H is a subgroup of G, then by the centralizer $C(H)$ of H we mean the set $\{x \in G \mid xh = hx \text{ all } h \in H\}$. Prove that $C(H)$ is a subgroup of G.

15. The *center* Z of a group G is defined by $Z = \{z \in G \mid zx = xz \text{ all } x \in G\}$. Prove that Z is a subgroup of G. Can you recognize Z as $C(T)$ for some subgroup T of G?

16. If H is a subgroup of G, let $N(H) = \{a \in G \mid aHa^{-1} = H\}$ [see Problem 4(a)]. Prove that
   (a) $N(H)$ is a subgroup of G. (b) $N(H) \supset H$.

17. Give an example of a group G and a subgroup H such that $N(H) \neq C(H)$. Is there any containing relation between $N(H)$ and $C(H)$?

18. If H is a subgroup of G let

$$N = \bigcap_{x \in G} xHx^{-1}.$$

Prove that N is a subgroup of G such that $aNa^{-1} = N$ for all $a \in G$.

*19. If H is a subgroup of finite index in G, prove that there is only a finite number of distinct subgroups in G of the form aHa^{-1}.

*20. If H is of finite index in G prove that there is a subgroup N of G, contained in H, and of finite index in G such that $aNa^{-1} = N$ for all $a \in G$. Can you give an upper bound for the index of this N in G?

21. Let the mapping τ_{ab} for a, b real numbers, map the reals into the reals by the rule $\tau_{ab}: x \to ax + b$. Let $G = \{\tau_{ab} \mid a \neq 0\}$. Prove that G is a group under the composition of mappings. Find the formula for $\tau_{ab}\tau_{cd}$.

22. In Problem 21, let $H = \{\tau_{ab} \in G \mid a \text{ is rational}\}$. Show that H is a subgroup of G. List all the right cosets of H in G, and all the left cosets of H in G. From this show that every left coset of H in G is a right coset of H in G.

23. In the group G of Problem 21, let $N = \{\tau_{1b} \in G\}$. Prove
(a) N is a subgroup of G.
(b) If $a \in G$, $n \in N$, then $ana^{-1} \in N$.

*24. Let G be a finite group whose order is *not* divisible by 3. Suppose that $(ab)^3 = a^3b^3$ for all $a, b \in G$. Prove that G must be abelian.

*25. Let G be an abelian group and suppose that G has elements of orders m and n, respectively. Prove that G has an element whose order is the least common multiple of m and n.

**26. If an abelian group has subgroups of orders m and n, respectively, then show it has a subgroup whose order is the least common multiple of m and n. (Don't be discouraged if you don't get this problem with what you know about group theory up to this stage. I don't know anybody, including myself, who has done it subject to the restriction of using material developed so far in the text. But it is fun to try. I've had more correspondence about this problem than about any other point in the whole book.)

27. Prove that any subgroup of a cyclic group is itself a cyclic group.

28. How many generators does a cyclic group of order n have? ($b \in G$ is a generator if $(b) = G$.)

Let U_n denote the integers relatively prime to n under multiplication mod n. In Problem 15(b), Section 2.3, it is indicated that U_n is a group.

In the next few problems we look at the nature of U_n as a group for some specific values of n.

29. Show that U_8 is not a cyclic group.
30. Show that U_9 is a cyclic group. What are all its generators?
31. Show that U_{17} is a cyclic group. What are all its generators?
32. Show that U_{18} is a cyclic group.
33. Show that U_{20} is not a cyclic group.
34. Show that both U_{25} and U_{27} are cyclic groups.
35. Hazard a guess at what all the n such that U_n is cyclic are. (You can verify your guess by looking in any reasonable book on number theory.)
36. If $a \in G$ and $a^m = e$, prove that $o(a) \mid m$.
37. If in the group G, $a^5 = e$, $aba^{-1} = b^2$ for some $a, b \in G$, find $o(b)$.
*38. Let G be a finite abelian group in which the number of solutions in G of the equation $x^n = e$ is at most n for every positive integer n. Prove that G must be a cyclic group.
39. Let G be a group and A, B subgroups of G. If $x, y \in G$ define $x \sim y$ if $y = axb$ for some $a \in A$, $b \in B$. Prove
    (a) The relation so defined is an equivalence relation.
    (b) The equivalence class of x is $AxB = \{axb \mid a \in A, b \in B\}$. ($AxB$ is called a *double coset* of A and B in G.)
40. If G is a finite group, show that the number of elements in the double coset AxB is
$$\frac{o(A)o(B)}{o(A \cap xBx^{-1})}.$$
41. If G is a finite group and A is a subgroup of G such that all double cosets AxA have the same number of elements, show that $gAg^{-1} = A$ for all $g \in G$.

2.6 Normal Subgroups and Quotient Groups

Let G be the group S_3 and let H be the subgroup $\{e, \phi\}$. Since the index of H in G is 3, there are three right cosets of H in G and three left cosets of H in G. We list them:

Right Cosets	Left Cosets
$H = \{e, \phi\}$	$H = \{e, \phi\}$
$H\psi = \{\psi, \phi\psi\}$	$\psi H = \{\psi, \psi\phi = \phi\psi^2\}$
$H\psi^2 = \{\psi^2, \phi\psi^2\}$	$\psi^2 H = \{\psi^2, \psi^2\phi = \phi\psi\}$

A quick inspection yields the interesting fact that the right coset $H\psi$ is not a left coset. Thus, at least for this subgroup, the notions of left and right coset need not coincide.

In $G = S_3$ let us consider the subgroup $N = \{e, \psi, \psi^2\}$. Since the index of N in G is 2 there are two left cosets and two right cosets of N in G. We list these:

Right Cosets	Left Cosets
$N = \{e, \psi, \psi^2\}$	$N = \{e, \psi, \psi^2\}$
$N\phi = \{\phi, \psi\phi, \psi^2\phi\}$	$\phi N = \{\phi, \phi\psi, \phi\psi^2\}$
	$= \{\phi, \psi^2\phi, \psi\phi\}$

A quick inspection here reveals that every left coset of N in G is a right coset in G and conversely. Thus we see that for some subgroups the notion of left coset coincides with that of right coset, whereas for some subgroups these concepts differ.

It is a tribute to the genius of Galois that he recognized that those subgroups for which the left and right cosets coincide are distinguished ones. Very often in mathematics the crucial problem is to recognize and to discover what are the relevant concepts; once this is accomplished the job may be more than half done.

We shall define this special class of subgroups in a slightly different way, which we shall then show to be equivalent to the remarks in the above paragraph.

DEFINITION A subgroup N of G is said to be a *normal subgroup* of G if for every $g \in G$ and $n \in N$, $gng^{-1} \in N$.

Equivalently, if by gNg^{-1} we mean the set of all gng^{-1}, $n \in N$, then N is a normal subgroup of G if and only if $gNg^{-1} \subset N$ for every $g \in G$.

LEMMA 2.6.1 *N is a normal subgroup of G if and only if $gNg^{-1} = N$ for every $g \in G$.*

Proof. If $gNg^{-1} = N$ for every $g \in G$, certainly $gNg^{-1} \subset N$, so N is normal in G.

Suppose that N is normal in G. Thus if $g \in G$, $gNg^{-1} \subset N$ and $g^{-1}Ng = g^{-1}N(g^{-1})^{-1} \subset N$. Now, since $g^{-1}Ng \subset N$, $N = g(g^{-1}Ng)g^{-1} \subset gNg^{-1} \subset N$, whence $N = gNg^{-1}$.

In order to avoid a point of confusion here let us stress that Lemma 2.6.1 *does not* say that for every $n \in N$ and every $g \in G$, $gng^{-1} = n$. No! This can be false. Take, for instance, the group G to be S_3 and N to be the sub-

group $\{e, \psi, \psi^2\}$. If we compute $\phi N \phi^{-1}$ we obtain $\{e, \phi\psi\phi^{-1}, \phi\psi^2\phi^{-1}\} = \{e, \psi^2, \psi\}$, yet $\phi\psi\phi^{-1} \neq \psi$. All we require is that the *set* of elements gNg^{-1} be the same as the *set* of elements N.

We now can return to the question of the equality of left cosets and right cosets.

LEMMA 2.6.2 *The subgroup N of G is a normal subgroup of G if and only if every left coset of N in G is a right coset of N in G.*

Proof. If N is a normal subgroup of G, then for every $g \in G$, $gNg^{-1} = N$, whence $(gNg^{-1})g = Ng$; equivalently $gN = Ng$, and so the left coset gN is the right coset Ng.

Suppose, conversely, that every left coset of N in G is a right coset of N in G. Thus, for $g \in G$, gN, being a left coset, must be a right coset. What right coset can it be?

Since $g = ge \in gN$, whatever right coset gN turns out to be, it must contain the element g; however, g is in the right coset Ng, and two distinct right cosets have no element in common. (Remember the proof of Lagrange's theorem?) So this right coset is unique. Thus $gN = Ng$ follows. In other words, $gNg^{-1} = Ngg^{-1} = N$, and so N is a normal subgroup of G.

We have already defined what is meant by HK whenever H, K are subgroups of G. We can easily extend this definition to arbitrary subsets, and we do so by defining, for two subsets, A and B, of G, $AB = \{x \in G \mid x = ab, a \in A, b \in B\}$. As a special case, what can we say when $A = B = H$, a subgroup of G? $HH = \{h_1h_2 \mid h_1, h_2 \in H\} \subset H$ since H is closed under multiplication. But $HH \supset He = H$ since $e \in H$. Thus $HH = H$.

Suppose that N is a normal subgroup of G, and that $a, b \in G$. Consider $(Na)(Nb)$; since N is normal in G, $aN = Na$, and so

$$NaNb = N(aN)b = N(Na)b = NNab = Nab.$$

What a world of possibilities this little formula opens! But before we get carried away, for emphasis and future reference we record this as

LEMMA 2.6.3 *A subgroup N of G is a normal subgroup of G if and only if the product of two right cosets of N in G is again a right coset of N in G.*

Proof. If N is normal in G we have just proved the result. The proof of the other half is one of the problems at the end of this section.

Suppose that N is a normal subgroup of G. The formula $NaNb = Nab$, for $a, b \in G$ is highly suggestive; the product of right cosets is a right coset. Can we use this product to make the collection of right cosets into a group? Indeed we can! This type of construction, often occurring in mathematics and usually called forming a *quotient structure*, is of the utmost importance.

Let G/N denote the collection of right cosets of N in G (that is, the elements of G/N are certain subsets of G) and we use the product of subsets of G to yield for us a product in G/N.

For this product we claim

1. $X, Y \in G/N$ implies $XY \in G/N$; for $X = Na$, $Y = Nb$ for some $a, b \in G$, and $XY = NaNb = Nab \in G/N$.
2. $X, Y, Z \in G/N$, then $X = Na$, $Y = Nb$, $Z = Nc$ with $a, b, c \in G$, and so $(XY)Z = (NaNb)Nc = N(ab)Nc = N(ab)c = Na(bc)$ (since G is associative) $= Na(Nbc) = Na(NbNc) = X(YZ)$. Thus the product in G/N satisfies the associative law.
3. Consider the element $N = Ne \in G/N$. If $X \in G/N$, $X = Na$, $a \in G$, so $XN = NaNe = Nae = Na = X$, and similarly $NX = X$. Consequently, Ne is an identity element for G/N.
4. Suppose $X = Na \in G/N$ (where $a \in G$); thus $Na^{-1} \in G/N$, and $NaNa^{-1} = Naa^{-1} = Ne$. Similarly $Na^{-1}Na = Ne$. Hence Na^{-1} is the inverse of Na in G/N.

But a system which satisfies 1, 2, 3, 4 is exactly what we called a group. That is,

THEOREM 2.6.1 *If G is a group, N a normal subgroup of G, then G/N is also a group. It is called the quotient group or factor group of G by N.*

If, in addition, G is a finite group, what is the order of G/N? Since G/N has as its elements the right cosets of N in G, and since there are precisely $i_G(N) = o(G)/o(N)$ such cosets, we can say

LEMMA 2.6.4 *If G is a finite group and N is a normal subgroup of G, then $o(G/N) = o(G)/o(N)$.*

We close this section with an example.

Let G be the group of integers under addition and let N be the set of all multiplies of 3. Since the operation in G is addition we shall write the cosets of N in G as $N + a$ rather than as Na. Consider the three cosets N, $N + 1$, $N + 2$. We claim that these are all the cosets of N in G. For, given $a \in G$, $a = 3b + c$ where $b \in G$ and $c = 0$, 1, or 2 (c is the remainder of a on division by 3). Thus $N + a = N + 3b + c = (N + 3b) + c = N + c$ since $3b \in N$. Thus every coset is, as we stated, one of N, $N + 1$, or $N + 2$, and $G/N = \{N, N + 1, N + 2\}$. How do we add elements in G/N? Our formula $NaNb = Nab$ translates into: $(N + 1) + (N + 2) = N + 3 = N$ since $3 \in N$; $(N + 2) + (N + 2) = N + 4 = N + 1$ and so on. Without being specific one feels that G/N is closely related to the integers mod 3 under addition. Clearly what we did for 3 we could emulate

for any integer n, in which case the factor group should suggest a relation to the integers mod n under addition. This type of relation will be clarified in the next section.

Problems

1. If H is a subgroup of G such that the product of two right cosets of H in G is again a right coset of H in G, prove that H is normal in G.
2. If G is a group and H is a subgroup of index 2 in G, prove that H is a normal subgroup of G.
3. If N is a normal subgroup of G and H is any subgroup of G, prove that NH is a subgroup of G.
4. Show that the intersection of two normal subgroups of G is a normal subgroup of G.
5. If H is a subgroup of G and N is a normal subgroup of G, show that $H \cap N$ is a normal subgroup of H.
6. Show that every subgroup of an abelian group is normal.
*7. Is the converse of Problem 6 true? If yes, prove it, if no, give an example of a non-abelian group all of whose subgroups are normal.
8. Give an example of a group G, subgroup H, and an element $a \in G$ such that $aHa^{-1} \subset H$ but $aHa^{-1} \neq H$.
9. Suppose H is the only subgroup of order $o(H)$ in the finite group G. Prove that H is a normal subgroup of G.
10. If H is a subgroup of G, let $N(H) = \{g \in G \mid gHg^{-1} = H\}$. Prove
    (a) $N(H)$ is a subgroup of G.
    (b) H is normal in $N(H)$.
    (c) If H is a normal subgroup of the subgroup K in G, then $K \subset N(H)$ (that is, $N(H)$ is the largest subgroup of G in which H is normal).
    (d) H is normal in G if and only if $N(H) = G$.
11. If N and M are normal subgroups of G, prove that NM is also a normal subgroup of G.
*12. Suppose that N and M are two normal subgroups of G and that $N \cap M = (e)$. Show that for any $n \in N$, $m \in M$, $nm = mn$.
13. If a cyclic subgroup T of G is normal in G, then show that every subgroup of T is normal in G.
*14. Prove, by an example, that we can find three groups $E \subset F \subset G$, where E is normal in F, F is normal in G, but E is *not* normal in G.
15. If N is normal in G and $a \in G$ is of order $o(a)$, prove that the order, m, of Na in G/N is a divisor of $o(a)$.

16. If N is a normal subgroup in the finite group such that $i_G(N)$ and $o(N)$ are relatively prime, show that any element $x \in G$ satisfying $x^{o(N)} = e$ must be in N.

17. Let G be defined as all formal symbols $x^i y^j$, $i = 0, i, j = 0, 1, 2, \ldots, n - 1$ where we assume

$$x^i y^j = x^{i'} y^{j'} \text{ if and only if } i = i', j = j'$$

$$x^2 = y^n = e, \quad n > 2$$

$$xy = y^{-1} x.$$

(a) Find the form of the product $(x^i y^j)(x^k y^l)$ as $x^\alpha y^\beta$.
(b) Using this, prove that G is a non-abelian group of order $2n$.
(c) If n is odd, prove that the center of G is (e), while if n is even the center of G is larger than (e).

This group is known as a *dihedral* group. A geometric realization of this is obtained as follows: let y be a rotation of the Euclidean plane about the origin through an angle of $2\pi/n$, and x the reflection about the vertical axis. G is the group of motions of the plane generated by y and x.

18. Let G be a group in which, for some integer $n > 1$, $(ab)^n = a^n b^n$ for all $a, b \in G$. Show that
(a) $G^{(n)} = \{x^n \mid x \in G\}$ is a normal subgroup of G.
(b) $G^{(n-1)} = \{x^{n-1} \mid x \in G\}$ is a normal subgroup of G.

19. Let G be as in Problem 18. Show
(a) $a^{n-1} b^n = b^n a^{n-1}$ for all $a, b \in G$.
(b) $(aba^{-1}b^{-1})^{n(n-1)} = e$ for all $a, b \in G$.

20. Let G be a group such that $(ab)^p = a^p b^p$ for all $a, b \in G$, where p is a prime number. Let $S = \{x \in G \mid x^{p^m} = e \text{ for some } m \text{ depending on } x\}$. Prove
(a) S is a normal subgroup of G.
(b) If $\bar{G} = G/S$ and if $\bar{x} \in \bar{G}$ is such that $\bar{x}^p = \bar{e}$ then $\bar{x} = \bar{e}$.

#21. Let G be the set of all real 2×2 matrices $\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$ where $ad \neq 0$, under matrix multiplication. Let $N = \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \right\}$. Prove that
(a) N is a normal subgroup of G.
(b) G/N is abelian.

2.7 Homomorphisms

The ideas and results in this section are closely interwoven with those of the preceding one. If there is one central idea which is common to all aspects of modern algebra it is the notion of homomorphism. By this one means

a mapping from one algebraic system to a like algebraic system which preserves structure. We make this precise, for groups, in the next definition.

DEFINITION A mapping ϕ from a group G into a group $\bar{G}$ is said to be a *homomorphism* if for all $a, b \in G$, $\phi(ab) = \phi(a)\phi(b)$.

Notice that on the left side of this relation, namely, in the term $\phi(ab)$, the product ab is computed in G using the product of elements of G, whereas on the right side of this relation, namely, in the term $\phi(a)\phi(b)$, the product is that of elements in $\bar{G}$.

Example 2.7.0 $\phi(x) = e$ all $x \in G$. This is trivially a homomorphism. Likewise $\phi(x) = x$ for every $x \in G$ is a homomorphism.

Example 2.7.1 Let G be the group of all real numbers under addition (i.e., ab for $a, b \in G$ is really the real number $a + b$) and let $\bar{G}$ be the group of nonzero real numbers with the product being ordinary multiplication of real numbers. Define $\phi: G \to \bar{G}$ by $\phi(a) = 2^a$. In order to verify that this mapping is a homomorphism we must check to see whether $\phi(ab) = \phi(a)\phi(b)$, remembering that by the product on the left side we mean the operation in G (namely, addition), that is, we must check if $2^{a+b} = 2^a 2^b$, which indeed is true. Since 2^a is always positive, the image of ϕ is not all of $\bar{G}$, so ϕ is a homomorphism of G into $\bar{G}$, but not onto $\bar{G}$.

Example 2.7.2 Let $G = S_3 = \{e, \phi, \psi, \psi^2, \phi\psi, \phi\psi^2\}$ and $\bar{G} = \{e, \phi\}$. Define the mapping $f: G \to \bar{G}$ by $f(\phi^i\psi^j) = \phi^i$. Thus $f(e) = e, f(\phi) = \phi, f(\psi) = e, f(\psi^2) = e, f(\phi\psi) = \phi, f(\phi\psi^2) = \phi$. The reader should verify that f so defined is a homomorphism.

Example 2.7.3 Let G be the group of integers under addition and let $\bar{G} = G$. For the integer $x \in G$ define ϕ by $\phi(x) = 2x$. That ϕ is a homomorphism then follows from $\phi(x + y) = 2(x + y) = 2x + 2y = \phi(x) + \phi(y)$.

Example 2.7.4 Let G be the group of nonzero real numbers under multiplication, $\bar{G} = \{1, -1\}$, where $1.1 = 1$, $(-1)(-1) = 1$, $1(-1) = (-1)1 = -1$. Define $\phi: G \to \bar{G}$ by $\phi(x) = 1$ if x is positive, $\phi(x) = -1$ if x is negative. The fact that ϕ is a homomorphism is equivalent to the statements: positive times positive is positive, positive times negative is negative, negative times negative is positive.

Example 2.7.5 Let G be the group of integers under addition, let $\bar{G}_n$ be the group of integers under addition modulo n. Define ϕ by $\phi(x) =$ remainder of x on division by n. One can easily verify this is a homomorphism.

Example 2.7.6 Let G be the group of positive real numbers under multiplication and let $\bar{G}$ be the group of all real numbers under addition. Define $\phi: G \to \bar{G}$ by $\phi(x) = \log_{10} x$. Thus

$$\phi(xy) = \log_{10}(xy) = \log_{10}(x) + \log_{10}(y) = \phi(x)\phi(y)$$

since the operation, on the right side, in $\bar{G}$ is in fact addition. Thus ϕ is a homomorphism of G into $\bar{G}$. In fact, not only is ϕ a homomorphism but, in addition, it is one-to-one and onto.

#Example 2.7.7 Let G be the group of all real 2×2 matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $ad - bc \neq 0$, under matrix multiplication. Let $\bar{G}$ be the group of all nonzero real numbers under multiplication. Define $\phi: G \to \bar{G}$ by $\phi\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$.

We leave it to the reader to check that ϕ is a homomorphism of G onto $\bar{G}$.

The result of the following lemma yields, for us, an infinite class of examples of homomorphisms. When we prove Theorem 2.7.1 it will turn out that in some sense this provides us with the most general example of a homomorphism.

LEMMA 2.7.1 *Suppose G is a group, N a normal subgroup of G; define the mapping ϕ from G to G/N by $\phi(x) = Nx$ for all $x \in G$. Then ϕ is a homomorphism of G onto G/N.*

Proof. In actuality, there is nothing to prove, for we already have proved this fact several times. But for the sake of emphasis we repeat it.

That ϕ is onto is trivial, for every element $X \in G/N$ is of the form $X = Ny$, $y \in G$, so $X = \phi(y)$. To verify the multiplicative property required in order that ϕ be a homomorphism, one just notes that if $x, y \in G$,

$$\phi(xy) = Nxy = NxNy = \phi(x)\phi(y).$$

In Lemma 2.7.1 and in the examples preceding it, a fact which comes through is that a homomorphism need not be one-to-one; but there is a certain uniformity in this process of deviating from one-to-oneness. This will become apparent in a few lines.

DEFINITION If ϕ is a homomorphism of G into $\bar{G}$, the *kernel of* ϕ, K_ϕ, is defined by $K_\phi = \{x \in G \mid \phi(x) = \bar{e}, \bar{e} = \text{identity element of } \bar{G}\}$.

Before investigating any properties of K_ϕ it is advisable to establish that, as a set, K_ϕ is not empty. This is furnished us by the first part of

LEMMA 2.7.2 *If ϕ is a homomorphism of G into $\bar{G}$, then*

1. $\phi(e) = \bar{e}$, *the unit element of* $\bar{G}$.
2. $\phi(x^{-1}) = \phi(x)^{-1}$ *for all* $x \in G$.

Proof. To prove (1) we merely calculate $\phi(x)\bar{e} = \phi(x) = \phi(xe) = \phi(x)\phi(e)$, so by the cancellation property in $\bar{G}$ we have that $\phi(e) = \bar{e}$.

To establish (2) one notes that $\bar{e} = \phi(e) = \phi(xx^{-1}) = \phi(x)\phi(x^{-1})$, so by the very definition of $\phi(x)^{-1}$ in $\bar{G}$ we obtain the result that $\phi(x^{-1}) = \phi(x)^{-1}$.

The argument used in the proof of Lemma 2.7.2 should remind any reader who has been exposed to a development of logarithms of the argument used in proving the familiar results that $\log 1 = 0$ and $\log (1/x) = -\log x$; this is no coincidence, for the mapping $\phi: x \to \log x$ is a homomorphism of the group of positive real numbers under multiplication into the group of real numbers under addition, as we have seen in Example 2.7.6.

Lemma 2.7.2 shows that e is in the kernel of any homomorphism, so any such kernel is not empty. But we can say even more.

LEMMA 2.7.3 *If ϕ is a homomorphism of G into $\bar{G}$ with kernel K, then K is a normal subgroup of G.*

Proof. First we must check whether K is a subgroup of G. To see this one must show that K is closed under multiplication and has inverses in it for every element belonging to K.

If $x, y \in K$, then $\phi(x) = \bar{e}$, $\phi(y) = \bar{e}$, where $\bar{e}$ is the identity element of $\bar{G}$, and so $\phi(xy) = \phi(x)\phi(y) = \bar{e}\bar{e} = \bar{e}$, whence $xy \in K$. Also, if $x \in K$, $\phi(x) = \bar{e}$, so, by Lemma 2.7.2, $\phi(x^{-1}) = \phi(x)^{-1} = \bar{e}^{-1} = \bar{e}$; thus $x^{-1} \in K$. K is, accordingly, a subgroup of G.

To prove the normality of K one must establish that for any $g \in G$, $k \in K$, $gkg^{-1} \in K$; in other words, one must prove that $\phi(gkg^{-1}) = \bar{e}$ whenever $\phi(k) = \bar{e}$. But $\phi(gkg^{-1}) = \phi(g)\phi(k)\phi(g^{-1}) = \phi(g)\bar{e}\phi(g)^{-1} = \phi(g)\phi(g)^{-1} = \bar{e}$. This completes the proof of Lemma 2.7.3.

Let ϕ now be a homomorphism of the group G *onto* the group $\bar{G}$, and suppose that K is the kernel of ϕ. If $\bar{g} \in \bar{G}$, we say an element $x \in G$ is an *inverse image* of $\bar{g}$ under ϕ if $\phi(x) = \bar{g}$. What are all the inverse images of $\bar{g}$? For $\bar{g} = \bar{e}$ we have the answer, namely (by its very definition) K. What about elements $\bar{g} \neq \bar{e}$? Well, suppose $x \in G$ is one inverse image of $\bar{g}$; can we write down others? Clearly yes, for if $k \in K$, and if $y = kx$, then $\phi(y) = \phi(kx) = \phi(k)\phi(x) = \bar{e}\bar{g} = \bar{g}$. Thus all the elements Kx are in the inverse image of $\bar{g}$ whenever x is. Can there be others? Let us suppose that $\phi(z) = \bar{g} = \phi(x)$. Ignoring the middle term we are left with $\phi(z) = \phi(x)$, and so $\phi(z)\phi(x)^{-1} = \bar{e}$. But $\phi(x)^{-1} = \phi(x^{-1})$, whence

$\bar{e} = \phi(z)\phi(x)^{-1} = \phi(z)\phi(x^{-1}) = \phi(zx^{-1})$, in consequence of which $zx^{-1} \in K$; thus $z \in Kx$. In other words, we have shown that Kx accounts for exactly all the inverse images of $\bar{g}$ whenever x is a single such inverse image. We record this as

LEMMA 2.7.4 *If ϕ is a homomorphism of G onto $\bar{G}$ with kernel K, then the set of all inverse images of $\bar{g} \in \bar{G}$ under ϕ in G is given by Kx, where x is any particular inverse image of $\bar{g}$ in G.*

A special case immediately presents itself, namely, the situation when $K = (e)$. But here, by Lemma 2.7.4, any $\bar{g} \in \bar{G}$ has exactly one inverse image. That is, ϕ is a one-to-one mapping. The converse is trivially true, namely, if ϕ is a one-to-one homomorphism of G into (not even onto) G, its kernel must consist exactly of e.

DEFINITION A homomorphism ϕ from G into $\bar{G}$ is said to be an *isomorphism* if ϕ is one-to-one.

DEFINITION Two groups G, G^* are said to be *isomorphic* if there is an isomorphism of G *onto* G^*. In this case we write $G \approx G^*$.

We leave to the reader to verify the following three facts:

1. $G \approx G$.
2. $G \approx G^*$ implies $G^* \approx G$.
3. $G \approx G^*$, $G^* \approx G^{**}$ implies $G \approx G^{**}$.

When two groups are isomorphic, then, in some sense, they are equal. They differ in that their elements are labeled differently. The isomorphism gives us the key to the labeling, and with it, knowing a given computation in one group, we can carry out the analogous computation in the other. The isomorphism is like a dictionary which enables one to translate a sentence in one language into a sentence, of the same meaning, in another language. (Unfortunately no such perfect dictionary exists, for in languages words do not have single meanings, and nuances do not come through in a literal translation.) But merely to say that a given sentence in one language can be expressed in another is of little consequence; one needs the dictionary to carry out the translation. Similarly it might be of little consequence to know that two groups are isomorphic; the object of interest might very well be the isomorphism itself. So, whenever we prove two groups to be isomorphic, we shall endeavor to exhibit the precise mapping which yields this isomorphism.

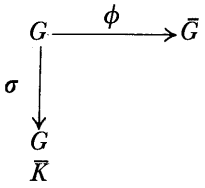
Returning to Lemma 2.7.4 for a moment, we see in it a means of characterizing in terms of the kernel when a homomorphism is actually an isomorphism.

COROLLARY *A homomorphism ϕ of G into $\bar{G}$ with kernel K_ϕ is an isomorphism of G into $\bar{G}$ if and only if $K_\phi = (e)$.*

This corollary provides us with a standard technique for proving two groups to be isomorphic. First we find a homomorphism of one onto the other, and then prove the kernel of this homomorphism consists only of the identity element. This method will be illustrated for us in the proof of the very important

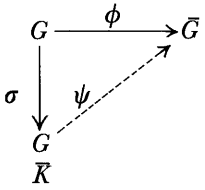
THEOREM 2.7.1 *Let ϕ be a homomorphism of G onto $\bar{G}$ with kernel K. Then $G/K \approx \bar{G}$.*

Proof. Consider the diagram

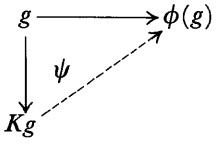


where $\sigma(g) = Kg$.

We should like to complete this to



It seems clear that, in order to construct the mapping ψ from G/K to $\bar{G}$, we should use G as an intermediary, and also that this construction should be relatively uncomplicated. What is more natural than to complete the diagram using



With this preamble we formally define the mapping ψ from G/K to $\bar{G}$ by: if $X \in G/K$, $X = Kg$, then $\psi(X) = \phi(g)$. A problem immediately arises: is this mapping well defined? If $X \in G/K$, it can be written as Kg in several ways (for instance, $Kg = Kkg$, $k \in K$); but if $X = Kg = Kg'$, $g, g' \in G$, then on one hand $\psi(X) = \phi(g)$, and on the other, $\psi(X) = \phi(g')$. For the mapping ψ to make sense it had better be true that $\phi(g) = \phi(g')$. So, suppose $Kg = Kg'$; then $g = kg'$, where $k \in K$, hence $\phi(g) = \phi(kg') = \phi(k)\phi(g') = \bar{e}\phi(g') = \phi(g')$ since $k \in K$, the kernel of ϕ.

We next determine that ψ is onto. For, if $\bar{x} \in \bar{G}$, $\bar{x} = \phi(g)$, $g \in G$ (since ϕ is onto) so $\bar{x} = \phi(g) = \psi(Kg)$.

If $X, Y \in G/K$, $X = Kg$, $Y = Kf$, $g, f \in G$, then $XY = KgKf = Kgf$, so that $\psi(XY) = \psi(Kgf) = \phi(gf) = \phi(g)\phi(f)$ since ϕ is a homomorphism of G onto $\bar{G}$. But $\psi(X) = \psi(Kg) = \phi(g)$, $\psi(Y) = \psi(Kf) = \phi(f)$, so we see that $\psi(XY) = \psi(X)\psi(Y)$, and ψ is a homomorphism of G/K onto $\bar{G}$.

To prove that ψ is an isomorphism of G/K onto $\bar{G}$ all that remains is to demonstrate that the kernel of ψ is the unit element of G/K. Since the unit element of G/K is $K = Ke$, we must show that if $\psi(Kg) = \bar{e}$, then $Kg = Ke = K$. This is now easy, for $\bar{e} = \psi(Kg) = \phi(g)$, so that $\phi(g) = \bar{e}$, whence g is in the kernel of ϕ, namely K. But then $Kg = K$ since K is a subgroup of G. All the pieces have been put together. We have exhibited a one-to-one homomorphism of G/K onto $\bar{G}$. Thus $G/K \approx \bar{G}$, and Theorem 2.7.1 is established.

Theorem 2.7.1 is important, for it tells us precisely what groups can be expected to arise as homomorphic images of a given group. These must be expressible in the form G/K, where K is normal in G. But, by Lemma 2.7.1, for any normal subgroup N of G, G/N is a homomorphic image of G. Thus there is a one-to-one correspondence between homomorphic images of G and normal subgroups of G. If one were to seek all homomorphic images of G one could do it by never leaving G as follows: find all normal subgroups N of G and construct all groups G/N. The set of groups so constructed yields all homomorphic images of G (up to isomorphisms).

A group is said to be *simple* if it has no nontrivial homomorphic images, that is, if it has no nontrivial normal subgroups. A famous, long-standing conjecture was that a non-abelian simple group of finite order has an even number of elements. This important result has been proved by the two American mathematicians, Walter Feit and John Thompson.

We have stated that the concept of a homomorphism is a very important one. To strengthen this statement we shall now show how the methods and results of this section can be used to prove nontrivial facts about groups. When we construct the group G/N, where N is normal in G, if we should happen to know the structure of G/N we would know that of G "up to N." True, we blot out a certain amount of information about G, but often

enough is left so that from facts about G/N we can ascertain certain ones about G. When we photograph a certain scene we transfer a three-dimensional object to a two-dimensional representation of it. Yet, looking at the picture we can derive a great deal of information about the scene photographed.

In the two applications of the ideas developed so far, which are given below, the proofs given are not the best possible. In fact, a little later in this chapter these results will be proved in a more general situation in an easier manner. We use the presentation here because it does illustrate effectively many group-theoretic concepts.

APPLICATION 1 (CAUCHY'S THEOREM FOR ABELIAN GROUPS) *Suppose G is a finite abelian group and $p \mid o(G)$, where p is a prime number. Then there is an element $a \neq e \in G$ such that $a^p = e$.*

Proof. We proceed by induction over $o(G)$. In other words, we assume that the theorem is true for all abelian groups having fewer elements than G. From this we wish to prove that the result holds for G. To start the induction we note that the theorem is vacuously true for groups having a single element.

If G has no subgroups $H \neq (e), G$, by the result of a problem earlier in the chapter, G must be cyclic of prime order. This prime must be p, and G certainly has $p - 1$ elements $a \neq e$ satisfying $a^p = a^{o(G)} = e$.

So suppose G has a subgroup $N \neq (e), G$. If $p \mid o(N)$, by our induction hypothesis, since $o(N) < o(G)$ and N is abelian, there is an element $b \in N$, $b \neq e$, satisfying $b^p = e$; since $b \in N \subset G$ we would have exhibited an element of the type required. So we may assume that $p \nmid o(N)$. Since G is abelian, N is a normal subgroup of G, so G/N is a group. Moreover, $o(G/N) = o(G)/o(N)$, and since $p \nmid o(N)$,

$$p \left| \frac{o(G)}{o(N)} < o(G). \right.$$

Also, since G is abelian, G/N is abelian. Thus by our induction hypothesis there is an element $X \in G/N$ satisfying $X^p = e_1$, the unit element of G/N, $X \neq e_1$. By the very form of the elements of G/N, $X = Nb$, $b \in G$, so that $X^p = (Nb)^p = Nb^p$. Since $e_1 = Ne$, $X^p = e_1$, $X \neq e_1$ translates into $Nb^p = N$, $Nb \neq N$. Thus $b^p \in N$, $b \notin N$. Using one of the corollaries to Lagrange's theorem, $(b^p)^{o(N)} = e$. That is, $b^{o(N)p} = e$. Let $c = b^{o(N)}$. Certainly $c^p = e$. In order to show that c is an element that satisfies the conclusion of the theorem we must finally show that $c \neq e$. However, if $c = e$, $b^{o(N)} = e$, and so $(Nb)^{o(N)} = N$. Combining this with $(Nb)^p = N$, $p \nmid o(N)$, p a prime number, we find that $Nb = N$, and so $b \in N$, a contradiction. Thus $c \neq e$, $c^p = e$, and we have completed the induction. This proves the result.

APPLICATION 2 (SYLOW'S THEOREM FOR ABELIAN GROUPS) *If G is an abelian group of order $o(G)$, and if p is a prime number, such that $p^\alpha \mid o(G)$, $p^{\alpha+1} \nmid o(G)$, then G has a subgroup of order p^α.*

Proof. If $\alpha = 0$, the subgroup (e) satisfies the conclusion of the result. So suppose $\alpha \neq 0$. Then $p \mid o(G)$. By Application 1, there is an element $a \neq e \in G$ satisfying $a^p = e$. Let $S = \{x \in G \mid x^{p^n} = e \text{ some integer } n\}$. Since $a \in S$, $a \neq e$, it follows that $S \neq (e)$. We now assert that S is a subgroup of G. Since G is finite we must only verify that S is closed. If $x, y \in S$, $x^{p^n} = e$, $y^{p^m} = e$, so that $(xy)^{p^{n+m}} = x^{p^{n+m}} y^{p^{n+m}} = e$ (we have used that G is abelian), proving that $xy \in S$.

We next claim that $o(S) = p^\beta$ with β an integer $0 < \beta \leq \alpha$. For, if some prime $q \mid o(S)$, $q \neq p$, by the result of Application 1 there is an element $c \in S$, $c \neq e$, satisfying $c^q = e$. However, $c^{p^n} = e$ for some n since $c \in S$. Since p^n, q are relatively prime, we can find integers λ, μ such that $\lambda q + \mu p^n = 1$, so that $c = c^1 = c^{\lambda q + \mu p^n} = (c^q)^\lambda (c^{p^n})^\mu = e$, contradicting $c \neq e$. By Lagrange's theorem $o(S) \mid o(G)$, so that $\beta \leq \alpha$. Suppose that $\beta < \alpha$; consider the abelian group G/S. Since $\beta < \alpha$ and $o(G/S) = o(G)/o(S)$, $p \mid o(G/S)$, there is an element Sx, $(x \in G)$ in G/S satisfying $Sx \neq S$, $(Sx)^{p^n} = S$ for some integer $n > 0$. But $S = (Sx)^{p^n} = Sx^{p^n}$, and so $x^{p^n} \in S$; consequently $e = (x^{p^n})^{o(S)} = (x^{p^n})^{p^\beta} = x^{p^{n+\beta}}$. Therefore, x satisfies the exact requirements needed to put it in S; in other words, $x \in S$. Consequently $Sx = S$ contradicting $Sx \neq S$. Thus $\beta < \alpha$ is impossible and we are left with the only alternative, namely, that $\beta = \alpha$. S is the required subgroup of order p^α.

We strengthen the application slightly. Suppose T is another subgroup of G of order p^α, $T \neq S$. Since G is abelian $ST = TS$, so that ST is a subgroup of G. By Theorem 2.5.1

$$o(ST) = \frac{o(S)o(T)}{o(S \cap T)} = \frac{p^\alpha p^\alpha}{o(S \cap T)}$$

and since $S \neq T$, $o(S \cap T) < p^\alpha$, leaving us with $o(ST) = p^\gamma$, $\gamma > \alpha$. Since ST is a subgroup of G, $o(ST) \mid o(G)$; thus $p^\gamma \mid o(G)$ violating the fact that α is the largest power of p which divides $o(G)$. Thus no such subgroup T exists, and S is the unique subgroup of order p^α. We have proved the

COROLLARY *If G is abelian of order $o(G)$ and $p^\alpha \mid o(G)$, $p^{\alpha+1} \nmid o(G)$, there is a unique subgroup of G of order p^α.*

If we look at $G = S_3$, which is non-abelian, $o(G) = 2.3$, we see that G has 3 distinct subgroups of order 2, namely, $\{e, \phi\}$, $\{e, \phi\psi\}$, $\{e, \phi\psi^2\}$, so that the corollary asserting the uniqueness does not carry over to non-abelian groups. But Sylow's theorem holds for all finite groups.

We leave the application and return to the general development. Suppose ϕ is a homomorphism of G onto $\bar{G}$ with kernel K, and suppose that $\bar{H}$ is a subgroup of $\bar{G}$. Let $H = \{x \in G \mid \phi(x) \in \bar{H}\}$. We assert that H is a subgroup of G and that $H \supset K$. That $H \supset K$ is trivial, for if $x \in K$, $\phi(x) = \bar{e}$ is in $\bar{H}$, so that $K \subset H$ follows. Suppose now that $x, y \in H$; hence $\phi(x) \in \bar{H}$, $\phi(y) \in \bar{H}$ from which we deduce that $\phi(xy) = \phi(x)\phi(y) \in \bar{H}$. Therefore, $xy \in H$ and H is closed under the product in G. Furthermore, if $x \in H$, $\phi(x) \in \bar{H}$ and so $\phi(x^{-1}) = \phi(x)^{-1} \in \bar{H}$ from which it follows that $x^{-1} \in H$. All in all, our assertion has been established. What can we say in addition in case $\bar{H}$ is normal in $\bar{G}$? Let $g \in G$, $h \in H$; then $\phi(h) \in \bar{H}$, whence $\phi(ghg^{-1}) = \phi(g)\phi(h)\phi(g)^{-1} \in \bar{H}$, since $\bar{H}$ is normal in G. Otherwise stated, $ghg^{-1} \in H$, from which it follows that H is normal in G. One other point should be noted, namely, that the homomorphism ϕ from G onto $\bar{G}$, when just considered on elements of H, induces a homomorphism of H onto $\bar{H}$, with kernel exactly K, since $K \subset H$; by Theorem 2.7.1 we have that $\bar{H} \approx H/K$.

Suppose, conversely, that L is a subgroup of G and $K \subset L$. Let $\bar{L} = \{\bar{x} \in \bar{G} \mid \bar{x} = \phi(l), l \in L\}$. The reader should verify that $\bar{L}$ is a subgroup of $\bar{G}$. Can we explicitly describe the subgroup $T = \{y \in G \mid \phi(y) \in \bar{L}\}$? Clearly $L \subset T$. Is there any element $t \in T$ which is not in L? So, suppose $t \in T$; thus $\phi(t) \in \bar{L}$, so by the very definition of $\bar{L}$, $\phi(t) = \phi(l)$ for some $l \in L$. Thus $\phi(tl^{-1}) = \phi(t)\phi(l)^{-1} = \bar{e}$, whence $tl^{-1} \in K \subset L$, thus t is in $Ll = L$. Equivalently we have proved that $T \subset L$, which, combined with $L \subset T$, yields that $L = T$.

Thus we have set up a one-to-one correspondence between the set of all subgroups of $\bar{G}$ and the set of all subgroups of G which contain K. Moreover, in this correspondence, a normal subgroup of G corresponds to a normal subgroup of $\bar{G}$.

We summarize these few paragraphs in

LEMMA 2.7.5 *Let ϕ be a homomorphism of G onto $\bar{G}$ with kernel K. For $\bar{H}$ a subgroup of $\bar{G}$ let H be defined by $H = \{x \in G \mid \phi(x) \in \bar{H}\}$. Then H is a subgroup of G and $H \supset K$; if $\bar{H}$ is normal in $\bar{G}$, then H is normal in G. Moreover, this association sets up a one-to-one mapping from the set of all subgroups of $\bar{G}$ onto the set of all subgroups of G which contain K.*

We wish to prove one more general theorem about the relation of two groups which are homomorphic.

THEOREM 2.7.2. *Let ϕ be a homomorphism of G onto $\bar{G}$ with kernel K, and let $\bar{N}$ be a normal subgroup of $\bar{G}$, $N = \{x \in G \mid \phi(x) \in \bar{N}\}$. Then $G/N \approx \bar{G}/\bar{N}$. Equivalently, $G/N \approx (G/K)/(N/K)$.*

Proof. As we already know, there is a homomorphism θ of $\bar{G}$ onto $\bar{G}/\bar{N}$ defined by $\theta(\bar{g}) = \bar{N}\bar{g}$. We define the mapping $\psi: G \to \bar{G}/\bar{N}$ by $\psi(g) = \bar{N}\phi(g)$ for all $g \in G$. To begin with, ψ is onto, for if $\bar{g} \in \bar{G}$, $\bar{g} = \phi(g)$ for some $g \in G$, since ϕ is onto, so the typical element $\bar{N}\bar{g}$ in $\bar{G}/\bar{N}$ can be represented as $\bar{N}\phi(g) = \psi(g)$.

If $a, b \in G$, $\psi(ab) = \bar{N}\phi(ab)$ by the definition of the mapping ψ. However, since ϕ is a homomorphism, $\phi(ab) = \phi(a)\phi(b)$. Thus $\psi(ab) = \bar{N}\phi(a)\phi(b) = \bar{N}\phi(a)\bar{N}\phi(b) = \psi(a)\psi(b)$. So far we have shown that ψ is a homomorphism of G onto $\bar{G}/\bar{N}$. What is the kernel, T, of ψ? Firstly, if $n \in N$, $\phi(n) \in \bar{N}$, so that $\psi(n) = \bar{N}\phi(n) = \bar{N}$, the identity element of $\bar{G}/\bar{N}$, proving that $N \subset T$. On the other hand, if $t \in T$, $\psi(t) =$ identity element of $\bar{G}/\bar{N} = \bar{N}$; but $\psi(t) = \bar{N}\phi(t)$. Comparing these two evaluations of $\psi(t)$, we arrive at $\bar{N} = \bar{N}\phi(t)$, which forces $\phi(t) \in \bar{N}$; but this places t in N by definition of N. That is, $T \subset N$. The kernel of ψ has been proved to be equal to N. But then ψ is a homomorphism of G onto $\bar{G}/\bar{N}$ with kernel N. By Theorem 2.7.1 $G/N \approx \bar{G}/\bar{N}$, which is the first part of the theorem. The last statement in the theorem is immediate from the observation (following as a consequence of Theorem 2.7.1) that $\bar{G} \approx G/K$, $\bar{N} \approx N/K$, $\bar{G}/\bar{N} \approx (G/K)/(N/K)$.

Problems

1. In the following, verify if the mappings defined are homomorphisms, and in those cases in which they are homomorphisms, determine the kernel.
   (a) G is the group of nonzero real numbers under multiplication, $\bar{G} = G$, $\phi(x) = x^2$ all $x \in G$.
   (b) G, $\bar{G}$ as in (a), $\phi(x) = 2^x$.
   (c) G is the group of real numbers under addition, $\bar{G} = G$, $\phi(x) = x + 1$ all $x \in G$.
   (d) G, $\bar{G}$ as in (c), $\phi(x) = 13x$ for $x \in G$.
   (e) G is any abelian group, $\bar{G} = G$, $\phi(x) = x^5$ all $x \in G$.
2. Let G be any group, g a fixed element in G. Define $\phi: G \to G$ by $\phi(x) = gxg^{-1}$. Prove that ϕ is an isomorphism of G onto G.
3. Let G be a finite abelian group of order $o(G)$ and suppose the integer n is relatively prime to $o(G)$. Prove that every $g \in G$ can be written as $g = x^n$ with $x \in G$. (*Hint:* Consider the mapping $\phi: G \to G$ defined by $\phi(y) = y^n$, and prove this mapping is an isomorphism of G onto G.)
4. (a) Given any group G and a subset U, let $\hat{U}$ be the smallest subgroup of G which contains U. Prove there is such a subgroup $\hat{U}$ in G. ($\hat{U}$ is called the *subgroup generated by* U.)

(b) If $gug^{-1} \in U$ for all $g \in G$, $u \in U$, prove that $\hat{U}$ is a normal subgroup of G.

5. Let $U = \{xyx^{-1}y^{-1} \mid x, y \in G\}$. In this case $\hat{U}$ is usually written as G' and is called the *commutator subgroup of* G.
   (a) Prove that G' is normal in G.
   (b) Prove that G/G' is abelian.
   (c) If G/N is abelian, prove that $N \supset G'$.
   (d) Prove that if H is a subgroup of G and $H \supset G'$, then H is normal in G.

6. If N, M are normal subgroups of G, prove that $NM/M \approx N/N \cap M$.

7. Let V be the set of real numbers, and for a, b real, $a \neq 0$ let $\tau_{ab}: V \to V$ defined by $\tau_{ab}(x) = ax + b$. Let $G = \{\tau_{ab} \mid a, b \text{ real}, a \neq 0\}$ and let $N = \{\tau_{1b} \in G\}$. Prove that N is a normal subgroup of G and that $G/N \approx$ group of nonzero real numbers under multiplication.

8. Let G be the dihedral group defined as the set of all formal symbols $x^i y^j$, $i = 0, 1$, $j = 0, 1, \ldots, n-1$, where $x^2 = e$, $y^n = e$, $xy = y^{-1}x$. Prove
   (a) The subgroup $N = \{e, y, y^2, \ldots, y^{n-1}\}$ is normal in G.
   (b) That $G/N \approx W$, where $W = \{1, -1\}$ is the group under the multiplication of the real numbers.

9. Prove that the center of a group is always a normal subgroup.

10. Prove that a group of order 9 is abelian.

11. If G is a non-abelian group of order 6, prove that $G \approx S_3$.

12. If G is abelian and if N is any subgroup of G, prove that G/N is abelian.

13. Let G be the dihedral group defined in Problem 8. Find the center of G.

14. Let G be as in Problem 13. Find G', the commutator subgroup of G.

15. Let G be the group of nonzero complex numbers under multiplication and let N be the set of complex numbers of absolute value 1 (that is, $a + bi \in N$ if $a^2 + b^2 = 1$). Show that G/N is isomorphic to the group of all positive real numbers under multiplication.

#16. Let G be the group of all nonzero complex numbers under multiplication and let $\bar{G}$ be the group of all real 2×2 matrices of the form $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, where not both a and b are 0, under matrix multiplication. Show that G and $\bar{G}$ are isomorphic by exhibiting an isomorphism of G onto $\bar{G}$.

*17. Let G be the group of real numbers under addition and let N be the subgroup of G consisting of all the integers. Prove that G/N is isomorphic to the group of all complex numbers of absolute value 1 under multiplication.

#18. Let G be the group of all real 2×2 matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, with $ad - bc \neq 0$, under matrix multiplication, and let

$$N = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G \mid ad - bc = 1 \right\}.$$

Prove that $N \supset G'$, the commutator subgroup of G.

*#19. In Problem 18 show, in fact, that $N = G'$.

#20. Let G be the group of all real 2×2 matrices of the form $\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$, where $ad \neq 0$, under matrix multiplication. Show that G' is precisely the set of all matrices of the form $\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$.

21. Let S_1 and S_2 be two sets. Suppose that there exists a one-to-one mapping ψ of S_1 into S_2. Show that there exists an isomorphism of $A(S_1)$ into $A(S_2)$, where $A(S)$ means the set of all one-to-one mappings of S onto itself.

2.8 Automorphisms

In the preceding section the concept of an isomorphism of one group into another was defined and examined. The special case in which the isomorphism maps a given group into itself should obviously be of some importance. We use the word "into" advisedly, for groups G do exist which have isomorphisms mapping G into, and not onto, itself. The easiest such example is the following: Let G be the group of integers under addition and define $\phi: G \to G$ by $\phi: x \to 2x$ for every $x \in G$. Since $\phi: x + y \to 2(x + y) = 2x + 2y$, ϕ is a homomorphism. Also if the image of x and y under ϕ are equal, then $2x = 2y$ whence $x = y$. ϕ is thus an isomorphism. Yet ϕ is not onto, for the image of any integer under ϕ is an even integer, so, for instance, 1 does not appear an image under ϕ of any element of G. Of greatest interest to us will be the isomorphisms of a group *onto* itself.

DEFINITION By an *automorphism* of a group G we shall mean an isomorphism of G onto itself.

As we mentioned in Chapter 1, whenever we talk about mappings of a set into itself we shall write the mappings on the right side, thus if $T: S \to S$, $x \in S$, then xT is the image of x under T.

Let I be the mapping of G which sends every element onto itself, that is, $xI = x$ for all $x \in G$. Trivially I is an automorphism of G. Let $\mathscr{A}(G)$ denote the set of all automorphisms of G; being a subset of $A(G)$, the set of one-to-one mappings of G onto itself, for elements of $\mathscr{A}(G)$ we can use the product of $A(G)$, namely, composition of mappings. This product then satisfies the associative law in $A(G)$, and so, *a fortiori*, in $\mathscr{A}(G)$. Also I, the unit element of $A(G)$, is in $\mathscr{A}(G)$, so $\mathscr{A}(G)$ is not empty.

An obvious fact that we should try to establish is that $\mathscr{A}(G)$ is a subgroup of $A(G)$, and so, in its own rights, $\mathscr{A}(G)$ should be a group. If T_1, T_2 are in $\mathscr{A}(G)$ we already know that $T_1T_2 \in A(G)$. We want it to be in the smaller set $\mathscr{A}(G)$. We proceed to verify this. For all $x, y \in G$,

$$(xy)T_1 = (xT_1)(yT_1),$$
$$(xy)T_2 = (xT_2)(yT_2),$$

therefore

$$(xy)T_1T_2 = ((xy)T_1)T_2 = ((xT_1)(yT_1))T_2$$
$$= ((xT_1)T_2)((yT_1)T_2) = (xT_1T_2)(yT_1T_2).$$

That is, $T_1T_2 \in \mathscr{A}(G)$. There is only one other fact that needs verifying in order that $\mathscr{A}(G)$ be a subgroup of $A(G)$, namely, that if $T \in \mathscr{A}(G)$, then $T^{-1} \in \mathscr{A}(G)$. If $x, y \in G$, then

$$((xT^{-1})(yT^{-1}))T = ((xT^{-1})T)((yT^{-1})T) = (xI)(yI) = xy,$$

thus

$$(xT^{-1})(yT^{-1}) = (xy)T^{-1},$$

placing T^{-1} in $\mathscr{A}(G)$. Summarizing these remarks, we have proved

LEMMA 2.8.1 *If G is a group, then $\mathscr{A}(G)$, the set of automorphisms of G, is also a group.*

Of course, as yet, we have no way of knowing that $\mathscr{A}(G)$, in general, has elements other than I. If G is a group having only two elements, the reader should convince himself that $\mathscr{A}(G)$ consists only of I. For groups G with more than two elements, $\mathscr{A}(G)$ always has more than one element.

What we should like is a richer sample of automorphisms than the ones we have (namely, I). If the group G is abelian and there is some element $x_0 \in G$ satisfying $x_0 \neq x_0^{-1}$, we can write down an explicit automorphism, the mapping T defined by $xT = x^{-1}$ for all $x \in G$. For any group G, T is onto; for any abelian G, $(xy)T = (xy)^{-1} = y^{-1}x^{-1} = x^{-1}y^{-1} = (xT)(yT)$. Also $x_0T = x_0^{-1} \neq x_0$, so $T \neq I$.

However, the class of abelian groups is a little limited, and we should like to have some automorphisms of non-abelian groups. Strangely enough the task of finding automorphisms for such groups is easier than for abelian groups.

Let G be a group; for $g \in G$ define $T_g: G \to G$ by $xT_g = g^{-1}xg$ for all $x \in G$. We claim that T_g is an automorphism of G. First, T_g is onto, for given $y \in G$, let $x = gyg^{-1}$. Then $xT_g = g^{-1}(x)g = g^{-1}(gyg^{-1})g = y$, so T_g is onto. Now consider, for $x, y \in G$, $(xy)T_g = g^{-1}(xy)g = g^{-1}(xgg^{-1}y)g = (g^{-1}xg)(g^{-1}yg) = (xT_g)(yT_g)$. Consequently T_g is a homomorphism of G onto itself. We further assert that T_g is one-to-one, for if $xT_g = yT_g$, then $g^{-1}xg = g^{-1}yg$, so by the cancellation laws in G, $x = y$. T_g is called the *inner automorphism* corresponding to g. If G is non-abelian, there is a pair $a, b \in G$ such that $ab \neq ba$; but then $bT_a = a^{-1}ba \neq b$, so that $T_a \neq I$. Thus for a non-abelian group G there always exist nontrivial automorphisms.

Let $\mathscr{I}(G) = \{T_g \in \mathscr{A}(G) \mid g \in G\}$. The computation of T_{gh}, for $g, h \in G$, might be of some interest. So, suppose $x \in G$; by definition,

$$xT_{gh} = (gh)^{-1}x(gh) = h^{-1}g^{-1}xgh = (g^{-1}xg)T_h = (xT_g)T_h = xT_gT_h.$$

Looking at the start and finish of this chain of equalities we find that $T_{gh} = T_gT_h$. This little remark is both interesting and suggestive. It is of interest because it immediately yields that $\mathscr{I}(G)$ is a subgroup of $\mathscr{A}(G)$. (Verify!) $\mathscr{I}(G)$ is usually called *the group of inner automorphisms of* G. It is suggestive, for if we consider the mapping $\psi: G \to \mathscr{A}(G)$ defined by $\psi(g) = T_g$ for every $g \in G$, then $\psi(gh) = T_{gh} = T_gT_h = \psi(g)\psi(h)$. That is, ψ is a homomorphism of G into $\mathscr{A}(G)$ whose image is $\mathscr{I}(G)$. What is the kernel of ψ? Suppose we call it K, and suppose $g_0 \in K$. Then $\psi(g_0) = I$, or, equivalently, $T_{g_0} = I$. But this says that for any $x \in G$, $xT_{g_0} = x$; however, $xT_{g_0} = g_0^{-1}xg_0$, and so $x = g_0^{-1}xg_0$ for all $x \in G$. Thus $g_0x = g_0g_0^{-1}xg_0 = xg_0$; g_0 must commute with all elements of G. But the center of G, Z, was defined to be precisely all elements in G which commute with every element of G. (See Problem 15, Section 2.5.) Thus $K \subset Z$. However, if $z \in Z$, then $xT_z = z^{-1}xz = z^{-1}(zx)$ (since $zx = xz$) $= x$, whence $T_z = I$ and so $z \in K$. Therefore, $Z \subset K$. Having proved both $K \subset Z$ and $Z \subset K$ we have that $Z = K$. Summarizing, ψ is a homomorphism of G into $\mathscr{A}(G)$ with image $\mathscr{I}(G)$ and kernel Z. By Theorem 2.7.1 $\mathscr{I}(G) \approx G/Z$. In order to emphasize this general result we record it as

LEMMA 2.8.2 *$\mathscr{I}(G) \approx G/Z$, where $\mathscr{I}(G)$ is the group of inner automorphisms of G, and Z is the center of G.*

Suppose that ϕ is an automorphisms of a group G, and suppose that $a \in G$ has order n (that is, $a^n = e$ *but for no lower positive power*). Then $\phi(a)^n = \phi(a^n) = \phi(e) = e$, hence $\phi(a)^n = e$. If $\phi(a)^m = e$ for some $0 < m < n$, then $\phi(a^m) = \phi(a)^m = e$, which implies, since ϕ is one-to-one, that $a^m = e$, a contradiction. Thus

LEMMA 2.8.3 *Let G be a group and ϕ an automorphism of G. If $a \in G$ is of order $o(a) > 0$, then $o(\phi(a)) = o(a)$.*

Automorphisms of groups can be used as a means of constructing new groups from the original group. Before explaining this abstractly, we consider a particular example.

Let G be a cyclic group of order 7, that is, G consists of all a^i, where we assume $a^7 = e$. The mapping $\phi: a^i \to a^{2i}$, as can be checked trivially, is an automorphism of G of order 3, that is, $\phi^3 = I$. Let x be a symbol which we formally subject to the following conditions: $x^3 = e$, $x^{-1}a^i x = \phi(a^i) = a^{2i}$, and consider all formal symbols $x^i a^j$, where $i = 0, 1, 2$ and $j = 0, 1, 2, \ldots, 6$. We declare that $x^i a^j = x^k a^l$ if and only if $i \equiv k \bmod 3$ and $j \equiv l \bmod 7$. We multiply these symbols using the rules $x^3 = a^7 = e$, $x^{-1}ax = a^2$. For instance, $(xa)(xa^2) = x(ax)a^2 = x(xa^2)a^2 = x^2a^4$. The reader can verify that one obtains, in this way, a non-abelian group of order 21.

Generally, if G is a group, T an automorphism of order r of G which is not an inner automorphism, pick a symbol x and consider all elements $x^i g$, $i = 0, \pm 1, \pm 2, \ldots$, $g \in G$ subject to $x^i g = x^{i'} g'$ if and only if $i \equiv i' \bmod r$, $g = g'$ and $x^{-1} g^i x = gT^i$ for all i. This way we obtain a larger group $\{G, T\}$; G is normal in $\{G, T\}$ and $\{G, T\}/G \approx$ group generated by T = cyclic group of order r.

We close the section by determining $\mathscr{A}(G)$ for all cyclic groups.

Example 2.8.1 Let G be a finite cyclic group of order r, $G = (a)$, $a^r = e$. Suppose T is an automorphism of G. If aT is known, since $a^i T = (aT)^i$, $a^i T$ is determined, so gT is determined for all $g \in G = (a)$. Thus we need consider only possible images of a under T. Since $aT \in G$, and since every element in G is a power of a, $aT = a^t$ for some integer $0 < t < r$. However, since T is an automorphism, aT must have the same order as a (Lemma 2.8.3), and this condition, we claim, forces t to be relatively prime to r. For if $d \mid t$, $d \mid r$, then $(aT)^{r/d} = a^{t(r/d)} = a^{r(t/d)} = (a^r)^{t/d} = e$; thus aT has order a divisor of r/d, which, combined with the fact that aT has order r, leads us to $d = 1$. Conversely, for any $0 < s < r$ and relatively prime to r, the mapping $S: a^i \to a^{si}$ is an automorphism of G. Thus $\mathscr{A}(G)$ is in one-to-one correspondence with the group U_r of integers less than r and relatively prime to r under multiplication modulo r. We claim not only is there such a one-to-one correspondence, but there is one which furthermore is an isomorphism. Let us label the elements of $\mathscr{A}(G)$ as T_i where $T_i: a \to a^i$, $0 < i < r$ and relatively prime to r; $T_i T_j: a \to a^i \to (a^i)^j = a^{ij}$, thus $T_i T_j = T_{ij}$. The mapping $i \to T_i$ exhibits the isomorphism of U_r onto $\mathscr{A}(G)$. Here then, $\mathscr{A}(G) \approx U_r$.

Example 2.8.2 G is an infinite cyclic group. That is, G consists of all a^i, $i = 0, \pm 1, \pm 2, \ldots$, where we assume that $a^i = e$ if and only if $i = 0$. Suppose that T is an automorphism of G. As in Example 2.8.1, $aT = a^t$.

The question now becomes, What values of t are possible? Since T is an automorphism of G, it maps G onto itself, so that $a = gT$ for some $g \in G$. Thus $a = a^iT = (aT)^i$ for some integer i. Since $aT = a^t$, we must have that $a = a^{ti}$, so that $a^{ti-1} = e$. Hence $ti - 1 = 0$; that is, $ti = 1$. Clearly, since t and i are integers, this must force $t = \pm 1$, and each of these gives rise to an automorphism, $t = 1$ yielding the identity automorphism I, $t = -1$ giving rise to the automorphism $T: g \to g^{-1}$ for every g in the cyclic group G. Thus here, $\mathscr{A}(G) \approx$ cyclic group of order 2.

Problems

1. Are the following mappings automorphisms of their respective groups?
   (a) G group of integers under addition, $T: x \to -x$.
   (b) G group of positive reals under multiplication, $T: x \to x^2$.
   (c) G cyclic group of order 12, $T: x \to x^3$.
   (d) G is the group S_3, $T: x \to x^{-1}$.
2. Let G be a group, H a subgroup of G, T an automorphism of G. Let $(H)T = \{hT \mid h \in H\}$. Prove $(H)T$ is a subgroup of G.
3. Let G be a group, T an automorphism of G, N a normal subgroup of G. Prove that $(N)T$ is a normal subgroup of G.
4. For $G = S_3$ prove that $G \approx \mathscr{I}(G)$.
5. For any group G prove that $\mathscr{I}(G)$ is a normal subgroup of $\mathscr{A}(G)$ (the group $\mathscr{A}(G)/\mathscr{I}(G)$ is called the *group of outer automorphisms* of G).
6. Let G be a group of order 4, $G = \{e, a, b, ab\}$, $a^2 = b^2 = e$, $ab = ba$. Determine $\mathscr{A}(G)$.
7. (a) A subgroup C of G is said to be a *characteristic subgroup* of G if $(C)T \subset C$ for all automorphisms T of G. Prove a characteristic subgroup of G must be a normal subgroup of G.
   (b) Prove that the converse of (a) is false.
8. For any group G, prove that the commutator subgroup G' is a characteristic subgroup of G. (See Problem 5, Section 2.7).
9. If G is a group, N a normal subgroup of G, M a characteristic subgroup of N, prove that M is a normal subgroup of G.
10. Let G be a finite group, T an automorphism of G with the property that $xT = x$ for $x \in G$ if and only if $x = e$. Prove that every $g \in G$ can be represented as $g = x^{-1}(xT)$ for some $x \in G$.
11. Let G be a finite group, T an automorphism of G with the property that $xT = x$ if and only if $x = e$. Suppose further that $T^2 = I$. Prove that G must be abelian.

*12. Let G be a finite group and suppose the automorphism T sends more than three-quarters of the elements of G onto their inverses. Prove that $xT = x^{-1}$ for all $x \in G$ and that G is abelian.

13. In Problem 12, can you find an example of a finite group which is non-abelian and which has an automorphism which maps exactly three-quarters of the elements of G onto their inverses?

*14. Prove that every finite group having more than two elements has a nontrivial automorphism.

*15. Let G be a group of order $2n$. Suppose that half of the elements of G are of order 2, and the other half form a subgroup H of order n. Prove that H is of odd order and is an abelian subgroup of G.

*16. Let $\phi(n)$ be the Euler ϕ-function. If $a > 1$ is an integer, prove that $n \mid \phi(a^n - 1)$.

17. Let G be a group and Z the center of G. If T is any automorphism of G, prove that $(Z)T \subset Z$.

18. Let G be a group and T an automorphism of G. If, for $a \in G$, $N(a) = \{x \in G \mid xa = ax\}$, prove that $N(aT) = (N(a))T$.

19. Let G be a group and T an automorphism of G. If N is a normal subgroup of G such that $(N)T \subset N$, show how you could use T to define an automorphism of G/N.

20. Use the discussion following Lemma 2.8.3 to construct
(a) a non-abelian group of order 55.
(b) a non-abelian group of order 203.

21. Let G be the group of order 9 generated by elements a, b, where $a^3 = b^3 = e$. Find all the automorphisms of G.

2.9 Cayley's Theorem

When groups first arose in mathematics they usually came from some specific source and in some very concrete form. Very often it was in the form of a set of transformations of some particular mathematical object. In fact, most finite groups appeared as groups of permutations, that is, as subgroups of S_n. ($S_n = A(S)$ when S is a finite set with n elements.) The English mathematician Cayley first noted that every group could be realized as a subgroup of $A(S)$ for some S. Our concern, in this section, will be with a presentation of Cayley's theorem and some related results.

THEOREM 2.9.1 (CAYLEY) *Every group is isomorphic to a subgroup of $A(S)$ for some appropriate S.*

Proof. Let G be a group. For the set S we will use the elements of G; that is, put $S = G$. If $g \in G$, define $\tau_g : S(= G) \to S(= G)$ by $x\tau_g = xg$

for every $x \in G$. If $y \in G$, then $y = (yg^{-1})g = (yg^{-1})\tau_g$, so that τ_g maps S onto itself. Moreover, τ_g is one-to-one, for if $x, y \in S$ and $x\tau_g = y\tau_g$, then $xg = yg$, which, by the cancellation property of groups, implies that $x = y$. We have proved that for every $g \in G$, $\tau_g \in A(S)$.

If $g, h \in G$, consider τ_{gh}. For any $x \in S = G$, $x\tau_{gh} = x(gh) = (xg)h = (x\tau_g)\tau_h = x\tau_g\tau_h$. Note that we used the associative law in a very essential way here. From $x\tau_{gh} = x\tau_g\tau_h$ we deduce that $\tau_{gh} = \tau_g\tau_h$. Therefore, if $\psi: G \to A(S)$ is defined by $\psi(g) = \tau_g$, the relation $\tau_{gh} = \tau_g\tau_h$ tells us that ψ is a homomorphism. What is the kernel K of ψ? If $g_0 \in K$, then $\psi(g_0) = \tau_{g_0}$ is the identity map on S, so that for $x \in G$, and, in particular, for $e \in G$, $e\tau_{g_0} = e$. But $e\tau_{g_0} = eg_0 = g_0$. Thus comparing these two expressions for $e\tau_{g_0}$ we conclude that $g_0 = e$, whence $K = (e)$. Thus by the corollary to Lemma 2.7.4 ψ is an isomorphism of G into $A(S)$, proving the theorem.

The theorem enables us to exhibit any abstract group as a more concrete object, namely, as a group of mappings. However, it has its shortcomings; for if G is a finite group of order $o(G)$, then, using $S = G$, as in our proof, $A(S)$ has $o(G)!$ elements. Our group G of order $o(G)$ is somewhat lost in the group $A(S)$ which, with its $o(G)!$ elements, is huge in comparison to G. We ask: Can we find a more economical S, one for which $A(S)$ is smaller? This we now attempt to accomplish.

Let G be a group, H a subgroup of G. Let S be the set whose elements are the right cosets of H in G. That is, $S = \{Hg \mid g \in G\}$. S need not be a group itself, in fact, it would be a group only if H were a normal subgroup of G. However, we can make our group G act on S in the following natural way: for $g \in G$ let $t_g: S \to S$ be defined by $(Hx)t_g = Hxg$. Emulating the proof of Theorem 2.9.1 we can easily prove

1. $t_g \in A(S)$ for every $g \in G$.
2. $t_{gh} = t_g t_h$.

Thus the mapping $\theta: G \to A(S)$ defined by $\theta(g) = t_g$ is a homomorphism of G into $A(S)$. Can one always say that θ is an isomorphism? Suppose that K is the kernel of θ. If $g_0 \in K$, then $\theta(g_0) = t_{g_0}$ is the identity map on S, so that for every $X \in S$, $Xt_{g_0} = X$. Since every element of S is a right coset of H in G, we must have that $Hat_{g_0} = Ha$ for every $a \in G$, and using the definition of t_{g_0}, namely, $Hat_{g_0} = Hag_0$, we arrive at the identity $Hag_0 = Ha$ for every $a \in G$. On the other hand, if $b \in G$ is such that $Hxb = Hx$ for every $x \in G$, retracing our argument we could show that $b \in K$. Thus $K = \{b \in G \mid Hxb = Hx \text{ all } x \in G\}$. We claim that from this characterization of K, K must be the largest normal subgroup of G which is contained in H. We first explain the use of the word largest; by this we mean that if N is a normal subgroup of G which is contained in H, then N must be contained in K. We wish to show this is the case. That K is a normal subgroup

of G follows from the fact that it is the kernel of a homomorphism of G. Now we assert that $K \subset H$, for if $b \in K$, $Hab = Ha$ for every $a \in G$, so, in particular, $Hb = Heb = He = H$, whence $b \in H$. Finally, if N is a normal subgroup of G which is contained in H, if $n \in N$, $a \in G$, then $ana^{-1} \in N \subset H$, so that $Hana^{-1} = H$; thus $Han = Ha$ for all $a \in G$. Therefore, $n \in K$ by our characterization of K.

We have proved

THEOREM 2.9.2 *If G is a group, H a subgroup of G, and S is the set of all right cosets of H in G, then there is a homomorphism θ of G into $A(S)$ and the kernel of θ is the largest normal subgroup of G which is contained in H.*

The case $H = (e)$ just yields Cayley's theorem (Theorem 2.9.1). If H should happen to have no normal subgroup of G other than (e) in it, then θ must be an isomorphism of G into $A(S)$. In this case we would have cut down the size of the S used in proving Theorem 2.9.1. This is interesting mostly for finite groups. For we shall use this observation both as a means of proving certain finite groups have nontrivial normal subgroups, and also as a means of representing certain finite groups as permutation groups on small sets.

We examine these remarks a little more closely. Suppose that G has a subgroup H whose index $i(H)$ (that is, the number of right cosets of H in G) satisfies $i(H)! < o(G)$. Let S be the set of all right cosets of H in G. The mapping, θ, of Theorem 2.9.2 cannot be an isomorphism, for if it were, $\theta(G)$ would have $o(G)$ elements and yet would be a subgroup of $A(S)$ which has $i(H)! < o(G)$ elements. Therefore the kernel of θ must be larger than (e); this kernel being the largest normal subgroup of G which is contained in H, we can conclude that H contains a nontrivial normal subgroup of G.

However, the argument used above has implications even when $i(H)!$ is not less than $o(G)$. If $o(G)$ does not divide $i(H)!$ then by invoking Lagrange's theorem we know that $A(S)$ can have no subgroup of order $o(G)$, hence no subgroup isomorphic to G. However, $A(S)$ does contain $\theta(G)$, whence $\theta(G)$ cannot be isomorphic to G; that is, θ cannot be an isomorphism. But then, as above, H must contain a nontrivial normal subgroup of G.

We summarize this as

LEMMA 2.9.1 *If G is a finite group, and $H \neq G$ is a subgroup of G such that $o(G) \nmid i(H)!$ then H must contain a nontrivial normal subgroup of G. In particular, G cannot be simple.*

APPLICATIONS

1. Let G be a group of order 36. Suppose that G has a subgroup H of order 9 (we shall see later that this is always the case). Then $i(H) = 4$,

$4! = 24 < 36 = o(G)$ so that in H there must be a normal subgroup $N \neq (e)$, of G, of order a divisor of 9, that is, of order 3 or 9.

2. Let G be a group of order 99 and suppose that H is a subgroup of G of order 11 (we shall also see, later, that this must be true). Then $i(H) = 9$, and since $99 \nmid 9!$ there is a nontrivial normal subgroup $N \neq (e)$ of G in H. Since H is of order 11, which is a prime, its only subgroup other than (e) is itself, implying that $N = H$. That is, H itself is a normal subgroup of G.

3. Let G be a non-abelian group of order 6. By Problem 11, Section 2.3, there is an $a \neq e \in G$ satisfying $a^2 = e$. Thus the subgroup $H = \{e, a\}$ is of order 2, and $i(H) = 3$. Suppose, for the moment, that we know that H is not normal in G. Since H has only itself and (e) as subgroups, H has no nontrivial normal subgroups of G in it. Thus G is isomorphic to a subgroup T of order 6 in $A(S)$, where S is the set of right cosets of H in G. Since $o(A(S)) = i(H)! = 3! = 6$, $T = S$. In other words, $G \approx A(S) = S_3$. We would have proved that any non-abelian group of order 6 is isomorphic to S_3. All that remains is to show that H is not normal in G. Since it might be of some interest we go through a detailed proof of this. If $H = \{e, a\}$ were normal in G, then for every $g \in G$, since $gag^{-1} \in H$ and $gag^{-1} \neq e$, we would have that $gag^{-1} = a$, or, equivalently, that $ga = ag$ for every $g \in G$. Let $b \in G$, $b \notin H$, and consider $N(b) = \{x \in G \mid xb = bx\}$. By an earlier problem, $N(b)$ is a subgroup of G, and $N(b) \supset H$; $N(b) \neq H$ since $b \in N(b)$, $b \notin H$. Since H is a subgroup of $N(b)$, $o(H) \mid o(N(b)) \mid 6$. The only even number n, $2 < n \leq 6$ which divides 6 is 6. So $o(N(b)) = 6$; whence b commutes with all elements of G. Thus every element of G commutes with every other element of G, making G into an abelian group, contrary to assumption. Thus H could not have been normal in G. This proof is somewhat long-winded, but it illustrates some of the ideas already developed.

Problems

1. Let G be a group; consider the mappings of G into itself, λ_g, defined for $g \in G$ by $x\lambda_g = gx$ for all $x \in G$. Prove that λ_g is one-to-one and onto, and that $\lambda_{gh} = \lambda_h\lambda_g$.
2. Let λ_g be defined as in Problem 1, τ_g as in the proof of Theorem 2.9.1. Prove that for any $g, h \in G$, the mappings λ_g, τ_h satisfy $\lambda_g\tau_h = \tau_h\lambda_g$. (*Hint:* For $x \in G$ consider $x(\lambda_g\tau_h)$ and $x(\tau_h\lambda_g)$.)
3. If θ is a one-to-one mapping of G onto itself such that $\lambda_g\theta = \theta\lambda_g$ for all $g \in G$, prove that $\theta = \tau_h$ for some $h \in G$.
4. (a) If H is a subgroup of G show that for every $g \in G$, gHg^{-1} is a subgroup of G.

(b) Prove that W = intersection of all gHg^{-1} is a normal subgroup of G.

5. Using Lemma 2.9.1 prove that a group of order p^2, where p is a prime number, must have a normal subgroup of order p.

6. Show that in a group G of order p^2 any normal subgroup of order p must lie in the center of G.

7. Using the result of Problem 6, prove that any group of order p^2 is abelian.

8. If p is a prime number, prove that any group G of order $2p$ must have a subgroup of order p, and that this subgroup is normal in G.

9. If $o(G)$ is pq where p and q are distinct prime numbers and if G has a normal subgroup of order p and a normal subgroup of order q, prove that G is cyclic.

*10. Let $o(G)$ be pq, $p > q$ are primes, prove
(a) G has a subgroup of order p and a subgroup of order q.
(b) If $q \nmid p - 1$, then G is cyclic.
(c) Given two primes p, q, $q \mid p - 1$, there exists a non-abelian group of order pq.
(d) Any two non-abelian groups of order pq are isomorphic.

2.10 Permutation Groups

We have seen that every group can be represented isomorphically as a subgroup of $A(S)$ for some set S, and, in particular, a finite group G can be represented as a subgroup of S_n, for some n, where S_n is the symmetric group of degree n. This clearly shows that the groups S_n themselves merit closer examination.

Suppose that S is a finite set having n elements $x_1, x_2, \ldots, x_n$. If $\phi \in A(S) = S_n$, then ϕ is a one-to-one mapping of S onto itself, and we could write ϕ out by showing what it does to every element, e.g., $\phi: x_1 \to x_2$, $x_2 \to x_4$, $x_4 \to x_3$, $x_3 \to x_1$. But this is very cumbersome. One short cut might be to write ϕ out as

$$\begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ x_{i_1} & x_{i_2} & x_{i_3} & \cdots & x_{i_n} \end{pmatrix},$$

where x_{i_k} is the image of x_i under ϕ. Returning to our example just above, ϕ might be represented by

$$\begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ x_2 & x_4 & x_1 & x_3 \end{pmatrix}.$$

While this notation is a little handier there still is waste in it, for there seems to be no purpose served by the symbol x. We could equally well represent the permutation as

$$\begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix}.$$

Our specific example would read

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}.$$

Given two permutations θ, ψ in S_n, using this symbolic representation of θ and ψ, what would the representation of $\theta\psi$ be? To compute it we could start and see what $\theta\psi$ does to x_1 (henceforth written as 1). θ takes 1 into i_1, while ψ takes i_1 into k, say, then $\theta\psi$ takes 1 into k. Then repeat this procedure for 2, 3, . . . , n. For instance, if θ is the permutation represented by

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}$$

and ψ by

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix},$$

then $i_1 = 3$ and ψ takes 3 into 2, so $k = 2$ and $\theta\psi$ takes 1 into 2. Similarly $\theta\psi$:$2 \to 1$, $3 \to 3$, $4 \to 4$. That is, the representation for $\theta\psi$ is

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 4 \end{pmatrix}.$$

If we write

$$\theta = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}$$

and

$$\psi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix},$$

then

$$\theta\psi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 4 \end{pmatrix}.$$

This is the way we shall multiply the symbols of the form

$$\begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 & \cdots & n \\ k_1 & k_2 & \cdots & k_n \end{pmatrix}.$$

Let S be a set and $\theta \in A(S)$. Given two elements $a, b \in S$ we define $a \equiv {}_\theta b$ if and only if $b = a\theta^i$ for some integer i (i can be positive, negative, or 0). We claim this defines an equivalence relation on S. For

1. $a \equiv {}_\theta a$ since $a = a\theta^0 = ae$.
2. If $a \equiv {}_\theta b$, then $b = a\theta^i$, so that $a = b\theta^{-i}$, whence $b \equiv {}_\theta a$.
3. If $a \equiv {}_\theta b$, $b \equiv {}_\theta c$, then $b = a\theta^i$, $c = b\theta^j = (a\theta^i)\theta^j = a\theta^{i+j}$, which implies that $a \equiv {}_\theta c$.

This equivalence relation by Theorem 1.1.1 induces a decomposition of S into disjoint subsets, namely, the equivalence classes. We call the equivalence class of an element $s \in S$ the *orbit* of s under θ; thus the orbit of s under θ consists of all the elements $s\theta^i$, $i = 0, \pm 1, \pm 2, \ldots$.

In particular, if S is a finite set and $s \in S$, there is a smallest positive integer $l = l(s)$ depending on s such that $s\theta^l = s$. The orbit of s under θ then consists of the elements $s, s\theta, s\theta^2, \ldots, s\theta^{l-1}$. By a *cycle* of θ we mean the *ordered* set $(s, s\theta, s\theta^2, \ldots, s\theta^{l-1})$. If we know all the cycles of θ we clearly know θ since we would know the image of any element under θ. Before proceeding we illustrate these ideas with an example. Let

$$\theta = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 3 & 5 & 6 & 4 \end{pmatrix},$$

where S consists of the elements $1, 2, \ldots, 6$ (remember 1 stands for x_1, 2 for x_2, etc.). Starting with 1, then the orbit of 1 consists of $1 = 1\theta^0$, $1\theta^1 = 2$, $1\theta^2 = 2\theta = 1$, so the orbit of 1 is the set of elements 1 and 2. This tells us the orbit of 2 is the same set. The orbit of 3 consists just of 3; that of 4 consists of the elements 4, $4\theta = 5$, $4\theta^2 = 5\theta = 6$, $4\theta^3 = 6\theta = 4$. The cycles of θ are (1, 2), (3), (4, 5, 6).

We digress for a moment, leaving our particular θ. Suppose that by the cycle $(i_1, i_2, \ldots, i_r)$ we mean the permutation ψ which sends i_1 into i_2, i_2 into $i_3 \cdots i_{r-1}$ into i_r and i_r into i_1, and leaves all other elements of S fixed. Thus, for instance, if S consists of the elements $1, 2, \ldots, 9$, then the symbol (1, 3, 4, 2, 6) means the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 6 & 4 & 2 & 5 & 1 & 7 & 8 & 9 \end{pmatrix}.$$

We multiply cycles by multiplying the permutations they represent. Thus again, if S has 9 elements,

$$(1\ 2\ 3)(5\ 6\ 4\ 1\ 8)$$

$$= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 3 & 1 & 4 & 5 & 6 & 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 8 & 2 & 3 & 1 & 6 & 4 & 7 & 5 & 9 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 3 & 8 & 1 & 6 & 4 & 7 & 5 & 9 \end{pmatrix}.$$

Let us return to the ideas of the paragraph preceding the last one, and ask: Given the permutation

$$\theta = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 3 & 8 & 1 & 6 & 4 & 7 & 5 & 9 \end{pmatrix},$$

what are the cycles of θ? We first find the orbit of 1; namely, $1, 1\theta = 2, 1\theta^2 = 2\theta = 3, 1\theta^3 = 3\theta = 8, 1\theta^4 = 8\theta = 5, 1\theta^5 = 5\theta = 6, 1\theta^6 = 6\theta = 4, 1\theta^7 = 4\theta = 1$. That is, the orbit of 1 is the set $\{1, 2, 3, 8, 5, 6, 4\}$. The orbits of 7 and 9 can be found to be $\{7\}$, $\{9\}$, respectively. The cycles of θ thus are (7), (9), $(1, 1\theta, 1\theta^2, \ldots, 1\theta^6) = (1, 2, 3, 8, 5, 6, 4)$. The reader should now verify that if he takes the product (as defined in the last paragraph) of $(1, 2, 3, 8, 5, 6, 4)$, (7), (9) he will obtain θ. That is, at least in this case, θ is the product of its cycles.

But this is no accident for it is now trivial to prove

LEMMA 2.10.1 *Every permutation is the product of its cycles.*

Proof. Let θ be the permutation. Then its cycles are of the form $(s, s\theta, \ldots, s\theta^{l-1})$. By the multiplication of cycles, as defined above, and since the cycles of θ are disjoint, the image of $s' \in S$ under θ, which is $s'\theta$, is the same as the image of s' under the product, ψ, of all the distinct cycles of θ. So θ, ψ have the same effect on every element of S, hence $\theta = \psi$, which is what we sought to prove.

If the remarks above are still not transparent at this point, the reader should take a given permutation, find its cycles, take their product, and verify the lemma. In doing so the lemma itself will become obvious.

Lemma 2.10.1 is usually stated in the form *every permutation can be uniquely expressed as a product of disjoint cycles.*

Consider the m-cycle $(1, 2, \ldots, m)$. A simple computation shows that $(1, 2, \ldots, m) = (1, 2)(1, 3)\cdots(1, m)$. More generally the m-cycle $(a_1, a_2, \ldots, a_m) = (a_1, a_2)(a_1, a_3)\cdots(a_1, a_m)$. This decomposition is not unique; by this we mean that an m-cycle can be written as a product of 2-cycles in more than one way. For instance, $(1, 2, 3) = (1, 2)(1, 3) = (3, 1)(3, 2)$. Now, since every permutation is a product of disjoint cycles and every cycle is a product of 2-cycles, we have proved

LEMMA 2.10.2 *Every permutation is a product of 2-cycles.*

We shall refer to 2-cycles as *transpositions.*

DEFINITION A permutation $\theta \in S_n$ is said to be an *even permutation* if it can be represented as a product of an even number of transpositions.

The definition given just insists that θ have one representation as a product of an even number of transpositions. Perhaps it has other representations as a product of an odd number of transpositions. We first want to show that this cannot happen. Frankly, we are not happy with the proof we give of this fact for it introduces a polynomial which seems extraneous to the matter at hand.

Consider the polynomial in n-variables

$$p(x_1, \ldots, x_n) = \prod_{i<j} (x_i - x_j).$$

If $\theta \in S_n$ let θ act on the polynomial $p(x_1, \ldots, x_n)$ by

$$\theta{:}p(x_1, \ldots, x_n) = \prod_{i<j} (x_i - x_j) \to \prod_{i<j} (x_{\theta(i)} - x_{\theta(j)}).$$

It is clear that $\theta{:}p(x_1, \ldots, x_n) \to \pm p(x_1, \ldots, x_n)$. For instance, in S_5, $\theta = (134)(25)$ takes

$$p(x_1, \ldots, x_5) = (x_1 - x_2)(x_1 - x_3)(x_1 - x_4)(x_1 - x_5)(x_2 - x_3) \times (x_2 - x_4)(x_2 - x_5)(x_3 - x_4)(x_3 - x_5)(x_4 - x_5)$$

into

$$(x_3 - x_5)(x_3 - x_4)(x_3 - x_1)(x_3 - x_2)(x_5 - x_4)(x_5 - x_1) \times (x_5 - x_2)(x_4 - x_1)(x_4 - x_2)(x_1 - x_2),$$

which can easily be verified to be $-p(x_1, \ldots, x_5)$.

If, in particular, θ is a transposition, $\theta{:}p(x_1, \ldots, x_n) \to -p(x_1, \ldots, x_n)$. (Verify!) Thus if a permutation Π can be represented as a product of an even number of transpositions in one representation, Π must leave $p(x_1, \ldots, x_n)$ fixed, so that any representation of Π as a product of transposition must be such that it leaves $p(x_1, \ldots, x_n)$ fixed; that is, in any representation it is a product of an even number of transpositions. This establishes that the definition given for an even permutation is a significant one. We call a permutation *odd* if it is not an even permutation.

The following facts are now clear:

1. The product of two even permutations is an even permutation.
2. The product of an even permutation and an odd one is odd (likewise for the product of an odd and even permutation).
3. The product of two odd permutations is an even permutation.

The rule for combining even and odd permutations is like that of combining even and odd numbers under addition. This is not a coincidence since this latter rule is used in establishing 1, 2, and 3.

Let A_n be the subset of S_n consisting of all even permutations. Since the product of two even permutations is even, A_n must be a subgroup of S_n. We claim it is normal in S_n. Perhaps the best way of seeing this is as follows:

let W be the group of real numbers 1 and -1 under multiplication. Define $\psi: S_n \to W$ by $\psi(s) = 1$ if s is an even permutation, $\psi(s) = -1$ if s is an odd permutation. By the rules 1, 2, 3 above ψ is a homomorphism onto W. The kernel of ψ is precisely A_n; being the kernel of a homomorphism A_n is a normal subgroup of S_n. By Theorem 2.7.1 $S_n/A_n \approx W$, so, since

$$2 = o(W) = o\left(\frac{S_n}{A_n}\right) = \frac{o(S_n)}{o(A_n)},$$

we see that $o(A_n) = \frac{1}{2}n!$. A_n is called the *alternating group* of degree n. We summarize our remarks in

LEMMA 2.10.3 *S_n has as a normal subgroup of index 2 the alternating group, A_n, consisting of all even permutations.*

At the end of the next section we shall return to S_n again.

Problems

1. Find the orbits and cycles of the following permutations:
   (a) $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 3 & 4 & 5 & 1 & 6 & 7 & 9 & 8 \end{pmatrix}$.
   (b) $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 5 & 4 & 3 & 1 & 2 \end{pmatrix}$.
2. Write the permutations in Problem 1 as the product of disjoint cycles.
3. Express as the product of disjoint cycles:
   (a) $(1, 2, 3)(4, 5)(1, 6, 7, 8, 9)(1, 5)$.
   (b) $(1, 2)(1, 2, 3)(1, 2)$.
4. Prove that $(1, 2, \ldots, n)^{-1} = (n, n-1, n-2, \ldots, 2, 1)$.
5. Find the cycle structure of all the powers of $(1, 2, \ldots, 8)$.
6. (a) What is the order of an n-cycle?
   (b) What is the order of the product of the disjoint cycles of lengths $m_1, m_2, \ldots, m_k$?
   (c) How do you find the order of a given permutation?
7. Compute $a^{-1}ba$, where
   (1) $a = (1, 3, 5)(1, 2)$, $b = (1, 5, 7, 9)$.
   (2) $a = (5, 7, 9)$, $b = (1, 2, 3)$.
8. (a) Given the permutation $x = (1, 2)(3, 4)$, $y = (5, 6)(1, 3)$, find a permutation a such that $a^{-1}xa = y$.
   (b) Prove that there is no a such that $a^{-1}(1, 2, 3)a = (1, 3)(5, 7, 8)$.
   (c) Prove that there is no permutation a such that $a^{-1}(1, 2)a = (3, 4)(1, 5)$.
9. Determine for what m an m-cycle is an even permutation.

10. Determine which of the following are even permutations:
(a) $(1, 2, 3)(1, 2)$.
(b) $(1, 2, 3, 4, 5)(1, 2, 3)(4, 5)$.
(c) $(1, 2)(1, 3)(1, 4)(2, 5)$.

11. Prove that the smallest subgroup of S_n containing $(1, 2)$ and $(1, 2, \ldots, n)$ is S_n. (In other words, these generate S_n.)

*12. Prove that for $n \geq 3$ the subgroup generated by the 3-cycles is A_n.

*13. Prove that if a normal subgroup of A_n contains even a single 3-cycle it must be all of A_n.

*14. Prove that A_5 has no normal subgroups $N \neq (e)$, A_5.

15. Assuming the result of Problem 14, prove that any subgroup of A_5 has order at most 12.

16. Find all the normal subgroups in S_4.

*17. If $n \geq 5$ prove that A_n is the only nontrivial normal subgroup in S_n.

Cayley's theorem (Theorem 2.9.1) asserts that every group is isomorphic to a subgroup of $A(S)$ for some S. In particular, it says that every finite group can be realized as a group of permutations. Let us call the realization of the group as a group of permutations as given in the proof of Theorem 2.9.1 the *permutation representation* of G.

18. Find the permutation representation of a cyclic group of order n.

19. Let G be the group $\{e, a, b, ab\}$ of order 4, where $a^2 = b^2 = e$, $ab = ba$. Find the permutation representation of G.

20. Let G be the group S_3. Find the permutation representation of S_3. (*Note:* This gives an isomorphism of S_3 into S_6.)

21. Let G be the group $\{e, \theta, a, b, c, \theta a, \theta b, \theta c\}$, where $a^2 = b^2 = c^2 = \theta$, $\theta^2 = e$, $ab = \theta ba = c$, $bc = \theta cb = a$, $ca = \theta ac = b$.
(a) Show that θ is in the center Z of G, and that $Z = \{e, \theta\}$.
(b) Find the commutator subgroup of G.
(c) Show that every subgroup of G is normal.
(d) Find the permutation representation of G.
(*Note:* G is often called the group of *quaternion units*; it, and algebraic systems constructed from it, will reappear in the book.)

22. Let G be the dihedral group of order $2n$ (see Problem 17, Section 2.6). Find the permutation representation of G.

Let us call the realization of a group G as a set of permutations given in Problem 1, Section 2.9 the *second permutation representation* of G.

23. Show that if G is an abelian group, then the permutation representation of G coincides with the second permutation representation of G (i.e., in the notation of the previous section, $\lambda_g = \tau_g$ for all $g \in G$.)

24. Find the second permutation representation of S_3. Verify directly from the permutations obtained here and in Problem 20 that $\lambda_a\tau_b = \tau_b\lambda_a$ for all $a, b \in S_3$.
25. Find the second permutation representation of the group G defined in Problem 21.
26. Find the second permutation representation of the dihedral group of order $2n$.

If H is a subgroup of G, let us call the mapping $\{t_g \mid g \in G\}$ defined in the discussion preceding Theorem 2.9.2 the *coset representation* of G by H. This also realizes G as a group of permutations, but not necessarily isomorphically, merely homomorphically (see Theorem 2.9.2).

27. Let $G = (a)$ be a cyclic group of order 8 and let $H = (a^4)$ be its subgroup of order 2. Find the coset representation of G by H.
28. Let G be the dihedral group of order $2n$ generated by elements a, b such that $a^2 = b^n = e$, $ab = b^{-1}a$. Let $H = \{e, a\}$. Find the coset representation of G by H.
29. Let G be the group of Problem 21 and let $H = \{e, \theta\}$. Find the coset representation of G by H.
30. Let G be S_n, the symmetric group of order n, acting as permutations on the set $\{1, 2, \ldots, n\}$. Let $H = \{\sigma \in G \mid n\sigma = n\}$.
    (a) Prove that H is isomorphic to S_{n-1}.
    (b) Find a set of elements $a_1, \ldots, a_n \in G$ such that $Ha_1, \ldots, Ha_n$ give all the right cosets of H in G.
    (c) Find the coset representation of G by H.

2.11 Another Counting Principle

Mathematics is rich in technique and arguments. In this great variety one of the most basic tools is counting. Yet, strangely enough, it is one of the most difficult. Of course, by counting we do not mean the creation of tables of logarithms or addition tables; rather, we mean the process of precisely accounting for all possibilities in highly complex situations. This can sometimes be done by a brute force case-by-case exhaustion, but such a routine is invariably dull and violates a mathematician's sense of aesthetics. One prefers the light, deft, delicate touch to the hammer blow. But the most serious objection to case-by-case division is that it works far too rarely. Thus in various phases of mathematics we find neat counting devices which tell us exactly how many elements, in some fairly broad context, satisfy certain conditions. A great favorite with mathematicians is the process of counting up a given situation in two different ways; the comparison of the

two counts is then used as a means of drawing conclusions. Generally speaking, one introduces an equivalence relation on a finite set, measures the size of the equivalence classes under this relation, and then equates the number of elements in the set to the sum of the orders of these equivalence classes. This kind of an approach will be illustrated in this section. We shall introduce a relation, prove it is an equivalence relation, and then find a neat algebraic description for the size of each equivalence class. From this simple description there will flow a stream of beautiful and powerful results about finite groups.

DEFINITION If $a, b \in G$, then b is said to be a *conjugate* of a in G if there exists an element $c \in G$ such that $b = c^{-1}ac$.

We shall write, for this, $a \sim b$ and shall refer to this relation as *conjugacy*.

LEMMA 2.11.1 *Conjugacy is an equivalence relation on G.*

Proof. As usual, in order to establish this, we must prove that

1. $a \sim a$;
2. $a \sim b$ implies that $b \sim a$;
3. $a \sim b$, $b \sim c$ implies that $a \sim c$

for all a, b, c in G.

We prove each of these in turn.

1. Since $a = e^{-1}ae$, $a \sim a$, with $c = e$ serving as the c in the definition of conjugacy.
2. If $a \sim b$, then $b = x^{-1}ax$ for some $x \in G$, hence, $a = (x^{-1})^{-1}b(x^{-1})$, and since $y = x^{-1} \in G$ and $a = y^{-1}by$, $b \sim a$ follows.
3. Suppose that $a \sim b$ and $b \sim c$ where $a, b, c \in G$. Then $b = x^{-1}ax$, $c = y^{-1}by$ for some $x, y \in G$. Substituting for b in the expression for c we obtain $c = y^{-1}(x^{-1}ax)y = (xy)^{-1}a(xy)$; since $xy \in G$, $a \sim c$ is a consequence.

For $a \in G$ let $C(a) = \{x \in G \mid a \sim x\}$. $C(a)$, the equivalence class of a in G under our relation, is usually called the *conjugate class* of a in G; it consists of the set of all distinct elements of the form $y^{-1}ay$ as y ranges over G.

Our attention now narrows to the case in which G is a finite group. Suppose that $C(a)$ has c_a elements. We seek an alternative description of c_a. Before doing so, note that $o(G) = \sum c_a$ where the sum runs over a set of $a \in G$ using one a from each conjugate class. This remark is, of course, merely a restatement of the fact that our equivalence relation—conjugacy—

induces a decomposition of G into disjoint equivalence classes—the conjugate classes. Of paramount interest now is an evaluation of c_a.

In order to carry this out we recall a concept introduced in Problem 13, Section 2.5. Since this concept is important—far too important to leave to the off-chance that the student solved the particular problem—we go over what may very well be familiar ground to many of the readers.

DEFINITION If $a \in G$, then $N(a)$, the *normalizer of a in G*, is the set $N(a) = \{x \in G \mid xa = ax\}$.

$N(a)$ consists of precisely those elements in G which commute with a.

LEMMA 2.11.2 *$N(a)$ is a subgroup of G.*

Proof. In this result the order of G, whether it be finite or infinite, is of no relevance, and so we put no restrictions on the order of G.

Suppose that $x, y \in N(a)$. Thus $xa = ax$ and $ya = ay$. Therefore, $(xy)a = x(ya) = x(ay) = (xa)y = (ax)y = a(xy)$, in consequence of which $xy \in N(a)$. From $ax = xa$ it follows that $x^{-1}a = x^{-1}(ax)x^{-1} = x^{-1}(xa)x^{-1} = ax^{-1}$, so that x^{-1} is also in $N(a)$. But then $N(a)$ has been demonstrated to be a subgroup of G.

We are now in a position to enunciate our counting principle.

THEOREM 2.11.1 *If G is a finite group, then $c_a = o(G)/o(N(a))$; in other words, the number of elements conjugate to a in G is the index of the normalizer of a in G.*

Proof. To begin with, the conjugate class of a in G, $C(a)$, consists exactly of all the elements $x^{-1}ax$ as x ranges over G. c_a measures the number of distinct $x^{-1}ax$'s. Our method of proof will be to show that two elements in the same right coset of $N(a)$ in G yield the same conjugate of a whereas two elements in different right cosets of $N(a)$ in G give rise to different conjugates of a. In this way we shall have a one-to-one correspondence between conjugates of a and right cosets of $N(a)$.

Suppose that $x, y \in G$ are in the same right coset of $N(a)$ in G. Thus $y = nx$, where $n \in N(a)$, and so $na = an$. Therefore, since $y^{-1} = (nx)^{-1} = x^{-1}n^{-1}$, $y^{-1}ay = x^{-1}n^{-1}anx = x^{-1}n^{-1}nax = x^{-1}ax$, whence x and y result in the same conjugate of a.

If, on the other hand, x and y are in different right cosets of $N(a)$ in G we claim that $x^{-1}ax \neq y^{-1}ay$. Were this not the case, from $x^{-1}ax = y^{-1}ay$ we would deduce that $yx^{-1}a = ayx^{-1}$; this in turn would imply that $yx^{-1} \in N(a)$. However, this declares x and y to be in the same right coset of $N(a)$ in G, contradicting the fact that they are in different cosets. The proof is now complete.

COROLLARY

$$o(G) = \sum \frac{o(G)}{o(N(a))}$$

where this sum runs over one element a in each conjugate class.

Proof. Since $o(G) = \sum c_a$, using the theorem the corollary becomes immediate.

The equation in this corollary is usually referred to as the *class equation* of G.

Before going on to the applications of these results let us examine these concepts in some specific group. There is no point in looking at abelian groups because there two elements are conjugate if and only if they are equal (that is, $c_a = 1$ for every a). So we turn to our familiar friend, the group S_3. Its elements are e, $(1, 2)$, $(1, 3)$, $(2, 3)$, $(1, 2, 3)$, $(1, 3, 2)$. We enumerate the conjugate classes:

$$C(e) = \{e\}$$

$$\begin{aligned} C(1,2) &= \{(1,2), (1,3)^{-1}(1,2)(1,3), (2,3)^{-1}(1,2)(2,3), \\ &\quad (1,2,3)^{-1}(1,2)(1,2,3), (1,3,2)^{-1}(1,2)(1,3,2)\} \\ &= \{(1,2), (1,3), (2,3)\} \quad \text{(Verify!)} \end{aligned}$$

$$C(1,2,3) = \{(1,2,3), (1,3,2)\} \quad \text{(after another verification).}$$

The student should verify that $N((1, 2)) = \{e, (1, 2)\}$ and $N((1, 2, 3)) = \{e, (1, 2, 3), (1, 3, 2)\}$, so that $c_{(1,2)} = \frac{6}{2} = 3$, $c_{(1,2,3)} = \frac{6}{3} = 2$.

Applications of Theorem 2.11.1

Theorem 2.11.1 lends itself to immediate and powerful application. We need no artificial constructs to illustrate its use, for the results below which reveal the strength of the theorem are themselves theorems of stature and importance.

Let us recall that the center $Z(G)$ of a group G is the set of all $a \in G$ such that $ax = xa$ for all $x \in G$. Note the

SUBLEMMA *$a \in Z$ if and only if $N(a) = G$. If G is finite, $a \in Z$ if and only if $o(N(a)) = o(G)$.*

Proof. If $a \in Z$, $xa = ax$ for all $x \in G$, whence $N(a) = G$. If, conversely, $N(a) = G$, $xa = ax$ for all $x \in G$, so that $a \in Z$. If G is finite, $o(N(a)) = o(G)$ is equivalent to $N(a) = G$.

APPLICATION 1

THEOREM 2.11.2 *If $o(G) = p^n$ where p is a prime number, then $Z(G) \neq (e)$.*

Proof. If $a \in G$, since $N(a)$ is a subgroup of G, $o(N(a))$, being a divisor of $o(G) = p^n$, must be of the form $o(N(a)) = p^{n_a}$; $a \in Z(G)$ if and only if $n_a = n$. Write out the class equation for this G, letting $z = o(Z(G))$. We get $p^n = o(G) = \sum(p^n/p^{n_a})$; however, since there are exactly z elements such that $n_a = n$, we find that

$$p^n = z + \sum_{n_a < n} \frac{p^n}{p^{n_a}}.$$

Now look at this! p is a divisor of the left-hand side; since $n_a < n$ for each term in the $\sum$ of the right side,

$$p \left| \frac{p^n}{p^{n_a}} = p^{n-n_a} \right.$$

so that p is a divisor of each term of this sum, hence a divisor of this sum. Therefore,

$$p \left| \left(p^n - \sum_{n_a < n} \frac{p^n}{p^{n_a}} \right) = z. \right.$$

Since $e \in Z(G)$, $z \neq 0$; thus z is a positive integer divisible by the prime p. Therefore, $z > 1$! But then there must be an element, besides e, in $Z(G)$! This is the contention of the theorem.

Rephrasing, the theorem states that a group of prime-power order must always have a nontrivial center.

We can now simply prove, as a corollary for this, a result given in an earlier problem.

COROLLARY *If $o(G) = p^2$ where p is a prime number, then G is abelian.*

Proof. Our aim is to show that $Z(G) = G$. At any rate, we already know that $Z(G) \neq (e)$ is a subgroup of G so that $o(Z(G)) = p$ or p^2. If $o(Z(G)) = p^2$, then $Z(G) = G$ and we are done. Suppose that $o(Z(G)) = p$; let $a \in G$, $a \notin Z(G)$. Thus $N(a)$ is a subgroup of G, $Z(G) \subset N(a)$, $a \in N(a)$, so that $o(N(a)) > p$, yet by Lagrange's theorem $o(N(a)) \mid o(G) = p^2$. The only way out is for $o(N(a)) = p^2$, implying that $a \in Z(G)$, a contradiction. Thus $o(Z(G)) = p$ is not an actual possibility.

APPLICATION 2 We now use Theorem 2.11.1 to prove an important theorem due to Cauchy. The reader may remember that this theorem was already proved for abelian groups as an application of the results developed in the section on homomorphisms. In fact, we shall make use of this special

case in the proof below. But, to be frank, we shall prove, in the very next section, a much stronger result, due to Sylow, which has Cauchy's theorem as an immediate corollary, in a manner which completely avoids Theorem 2.11.1. To continue our candor, were Cauchy's theorem itself our ultimate and only goal, we could prove it, using the barest essentials of group theory, in a few lines. [The reader should look up the charming, one-paragraph proof of Cauchy's theorem found by McKay and published in the *American Mathematical Monthly*, Vol. 66 (1959), page 119.] Yet, despite all these counter-arguments we present Cauchy's theorem here as a striking illustration of Theorem 2.11.1.

THEOREM 2.11.3 (CAUCHY) *If p is a prime number and $p \mid o(G)$, then G has an element of order p.*

Proof. We seek an element $a \neq e \in G$ satisfying $a^p = e$. To prove its existence we proceed by induction on $o(G)$; that is, we assume the theorem to be true for all groups T such that $o(T) < o(G)$. We need not worry about starting the induction for the result is vacuously true for groups of order 1.

If for any subgroup W of G, $W \neq G$, were it to happen that $p \mid o(W)$, then by our induction hypothesis there would exist an element of order p in W, and thus there would be such an element in G. Thus we may assume that p is not a divisor of the order of any proper subgroup of G. In particular, if $a \notin Z(G)$, since $N(a) \neq G$, $p \nmid o(N(a))$. Let us write down the class equation:

$$o(G) = o(Z(G)) + \sum_{N(a) \neq G} \frac{o(G)}{o(N(a))}.$$

Since $p \mid o(G)$, $p \nmid o(N(a))$ we have that

$$p \left| \frac{o(G)}{o(N(a))}, \right.$$

and so

$$p \left| \sum_{N(a) \neq G} \frac{o(G)}{o(N(a))}; \right.$$

since we also have that $p \mid o(G)$, we conclude that

$$p \left| \left(o(G) - \sum_{N(a) \neq G} \frac{o(G)}{o(N(a))} \right) = o(Z(G)). \right.$$

$Z(G)$ is thus a subgroup of G whose order is divisible by p. But, after all, we have assumed that p is not a divisor of the order of any proper subgroup of G, so that $Z(G)$ cannot be a proper subgroup of G. We are forced to

accept the only possibility left us, namely, that $Z(G) = G$. But then G is abelian; now we invoke the result already established for abelian groups to complete the induction. This proves the theorem.

We conclude this section with a consideration of the conjugacy relation in a specific class of groups, namely, the symmetric groups S_n.

Given the integer n we say the sequence of positive integers $n_1, n_2, \ldots, n_r$, $n_1 \leq n_2 \leq \cdots \leq n_r$ constitute a *partition* of n if $n = n_1 + n_2 + \cdots + n_r$. Let $p(n)$ denote the number of partitions of n. Let us determine $p(n)$ for small values of n:

$$p(1) = 1 \text{ since } 1 = 1 \text{ is the only partition of 1,}$$

$$p(2) = 2 \text{ since } 2 = 2 \text{ and } 2 = 1 + 1,$$

$$p(3) = 3 \text{ since } 3 = 3,\ 3 = 1 + 2,\ 3 = 1 + 1 + 1,$$

$$p(4) = 5 \text{ since } 4 = 4,\ 4 = 1 + 3,\ 4 = 1 + 1 + 2,$$
$$4 = 1 + 1 + 1 + 1,\ 4 = 2 + 2.$$

Some others are $p(5) = 7$, $p(6) = 11$, $p(61) = 1{,}121{,}505$. There is a large mathematical literature on $p(n)$.

Every time we break a given permutation in S_n into a product of disjoint cycles we obtain a partition of n; for if the cycles appearing have lengths $n_1, n_2, \ldots, n_r$, respectively, $n_1 \leq n_2 \leq \cdots \leq n_r$, then $n = n_1 + n_2 + \cdots + n_r$. We shall say a permutation $\sigma \in S_n$ has the cycle decomposition $\{n_1, n_2, \ldots, n_r\}$ if it can be written as the product of disjoint cycles of lengths $n_1, n_2, \ldots, n_r$, $n_1 \leq n_2 \leq \cdots \leq n_r$. Thus in S_9

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 1 & 3 & 2 & 5 & 6 & 4 & 7 & 9 & 8 \end{pmatrix} = (1)(2.\ 3)(4,\ 5,\ 6)(7)(8,\ 9)$$

has cycle decomposition $\{1, 1, 2, 2, 3\}$; note that $1 + 1 + 2 + 2 + 3 = 9$. We now aim to prove that two permutations in S_n are conjugate if and only if they have the same cycle decomposition. Once this is proved, then S_n will have exactly $p(n)$ conjugate classes.

To reach our goal we exhibit a very simple rule for computing the conjugate of a given permutation. Suppose that $\sigma \in S_n$ and that σ sends $i \to j$. How do we find $\theta^{-1}\sigma\theta$ where $\theta \in S_n$? Suppose that θ sends $i \to s$ and $j \to t$; then $\theta^{-1}\sigma\theta$ sends $s \to t$. *In other words, to compute* $\theta^{-1}\sigma\theta$ *replace every symbol in* σ *by its image under* θ. For example, to determine $\theta^{-1}\sigma\theta$ where $\theta = (1, 2, 3)(4, 7)$ and $\sigma = (5, 6, 7)(3, 4, 2)$, then, since θ:$5 \to 5$, $6 \to 6$, $7 \to 4$, $3 \to 1$, $4 \to 7$, $2 \to 3$, $\theta^{-1}\sigma\theta$ is obtained from σ by replacing in σ, 5 by 5, 6 by 6, 7 by 4, 3 by 1, 4 by 7, and 2 by 3, so that $\theta^{-1}\sigma\theta = (5, 6, 4)(1, 7, 3)$.

With this algorithm for computing conjugates it becomes clear that two permutations having the same cycle decomposition are conjugate. For if

$\sigma = (a_1, a_2, \ldots, a_{n_1})(b_1, b_2, \ldots, b_{n_2}) \cdots (x_1, x_2, \ldots, x_{n_r})$ and $\tau = (\alpha_1, \alpha_2, \ldots, \alpha_{n_1})(\beta_1, \beta_2, \ldots, \beta_{n_2}) \cdots (\chi_1, \chi_2, \ldots, \chi_{n_r})$, then $\tau = \theta^{-1}\sigma\theta$, where one could use as θ the permutation

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_{n_1} & b_1 & \cdots & b_{n_2} & \cdots & x_1 & \cdots & x_{n_r} \\ \alpha_1 & \alpha_2 & \cdots & \alpha_{n_1} & \beta_1 & \cdots & \beta_{n_2} & \cdots & \chi_1 & \cdots & \chi_{n_r} \end{pmatrix}.$$

Thus, for instance, $(1, 2)(3, 4, 5)(6, 7, 8)$ and $(7, 5)(1, 3, 6)(2, 4, 8)$ can be exhibited as conjugates by using the conjugating permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 7 & 5 & 1 & 3 & 6 & 2 & 4 & 8 \end{pmatrix}.$$

That two conjugates have the same cycle decomposition is now trivial for, by our rule, to compute a conjugate, replace every element in a given cycle by its image under the conjugating permutation.

We restate the result proved in the previous discussion as

LEMMA 2.11.3 *The number of conjugate classes in S_n is $p(n)$, the number of partitions of n.*

Since we have such an explicit description of the conjugate classes in S_n we can find all the elements commuting with a given permutation. We illustrate this with a very special and simple case.

Given the permutation $(1, 2)$ in S_n, what elements commute with it? Certainly any permutation leaving both 1 and 2 fixed does. There are $(n - 2)!$ such. Also $(1, 2)$ commutes with itself. This way we get $2(n - 2)!$ elements in the group generated by $(1, 2)$ and the $(n - 2)!$ permutations leaving 1 and 2 fixed. Are there others? There are $n(n - 1)/2$ transpositions and these are precisely all the conjugates of $(1, 2)$. Thus the conjugate class of $(1, 2)$ has in it $n(n - 1)/2$ elements. If the order of the normalizer of $(1, 2)$ is r, then, by our counting principle,

$$\frac{n(n-1)}{2} = \frac{o(S_n)}{r} = \frac{n!}{r}.$$

Thus $r = 2(n - 2)!$. That is, the order of the normalizer of $(1, 2)$ is $2(n - 2)!$. But we exhibited $2(n - 2)!$ elements which commute with $(1, 2)$; thus the general element σ commuting with $(1, 2)$ is $\sigma = (1, 2)^i\tau$, where $i = 0$ or 1, τ is a permutation leaving both 1 and 2 fixed.

As another application consider the permutation $(1, 2, 3, \ldots, n) \in S_n$. We claim this element commutes only with its powers. Certainly it does commute with all its powers, and this gives rise to n elements. Now, any n-cycle is conjugate to $(1, 2, \ldots, n)$ and there are $(n - 1)!$ distinct n-cycles in S_n. Thus if u denotes the order of the normalizer of $(1, 2, \ldots, n)$

in S_n, since $o(S_n)/u$ = number of conjugates of $(1, 2, \ldots, n)$ in S_n = $(n-1)!$,

$$u = \frac{n!}{(n-1)!} = n.$$

So the order of the normalizer of $(1, 2, \ldots, n)$ in S_n is n. The powers of $(1, 2, \ldots, n)$ having given us n such elements, there is no room left for others and we have proved our contention.

Problems

1. List all the conjugate classes in S_3, find the c_a's, and verify the class equation.
2. List all the conjugate classes in S_4, find the c_a's and verify the class equation.
3. List all the conjugate classes in the group of quaternion units (see Problem 21, Section 2.10), find the c_a's and verify the class equation.
4. List all the conjugate classes in the dihedral group of order $2n$, find the c_a's and verify the class equation. Notice how the answer depends on the parity of n.
5. (a) In S_n prove that there are $\frac{1}{r}\frac{n!}{(n-r)!}$ distinct r cycles.
   (b) Using this, find the number of conjugates that the r-cycle $(1, 2, \ldots, r)$ has in S_n.
   (c) Prove that any element σ in S_n which commutes with $(1, 2, \ldots, r)$ is of the form $\sigma = (1, 2, \ldots, r)^i\tau$, where $i = 0, 1, 2, \ldots, r$, τ is a permutation leaving all of $1, 2, \ldots, r$ fixed.
6. (a) Find the number of conjugates of $(1, 2)(3, 4)$ in S_n, $n \geq 4$.
   (b) Find the form of all elements commuting with $(1, 2)(3, 4)$ in S_n.
7. If p is a prime number, show that in S_p there are $(p-1)! + 1$ elements x satisfying $x^p = e$.
8. If in a finite group G an element a has exactly two conjugates, prove that G has a normal subgroup $N \neq (e), G$.
9. (a) Find two elements in A_5, the alternating group of degree 5, which are conjugate in S_5 but not in A_5.
   (b) Find all the conjugate classes in A_5 and the number of elements in each conjugate class.
10. (a) If N is a normal subgroup of G and $a \in N$, show that every conjugate of a in G is also in N.
    (b) Prove that $o(N) = \sum c_a$ for some choices of a in N.

(c) Using this and the result for Problem 9(b), prove that in A_5 there is no normal subgroup N other than (e) and A_5.

11. Using Theorem 2.11.2 as a tool, prove that if $o(G) = p^n$, p a prime number, then G has a subgroup of order p^α for all $0 \leq \alpha \leq n$.

12. If $o(G) = p^n$, p a prime number, prove that there exist subgroups N_i, $i = 0, 1, \ldots, r$ (for some r) such that $G = N_0 \supset N_1 \supset N_2 \supset \cdots \supset N_r = (e)$ where N_i is a normal subgroup of N_{i-1} and where N_{i-1}/N_i is abelian.

13. If $o(G) = p^n$, p a prime number, and $H \neq G$ is a subgroup of G, show that there exists an $x \in G$, $x \notin H$ such that $x^{-1}Hx = H$.

14. Prove that any subgroup of order p^{n-1} in a group G of order p^n, p a prime number, is normal in G.

*15. If $o(G) = p^n$, p a prime number, and if $N \neq (e)$ is a normal subgroup of G, prove that $N \cap Z \neq (e)$, where Z is the center of G.

16. If G is a group, Z its center, and if G/Z is cyclic, prove that G must be abelian.

17. Prove that any group of order 15 is cyclic.

18. Prove that a group of order 28 has a normal subgroup of order 7.

19. Prove that if a group G of order 28 has a normal subgroup of order 4, then G is abelian.

2.12 Sylow's Theorem

Lagrange's theorem tells us that the order of a subgroup of a finite group is a divisor of the order of that group. The converse, however, is false. There are very few theorems which assert the existence of subgroups of prescribed order in arbitrary finite groups. The most basic, and widely used, is a classic theorem due to the Norwegian mathematician Sylow.

We present here three proofs of this result of Sylow. The first is a very elegant and elementary argument due to Wielandt. It appeared in the journal *Archiv der Matematik*, Vol. 10 (1959), pages 401–402. The basic elements in Wielandt's proof are number-theoretic and combinatorial. It has the advantage, aside from its elegance and simplicity, of producing the subgroup we are seeking. The second proof is based on an exploitation of induction in an interplay with the class equation. It is one of the standard classical proofs, and is a nice illustration of combining many of the ideals developed so far in the text to derive this very important cornerstone due to Sylow. The third proof is of a completely different philosophy. The basic idea there is to show that if a larger group than the one we are considering satisfies the conclusion of Sylow's theorem, then our group also must.

This forces us to prove Sylow's theorem for a special family of groups—the symmetric groups. By invoking Cayley's theorem (Theorem 2.9.1) we are then able to deduce Sylow's theorem for all finite groups. Apart from this strange approach—to prove something for a given group, first prove it for a much larger one—this third proof has its own advantages. Exploiting the ideas used, we easily derive the so-called second and third parts of Sylow's theorem.

One might wonder: why give three proofs of the same result when, clearly, one suffices? The answer is simple. Sylow's theorem is *that* important that it merits this multifront approach. Add to this the completely diverse nature of the three proofs and the nice application each gives of different things that we have learned, the justification for the whole affair becomes persuasive (at least to the author). Be that as it may, we state Sylow's theorem and get on with Wielandt's proof.

THEOREM 2.12.1 (SYLOW) *If p is a prime number and $p^\alpha \mid o(G)$, then G has a subgroup of order p^α.*

Before entering the first proof of the theorem we digress slightly to a brief number-theoretic and combinatorial discussion.

The number of ways of picking a subset of k elements from a set of n elements can easily be shown to be

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

If $n = p^\alpha m$ where p is a prime number, and if $p^r \mid m$ but $p^{r+1} \nmid m$, consider

$$\begin{aligned}\binom{p^\alpha m}{p^\alpha} &= \frac{(p^\alpha m)!}{(p^\alpha)!(p^\alpha m - p^\alpha)!} \\ &= \frac{p^\alpha m(p^\alpha m - 1)\cdots(p^\alpha m - i)\cdots(p^\alpha m - p^\alpha + 1)}{p^\alpha(p^\alpha - 1)\cdots(p^\alpha - i)\cdots(p^\alpha - p^\alpha + 1)}.\end{aligned}$$

The question is, What power of p divides $\binom{p^\alpha m}{p^\alpha}$? Looking at this number, written out as we have written it out, one can see that except for the term m in the numerator, the power of p dividing $(p^\alpha m - i)$ is the same as that dividing $p^\alpha - i$, so all powers of p cancel out except the power which divides m. Thus

$$p^r \mid \binom{p^\alpha m}{p^\alpha} \quad \text{but} \quad p^{r+1} \nmid \binom{p^\alpha m}{p^\alpha}.$$

First Proof of the Theorem. Let $\mathcal{M}$ be the set of all subsets of G which have p^α elements. Thus $\mathcal{M}$ has $\binom{p^\alpha m}{p^\alpha}$ elements. Given $M_1, M_2 \in \mathcal{M}$ (M is a subset of G having p^α elements, and likewise so is M_2) define $M_1 \sim M_2$ if there exists an element $g \in G$ such that $M_1 = M_2 g$. It is immediate to verify that this defines an equivalence relation on $\mathcal{M}$. We claim that there is at least one equivalence class of elements in $\mathcal{M}$ such that the number of elements in this class is not a multiple of p^{r+1}, for if p^{r+1} is a divisor of the size of each equivalence class, then p^{r+1} would be a divisor of the number of elements in $\mathcal{M}$. Since $\mathcal{M}$ has $\binom{p^\alpha m}{p^\alpha}$ elements and $p^{r+1} \nmid \binom{p^\alpha m}{p^\alpha}$, this cannot be the case. Let $\{M_1, \ldots, M_n\}$ be such an equivalence class in $\mathcal{M}$ where $p^{r+1} \nmid n$. By our very definition of equivalence in $\mathcal{M}$, if $g \in G$, for each $i = 1, \ldots, n$, $M_i g = M_j$ for some j, $1 \leq j \leq n$. We let $H = \{g \in G \mid M_1 g = M_1\}$. Clearly H is a subgroup of G, for if $a, b \in H$, then $M_1 a = M_1$, $M_1 b = M_1$ whence $M_1 ab = (M_1 a)b = M_1 b = M_1$. We shall be vitally concerned with $o(H)$. We claim that $no(H) = o(G)$; we leave the proof to the reader, but suggest the argument used in the counting principle in Section 2.11. Now $no(H) = o(G) = p^\alpha m$; since $p^{r+1} \nmid n$ and $p^{\alpha+r} \mid p^\alpha m = no(H)$, it must follow that $p^\alpha \mid o(H)$, and so $o(H) \geq p^\alpha$. However, if $m_1 \in M_1$, then for all $h \in H$, $m_1 h \in M_1$. Thus M_1 has at least $o(H)$ distinct elements. However, M_1 was a subset of G containing p^α elements. Thus $p^\alpha \geq o(H)$. Combined with $o(H) \geq p^\alpha$ we have that $o(H) = p^\alpha$. *But then we have exhibited a subgroup of G having exactly p^α elements, namely H.* This proves the theorem; it actually has done more—it has constructed the required subgroup before our very eyes!

What is usually known as Sylow's theorem is a special case of Theorem 2.12.1, namely that

COROLLARY *If $p^m \mid o(G)$, $p^{m+1} \nmid o(G)$, then G has a subgroup of order p^m.*

A subgroup of G of order p^m, where $p^m \mid o(G)$ but $p^{m+1} \nmid o(G)$, is called a *p-Sylow subgroup of G*. The corollary above asserts that a finite group has p-Sylow subgroups for every prime p dividing its order. Of course the conjugate of a p-Sylow subgroup is a p-Sylow subgroup. In a short while we shall see how any two p-Sylow subgroups of G—for the same prime p—are related. We shall also get some information on how many p-Sylow subgroups there are in G for a given prime p. Before passing to this, we want to give two other proofs of Sylow's theorem.

We begin with a remark. As we observed just prior to the corollary, the corollary is a special case of the theorem. However, we claim that the

theorem is easily derivable from the corollary. That is, if we know that G possesses a subgroup of order p^m, where $p^m \mid o(G)$ but $p^{m+1} \nmid o(G)$, then we know that G has a subgroup of order p^α for any α such that $p^\alpha \mid o(G)$. This follows from the result of Problem 11, Section 2.11. This result states that any group of order p^m, p a prime, has subgroups of order p^α for any $0 \leq \alpha \leq m$. Thus to prove Theorem 2.12.1—as we shall proceed to do, again, in two more ways—it is enough for us to prove the existence of p-Sylow subgroups of G, for every prime p dividing the order of G.

Second Proof of Sylow's Theorem. We prove, by induction on the order of the group G, that for every prime p dividing the order of G, G has a p-Sylow subgroup.

If the order of the group is 2, the only relevant prime is 2 and the group certainly has a subgroup of order 2, namely itself.

So we suppose the result to be correct for all groups of order less than $o(G)$. From this we want to show that the result is valid for G. Suppose, then, that $p^m \mid o(G)$, $p^{m+1} \nmid o(G)$, where p is a prime, $m \geq 1$. If $p^m \mid o(H)$ for any subgroup H of G, where $H \neq G$, then by the induction hypothesis, H would have a subgroup T of order p^m. However, since T is a subgroup of H, and H is a subgroup of G, T too is a subgroup of G. But then T would be the sought-after subgroup of order p^m.

We therefore may assume that $p^m \nmid o(H)$ for *any* subgroup H of G, where $H \neq G$. We restrict our attention to a limited set of such subgroups. Recall that if $a \in G$ then $N(a) = \{x \in G \mid xa = ax\}$ is a subgroup of G; moreover, if $a \notin Z$, the center of G, then $N(a) \neq G$. Recall, too, that the class equation of G states that

$$o(G) = \sum \frac{o(G)}{o(N(a))},$$

where this sum runs over one element a from each conjugate class. We separate this sum into two pieces: those a which lie in Z, and those which don't. This gives

$$o(G) = z + \sum_{a \notin Z} \frac{o(G)}{o(N(a))},$$

where $z = o(Z)$. Now invoke the reduction we have made, namely, that $p^m \nmid o(H)$ for any subgroup $H \neq G$ of G, to those subgroups $N(a)$ for $a \notin Z$. Since in this case, $p^m \mid o(G)$ and $p^m \nmid o(N(a))$, we must have that

$$p \left| \frac{o(G)}{o(N(a))} \right.$$

Restating this result,

$$p \left| \frac{o(G)}{o(N(a))} \right.$$

for every $a \in G$ where $a \notin Z$. Look at the class equation with this information in hand. Since $p^m \mid o(G)$, we have that $p \mid o(G)$; also

$$p \left| \sum_{a \notin Z} \frac{o(G)}{o(N(a))} \right..$$

Thus the class equation gives us that $p \mid z$. Since $p \mid z = o(Z)$, by Cauchy's theorem (Theorem 2.11.3), Z has an element $b \neq e$ of order p. Let $B = (b)$, the subgroup of G generated by b. B is of order p; moreover, *since* $b \in Z$, B *must be normal* in G. Hence we can form the quotient group $\bar{G} = G/B$. We look at $\bar{G}$. First of all, its order is $o(G)/o(B) = o(G)/p$, hence is certainly less than $o(G)$. Secondly, we have $p^{m-1} \mid o(\bar{G})$, but $p^m \nmid o(\bar{G})$. Thus, by the induction hypothesis, $\bar{G}$ has a subgroup $\bar{P}$ of order p^{m-1}. Let $P = \{x \in G \mid xB \in \bar{P}\}$; by Lemma 2.7.5, P is a subgroup of G. Moreover, $\bar{P} \approx P/B$ (Prove!); thus

$$p^{m-1} = o(\bar{P}) = \frac{o(P)}{o(B)} = \frac{o(P)}{p}.$$

This results in $o(P) = p^m$. Therefore P is the required p-Sylow subgroup of G. This completes the induction and so proves the theorem.

With this we have finished the second proof of Sylow's theorem. Note that this second proof can easily be adapted to prove that if $p^\alpha \mid o(G)$, then G has a subgroup of order p^α directly, without first passing to the existence of a p-Sylow subgroup. (This is Problem 1 of the problems at the end of this section.)

We now proceed to the third proof of Sylow's theorem.

Third Proof of Sylow's Theorem. Before going into the details of the proof proper, we outline its basic strategy. We will first show that the symmetric groups S_{p^r}, p a prime, all have p-Sylow subgroups. The next step will be to show that if G is contained in M and M has a p-Sylow subgroup, then G has a p-Sylow subgroup. Finally we will show, via Cayley's theorem, that we can use S_{p^k}, for large enough k, as our M. With this we will have all the pieces, and the theorem will drop out.

In carrying out this program in detail, we will have to know how large a p-Sylow subgroup of S_{p^r} should be. This will necessitate knowing what power of p divides $(p^r)!$. This will be easy. To produce the p-Sylow subgroup of S_{p^r} will be harder. To carry out another vital step in this rough sketch, it will be necessary to introduce a new equivalence relation in groups, and the corresponding equivalence classes known as *double cosets*. This will have several payoffs, not only in pushing through the proof of Sylow's theorem, but also in getting us the second and third parts of the full Sylow theorem.

So we get down to our first task, that of finding what power of a prime p exactly divides $(p^k)!$. Actually, it is quite easy to do this for $n!$ for any integer n (see Problem 2). But, for our purposes, it will be clearer and will suffice to do it only for $(p^k)!$.

Let $n(k)$ be defined by $p^{n(k)} \mid (p^k)!$ but $p^{n(k)+1} \nmid (p^k)!$.

LEMMA 2.12.1 $n(k) = 1 + p + \cdots + p^{k-1}$.

Proof. If $k = 1$ then, since $p! = 1 \cdot 2 \cdots (p-1) \cdot p$, it is clear that $p \mid p!$ but $p^2 \nmid p!$. Hence $n(1) = 1$, as it should be.

What terms in the expansion of $(p^k)!$ can contribute to powers of p dividing $(p^k)!$? Clearly, only the multiples of p; that is, $p, 2p, \ldots, p^{k-1}p$. In other words $n(k)$ must be the power of p which divides $p(2p)(3p)\cdots(p^{k-1}p) = p^{p^{k-1}}(p^{k-1})!$. But then $n(k) = p^{k-1} + n(k-1)$. Similarly, $n(k-1) = n(k-2) + p^{k-2}$, and so on. Write these out as

$$\begin{aligned} n(k) - n(k-1) &= p^{k-1}, \\ n(k-1) - n(k-2) &= p^{k-2}, \\ &\vdots \\ n(2) - n(1) &= p, \\ n(1) &= 1. \end{aligned}$$

Adding these up, with the cross-cancellation that we get, we obtain $n(k) = 1 + p + p^2 + \cdots + p^{k-1}$. This is what was claimed in the lemma, so we are done.

We are now ready to show that S_{p^k} has a p-Sylow subgroup; that is, we shall show (in fact, produce) a subgroup of order $p^{n(k)}$ in S_{p^k}.

LEMMA 2.12.2 *S_{p^k} has a p-Sylow subgroup.*

Proof. We go by induction on k. If $k = 1$, then the element $(1\ 2\ \ldots\ p)$, in S_p is of order p, so generated a subgroup of order p. Since $n(1) = 1$, the result certainly checks out for $k = 1$.

Suppose that the result is correct for $k-1$; we want to show that it then must follow for k. Divide the integers $1, 2, \ldots, p^k$ into p clumps, each with p^{k-1} elements as follows:

$$\{1, 2, \ldots, p^{k-1}\}, \{p^{k-1}+1, p^{k-1}+2, \ldots, 2p^{k-1}\}, \ldots, \{(p-1)p^{k-1}+1, \ldots, p^k\}.$$

The permutation σ defined by $\sigma = (1, p^{k-1}+1, 2p^{k-1}+1, \ldots, (p-1)p^{k-1}+1) \cdots (j, p^{k-1}+j, 2p^{k-1}+j, \ldots, (p-1)p^{k-1}+1+j) \cdots (p^{k-1}, 2p^{k-1}, \ldots, (p-1)p^{k-1}, p^k)$ has the following properties:

1. $\sigma^p = e$.

2. If τ is a permutation that leaves all i fixed for $i > p^{k-1}$ (hence, affects only $1, 2, \ldots, p^{k-1}$), then $\sigma^{-1}\tau\sigma$ moves only elements in $\{p^{k-1}+1, p^{k-1}+2, \ldots, 2p^{k-1}\}$, and more generally, $\sigma^{-j}\tau\sigma^j$ moves only elements in $\{jp^{k-1}+1, jp^{k-1}+2, \ldots, (j+1)p^{k-1}\}$.

Consider $A = \{\tau \in S_{p^k} \mid \tau(i) = i \text{ if } i > p^{k-1}\}$. A is a subgroup of S_{p^k} and elements in A can carry out any permutation on $1, 2, \ldots, p^{k-1}$. From this it follows easily that $A \approx S_{p^{k-1}}$. By induction, A has a subgroup P_1 of order $p^{n(k-1)}$.

Let $T = P_1(\sigma^{-1}P_1\sigma)(\sigma^{-2}P_1\sigma^2)\cdots(\sigma^{-(p-1)}P_1\sigma^{p-1}) = P_1P_2\cdots P_{n-1}$, where $P_i = \sigma^{-i}P_1\sigma^i$. Each P_i is isomorphic to P_1 so has order $p^{n(k-1)}$. Also elements in distinct P_i's influence nonoverlapping sets of integers, *hence commute.* Thus T is a subgroup of S_{p^k}. What is its order? Since $P_i \cap P_j = (e)$ if $0 \leq i \neq j \leq p-1$, we see that $o(T) = o(P_1)^p = p^{pn(k-1)}$. We are not quite there yet. T is not the p-Sylow subgroup we seek!

Since $\sigma^p = e$ and $\sigma^{-i}P_1\sigma^i = P_i$ we have $\sigma^{-1}T\sigma = T$. Let $P = \{\sigma^j t \mid t \in T, 0 \leq j \leq p-1\}$. Since $\sigma \notin T$ and $\sigma^{-1}T\sigma = T$ we have two things: firstly, T is a subgroup of S_{p^k} and, furthermore, $o(P) = p \cdot o(T) = p \cdot p^{n(k-1)p} = p^{n(k-1)p+1}$. Now we are finally there! P is the sought-after p-Sylow subgroup of S_{p^k}.

Why? Well, what is its order? It is $p^{n(k-1)p+1}$. But $n(k-1) = 1 + p + \cdots + p^{k-2}$, hence $pn(k-1) + 1 = 1 + p + \cdots + p^{k-1} = n(k)$. Since now $o(P) = p^{n(k)}$, P is indeed a p-Sylow subgroup of S_{p^k}.

Note something about the proof. Not only does it prove the lemma, it actually allows us to construct the p-Sylow subgroup inductively. We follow the procedure of the proof to construct a 2-Sylow subgroup in S_4.

Divide 1, 2, 3, 4 into $\{1, 2\}$ and $\{3, 4\}$. Let $P_1 = ((1\,2))$ and $\sigma = (1\,3)(2\,4)$. Then $P_2 = \sigma^{-1}P_1\sigma = (3\,4)$. Our 2-Sylow subgroup is then the group generated by $(1\,3)(2\,4)$ and

$$T = P_1P_2 = \{(1\,2), (3\,4), (1\,2)(3\,4), e\}.$$

In order to carry out the program of the third proof that we outlined, we now introduce a new equivalence relation in groups (see Problem 39, Section 2.5).

DEFINITION Let G be a group, A, B subgroups of G. If $x, y \in G$ define $x \sim y$ if $y = axb$ for some $a \in A$, $b \in B$.

We leave to the reader the verification—it is easy—of

LEMMA 2.12.3 *The relation defined above is an equivalence relation on G. The equivalence class of $x \in G$ is the set $AxB = \{axb \mid a \in A, b \in B\}$.*

We call the set AxB a *double coset* of A, B in G.

If A, B are finite subgroups of G, how many elements are there in the double coset AxB? To begin with, the mapping $T:AxB \to AxBx^{-1}$ given by $(axb)T = axbx^{-1}$ is one-to-one and onto (verify). Thus $o(AxB) = o(AxBx^{-1})$. Since xBx^{-1} is a subgroup of G, of order $o(B)$, by Theorem 2.5.1,

$$o(AxB) = o(AxBx^{-1}) = \frac{o(A)o(xBx^{-1})}{o(A \cap xBx^{-1})} = \frac{o(A)o(B)}{o(A \cap xBx^{-1})}.$$

We summarize this in

LEMMA 2.12.4 *If A, B are finite subgroups of G then*

$$o(AxB) = \frac{o(A)o(B)}{o(A \cap xBx^{-1})}.$$

We now come to the gut step in this third proof of Sylow's theorem.

LEMMA 2.12.5 *Let G be a finite group and suppose that G is a subgroup of the finite group M. Suppose further that M has a p-Sylow subgroup Q. Then G has a p-Sylow subgroup P. In fact, $P = G \cap xQx^{-1}$ for some $x \in M$.*

Proof. Before starting the details of the proof, we translate the hypotheses somewhat. Suppose that $p^m \mid o(M)$, $p^{m+1} \nmid o(M)$, Q is a subgroup of M of order p^m. Let $o(G) = p^n t$ where $p \nmid t$. We want to produce a subgroup P in G of order p^n.

Consider the double coset decomposition of M given by G and Q; $M = \bigcup GxQ$. By Lemma 2.12.4,

$$o(GxQ) = \frac{o(G)o(Q)}{o(G \cap xQx^{-1})} = \frac{p^n t p^m}{o(G \cap xQx^{-1})}.$$

Since $G \cap xQx^{-1}$ is a subgroup of xQx^{-1}, its order is p^{m_x}. We claim that $m_x = n$ for some $x \in M$. If not, then

$$o(GxQ) = \frac{p^n t p^m}{p^{m_x}} = tp^{m+n-m_x},$$

so is divisible by p^{m+1}. Now, since $M = \bigcup GxQ$, and this is disjoint union, $o(M) = \sum o(GxQ)$, the sum running over one element from each double coset. But $p^{m+1} \mid o(GxQ)$; hence $p^{m+1} \mid o(M)$. This contradicts $p^{m+1} \nmid o(M)$. Thus $m_x = n$ for some $x \in M$. But then $o(G \cap xQx^{-1}) = p^n$. Since $G \cap xQx^{-1} = P$ is a subgroup of G and has order p^n, the lemma is proved.

We now can easily prove Sylow's theorem. By Cayley's theorem (Theorem 2.9.1) we can isomorphically embed our finite group G in S_n, the symmetric group of degree n. Pick k so that $n < p^k$; then we can isomorphically embed S_n in S_{p^k} (by acting on $1, 2, \ldots, n$ only in the set

$1, 2, \ldots, n, \ldots, p^k$), hence G is isomorphically embedded in S_{p^k}. By Lemma 2.12.2, S_{p^k} has a p-Sylow subgroup. Hence, by Lemma 2.12.5, G must have a p-Sylow subgroup. This finishes the third proof of Sylow's theorem.

This third proof has given us quite a bit more. From it we have the machinery to get the other parts of Sylow's theorem.

THEOREM 2.12.2 (SECOND PART OF SYLOW'S THEOREM) *If G is a finite group, p a prime and $p^n \mid o(G)$ but $p^{n+1} \nmid o(G)$, then any two subgroups of G of order p^n are conjugate.*

Proof. Let A, B be subgroups of G, each of order p^n. We want to show that $A = gBg^{-1}$ for some $g \in G$.

Decompose G into double cosets of A and B; $G = \bigcup AxB$. Now, by Lemma 2.12.4,

$$o(AxB) = \frac{o(A)o(B)}{o(A \cap xBx^{-1})}.$$

If $A \neq xBx^{-1}$ for every $x \in G$ then $o(A \cap xBx^{-1}) = p^m$ where $m < n$. Thus

$$o(AxB) = \frac{o(A)o(B)}{p^m} = \frac{p^{2n}}{p^m} = p^{2n-m}$$

and $2n - m \geq n + 1$. Since $p^{n+1} \mid o(AxB)$ for every x and since $o(G) = \sum o(AxB)$, we would get the contradiction $p^{n+1} \mid o(G)$. Thus $A = gBg^{-1}$ for some $g \in G$. This is the assertion of the theorem.

Knowing that for a given prime p all p-Sylow subgroups of G are conjugate allows us to count up precisely how many such p-Sylow subgroups there are in G. The argument is exactly as that given in proving Theorem 2.11.1. In some earlier problems (see, in particular, Problem 16, Section 2.5) we discussed the normalizer $N(H)$, of a subgroup, defined by $N(H) = \{x \in G \mid xHx^{-1} = H\}$. Then, as in the proof of Theorem 2.11.1, we have that *the number of distinct conjugates, xHx^{-1}, of H in G is the index of $N(H)$ in G.* Since all p-Sylow subgroups are conjugate we have

LEMMA 2.12.6 *The number of p-Sylow subgroups in G equals $o(G)/o(N(P))$, where P is any p-Sylow subgroup of G. In particular, this number is a divisor of $o(G)$.*

However, much more can be said about the number of p-Sylow subgroups there are, for a given prime p, in G. We go into this now. The technique will involve double cosets again.

THEOREM 2.12.3 (THIRD PART OF SYLOW'S THEOREM) *The number of p-Sylow subgroups in G, for a given prime, is of the form $1 + kp$.*

Proof. Let P be a p-Sylow subgroup of G. We decompose G into double cosets of P and P. Thus $G = \bigcup PxP$. We now ask: How many elements are there in PxP? By Lemma 2.12.4 we know the answer:

$$o(PxP) = \frac{o(P)^2}{o(P \cap xPx^{-1})}.$$

Thus, if $P \cap xPx^{-1} \neq P$ then $p^{n+1} \mid o(PxP)$, where $p^n = o(P)$. Paraphrasing this: if $x \notin N(P)$ then $p^{n+1} \mid o(PxP)$. Also, if $x \in N(P)$, then $PxP = P(Px) = P^2x = Px$, so $o(PxP) = p^n$ in this case.

Now

$$o(G) = \sum_{x \in N(P)} o(PxP) + \sum_{x \notin N(P)} o(PxP),$$

where each sum runs over one element from each double coset. However, if $x \in N(P)$, since $PxP = Px$, the first sum is merely $\sum_{x \in N(P)} o(Px)$ over the *distinct cosets* of P in $N(P)$. Thus this first sum is just $o(N(P))$. What about the second sum? We saw that each of its constituent terms is divisible by p^{n+1}, hence

$$p^{n+1} \Bigg| \sum_{x \notin N(P)} o(PxP).$$

We can thus write this second sum as

$$\sum_{x \notin N(P)} o(PxP) = p^{n+1}u.$$

Therefore $o(G) = o(N(P)) + p^{n+1}u$, so

$$\frac{o(G)}{o(N(P))} = 1 + \frac{p^{n+1}u}{o(N(P))}.$$

Now $o(N(P)) \mid o(G)$ since $N(P)$ is a subgroup of G, hence $p^{n+1}u/o(N(P))$ is an integer. Also, since $p^{n+1} \nmid o(G)$, p^{n+1} can't divide $o(N(P))$. But then $p^{n+1}u/o(N(P))$ *must be divisible by* p, so we can write $p^{n+1}u/o(N(P))$ as kp, where k is an integer. Feeding this information back into our equation above, we have

$$\frac{o(G)}{o(N(P))} = 1 + kp.$$

Recalling that $o(G)/o(N(P))$ is the number of p-Sylow subgroups in G, we have the theorem.

In Problems 20–24 in the Supplementary Problems at the end of this chapter, there is outlined another approach to proving the second and third parts of Sylow's theorem.

We close this section by demonstrating how the various parts of Sylow's theorem can be used to gain a great deal of information about finite groups.

Let G be a group of order $11^2 \cdot 13^2$. We want to determine how many 11-Sylow subgroups and how many 13-Sylow subgroups there are in G. The number of 11-Sylow subgroups, by Theorem 2.12.13, is of the form $1 + 11k$. By Lemma 2.12.5, this must divide $11^2 \cdot 13^2$; being prime to 11, it must divide 13^2. Can 13^2 have a factor of the form $1 + 11k$? Clearly no, other than 1 itself. Thus $1 + 11k = 1$, and so there must be only one 11-Sylow subgroup in G. Since all 11-Sylow subgroups are conjugate (Theorem 2.12.2) we conclude that the 11-Sylow subgroup is *normal* in G.

What about the 13-Sylow subgroups? Their number is of the form $1 + 13k$ and must divide $11^2 \cdot 13^2$, hence must divide 11^2. Here, too, we conclude that there can be only one 13-Sylow subgroup in G, and it must be normal.

We now know that G has a normal subgroup A of order 11^2 and a normal subgroup B of order 13^2. By the corollary to Theorem 2.11.2, any group of order p^2 is abelian; hence A and B are both abelian. Since $A \cap B = (e)$, we easily get $AB = G$. Finally, if $a \in A$, $b \in B$, then $aba^{-1}b^{-1} = a(ba^{-1}b^{-1}) \in A$ since A is normal, and $aba^{-1}b^{-1} = (aba^{-1})b^{-1} \in B$ since B is normal. Thus $aba^{-1}b^{-1} \in A \cap B = (e)$. This gives us $aba^{-1}b^{-1} = e$, and so $ab = ba$ for $a \in A$, $b \in B$. This, together with $AB = G$, A, B abelian, allows us to conclude that G *is abelian.* Hence any group of order $11^2 \cdot 13^2$ must be abelian.

We give one other illustration of the use of the various parts of Sylow's theorem. Let G be a group of order 72; $o(G) = 2^3 3^2$. How many 3-Sylow subgroups can there be in G? If this number is t, then, according to Theorem 2.12.3, $t = 1 + 3k$. According to Lemma 2.12.5, $t \mid 72$, and since t is prime to 3, we must have $t \mid 8$. The only factors of 8 of the form $1 + 3k$ are 1 and 4; hence $t = 1$ or $t = 4$ are the only possibilities. In other words G has either one 3-Sylow subgroup or 4 such.

If G has only one 3-Sylow subgroup, since all 3-Sylow subgroups are conjugate, this 3-Sylow subgroup must be normal in G. In this case G would certainly contain a nontrivial normal subgroup. On the other hand if the number of 3-Sylow subgroups of G is 4, by Lemma 2.12.5 the index of N in G is 4, where N is the normalizer of a 3-Sylow subgroup. But $72 \nmid 4! = (i(N))!$. By Lemma 2.9.1 N must contain a nontrivial normal subgroup of G (of order at least 3). Thus here again we can conclude that G contains a nontrivial normal subgroup. The upshot of the discussion is that any group of order 72 must have a nontrivial normal subgroup, hence cannot be simple.

Problems

1. Adapt the second proof given of Sylow's theorem to prove directly that if p is a prime and $p^\alpha \mid o(G)$, then G has a subgroup of order p^α.

2. If $x > 0$ is a real number, define $[x]$ to be m, where m is that integer such that $m \leq x < m + 1$. If p is a prime, show that the power of p which exactly divides $n!$ is given by

$$\left[\frac{n}{p}\right] + \left[\frac{n}{p^2}\right] + \cdots + \left[\frac{n}{p^r}\right] + \cdots.$$

3. Use the method for constructing the p-Sylow subgroup of S_{p^k} to find generators for
   (a) a 2-Sylow subgroup in S_8. (b) a 3-Sylow subgroup in S_9.

4. Adopt the method used in Problem 3 to find generators for
   (a) a 2-Sylow subgroup of S_6. (b) a 3-Sylow subgroup of S_6.

5. If p is a prime number, give explicit generators for a p-Sylow subgroup of S_{p^2}.

6. Discuss the number and nature of the 3-Sylow subgroups and 5-Sylow subgroups of a group of order $3^2 \cdot 5^2$.

7. Let G be a group of order 30.
   (a) Show that a 3-Sylow subgroup or a 5-Sylow subgroup of G must be normal in G.
   (b) From part (a) show that every 3-Sylow subgroup and every 5-Sylow subgroup of G must be normal in G.
   (c) Show that G has a normal subgroup of order 15.
   (d) From part (c) classify all groups of order 30.
   (e) How many different nonisomorphic groups of order 30 are there?

8. If G is a group of order 231, prove that the 11-Sylow subgroup is in the center of G.

9. If G is a group of order 385 show that its 11-Sylow subgroup is normal and its 7-Sylow subgroup is in the center of G.

10. If G is of order 108 show that G has a normal subgroup of order 3^k, where $k \geq 2$.

11. If $o(G) = pq$, p and q distinct primes, $p < q$, show
    (a) if $p \nmid (q - 1)$, then G is cyclic.
    *(b) if $p \mid (q - 1)$, then there exists a unique non-abelian group of order pq.

*12. Let G be a group of order pqr, $p < q < r$ primes. Prove
    (a) the r-Sylow subgroup is normal in G.
    (b) G has a normal subgroup of order qr.
    (c) if $q \nmid (r - 1)$, the q-Sylow subgroup of G is normal in G.

13. If G is of order p^2q, p, q primes, prove that G has a nontrivial normal subgroup.

*14. If G is of order p^2q, p, q primes, prove that either a p-Sylow subgroup or a q-Sylow subgroup of G must be normal in G.

15. Let G be a finite group in which $(ab)^p = a^p b^p$ for every $a, b \in G$, where p is a prime dividing $o(G)$. Prove
   (a) The p-Sylow subgroup of G is normal in G.
   *(b) If P is the p-Sylow subgroup of G, then there exists a normal subgroup N of G with $P \cap N = (e)$ and $PN = G$.
   (c) G has a nontrivial center.

**16. If G is a finite group and its p-Sylow subgroup P lies in the center of G, prove that there exists a normal subgroup N of G with $P \cap N = (e)$ and $PN = G$.

*17. If H is a subgroup of G, recall that $N(H) = \{x \in G \mid xHx^{-1} = H\}$. If P is a p-Sylow subgroup of G, prove that $N(N(P)) = N(P)$.

*18. Let P be a p-Sylow subgroup of G and suppose a, b are in the center of P. Suppose further that $a = xbx^{-1}$ for some $x \in G$. Prove that there exists a $y \in N(P)$ such that $a = yby^{-1}$.

**19. Let G be a finite group and suppose that ϕ is an automorphism of G such that ϕ^3 is the identity automorphism. Suppose further that $\phi(x) = x$ implies that $x = e$. Prove that for every prime p which divides $o(G)$, the p-Sylow subgroup is normal in G.

#20. Let G be the group of $n \times n$ matrices over the integers modulo p, p a prime, which are invertible. Find a p-Sylow subgroup of G.

21. Find the possible number of 11-Sylow subgroups, 7-Sylow subgroups, and 5-Sylow subgroups in a group of order $5^2 \cdot 7 \cdot 11$.

22. If G is S_3 and $A = ((1\ 2))$ in G, find all the double cosets AxA of A in G.

23. If G is S_4 and $A = ((1\ 2\ 3\ 4))$, $B = ((1\ 2))$, find all the double cosets AxB of A, B in G.

24. If G is the dihedral group of order 18 generated by $a^2 = b^9 = e$, $ab = b^{-1}a$, find the double cosets for H, K in G, where $H = (a)$ and $K = (b^3)$.

2.13 Direct Products

On several occasions in this chapter we have had a need for constructing a new group from some groups we already had on hand. For instance, towards the end of Section 2.8, we built up a new group using a given group and one of its automorphisms. A special case of this type of construction has been seen earlier in the recurring example of the dihedral group.

However, no attempt had been made for some systematic device for

constructing new groups from old. We shall do so now. The method represents the most simple-minded, straightforward way of combining groups to get other groups.

We first do it for two groups—not that two is sacrosanct. However, with this experience behind us, we shall be able to handle the case of any finite number easily and with dispatch. Not that any finite number is sacrosanct either; we could equally well carry out the discussion in the wider setting of any number of groups. However, we shall have no need for so general a situation here, so we settle for the case of any finite number of groups as our ultimate goal.

Let A and B be any two groups and consider the Cartesian product (which we discussed in Chapter 1) $G = A \times B$ of A and B. G consists of all ordered pairs (a, b), where $a \in A$ and $b \in B$. Can we use the operations in A and B to endow G with a product in such a way that G is a group? Why not try the obvious? Multiply componentwise. That is, let us define, for (a_1, b_1) and (a_2, b_2) in G, their product via $(a_1, b_1)(a_2, b_2) = (a_1a_2, b_1b_2)$. Here, the product a_1a_2 in the first component is the product of the elements a_1 and a_2 as calculated in the group A. The product b_1b_2 in the second component is that of b_1 and b_2 as elements in the group B.

With this definition we at least have a product defined in G. Is G a group relative to this product? The answer is yes, and is easy to verify. We do so now.

First we do the associative law. Let (a_1, b_1), (a_2, b_2), and (a_3, b_3) be three elements of G. Then $((a_1, b_1)(a_2, b_2))(a_3, b_3) = (a_1a_2, b_1b_2)(a_3, b_3) = ((a_1a_2)a_3, (b_1b_2)b_3)$, while $(a_1, b_1)((a_2, b_2)(a_3, b_3)) = (a_1, b_1)(a_2a_3, b_2b_3) = (a_1(a_2a_3), b_1(b_2b_3))$. The associativity of the product in A and in B then show us that our product in G is indeed associative.

Now to the unit element. What would be more natural than to try (e, f), where e is the unit element of A and f that of B, as the proposed unit element for G? We have $(a, b)(e, f) = (ae, bf) = (a, b)$ and $(e, f)(a, b) = (ea, fb) = (a, b)$. Thus (e, f) acts as a unit element in G.

Finally, we need the inverse in G for any element of G. Here, too, why not try the obvious? Let $(a, b) \in G$; try (a^{-1}, b^{-1}) as its inverse. Now $(a, b)(a^{-1}, b^{-1}) = (aa^{-1}, bb^{-1}) = (e, f)$ and $(a^{-1}, b^{-1})(a, b) = (a^{-1}a, b^{-1}b) = (e, f)$, so that (a^{-1}, b^{-1}) does serve as the inverse for (a, b).

With this we have verified that $G = A \times B$ is a group. We call it the *external direct product* of A and B.

Since $G = A \times B$ has been built up from A and B in such a trivial manner, we would expect that the structure of A and B would reflect heavily in that of G. This is indeed the case. Knowing A and B completely gives us complete information, structurally, about $A \times B$.

The construction of $G = A \times B$ has been from the outside, external. Now we want to turn the affair around and try to carry it out internally in G.

Consider $\bar{A} = \{(a,f) \in G \mid a \in A\} \subset G = A \times B$, where f is the unit element of B. What would one expect of $\bar{A}$? Answer: $\bar{A}$ is a subgroup of G and is isomorphic to A. To effect this isomorphism, define $\phi: A \to \bar{A}$ by $\phi(a) = (a,f)$ for $a \in A$. It is trivial that ϕ is an isomorphism of A onto $\bar{A}$. It is equally trivial that $\bar{A}$ is a subgroup of G. Furthermore, $\bar{A}$ is normal in G. For if $(a,f) \in \bar{A}$ and $(a_1, b_1) \in G$, then $(a_1, b_1)(a,f)(a_1, b_1)^{-1} = (a_1, b_1)(a,f)(a_1^{-1}, b_1^{-1}) = (a_1 a a_1^{-1}, b_1 f b^{1-1}) = (a_1 a a_1^{-1}, f) \in \bar{A}$. So we have an isomorphic copy, $\bar{A}$, of A in G which is a normal subgroup of G.

What we did for A we can also do for B. If $\bar{B} = \{(e, b) \in G \mid b \in B\}$, then $\bar{B}$ is isomorphic to B and is a normal subgroup of G.

We claim a little more, namely $G = \bar{A}\bar{B}$ and every $g \in G$ has a unique decomposition in the form $g = \bar{a}\bar{b}$ with $\bar{a} \in \bar{A}$ and $\bar{b} \in \bar{B}$. For, $g = (a, b) = (a,f)(e, b)$ and, since $(a,f) \in \bar{A}$ and $(e, b) \in \bar{B}$, we do have $g = \bar{a}\bar{b}$ with $\bar{a} = (a,f)$ and $\bar{b} = (e, b)$. Why is this unique? If $(a, b) = \bar{x}\bar{y}$, where $\bar{x} \in \bar{A}$ and $\bar{y} \in \bar{B}$, then $\bar{x} = (x,f)$, $x \in A$ and $\bar{y} = (e,y)$, $y \in B$; thus $(a, b) = \bar{x}\bar{y} = (x,f)(e,y) = (x,y)$. This gives $x = a$ and $y = b$, and so $\bar{x} = \bar{a}$ and $\bar{y} = \bar{b}$.

Thus we have realized G as an internal product $\bar{A}\bar{B}$ of two normal subgroups, $\bar{A}$ isomorphic to A, $\bar{B}$ to B in such a way that every element $g \in G$ has a unique representation in the form $g = \bar{a}\bar{b}$, with $\bar{a} \in \bar{A}$ and $\bar{b} \in \bar{B}$.

We leave the discussion of the product of two groups and go to the case of n groups, $n > 1$ any integer.

Let $G_1, G_2, \ldots, G_n$ be any n groups. Let $G = G_1 \times G_2 \times \cdots \times G_n = \{(g_1, g_2, \ldots, g_n) \mid g_i \in G_i\}$ be the set of all ordered n-tuples, that is, the Cartesian product of $G_1, G_2, \ldots, G_n$. We define a product in G via $(g_1, g_2, \ldots, g_n)(g_1', g_2', \ldots, g_n') = (g_1 g_1', g_2 g_2', \ldots, g_n g_n')$, that is, via componentwise multiplication. The product in the ith component is carried in the group G_i. Then G is a group in which $(e_1, e_2, \ldots, e_n)$ is the unit element, where each e_i is the unit element of G_i, and where $(g_1, g_2, \ldots, g_n)^{-1} = (g_1^{-1}, g_2^{-1}, \ldots, g_n^{-1})$. We call this group G the *external direct product* of $G_1, G_2, \ldots, G_n$.

In $G = G_1 \times G_2 \times \cdots \times G_n$ let $\bar{G}_i = \{(e_1, e_2, \ldots, e_{i-1}, g_i, e_{i+1}, \ldots, e_n) \mid g_i \in G_i\}$. Then $\bar{G}_i$ is a normal subgroup of G and is isomorphic to G_i. Moreover, $G = \bar{G}_1 \bar{G}_2 \cdots \bar{G}_n$ and every $g \in G$ has a unique decomposition $g = \bar{g}_1 \bar{g}_2 \cdots \bar{g}_n$, where $\bar{g}_1 \in \bar{G}_1, \ldots, \bar{g}_n \in \bar{G}_n$. We leave the verification of these facts to the reader.

Here, too, as in the case $A \times B$, we have realized the group G *internally* as the product of normal subgroups $\bar{G}_1, \ldots, \bar{G}_n$ in such a way that every element is uniquely representable as a product of elements $\bar{g}_1 \cdots \bar{g}_n$, where each $\bar{g}_i \in \bar{G}_i$. With this motivation we make the

DEFINITION Let G be a group and $N_1, N_2, \ldots, N_n$ *normal* subgroups of G such that

1. $G = N_1N_2 \cdots N_n$.
2. Given $g \in G$ then $g = m_1m_2 \cdots m_n$, $m_i \in N_i$ in a unique way.

We then say that G is the *internal direct product* of $N_1, N_2, \ldots, N_n$.

Before proceeding let's look at an example of a group G which is the internal direct product of some of its subgroups. Let G be a finite *abelian* group of order $p_1^{\alpha_1}p_2^{\alpha_2} \cdots p_k^{\alpha_k}$ where $p_1, p_2, \ldots, p_k$ are distinct primes and each $\alpha_i > 0$. If $P_1, \ldots, P_k$ are the p_1-Sylow subgroup, $\ldots$, p_k-Sylow subgroup respectively of G, then G is the internal direct product of $P_1, P_2, \ldots, P_k$ (see Problem 5).

We continue with the general discussion. Suppose that G is the internal direct product of the normal subgroups $N_1, \ldots, N_n$. The $N_1, \ldots, N_n$ are groups in their own right—forget that they are normal subgroups of G for the moment. Thus we can form the group $T = N_1 \times N_2 \times \cdots \times N_n$, the external direct product of $N_1, \ldots, N_n$. One feels that G and T should be related. Our aim, in fact, is to show that G is isomorphic to T. If we could establish this then we could abolish the prefix external and internal in the phrases external direct product, internal direct product—after all these would be the same group up to isomorphism—and just talk about the direct product.

We start with

LEMMA 2.13.1 *Suppose that G is the internal direct product of $N_1, \ldots, N_n$. Then for $i \neq j$, $N_i \cap N_j = (e)$, and if $a \in N_i$, $b \in N_j$ then $ab = ba$.*

Proof. Suppose that $x \in N_i \cap N_j$. Then we can write x as

$$x = e_1 \cdots e_{i-1}xe_{i+1} \cdots e_j \cdots e_n,$$

where $e_t = e$, viewing x as an element in N_i. Similarly, we can write x as

$$x = e_1 \cdots e_i \cdots e_{j-1}xe_{j+1} \cdots e_n,$$

where $e_t = e$, viewing x as an element of N_j. But every element—and so, in particular x—has a unique representation in the form $m_1m_2 \cdots m_n$, where $m_i \in N_1, \ldots, m_n \in N_n$. Since the two decompositions in this form for x must coincide, the entry from N_i in each must be equal. In our first decomposition this entry is x, in the other it is e; hence $x = e$. Thus $N_i \cap N_j = (e)$ for $i \neq j$.

Suppose $a \in N_i$, $b \in N_j$, and $i \neq j$. Then $aba^{-1} \in N_j$ since N_j is normal; thus $aba^{-1}b^{-1} \in N_j$. Similarly, since $a^{-1} \in N_i$, $ba^{-1}b^{-1} \in N_i$, whence $aba^{-1}b^{-1} \in N_i$. But then $aba^{-1}b^{-1} \in N_i \cap N_j = (e)$. Thus $aba^{-1}b^{-1} = e$; this gives the desired result $ab = ba$.

One should point out that if $K_1, \ldots, K_n$ are normal subgroups of G such that $G = K_1K_2 \cdots K_n$ and $K_i \cap K_j = (e)$ for $i \neq j$ it *need not be*

true that G is the internal direct product of $K_1, \ldots, K_n$. A more stringent condition is needed (see Problems 8 and 9).

We now can prove the desired isomorphism between the external and internal direct products that was stated earlier.

THEOREM 2.13.1 *Let G be a group and suppose that G is the internal direct product of $N_1, \ldots, N_n$. Let $T = N_1 \times N_2 \times \cdots \times N_n$. Then G and T are isomorphic.*

Proof. Define the mapping $\psi: T \to G$ by

$$\psi((b_1, b_2, \ldots, b_n)) = b_1 b_2 \cdots b_n,$$

where each $b_i \in N_i$, $i = 1, \ldots, n$. We claim that ψ is an isomorphism of T onto G.

To begin with, ψ is certainly onto; for, since G is the internal direct product of $N_1, \ldots, N_n$, if $x \in G$ then $x = a_1 a_2 \cdots a_n$ for some $a_1 \in N_1, \ldots, a_n \in N_n$. But then $\psi((a_1, a_2, \ldots, a_n)) = a_1 a_2 \cdots a_n = x$. The mapping ψ is one-to-one by the *uniqueness* of the representation of every element as a product of elements from $N_1, \ldots, N_n$. For, if $\psi((a_1, \ldots, a_n)) = \psi((c_1, \ldots, c_n))$, where $a_i \in N_i$, $c_i \in N_i$, for $i = 1, 2, \ldots, n$, then, by the definition of ψ, $a_1 a_2 \cdots a_n = c_1 c_2 \cdots c_n$. The uniqueness in the definition of internal direct product forces $a_1 = c_1$, $a_2 = c_2, \ldots, a_n = c_n$. Thus ψ is one-to-one.

All that remains is to show that ψ is a homomorphism of T onto G. If $X = (a_1, \ldots, a_n)$, $Y = (b_1, \ldots, b_n)$ are elements of T then

$$\begin{aligned}\psi(XY) &= \psi((a_1, \ldots, a_n)(b_1, \ldots, b_n)) \\ &= \psi(a_1 b_1, a_2 b_2, \ldots, a_n b_n) \\ &= a_1 b_1 a_2 b_2 \cdots a_n b_n.\end{aligned}$$

However, by Lemma 2.13.1, $a_i b_j = b_j a_i$ if $i \neq j$. This tells us that $a_1 b_1 a_2 b_2 \cdots a_n b_n = a_1 a_2 \cdots a_n b_1 b_2 \cdots b_n$. Thus $\psi(XY) = a_1 a_2 \cdots a_n b_1 b_2 \cdots b_n$. But we can recognize $a_1 a_2 \cdots a_n$ as $\psi((a_1, a_2, \ldots, a_n)) = \psi(X)$ and $b_1 b_2 \cdots b_n$ as $\psi(Y)$. We therefore have $\psi(XY) = \psi(X)\psi(Y)$. In short, we have shown that ψ is an isomorphism of T onto G. This proves the theorem.

Note one particular thing that the theorem proves. If a group G is isomorphic to an external direct product of certain groups G_i, then G *is*, in fact, the internal direct product of groups $\bar{G}_i$ isomorphic to the G_i. We simply say that G is the direct product of the $\bar{G}_i$ (or G_i).

In the next section we shall see that every finite abelian group is a direct product of cyclic groups. Once we have this, we have the structure of all finite abelian groups pretty well under our control.

One should point out that the analog of the direct product of groups exists in the study of almost all algebraic structures. We shall see this later

for vector-spaces, rings, and modules. Theorems that describe such an algebraic object in terms of direct products of more describable algebraic objects of the same kind (for example, the case of abelian groups above) are important theorems in general. Through such theorems we can reduce the study of a fairly complex algebraic situation to a much simpler one.

Problems

1. If A and B are groups, prove that $A \times B$ is isomorphic to $B \times A$.
2. If G_1, G_2, G_3 are groups, prove that $(G_1 \times G_2) \times G_3$ is isomorphic to $G_1 \times G_2 \times G_3$. Care to generalize?
3. If $T = G_1 \times G_2 \times \cdots \times G_n$ prove that for each $i = 1, 2, \ldots, n$ there is a homomorphism ϕ_i of T onto G_i. Find the kernel of ϕ_i.
4. Let G be a group and let $T = G \times G$.
   (a) Show that $D = \{(g, g) \in G \times G \mid g \in G\}$ is a group isomorphic to G.
   (b) Prove that D is normal in T if and only if G is abelian.
5. Let G be a finite abelian group. Prove that G is isomorphic to the direct product of its Sylow subgroups.
6. Let A, B be cyclic groups of order m and n, respectively. Prove that $A \times B$ is cyclic if and only if m and n are relatively prime.
7. Use the result of Problem 6 to prove the Chinese Remainder Theorem; namely, if m and n are relatively prime integers and u, v any two integers, then we can find an integer x such that $x \equiv u \bmod m$ and $x \equiv v \bmod n$.
8. Give an example of a group G and normal subgroups $N_1, \ldots, N_n$ such that $G = N_1 N_2 \cdots N_n$ and $N_i \cap N_j = (e)$ for $i \neq j$ and yet G is *not* the internal direct product of $N_1, \ldots, N_n$.
9. Prove that G is the internal direct product of the normal subgroups $N_1, \ldots, N_n$ if and only if
   1. $G = N_1 \cdots N_n$.
   2. $N_i \cap (N_1 N_2 \cdots N_{i-1} N_{i+1} \cdots N_n) = (e)$ for $i = 1, \ldots, n$.
10. Let G be a group, $K_1, \ldots, K_n$ normal subgroups of G. Suppose that $K_1 \cap K_2 \cap \cdots \cap K_n = (e)$. Let $V_i = G/K_i$. Prove that there is an isomorphism of G into $V_1 \times V_2 \times \cdots \times V_n$.

*11. Let G be a finite abelian group such that it contains a subgroup $H_0 \neq (e)$ which lies in *every* subgroup $H \neq (e)$. Prove that G must be cyclic. What can you say about $o(G)$?

12. Let G be a finite abelian group. Using Problem 11 show that G is isomorphic to a subgroup of a direct product of a finite number of finite cyclic groups.

13. Give an example of a finite non-abelian group G which contains a subgroup $H_0 \neq (e)$ such that $H_0 \subset H$ for *all* subgroups $H \neq (e)$ of G.
14. Show that every group of order p^2, p a prime, is either cyclic or is isomorphic to the direct product of two cyclic groups each of order p.
*15. Let $G = A \times A$ where A is cyclic of order p, p a prime. How many automorphisms does G have?
16. If $G = K_1 \times K_2 \times \cdots \times K_n$ describe the center of G in terms of those of the K_i.
17. If $G = K_1 \times K_2 \times \cdots \times K_n$ and $g \in G$, describe

$$N(g) = \{x \in G \mid xg = gx\}.$$

18. If G is a finite group and $N_1, \ldots, N_n$ are normal subgroups of G such that $G = N_1 N_2 \cdots N_n$ and $o(G) = o(N_1)o(N_2) \cdots o(N_n)$, prove that G is the direct product of $N_1, N_2, \ldots, N_n$.

2.14 Finite Abelian Groups

We close this chapter with a discussion (and description) of the structure of an arbitrary finite abelian group. The result which we shall obtain is a famous classical theorem, often referred to as the Fundamental Theorem on Finite Abelian Groups. It is a highly satisfying result because of its decisiveness. Rarely do we come out with so compact, succinct, and crisp a result. In it the structure of a finite abelian group is completely revealed, and by means of it we have a ready tool for attacking any structural problem about finite abelian groups. It even has some arithmetic consequences. For instance, one of its by-products is a precise count of how many non-isomorphic abelian groups there are of a given order.

In all fairness one should add that this description of finite abelian groups is not as general as we can go and still get so sharp a theorem. As you shall see in Section 4.5, we completely describe all abelian groups generated by a finite set of elements—a situation which not only covers the finite abelian group case, but much more.

We now state this very fundamental result.

THEOREM 2.14.1 *Every finite abelian group is the direct product of cyclic groups.*

Proof. Our first step is to reduce the problem to a slightly easier one. We have already indicated in the preceding section (see Problem 5 there) that any finite abelian group G is the direct product of its Sylow subgroups. If we knew that each such Sylow subgroup was a direct product of cyclic groups we could put the results together for these Sylow subgroups to

realize G as a direct product of cyclic groups. Thus it suffices to prove the theorem for abelian groups of order p^n where p is a prime.

So suppose that G is an abelian group of order p^n. Our objective is to find elements $a_1, \ldots, a_k$ in G such that every element $x \in G$ can be written in a unique fashion as $x = a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_k^{\alpha_k}$. Note that if this were true and $a_1, \ldots, a_k$ were of order $p^{n_1}, \ldots, p^{n_k}$, where $n_1 \geq n_2 \geq \cdots \geq n_k$, then the maximal order of any element in G would be p^{n_1} (Prove!). This gives us a cue of how to go about finding the elements $a_1, \ldots, a_k$ that we seek.

The procedure suggested by this is: let a_1 be an element of maximal order in G. How shall we pick a_2? Well, if $A_1 = (a_1)$ the subgroup generated by a_1, then a_2 maps into an element of highest order in G/A_1. If we can successfully exploit this to find an appropriate a_2, and if $A_2 = (a_2)$, then a_3 would map into an element of maximal order in G/A_1A_2, and so on. With this as guide we can now get down to the brass tacks of the proof.

Let a_1 be an element in G of highest possible order, p^{n_1}, and let $A_1 = (a_1)$. Pick b_2 in G such that $\bar{b}_2$, the image of b_2 in $\bar{G} = G/A_1$, has maximal order p^{n_2}. Since the order of $\bar{b}_2$ divides that of b_2, and since the order of a_1 is maximal, we must have that $n_1 \geq n_2$. In order to get a direct product of A_1 with (b_2) we would need $A_1 \cap (b_2) = (e)$; this might not be true for the initial choice of b_2, so we may have to adapt the element b_2. Suppose that $A_1 \cap (b_2) \neq (e)$; then, since ${b_2}^{p^{n_2}} \in A_1$ and is the first power of b_2 to fall in A_1 (by our mechanism of choosing b_2) we have that ${b_2}^{p^{n_2}} = {a_1}^i$. Therefore $({a_1}^i)^{p^{n_1-n_2}} = ({b_2}^{p^{n_2}})^{p^{n_1-n_2}} = {b_2}^{p^{n_1}} = e$, whence ${a_1}^{ip^{n_1-n_2}} = e$. Since a_1 is of order p^{n_1} we must have that $p^{n_1} \mid ip^{n_1-n_2}$, and so $p^{n_2} \mid i$. Thus, recalling what i is, we have ${b_2}^{p^{n_2}} = {a_1}^i = {a_1}^{jp^{n_2}}$. This tells us that if $a_2 = {a_1}^{-j} b_2$ then ${a_2}^{p^{n_2}} = e$. The element a_2 is indeed the element we seek. Let $A_2 = (a_2)$. We claim that $A_1 \cap A_2 = (e)$. For, suppose that ${a_2}^t \in A_1$; since $a_2 = {a_1}^{-j} b_2$, we get $({a_1}^{-j} b_2)^t \in A_1$ and so ${b_2}^t \in A_1$. By choice of b_2, this last relation forces $p^{n_2} \mid t$, and since ${a_2}^{p^{n_2}} = e$ we must have that ${a_2}^t = e$. In short $A_1 \cap A_2 = (e)$.

We continue one more step in the program we have outlined. Let $b_3 \in G$ map into an element of maximal order in $G/(A_1A_2)$. If the order of the image of b_3 in $G/(A_1A_2)$ is p^{n_3}, we claim that $n_3 \leq n_2 \leq n_1$. Why? By the choice of n_2, ${b_3}^{p^{n_2}} \in A_1$ so is certainly in A_1A_2. Thus $n_3 \leq n_2$. Since ${b_3}^{p^{n_3}} \in A_1A_2$, ${b_3}^{p^{n_3}} = {a_1}^{i_1} {a_2}^{i_2}$. We claim that $p^{n_3} \mid i_1$ and $p^{n_3} \mid i_2$. For, ${b_3}^{p^{n_2}} \in A_1$ hence $({a_1}^{i_1} {a_2}^{i_2})^{p^{n_2-n_3}} = ({b_3}^{p^{n_3}})^{p^{n_2-n_3}} \equiv {b_3}^{p^{n_2}} \in A_1$. This tells us that ${a_2}^{i_2 p^{n_2-n_3}} \in A_1$ and so $p^{n_2} \mid i_2 p^{n_2-n_3}$, which is to say, $p^{n_3} \mid i_2$. Also ${b_3}^{p^{n_1}} = e$, hence $({a_1}^{i_1} {a_2}^{i_2})^{p^{n_1-n_3}} = {b_3}^{p^{n_1}} = e$; this says that ${a_1}^{i_1 p^{n_1-n_3}} \in A_2 \cap A_1 = (e)$, that is, ${a_1}^{i_1 p^{n_1-n_3}} = e$. This yields that $p^{n_3} \mid i_1$. Let $i_1 = j_1 p^{n_3}$, $i_2 = j_2 p^{n_3}$; thus ${b_3}^{p^{n_3}} = {a_1}^{j_1 p^{n_3}} {a_2}^{j_2 p^{n_3}}$. Let $a_3 = {a_1}^{-j_1} {a_2}^{-j_2} b_3$, $A_3 = (a_3)$; note that ${a_3}^{p^{n_3}} = e$. We claim that $A_3 \cap (A_1A_2) = (e)$. For if ${a_3}^t \in A_1A_2$ then $({a_1}^{-i_1} {a_2}^{-j_2} b_3)^t \in A_1A_2$, giving us ${b_3}^t \in A_1A_2$. But then $p^{n_3} \mid t$, whence, since ${a_3}^{p^{n_3}} = e$, we have ${a_3}^t = e$. In other words, $A_3 \cap (A_1A_2) = (e)$.

Continuing this way we get cyclic subgroups $A_1 = (a_1)$, $A_2 = (a_2), \ldots, A_k = (a_k)$ of order $p^{n_1}, p^{n_2}, \ldots, p^{n_k}$, respectively, with $n_1 \geq n_2 \geq \cdots \geq n_k$ such that $G = A_1A_2 \cdots A_k$ and such that, for each i, $A_i \cap (A_1A_2 \cdots A_{i-1}) = (e)$. This tells us that every $x \in G$ has a unique representation as $x = a_1'a_2' \cdots a_k'$ where $a_1' \in A_1, \ldots, a_k' \in A_k$. In other words, G is the direct product of the cyclic subgroups $A_1, A_2, \ldots, A_k$. The theorem is now proved.

DEFINITION If G is an abelian group of order p^n, p a prime, and $G = A_1 \times A_2 \times \cdots \times A_k$ where each A_i is cyclic of order p^{n_i} with $n_1 \geq n_2 \geq \cdots \geq n_k > 0$, then the integers $n_1, n_2, \ldots, n_k$ are called the *invariants* of G.

Just because we called the integers above the invariants of G does not mean that they *are* really *the* invariants of G. That is, it is possible that we can assign different sets of invariants to G. We shall soon show that the invariants of G are indeed unique and completely describe G.

Note one other thing about the invariants of G. If $G = A_1 \times \cdots \times A_k$, where A_i is cyclic of order p^{n_i}, $n_1 \geq n_2 \geq \cdots \geq n_k > 0$, then $o(G) = o(A_1)o(A_2) \cdots o(A_k)$, hence $p^n = p^{n_1}p^{n_2} \cdots p^{n_k} = p^{n_1+n_2+\cdots+n_k}$, whence $n = n_1 + n_2 + \cdots + n_k$. In other words, $n_1, n_2, \ldots, n_k$ give us a *partition* of n. We have already run into this concept earlier in studying the conjugate classes in the symmetric group.

Before discussing the uniqueness of the invariants of G, one thing should be made absolutely clear: the elements $a_1, \ldots, a_k$ and the subgroups $A_1, \ldots, A_k$ which they generate, which arose above to give the decomposition of G into a direct product of cyclic groups, are *not* unique. Let's see this in a very simple example. Let $G = \{e, a, b, ab\}$ be an abelian group of order 4 where $a^2 = b^2 = e$, $ab = ba$. Then $G = A \times B$ where $A = (a)$, $B = (b)$ are cyclic groups of order 2. But we have another decomposition of G as a direct product, namely, $G = C \times B$ where $C = (ab)$ and $B = (b)$. So, even in this group of very small order, we can get distinct decompositions of the group as the direct product of cyclic groups. Our claim—which we now want to substantiate—is that while these cyclic subgroups are not unique, their *orders are*

DEFINITION If G is an abelian group and s is any integer, then $G(s) = \{x \in G \mid x^s = e\}$.

Because G is abelian it is evident that $G(s)$ is a subgroup of G. We now prove

LEMMA 2.14.1 *If G and G' are isomorphic abelian groups, then for every integer s, $G(s)$, and $G'(s)$ are isomorphic.*

Proof. Let ϕ be an isomorphism of G onto G'. We claim that ϕ maps $G(s)$ isomorphically onto $G'(s)$. First we show that $\phi(G(s)) \subset G'(s)$. For, if $x \in G(s)$ then $x^s = e$, hence $\phi(x^s) = \phi(e) = e'$. But $\phi(x^s) = \phi(x)^s$; hence $\phi(x)^s = e'$ and so $\phi(x)$ is in $G'(s)$. Thus $\phi(G(s)) \subset G'(s)$.

On the other hand, if $u' \in G'(s)$ then $(u')^s = e'$. But, since ϕ is onto, $u' = \phi(y)$ for some $y \in G$. Therefore $e' = (u')^s = \phi(y)^s = \phi(y^s)$. Because ϕ is one-to-one, we have $y^s = e$ and so $y \in G(s)$. Thus ϕ maps $G(s)$ *onto* $G'(s)$.

Therefore since ϕ is one-to-one, onto, and a homomorphism from $G(s)$ to $G'(s)$, we have that $G(s)$ and $G'(s)$ are isomorphic.

We continue with

LEMMA 2.14.2 *Let G be an abelian group of order p^n, p a prime. Suppose that $G = A_1 \times A_2 \times \cdots \times A_k$, where each $A_i = (a_i)$ is cyclic of order p^{n_i}, and $n_1 \geq n_2 \geq \cdots \geq n_k > 0$. If m is an integer such that $n_t > m \geq n_{t+1}$ then $G(p^m) = B_1 \times \cdots \times B_t \times A_{t+1} \times \cdots \times A_k$ where B_i is cyclic of order p^m, generated by $a_i^{p^{n_i - m}}$, for $i \leq t$. The order of $G(p^m)$ is p^u, where*

$$u = mt + \sum_{i=t+1}^{k} n_i.$$

Proof. First of all, we claim that $A_{t+1}, \ldots, A_k$ are all in $G(p^m)$. For, since $m \geq n_{t+1} \geq \cdots \geq n_k > 0$, if $j \geq t+1$, $a_j^{p^m} = (a_j^{p^{n_j}})^{p^{m-n_j}} = e$. Hence A_j, for $j \geq t+1$ lies in $G(p^m)$.

Secondly, if $i \leq t$ then $n_i > m$ and $(a_i^{p^{n_i - m}})^{p^m} = a_i^{p^{n_i}} = e$, whence each such $a_i^{p^{n_i - m}}$ is in $G(p^m)$ and so the subgroup it generates, B_i, is also in $G(p^m)$.

Since $B_1, \ldots, B_t, A_{t+1}, \ldots, A_k$ are all in $G(p^m)$, their product (which is direct, since the product $A_1 A_2 \cdots A_k$ is direct) is in $G(p^m)$. Hence $G(p^m) \supset B_1 \times \cdots \times B_t \times A_{t+1} \times \cdots \times A_k$.

On the other hand, if $x = a_1^{\lambda_1} a_2^{\lambda_2} \cdots a_k^{\lambda_k}$ is in $G(p^m)$, since it then satisfies $x^{p^m} = e$, we set $e = x^{p^m} = a_1^{\lambda_1 p^m} \cdots a_k^{\lambda_k p^m}$. However, the product of the subgroups $A_1, \ldots, A_k$ is direct, so we get

$$a_1^{\lambda_1 p^m} = e, \ldots, a_k^{\lambda_k p^m} = e.$$

Thus the order of a_i, that is, p^{n_i} must divide $\lambda_i p^m$ for $i = 1, 2, \ldots, k$. If $i \geq t+1$ this is automatically true whatever be the choice of $\lambda_{t+1}, \ldots, \lambda_k$ since $m \geq n_{t+1} \geq \cdots \geq n_k$, hence $p^{n_i} \mid p^m$, $i \geq t+1$. However, for $i \leq t$, we get from $p^{n_i} \mid \lambda_i p^m$ that $p^{n_i - m} \mid \lambda_i$. Therefore $\lambda_i = v_i p^{n_i - m}$ for some integer v_i. Putting all this information into the values of the λ_i's in the expression for x as $x = a_1^{\lambda_1} \cdots a_k^{\lambda_k}$ we see that

$$x = a_1^{v_1 p^{n_1 - m}} \cdots a_t^{v_t p^{n_t - m}} a_{t+1}^{\lambda_{t+1}} \cdots a_k^{\lambda_k}.$$

This says that $x \in B_1 \times \cdots \times B_t \times A_{t+1} \times \cdots \times A_k$.

Now since each B_i is of order p^m and since $o(A_i) = p^{n_i}$ and since $G = B_1 \times \cdots \times B_t \times A_{t+1} \times \cdots \times A_k$,

$$o(G) = o(B_1)o(B_2)\cdots o(B_t)o(A_{t+1})\cdots o(A_k) = \underbrace{p^m p^m \cdots p^m}_{t\text{-times}} p^{n_{t+1}} \cdots p^{n_k}.$$

Thus, if we write $o(G) = p^u$, then

$$u = mt + \sum_{i=t+1}^{k} n_i.$$

The lemma is proved.

COROLLARY *If G is as in Lemma 2.14.2, then $o(G(p)) = p^k$.*

Proof. Apply the lemma to the case $m = 1$. Then $t = k$, hence $u = 1k = k$ and so $o(G) = p^k$.

We now have all the pieces required to prove the uniqueness of the invariants of an abelian group of order p^n.

THEOREM 2.14.2 *Two abelian groups of order p^n are isomorphic if and only if they have the same invariants.*

In other words, if G and G' are abelian groups of order p^n and $G = A_1 \times \cdots \times A_k$, where each A_i is a cyclic group of order p^{n_i}, $n_1 \geq \cdots \geq n_k > 0$, and $G' = B_1' \times \cdots \times B_s'$, where each B_i' is a cyclic group of order p^{h_i}, $h_1 \geq \cdots \geq h_s > 0$, then G and G' are isomorphic if and only if $k = s$ and for each i, $n_i = h_i$.

Proof. One way is very easy, namely, if G and G' have the same invariants then they are isomorphic. For then $G = A_1 \times \cdots \times A_k$ where $A_i = (a_i)$ is cyclic of order p^{n_i}, and $G' = B_1' \times \cdots \times B_k'$ where $B_i' = (b_i')$ is cyclic of order p^{n_i}. Map G onto G' by the map $\phi(a_1^{\alpha_1} \cdots a_k^{\alpha_k}) = (b_1')^{\alpha_1} \cdots (b_k')^{\alpha_k}$. We leave it to the reader to verify that this defines an isomorphism of G onto G'.

Now for the other direction. Suppose that $G = A_1 \times \cdots \times A_k$, $G' = B_1' \times \cdots \times B_s'$, A_i, B_i' as described above, cyclic of orders p^{n_i}, p^{h_i}, respectively, where $n_1 \geq \cdots \geq n_k > 0$ and $h_1 \geq \cdots \geq h_s > 0$. We want to show that if G and G' are isomorphic then $k = s$ and each $n_i = h_i$.

If G and G' are isomorphic then, by Lemma 2.14.1, $G(p^m)$ and $G'(p^m)$ must be isomorphic for any integer $m \geq 0$, hence must have the same order. Let's see what this gives us in the special case $m = 1$; that is, what information can we garner from $o(G(p)) = o(G'(p))$. According to the corollary to Lemma 2.14.2, $o(G(p)) = p^k$ and $o(G'(p)) = p^s$. Hence $p^k = p^s$ and so $k = s$. At least we now know that the *number* of invariants for G and G' is the same.

If $n_i \neq h_i$ for some i, let t be the first i such that $n_t \neq h_t$; we may suppose that $n_t > h_t$. Let $m = h_t$. Consider the subgroups, $H = \{x^{p^m} \mid x \in G\}$ and $H' = \{(x')^{p^m} \mid x' \in G\}$, of G and G', respectively. Since G and G' are isomorphic, it follows easily that H and H' are isomorphic. We now examine the invariants of H and H'.

Because $G = A_1 \times \cdots \times A_k$, where $A_i = (a_i)$ is of order p^{n_i}, we get that

$$H = C_1 \times \cdots \times C_t \times \cdots \times C_r,$$

where $C_i = (a_i^{p^m})$ is of order $p^{n_i - m}$, and where r is such that $n_r > m = h_t \geq n_{r-1}$. Thus the invariants of H are $n_1 - m$, $n_2 - m, \ldots, n_r - m$ and the number of invariants of H is $r \geq t$.

Because $G' = B_1' \times \cdots \times B_k'$, where $B_i = (b_i')$ is cyclic of order p^{h_i}, we get that $H' = D_1' \times \cdots \times D_{t-1}'$, where $D_i' = ((b_i')^{p^m})$ is cyclic of order $p^{h_i - m}$. Thus the invariants of H' are $h_1 - m, \ldots, h_{t-1} - m$ and so the number of invariants of H' is $t - 1$.

But H and H' are isomorphic; as we saw above this forces them to have the same number of invariants. But we saw that assuming that $n_i \neq h_i$ for some i led to a discrepancy in the number of their invariants. In consequence each $n_i = h_i$, and the theorem is proved.

An immediate consequence of this last theorem is that an abelian group of order p^n can be decomposed in only one way—*as far as the orders of the cyclic subgroups is concerned*—as a direct product of cyclic subgroups. Hence the invariants are indeed *the* invariants of G and completely determine G.

If $n_1 \geq \cdots \geq n_k > 0$, $n = n_1 + \cdots + n_k$, is any partition of n, then we can easily construct an abelian group of order p^n whose invariants are $n_1 \geq \cdots \geq n_k > 0$. To do this, let A_i be a cyclic group of order p^{n_i} and let $G = A_1 \times \cdots \times A_k$ be the external direct product of $A_1, \ldots, A_k$. Then, by the very definition, the invariants of G are $n_1 \geq \cdots \geq n_k > 0$. Finally, two different partitions of n give rise to nonisomorphic abelian groups of order p^n. This, too, comes from Theorem 2.14.2. Hence we have

THEOREM 2.14.3 *The number of nonisomorphic abelian groups of order p^n, p a prime, equals the number of partitions of n.*

Note that the answer given in Theorem 2.14.3 does *not* depend on the prime p; it only depends on the exponent n. Hence, for instance, the number of nonisomorphic abelian groups of order 2^4 equals that of orders 3^4, or 5^4, etc. Since there are five partitions of 4, namely: $4 = 4$, $3 + 1$, $2 + 2$, $2 + 1 + 1$, $1 + 1 + 1 + 1$, then there are five nonisomorphic abelian groups of order p^4 for any prime p.

Since any finite abelian group is a direct product of its Sylow subgroups, and two abelian groups are isomorphic if and only if their corresponding Sylow subgroups are isomorphic, we have the

COROLLARY *The number of nonisomorphic abelian groups of order* $p_1^{\alpha_1} \cdots p_r^{\alpha_r}$, *where the* p_i *are distinct primes and where each* $\alpha_i > 0$, *is* $p(\alpha_1)p(\alpha_2) \cdots p(\alpha_r)$, *where* $p(u)$ *denotes the number of partitions of* u.

Problems

1. If G is an abelian group of order p^n, p a prime and $n_1 \geq n_2 \geq \cdots \geq n_k > 0$, are the invariants of G, show that the maximal order of any element in G is p^{n_1}.
2. If G is a group, $A_1, \ldots, A_k$ normal subgroups of G such that $A_i \cap (A_1A_2 \cdots A_{i-1}) = (e)$ for all i, show that G is the direct product of $A_1, \ldots, A_k$ if $G = A_1A_2 \cdots A_k$.
3. Using Theorem 2.14.1, prove that if a finite abelian group has subgroups of orders m and n, then it has a subgroup whose order is the least common multiple of m and n.
4. Describe all finite abelian groups of order
   (a) 2^6. (b) 11^6. (c) 7^5. (d) $2^4 \cdot 3^4$.
5. Show how to get all abelian groups of order $2^3 \cdot 3^4 \cdot 5$.
6. If G is an abelian group of order p^n with invariants $n_1 \geq \cdots \geq n_k > 0$ and $H \neq (e)$ is a subgroup of G, show that if $h_1 \geq \cdots \geq h_s > 0$ are the invariants of H, then $k \geq s$ and for each i, $h_i \leq n_i$ for $i = 1, 2, \ldots, s$.

   If G is an abelian group, let $\hat{G}$ be the set of all homomorphisms of G into the group of nonzero complex numbers under multiplication. If $\phi_1, \phi_2 \in \hat{G}$, define $\phi_1 \cdot \phi_2$ by $(\phi_1 \cdot \phi_2)(g) = \phi_1(g)\phi_2(g)$ for all $g \in G$.
7. Show that $\hat{G}$ is an abelian group under the operation defined.
8. If $\phi \in \hat{G}$ and G is finite, show that $\phi(g)$ is a root of unity for every $g \in G$.
9. If G is a finite cyclic group, show that $\hat{G}$ is cyclic and $o(\hat{G}) = o(G)$, hence G and $\hat{G}$ are isomorphic.
10. If $g_1 \neq g_2$ are in G, G a finite abelian group, prove that there is a $\phi \in \hat{G}$ with $\phi(g_1) \neq \phi(g_2)$.
11. If G is a finite abelian group prove that $o(G) = o(\hat{G})$ and G is isomorphic to $\hat{G}$.
12. If $\phi \neq 1 \in \hat{G}$ where G is an abelian group, show that $\sum_{g \in G} \phi(g) = 0$.

Supplementary Problems

There is no relation between the order in which the problems appear and the order of appearance of the sections, in this chapter, which might be relevant to their solutions. No hint is given regarding the difficulty of any problem.

1. (a) If G is a finite abelian group with elements $a_1, a_2, \ldots, a_n$, prove that $a_1a_2 \cdots a_n$ is an element whose square is the identity.
   (b) If the G in part (a) has no element of order 2 or more than one element of order 2, prove that $a_1a_2 \cdots a_n = e$.
   (c) If G has one element, y, of order 2, prove that $a_1a_2 \cdots a_n = y$.
   (d) (*Wilson's theorem*) If p is a prime number show that $(p-1)! \equiv -1(p)$.
2. If p is an odd prime and if
$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{p-1} = \frac{a}{b},$$
where a and b are integers, prove that $p \mid a$. If $p > 3$, prove that $p^2 \mid a$.
3. If p is an odd prime, $a \not\equiv 0\ (p)$ is said to be a *quadratic residue of* p if there exists an integer x such that $x^2 \equiv a(p)$. Prove
   (a) The quadratic residues of p form a subgroup Q of the group of nonzero integers mod p under multiplication.
   (b) $o(Q) = (p-1)/2$.
   (c) If $q \in Q$, $n \notin Q$ (n is called a *nonresidue*), then nq is a nonresidue.
   (d) If n_1, n_2 are nonresidues, then n_1n_2 is a residue.
   (e) If a is a quadratic residue of p, then $a^{(p-1)/2} \equiv +1(p)$.
4. Prove that in the integers mod p, p a prime, there are at most n solutions of $x^n \equiv 1(p)$ for every integer n.
5. Prove that the nonzero integers mod p under multiplication form a cyclic group if p is a prime.
6. Give an example of a non-abelian group in which $(xy)^3 = x^3y^3$ for all x and y.
7. If G is a finite abelian group, prove that the number of solutions of $x^n = e$ in G, where $n \mid o(G)$ is a multiple of n.
8. Same as Problem 7, but do not assume the group to be abelian.
9. Find all automorphisms of S_3 and S_4, the symmetric groups of degree 3 and 4.

DEFINITION *A group G is said to be* solvable *if there exist subgroups $G = N_0 \supset N_1 \supset N_2 \supset \cdots \supset N_r = (e)$ such that each N_i is normal in N_{i-1} and N_{i-1}/N_i is abelian.*

10. Prove that a subgroup of a solvable group and the homomorphic image of a solvable group must be solvable.
11. If G is a group and N is a normal subgroup of G such that both N and G/N are solvable, prove that G is solvable.
12. If G is a group, A a subgroup of G and N a normal subgroup of G, prove that if both A and N are solvable then so is AN.

13. If G is a group, define the sequence of subgroups $G^{(i)}$ of G by
    (1) $G^{(1)}$ = commutator subgroup of G = subgroup of G generated by all $aba^{-1}b^{-1}$ where $a, b \in G$.
    (2) $G^{(i)}$ = commutator subgroup of $G^{(i-1)}$ if $i > 1$.
    Prove
    (a) Each $G^{(i)}$ is a normal subgroup of G.
    (b) G is solvable if and only if $G^{(k)} = (e)$ for some $k \geq 1$.

14. Prove that a solvable group always has an abelian normal subgroup $M \neq (e)$.
    If G is a group, define the sequence of subgroups $G_{(i)}$ by
    (a) $G_{(1)}$ = commutator subgroup of G.
    (b) $G_{(i)}$ = subgroup of G generated by all $aba^{-1}b^{-1}$ where $a \in G$, $b \in G_{(i-1)}$.

G is said to be *nilpotent* if $G_{(k)} = (e)$ for some $k \geq 1$.

15. (a) Show that each $G_{(i)}$ is a normal subgroup of G and $G_{(i)} \supset G^{(i)}$.
    (b) If G is nilpotent, prove it must be solvable.
    (c) Give an example of a group which is solvable but not nilpotent.

16. Show that any subgroup and homomorphic image of a nilpotent group must be nilpotent.

17. Show that every homomorphic image, different from (e), of a nilpotent group has a nontrivial center.

18. (a) Show that any group of order p^n, p a prime, must be nilpotent.
    (b) If G is nilpotent, and $H \neq G$ is a subgroup of G, prove that $N(H) \neq H$ where $N(H) = \{x \in G \mid xHx^{-1} = H\}$.

19. If G is a finite group, prove that G is nilpotent if and only if G is the direct product of its Sylow subgroups.

20. Let G be a finite group and H a subgroup of G. For A, B subgroups of G, define A to be conjugate to B relative to H if $B = x^{-1}Ax$ for some $x \in H$. Prove
    (a) This defines an equivalence relation on the set of subgroups of G.
    (b) The number of subgroups of G conjugate to A relative to H equals the index of $N(A) \cap H$ in H.

21. (a) If G is a finite group and if P is a p-Sylow subgroup of G, prove that P is the only p-Sylow subgroup in $N(P)$.
    (b) If P is a p-Sylow subgroup of G and if $a^{p^k} = e$ then, if $a \in N(P)$, a must be in P.
    (c) Prove that $N(N(P)) = N(P)$.

22. (a) If G is a finite group and P is a p-Sylow subgroup of G, prove that the number of conjugates of P in G is *not* a multiple of p.

(b) Breaking up the conjugate class of P further by using conjugacy relative to P, prove that the conjugate class of P has $1 + kp$ distinct subgroups. (*Hint:* Use part (b) of Problem 20 and Problem 21. Note that together with Problem 23 this gives an alternative proof of Theorem 2.12.3, the third part of Sylow's theorem.)

23. (a) If P is a p-Sylow subgroup of G and B is a subgroup of G of order p^k, prove that if B is not contained in some conjugate of P, then the number of conjugates of P in G is a multiple of p.
    (b) Using part (a) and Problem 22, prove that B must be contained in some conjugate of P.
    (c) Prove that any two p-Sylow subgroups of G are conjugate in G. (This gives another proof of Theorem 2.12.2, the second part of Sylow's theorem.)
24. Combine Problems 22 and 23 to give another proof of all parts of Sylow's theorem.
25. Making a case-by-case discussion using the results developed in this chapter, prove that any group of order less than 60 either is of prime order or has a nontrivial normal subgroup.
26. Using the result of Problem 25, prove that any group of order less than 60 is solvable.
27. Show that the equation $x^2ax = a^{-1}$ is solvable for x in the group G if and only if a is the cube of some element in G.
28. Prove that (1 2 3) is not a cube of any element in S_n.
29. Prove that $xax = b$ is solvable for x in G if and only if ab is the square of some element in G.
30. If G is a group and $a \in G$ is of finite order and has only a finite number of conjugates in G, prove that these conjugates of a generate a finite normal subgroup of G.
31. Show that a group cannot be written as the set-theoretic union of two proper subgroups.
32. Show that a group G is the set-theoretic union of three proper subgroups if and only if G has, as a homomorphic image, a noncyclic group of order 4.

#33. Let p be a prime and let Z_p be the integers mod p under addition and multiplication. Let G be the group $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ where $a, b, c, d \in Z_p$ are such that $ad - bc = 1$. Let

$$C = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \right\}$$

and let $LF(2, p) = G/C$.

(a) Find the order of $LF(2, p)$.
(b) Prove that $LF(2, p)$ is simple if $p \geq 5$.

#34. Prove that $LF(2, 5)$ is isomorphic to A_5, the alternating group of degree 5.

#35. Let $G = LF(2, 7)$; according to Problem 33, G is a simple group of order 168. Determine exactly how many 2-Sylow, 3-Sylow, and 7-Sylow subgroups there are in G.

Supplementary Reading

BURNSIDE, W., *Theory of Groups of Finite Order*, 2nd ed. Cambridge, England: Cambridge University Press, 1911; New York: Dover Publications, 1955.
HALL, MARSHALL, *Theory of Groups*. New York: The Macmillan Company, 1961.

Topics for Class Discussion

ALPERIN, J. L., "A classification of n-abelian groups," *Canadian Journal of Mathematics*, Vol. XXI (1969), pages 1238–1244.
MCKAY, JAMES, H., "Another proof of Cauchy's group theorem," *American Mathematical Monthly*, Vol. 66 (1959), page 119.
SEGAL, I. E., "The automorphisms of the symmetric group," *Bulletin of the American Mathematical Society*, Vol. 46 (1940), page 565.

3

Ring Theory

3.1 Definition and Examples of Rings

As we indicated in Chapter 2, there are certain algebraic systems which serve as the building blocks for the structures comprising the subject which is today called modern algebra. At this stage of the development we have learned something about one of these, namely groups. It is our purpose now to introduce and to study a second such, namely rings. The abstract concept of a group has its origins in the set of mappings, or permutations, of a set onto itself. In contrast, rings stem from another and more familiar source, the set of integers. We shall see that they are patterned after, and are generalizations of, the algebraic aspects of the ordinary integers.

In the next paragraph it will become clear that a ring is quite different from a group in that it is a two-operational system; these operations are usually called addition and multiplication. Yet, despite the differences, the analysis of rings will follow the pattern already laid out for groups. We shall require the appropriate analogs of homomorphism, normal subgroups, factor groups, etc. With the experience gained in our study of groups we shall be able to make the requisite definitions, intertwine them with meaningful theorems, and end up proving results which are both interesting and important about mathematical objects with which we have had long acquaintance. To cite merely one instance, later on in the book, using the tools developed here, we shall prove that it is impossible to trisect an angle of 60° using only a straight-edge and compass.

DEFINITION A nonempty set R is said to be an *associative ring* if in R there are defined two operations, denoted by $+$ and $\cdot$ respectively, such that for all a, b, c in R:

1. $a + b$ is in R.
2. $a + b = b + a$.
3. $(a + b) + c = a + (b + c)$.
4. There is an element 0 in R such that $a + 0 = a$ (for every a in R).
5. There exists an element $-a$ in R such that $a + (-a) = 0$.
6. $a \cdot b$ is in R.
7. $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
8. $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(b + c) \cdot a = b \cdot a + c \cdot a$ (the two distributive laws).

Axioms 1 through 5 merely state that R is an abelian group under the operation $+$, which we call addition. Axioms 6 and 7 insist that R be closed under an associative operation $\cdot$, which we call multiplication. Axiom 8 serves to interrelate the two operations of R.

Whenever we speak of ring it will be understood we mean associative ring. Nonassociative rings, that is, those in which axiom 7 may fail to hold, do occur in mathematics and are studied, but we shall have no occasion to consider them.

It may very well happen, or not happen, that there is an element 1 in R such that $a \cdot 1 = 1 \cdot a = a$ for every a in R; if there is such we shall describe R as a *ring with unit element.*

If the multiplication of R is such that $a \cdot b = b \cdot a$ for every a, b in R, then we call R a *commutative ring.*

Before going on to work out some properties of rings, we pause to examine some examples. Motivated by these examples we shall define various special types of rings which are of importance.

Example 3.1.1 R is the set of integers, positive, negative, and 0; $+$ is the usual addition and $\cdot$ the usual multiplication of integers. R is a commutative ring with unit element.

Example 3.1.2 R is the set of even integers under the usual operations of addition and multiplication. R is a commutative ring but has no unit element.

Example 3.1.3 R is the set of rational numbers under the usual addition and multiplication of rational numbers. R is a commutative ring with unit element. But even more than that, note that the elements of R different from 0 form an abelian group under multiplication. A ring with this latter property is called a *field.*

Example 3.1.4 R is the set of integers mod 7 under the addition and multiplication mod 7. That is, the elements of R are the seven symbols $\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4}, \bar{5}, \bar{6}$, where

1. $\bar{i} + \bar{j} = \bar{k}$ where k is the remainder of $i + j$ on division by 7 (thus, for instance, $\bar{4} + \bar{5} = \bar{2}$ since $4 + 5 = 9$, which, when divided by 7, leaves a remainder of 2).
2. $\bar{i} \cdot \bar{j} = \bar{m}$ where m is the remainder of ij on division by 7 (thus, $\bar{5} \cdot \bar{3} = \bar{1}$ since $5 \cdot 3 = 15$ has 1 as a remainder on division by 7).

The student should verify that R is a commutative ring with unit element. However, much more can be shown; namely, since

$$\bar{1} \cdot \bar{1} = \bar{1} = \bar{6} \cdot \bar{6},$$
$$\bar{2} \cdot \bar{4} = \bar{1} = \bar{4} \cdot \bar{2},$$
$$\bar{3} \cdot \bar{5} = \bar{1} = \bar{5} \cdot \bar{3},$$

the nonzero elements of R form an abelian group under multiplication. R is thus a field. Since it only has a finite number of elements it is called a *finite field.*

Example 3.1.5 R is the set of integers mod 6 under addition and multiplication mod 6. If we denote the elements in R by $\bar{0}, \bar{1}, \bar{2}, \ldots, \bar{5}$, one sees that $\bar{2} \cdot \bar{3} = \bar{0}$, yet $\bar{2} \neq \bar{0}$ and $\bar{3} \neq \bar{0}$. *Thus it is possible in a ring R that $a \cdot b = 0$ with neither $a = 0$ nor $b = 0$.* This cannot happen in a field (see Problem 10, end of Section 3.2), thus the ring R in this example is certainly not a field.

Every example given so far has been a commutative ring. We now present a noncommutative ring.

Example 3.1.6 R will be the set of all symbols

$$\alpha_{11}e_{11} + \alpha_{12}e_{12} + \alpha_{21}e_{21} + \alpha_{22}e_{22} = \sum_{i,j=1}^{2} \alpha_{ij}e_{ij},$$

where all the α_{ij} are rational numbers and where we decree

$$\sum_{i,j=1}^{2} \alpha_{ij}e_{ij} = \sum_{i,j=1}^{2} \beta_{ij}e_{ij} \tag{1}$$

if and only if for all $i, j = 1, 2$, $\alpha_{ij} = \beta_{ij}$,

$$\sum_{i,j=1}^{2} \alpha_{ij}e_{ij} + \sum_{i,j=1}^{2} \beta_{ij}e_{ij} = \sum_{i,j=1}^{2} (\alpha_{ij} + \beta_{ij})e_{ij}. \tag{2}$$

$$\left(\sum_{i,j=1}^{2} \alpha_{ij}e_{ij}\right) \cdot \left(\sum_{i,j=1}^{2} \beta_{ij}e_{ij}\right) = \sum_{i,j=1}^{2} \gamma_{ij}e_{ij}, \tag{3}$$

where

$$\gamma_{ij} = \sum_{v=1}^{2} \alpha_{iv}\beta_{vj} = \alpha_{i1}\beta_{1j} + \alpha_{i2}\beta_{2j}.$$

This multiplication, when first seen, looks rather complicated. However, it is founded on relatively simple rules, namely, multiply $\sum \alpha_{ij}e_{ij}$ by $\sum \beta_{ij}e_{ij}$ formally, multiplying out term by term, and collecting terms, and using the relations $e_{ij} \cdot e_{kl} = 0$ for $j \neq k$, $e_{ij} \cdot e_{jl} = e_{il}$ in this term-by-term collecting. (Of course those of the readers who have already encountered some linear algebra will recognize this example as the ring of all 2×2 matrices over the field of rational numbers.)

To illustrate the multiplication, if $a = e_{11} - e_{21} + e_{22}$ and $b = e_{22} + 3e_{12}$, then

$$\begin{aligned} a \cdot b &= (e_{11} - e_{21} + e_{22}) \cdot (e_{22} + 3e_{12}) \\ &= e_{11} \cdot e_{22} + 3e_{11} \cdot e_{12} - e_{21} \cdot e_{22} - 3e_{21} \cdot e_{12} + e_{22} \cdot e_{22} + 3e_{22} \cdot e_{12} \\ &= 0 + 3e_{12} - 0 - 3e_{22} + e_{22} + 0 \\ &= 3e_{12} - 3e_{22} + e_{22} = 3e_{12} - 2e_{22}. \end{aligned}$$

Note that $e_{11} \cdot e_{12} = e_{12}$ whereas $e_{12} \cdot e_{11} = 0$. Thus the multiplication in R is not commutative. Also it is possible for $u \cdot v = 0$ with $u \neq 0$ and $v \neq 0$.

The student should verify that R is indeed a ring. It is called the ring of 2×2 rational matrices. It, and its relative, will occupy a good deal of our time later on in the book.

Example 3.1.7 Let C be the set of all symbols (α, β) where α, β are real numbers. We define

$$(\alpha, \beta) = (\gamma, \delta) \text{ if and only if } \alpha = \gamma \text{ and } \beta = \delta. \tag{1}$$

In C we introduce an addition by defining for $x = (\alpha, \beta), y = (\gamma, \delta)$

$$x + y = (\alpha, \beta) + (\gamma, \delta) = (\alpha + \gamma, \beta + \delta). \tag{2}$$

Note that $x + y$ is again in C. We assert that C is an abelian group under this operation with $(0, 0)$ serving as the identity element for addition, and $(-\alpha, -\beta)$ as the inverse, under addition, of (α, β).

Now that C is endowed with an addition, in order to make of C a ring we still need a multiplication. We achieve this by defining

$$\text{for } X = (\alpha, \beta), \qquad Y = (\gamma, \delta) \text{ in } C,$$

$$X \cdot Y = (\alpha, \beta) \cdot (\gamma, \delta) = (\alpha\gamma - \beta\delta, \alpha\delta + \beta\gamma). \tag{3}$$

Note that $X \cdot Y = Y \cdot X$. Also $X \cdot (1, 0) = (1, 0) \cdot X = X$ so that $(1, 0)$ is a unit element for C.

Again we notice that $X \cdot Y \in C$. Also, if $X = (\alpha, \beta) \neq (0, 0)$ then, since α, β are real and not both 0, $\alpha^2 + \beta^2 \neq 0$; thus

$$Y = \left(\frac{\alpha}{\alpha^2 + \beta^2}, \frac{-\beta}{\alpha^2 + \beta^2} \right)$$

is in C. Finally we see that

$$(\alpha, \beta) \cdot \left(\frac{\alpha}{\alpha^2 + \beta^2}, \frac{-\beta}{\alpha^2 + \beta^2} \right) = (1, 0).$$

All in all we have shown that C is a field. If we write (α, β) as $\alpha + \beta i$, the reader may verify that C is merely a disguised form of the familiar complex numbers.

Example 3.1.8 This last example is often called the ring of *real quaternions*. This ring was first described by the Irish mathematician Hamilton. Initially it was extensively used in the study of mechanics; today its primary interest is that of an important example, although it still plays key roles in geometry and number theory.

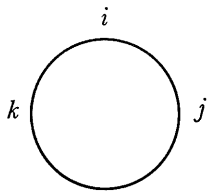
Let Q be the set of all symbols $\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$, where all the numbers α_0, α_1, α_2, and α_3 are real numbers. We declare two such symbols, $\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$ and $\beta_0 + \beta_1 i + \beta_2 j + \beta_3 k$, to be equal if and only if $\alpha_t = \beta_t$ for $t = 0, 1, 2, 3$. In order to make Q into a ring we must define a $+$ and a $\cdot$ for its elements. To this end we define

1. For any $X = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$, $Y = \beta_0 + \beta_1 i + \beta_2 j + \beta_3 k$ in Q, $X + Y = (\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k) + (\beta_0 + \beta_1 i + \beta_2 j + \beta_3 k) = (\alpha_0 + \beta_0) + (\alpha_1 + \beta_1) i + (\alpha_2 + \beta_2) j + (\alpha_3 + \beta_3) k$

and

2. $X \cdot Y = (\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k) \cdot (\beta_0 + \beta_1 i + \beta_2 j + \beta_3 k) = (\alpha_0\beta_0 - \alpha_1\beta_1 - \alpha_2\beta_2 - \alpha_3\beta_3) + (\alpha_0\beta_1 + \alpha_1\beta_0 + \alpha_2\beta_3 - \alpha_3\beta_2) i + (\alpha_0\beta_2 + \alpha_2\beta_0 + \alpha_3\beta_1 - \alpha_1\beta_3) j + (\alpha_0\beta_3 + \alpha_3\beta_0 + \alpha_1\beta_2 - \alpha_2\beta_1) k.$

Admittedly this formula for the product seems rather formidable; however, it looks much more complicated than it actually is. It comes from multiplying out two such symbols formally and collecting terms using the relations $i^2 = j^2 = k^2 = ijk = -1$, $ij = -ji = k$, $jk = -kj = i$, $ki = -ik = j$. The latter part of these relations, called the multiplication table of the quaternion units, can be remembered by the little diagram on page 125. As you go around clockwise you read off the product, e.g., $ij = k$, $jk = i$, $ki = j$; while going around counterclockwise you read off the negatives.



Notice that the elements $\pm 1, \pm i, \pm j, \pm k$ form a non-abelian group of order 8 under this product. In fact, this is the group we called the group of quaternion units in Chapter 2.

The reader may prove that Q is a noncommutative ring in which $0 = 0 + 0i + 0j + 0k$ and $1 = 1 + 0i + 0j + 0k$ serve as the zero and unit elements respectively. Now if $X = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$ is not 0, then not all of $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ are 0; since they are real, $\beta = \alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2 \neq 0$ follows. Thus

$$Y = \frac{\alpha_0}{\beta} - \frac{\alpha_1}{\beta} i - \frac{\alpha_2}{\beta} j - \frac{\alpha_3}{\beta} k \in Q.$$

A simple computation now shows that $X \cdot Y = 1$. Thus the nonzero elements of Q form a non-abelian group under multiplication. A ring in which the nonzero elements form a group is called a *division ring* or *skew-field*. Of course, a commutative division ring is a field. Q affords us a division ring which is not a field. Many other examples of noncommutative division rings exist, but we would be going too far afield to present one here. The investigation of the nature of division rings and the attempts to classify them form an important part of algebra.

3.2 Some Special Classes of Rings

The examples just discussed in Section 3.1 point out clearly that although rings are a direct generalization of the integers, certain arithmetic facts to which we have become accustomed in the ring of integers need not hold in general rings. For instance, we have seen the possibility of $a \cdot b = 0$ with neither a nor b being zero. Natural examples exist where $a \cdot b \neq b \cdot a$. All these run counter to our experience heretofore.

For simplicity of notation we shall henceforth drop the dot in $a \cdot b$ and merely write this product as ab.

DEFINITION If R is a commutative ring, then $a \neq 0 \in R$ is said to be a *zero-divisor* if there exists a $b \in R$, $b \neq 0$, such that $ab = 0$.

DEFINITION A commutative ring is an *integral domain* if it has no zero-divisors.

The ring of integers, naturally enough, is an example of an integral domain.

DEFINITION A ring is said to be a *division ring* if its nonzero elements form a group under multiplication.

The unit element under multiplication will be written as 1, and the inverse of an element a under multiplication will be denoted by a^{-1}.

Finally we make the definition of the ultra-important object known as a field.

DEFINITION A *field* is a commutative division ring.

In our examples in Section 3.1, we exhibited the noncommutative division ring of real quaternions and the following fields: the rational numbers, complex numbers, and the integers mod 7. Chapter 5 will concern itself with fields and their properties.

We wish to be able to compute in rings in much the same manner in which we compute with real numbers, keeping in mind always that there are differences—it may happen that $ab \neq ba$, or that one cannot divide. To this end we prove the next lemma, which asserts that certain things we should like to be true in rings are indeed true.

LEMMA 3.2.1 *If R is a ring, then for all $a, b \in R$*

1. $a0 = 0a = 0$.
2. $a(-b) = (-a)b = -(ab)$.
3. $(-a)(-b) = ab$.

If, in addition, R has a unit element 1, *then*

4. $(-1)a = -a$.
5. $(-1)(-1) = 1$.

Proof.

1. If $a \in R$, then $a0 = a(0 + 0) = a0 + a0$ (using the right distributive law), and since R is a group under addition, this equation implies that $a0 = 0$.

   Similarly, $0a = (0 + 0)a = 0a + 0a$, using the left distributive law, and so here too, $0a = 0$ follows.
2. In order to show that $a(-b) = -(ab)$ we must demonstrate that $ab + a(-b) = 0$. But $ab + a(-b) = a(b + (-b)) = a0 = 0$ by use of

the distributive law and the result of part 1 of this lemma. Similarly $(-a)b = -(ab)$.

3. That $(-a)(-b) = ab$ is really a special case of part 2; we single it out since its analog in the case of real numbers has been so stressed in our early education. So on with it:

$$\begin{aligned}(-a)(-b) &= -(a(-b)) \quad \text{(by part 2)}\\ &= -(-(ab)) \quad \text{(by part 2)}\\ &= ab\end{aligned}$$

since $-(-x) = x$ is a consequence of the fact that in any group $(u^{-1})^{-1} = u$.

4. Suppose that R has a unit element 1; then $a + (-1)a = 1a + (-1)a = (1 + (-1))a = 0a = 0$, whence $(-1)a = -a$. In particular, if $a = -1$, $(-1)(-1) = -(-1) = 1$, which establishes part 5.

With this lemma out of the way we shall, from now on, feel free to compute with negatives and 0 as we always have in the past. The result of Lemma 3.2.1 is our permit to do so. For convenience, $a + (-b)$ will be written $a - b$.

The lemma just proved, while it is very useful and important, is not very exciting. So let us proceed to results of greater interest. Before we do so, we enunciate a principle which, though completely trivial, provides a mighty weapon when wielded properly. This principle says no more or less than the following: if a postman distributes 101 letters to 100 mailboxes then some mailbox must receive at least two letters. It does not sound very promising as a tool, does it? Yet it will surprise us! Mathematical ideas can often be very difficult and obscure, but no such argument can be made against this very simple-minded principle given above. We formalize it and even give it a name.

THE PIGEONHOLE PRINCIPLE *If n objects are distributed over m places, and if $n > m$, then some place receives at least two objects.*

An equivalent formulation, and one which we shall often use is: If n objects are distributed over n places in such a way that no place receives more than one object, then each place receives *exactly* one object.

We immediately make use of this idea in proving

LEMMA 3.2.2 *A finite integral domain is a field.*

Proof. As we may recall, an integral domain is a commutative ring such that $ab = 0$ if and only if at least one of a or b is itself 0. A field, on the other hand, is a commutative ring with unit element in which every nonzero element has a multiplicative inverse in the ring.

Let D be a finite integral domain. In order to prove that D is a field we must

1. Produce an element $1 \in D$ such that $a1 = a$ for every $a \in D$.
2. For every element $a \neq 0 \in D$ produce an element $b \in D$ such that $ab = 1$.

Let $x_1, x_2, \ldots, x_n$ be all the elements of D, and suppose that $a \neq 0 \in D$. Consider the elements $x_1a, x_2a, \ldots, x_na$; they are all in D. We claim that they are all distinct! For suppose that $x_ia = x_ja$ for $i \neq j$; then $(x_i - x_j)a = 0$. Since D is an integral domain and $a \neq 0$, this forces $x_i - x_j = 0$, and so $x_i = x_j$, contradicting $i \neq j$. Thus $x_1a, x_2a, \ldots, x_na$ are n distinct elements lying in D, which has exactly n elements. By the pigeonhole principle these must account for all the elements of D; stated otherwise, every element $y \in D$ can be written as x_ia for some x_i. In particular, since $a \in D$, $a = x_{i_0}a$ for some $x_{i_0} \in D$. Since D is commutative, $a = x_{i_0}a = ax_{i_0}$. We propose to show that x_{i_0} acts as a unit element for every element of D. For, if $y \in D$, as we have seen, $y = x_ia$ for some $x_i \in D$, and so $yx_{i_0} = (x_ia)x_{i_0} = x_i(ax_{i_0}) = x_ia = y$. Thus x_{i_0} is a unit element for D and we write it as 1. Now $1 \in D$, so by our previous argument, it too is realizable as a multiple of a; that is, there exists a $b \in D$ such that $1 = ba$. The lemma is now completely proved.

COROLLARY *If p is a prime number then J_p, the ring of integers* mod p, *is a field.*

Proof. By the lemma it is enough to prove that J_p is an integral domain, since it only has a finite number of elements. If $a, b \in J_p$ and $ab \equiv 0$, then p must divide the ordinary integer ab, and so p, being a prime, must divide a or b. But then either $a \equiv 0 \bmod p$ or $b \equiv 0 \bmod p$, hence in J_p one of these is 0.

The corollary above assures us that we can find an infinity of fields having a finite number of elements. Such fields are called *finite fields*. The fields J_p do not give all the examples of finite fields; there are others. In fact, in Section 7.1 we give a complete description of all finite fields.

We point out a striking difference between finite fields and fields such as the rational numbers, real numbers, or complex numbers, with which we are more familiar.

Let F be a finite field having q elements (if you wish, think of J_p with its p elements). Viewing F merely as a group under addition, since F has q elements, by Corollary 2 to Theorem 2.4.1,

$$\underbrace{a + a + \cdots + a}_{q\text{-times}} = qa = 0$$

for any $a \in F$. Thus, in F, we have $qa = 0$ for some positive integer q, even if $a \neq 0$. This certainly cannot happen in the field of rational numbers, for instance. We formalize this distinction in the definitions we give below. In these definitions, instead of talking just about fields, we choose to widen the scope a little and talk about integral domains.

DEFINITION An integral domain D is said to be of *characteristic* 0 if the relation $ma = 0$, where $a \neq 0$ is in D, and where m is an integer, can hold only if $m = 0$.

The ring of integers is thus of characteristic 0, as are other familiar rings such as the even integers or the rationals.

DEFINITION An integral domain D is said to be of *finite characteristic* if there exists a *positive* integer m such that $ma = 0$ for all $a \in D$.

If D is of finite characteristic, then we define the *characteristic* of D to be the smallest positive integer p such that $pa = 0$ for all $a \in D$. It is not too hard to prove that if D *is of finite characteristic, then its characteristic is a prime number* (see Problem 6 below).

As we pointed out, any finite field is of finite characteristic. However, an integral domain may very well be infinite yet be of finite characteristic (see Problem 7).

One final remark on this question of characteristic: Why define it for integral domains, why not for arbitrary rings? The question is perfectly reasonable. Perhaps the example we give now points out what can happen if we drop the assumption "integral domain."

Let R be the set of all triples (a, b, c), where $a \in J_2$, the integers mod 2, $b \in J_3$, the integers mod 3, and c is any integer. We introduce a $+$ and a $\cdot$ to make of R a ring. We do so by defining $(a_1, b_1, c_1) + (a_2, b_2, c_2) = (a_1 + a_2, b_1 + b_2, c_1 + c_2)$ and $(a_1, b_1, c_1) \cdot (a_2, b_2, c_2) = (a_1a_2, b_1b_2, c_1c_2)$. It is easy to verify that R is a commutative ring. It is not an integral domain since $(1, 2, 0) \cdot (0, 0, 7) = (0, 0, 0)$, the zero-element of R. Note that in R, $2(1, 0, 0) = (1, 0, 0) + (1, 0, 0) = (2, 0, 0) = (0, 0, 0)$ since addition in the first component is in J_2. Similarly $3(0, 1, 0) = (0, 0, 0)$. Finally, for *no* positive integer m is $m(0, 0, 1) = (0, 0, 0)$.

Thus, from the point of view of the definition we gave above for characteristic, the ring R, which we just looked at, is neither fish nor fowl. The definition just doesn't have any meaning for R. We could generalize the notion of characteristic to arbitrary rings by doing it locally, defining it relative to given elements, rather than globally for the ring itself. We say that R has n-torsion, $n > 0$, if there is an element $a \neq 0$ in R such that $na = 0$, and $ma \neq 0$ for $0 < m < n$. For an integral domain D, it turns

out that if D has n-torsion, even for one $n > 0$, then it must be of finite characteristic (see Problem 8).

Problems

R is a ring in all the problems.

1. If $a, b, c, d \in R$, evaluate $(a + b)(c + d)$.
2. Prove that if $a, b \in R$, then $(a + b)^2 = a^2 + ab + ba + b^2$, where by x^2 we mean xx.
3. Find the form of the binomial theorem in a general ring; in other words, find an expression for $(a + b)^n$, where n is a positive integer.
4. If every $x \in R$ satisfies $x^2 = x$, prove that R must be commutative. (A ring in which $x^2 = x$ for all elements is called a *Boolean* ring.)
5. If R is a ring, merely considering it as an abelian group under its addition, we have defined, in Chapter 2, what is meant by na, where $a \in R$ and n is an integer. Prove that if $a, b \in R$ and n, m are integers, then $(na)(mb) = (nm)(ab)$.
6. If D is an integeral domain and D is of finite characteristic, prove that the characteristic of D is a prime number.
7. Give an example of an integral domain which has an infinite number of elements, yet is of finite characteristic.
8. If D is an integral domain and if $na = 0$ for some $a \neq 0$ in D and some integer $n \neq 0$, prove that D is of finite characteristic.
9. If R is a system satisfying all the conditions for a ring with unit element with the possible exception of $a + b = b + a$, prove that the axiom $a + b = b + a$ must hold in R and that R is thus a ring. (*Hint:* Expand $(a + b)(1 + 1)$ in two ways.)
10. Show that the commutative ring D is an integral domain if and only if for $a, b, c \in D$ with $a \neq 0$ the relation $ab = ac$ implies that $b = c$.
11. Prove that Lemma 3.2.2 is false if we drop the assumption that the integral domain is finite.
12. Prove that any field is an integral domain.
13. Useing the pigeonhole principle, prove that if m and n are relatively prime integers and a and b are any integers, there exists an integer x such that $x \equiv a \bmod m$ and $x \equiv b \bmod n$. (*Hint:* Consider the remainders of $a, a + m, a + 2m, \ldots, a + (n - 1)m$ on division by n.)
14. Using the pigeonhole principle, prove that the decimal expansion of a rational number must, after some point, become repeating.

3.3 Homomorphisms

In studying groups we have seen that the concept of a homomorphism turned out to be a fruitful one. This suggests that the appropriate analog for rings could also lead to important ideas. To recall, for groups a homomorphism was defined as a mapping such that $\phi(ab) = \phi(a)\phi(b)$. Since a ring has two operations, what could be a more natural extension of this type of formula than the

DEFINITION A mapping ϕ from the ring R into the ring R' is said to be a *homomorphism* if

1. $\phi(a + b) = \phi(a) + \phi(b)$,
2. $\phi(ab) = \phi(a)\phi(b)$,

for all $a, b \in R$.

As in the case of groups, let us again stress here that the $+$ and $\cdot$ occurring on the left-hand sides of the relations in 1 and 2 are those of R, whereas the $+$ and $\cdot$ occurring on the right-hand sides are those of R'.

A useful observation to make is that a homomorphism of one ring, R, into another, R', is, if we totally ignore the multiplications in both these rings, at least a homomorphism of R into R' when we consider them as abelian groups under their respective additions. Therefore, as far as addition is concerned, all the properties about homomorphisms of groups proved in Chapter 2 carry over. In particular, merely restating Lemma 2.7.2 for the case of the additive group of a ring yields for us

LEMMA 3.3.1 *If ϕ is a homomorphism of R into R', then*

1. $\phi(0) = 0$.
2. $\phi(-a) = -\phi(a)$ *for every* $a \in R$.

A word of caution: if both R and R' have the respective unit elements 1 and 1′ for their multiplications it need not follow that $\phi(1) = 1'$. However, if R' is an integral domain, or if R' is arbitrary but ϕ is onto, then $\phi(1) = 1'$ is indeed true.

In the case of groups, given a homomorphism we associated with this homomorphism a certain subset of the group which we called the kernel of the homomorphism. What should the appropriate definition of the kernel of a homomorphism be for rings? After all, the ring has two operations, addition and multiplication, and it might be natural to ask which of these should be singled out as the basis for the definition. However, the choice is clear. Built into the definition of an arbitrary ring is the condition that the ring forms an abelian group under addition. The ring multiplication

was left much more unrestricted, and so, in a sense, much less under our control than is the addition. For this reason the emphasis is given to the operation of addition in the ring, and we make the

DEFINITION If ϕ is a homomorphism of R into R' then the *kernel of* ϕ, $I(\phi)$, is the set of all elements $a \in R$ such that $\phi(a) = 0$, the zero-element of R'.

LEMMA 3.3.2 *If ϕ is a homomorphism of R into R' with kernel $I(\phi)$, then*

1. *$I(\phi)$ is a subgroup of R under addition.*
2. *If $a \in I(\phi)$ and $r \in R$ then both ar and ra are in $I(\phi)$.*

Proof. Since ϕ is, in particular, a homomorphism of R, as an additive group, into R', as an additive group, (1) follows directly from our results in group theory.

To see (2), suppose that $a \in I(\phi)$, $r \in R$. Then $\phi(a) = 0$ so that $\phi(ar) = \phi(a)\phi(r) = 0\phi(r) = 0$ by Lemma 3.2.1. Similarly $\phi(ra) = 0$. Thus by defining property of $I(\phi)$ both ar and ra are in $I(\phi)$.

Before proceeding we examine these concepts for certain examples.

Example 3.3.1 Let R and R' be two arbitrary rings and define $\phi(a) = 0$ for all $a \in R$. Trivially ϕ is a homomorphism and $I(\phi) = R$. ϕ is called the zero-homomorphism.

Example 3.3.2 Let R be a ring, $R' = R$ and define $\phi(x) = x$ for every $x \in R$. Clearly ϕ is a homomorphism and $I(\phi)$ consists only of 0.

Example 3.3.3 Let $J(\sqrt{2})$ be all real numbers of the form $m + n\sqrt{2}$ where m, n are integers; $J(\sqrt{2})$ forms a ring under the usual addition and multiplication of real numbers. (Verify!) Define $\phi: J(\sqrt{2}) \to J(\sqrt{2})$ by $\phi(m + n\sqrt{2}) = m - n\sqrt{2}$. ϕ is a homomorphism of $J(\sqrt{2})$ onto $J(\sqrt{2})$ and its kernel $I(\phi)$, consists only of 0. (Verify!)

Example 3.3.4 Let J be the ring of integers, J_n, the ring of integers modulo n. Define $\phi: J \to J_n$ by $\phi(a)$ = remainder of a on division by n. The student should verify that ϕ is a homomorphism of J onto J_n and that the kernel, $I(\phi)$, of ϕ consists of all multiples of n.

Example 3.3.5 Let R be the set of all continuous, real-valued functions on the closed unit interval. R is made into a ring by the usual addition and multiplication of functions; that it is a ring is a consequence of the fact that the sum and product of two continuous functions are continuous

functions. Let F be the ring of real numbers and define $\phi: R \to F$ by $\phi(f(x)) = f(\frac{1}{2})$. ϕ is then a homomorphism of R onto F and its kernel consists of all functions in R vanishing at $x = \frac{1}{2}$.

All the examples given here have used commutative rings. Many beautiful examples exist where the rings are noncommutative but it would be premature to discuss such an example now.

DEFINITION A homomorphism of R into R' is said to be an *isomorphism* if it is a one-to-one mapping.

DEFINITION Two rings are said to be *isomorphic* if there is an isomorphism of one *onto* the other.

The remarks made in Chapter 2 about the meaning of an isomorphism and of the statement that two groups are isomorphic carry over verbatim to rings. Likewise, the criterion given in Lemma 2.7.4 that a homomorphism be an isomorphism translates directly from groups to rings in the form

LEMMA 3.3.3 *The homomorphism ϕ of R into R' is an isomorphism if and only if $I(\phi) = (0)$.*

3.4 Ideals and Quotient Rings

Once the idea of a homomorphism and its kernel have been set up for rings, based on our experience with groups, it should be fruitful to carry over some analog to rings of the concept of normal subgroup. Once this is achieved, one would hope that this analog would lead to a construction in rings like that of the quotient group of a group by a normal subgroup. Finally, if one were an optimist, one would hope that the homomorphism theorems for groups would come over in their entirety to rings.

Fortunately all this can be done, thereby providing us with an incisive technique for analyzing rings.

The first business at hand, then, seems to be to define a suitable "normal subgroup" concept for rings. With a little hindsight this is not difficult. If you recall, normal subgroups eventually turned out to be nothing else than kernels of homomorphisms, even though their primary defining conditions did not involve homomorphisms. Why not use this observation as the keystone to our definition for rings? Lemma 3.3.2 has already provided us with some conditions that a subset of a ring be the kernel of a homomorphism. We now take the point of view that, since no other information is at present available to us, we shall make the conclusions of Lemma 3.3.2 as the starting point of our endeavor, and so we define

DEFINITION A nonempty subset U of R is said to be a (two-sided) *ideal* of R if

1. U is a subgroup of R under addition.
2. For every $u \in U$ and $r \in R$, both ur and ru are in U.

Condition 2 asserts that U "swallows up" multiplication from the right and left by arbitrary ring elements. For this reason U is usually called a two-sided ideal. Since we shall have no occasion, other than in some of the problems, to use any other derivative concept of ideal, we shall merely use the word ideal, rather than two-sided ideal, in all that follows.

Given an ideal U of a ring R, let R/U be the set of all the distinct cosets of U in R which we obtain by considering U as a subgroup of R under addition. We note that we merely say coset, rather than right coset or left coset; this is justified since R is an abelian group under addition. To restate what we have just said, R/U consists of all the cosets, $a + U$, where $a \in R$. By the results of Chapter 2, R/U is automatically a group under addition; this is achieved by the composition law $(a + U) + (b + U) = (a + b) + U$. In order to impose a ring structure on R/U we must define, in it, a multiplication. What is more natural than to define $(a + U)(b + U) = ab + U$? However, we must make sure that this is meaningful. Otherwise put, we are obliged to show that if $a + U = a' + U$ and $b + U = b' + U$, then under our definition of the multiplication, $(a + U)(b + U) = (a' + U)(b' + U)$. Equivalently, it must be established that $ab + U = a'b' + U$. To this end we first note that since $a + U = a' + U$, $a = a' + u_1$, where $u_1 \in U$; similarly $b = b' + u_2$ where $u_2 \in U$. Thus $ab = (a' + u_1)(b + u_2) = a'b' + u_1b' + a'u_2 + u_1u_2$; *since U is an ideal of R*, $u_1b' \in U$, $a'u_2 \in U$, and $u_1u_2 \in U$. Consequently $u_1b' + a'u_2 + u_1u_2 = u_3 \in U$. But then $ab = a'b' + u_3$, from which we deduce that $ab + U = a'b' + u_3 + U$, and since $u_3 \in U$, $u_3 + U = U$. The net consequence of all this is that $ab + U = a'b' + U$. We at least have achieved the principal step on the road to our goal, namely of introducing a well-defined multiplication. The rest now becomes routine. To establish that R/U is a ring we merely have to go through the various axioms which define a ring and check whether they hold in R/U. All these verifications have a certain sameness to them, so we pick one axiom, the right distributive law, and prove it holds in R/U. The rest we leave to the student as informal exercises. If $X = a + U$, $Y = b + U$, $Z = c + U$ are three elements of R/U, where $a, b, c \in R$, then $(X + Y)Z = ((a + U) + (b + U))(c + U) = ((a + b) + U)(c + U) = (a + b)c + U = ac + bc + U = (ac + U) + (bc + U) = (a + U)(c + U) + (b + U)(c + U) = XZ + YZ$.

R/U has now been made into a ring. Clearly, if R is commutative then so is R/U, for $(a + U)(b + U) = ab + U = ba + U = (b + U)(a + U)$. (The converse to this is false.) If R has a unit element 1, then R/U has a

unit element $1 + U$. We might ask: In what relation is R/U to R? With the experience we now have in hand this is easy to answer. There is a homomorphism ϕ of R *onto* R/U given by $\phi(a) = a + U$ for every $a \in R$, whose kernel is exactly U. (The reader should verify that ϕ so defined is a homomorphism of R onto R/U with kernel U.)

We summarize these remarks in

LEMMA 3.4.1 *If U is an ideal of the ring R, then R/U is a ring and is a homomorphic image of R.*

With this construction of the *quotient ring* of a ring by an ideal satisfactorily accomplished, we are ready to bring over to rings the homomorphism theorems of groups. Since the proof is an exact verbatim translation of that for groups into the language of rings we merely state the theorem without proof, referring the reader to Chapter 2 for the proof.

THEOREM 3.4.1 *Let R, R' be rings and ϕ a homomorphism of R onto R' with kernel U. Then R' is isomorphic to R/U. Moreover there is a one-to-one correspondence between the set of ideals of R' and the set of ideals of R which contain U. This correspondence can be achieved by associating with an ideal W' in R' the ideal W in R defined by $W = \{x \in R \mid \phi(x) \in W'\}$. With W so defined, R/W is isomorphic to R'/W'.*

Problems

1. If U is an ideal of R and $1 \in U$, prove that $U = R$.
2. If F is a field, prove its only ideals are (0) and F itself.
3. Prove that any homomorphism of a field is either an isomorphism or takes each element into 0.
4. If R is a commutative ring and $a \in R$,
   (a) Show that $aR = \{ar \mid r \in R\}$ is a two-sided ideal of R.
   (b) Show by an example that this may be false if R is not commutative.
5. If U, V are ideals of R, let $U + V = \{u + v \mid u \in U, v \in V\}$. Prove that $U + V$ is also an ideal.
6. If U, V are ideals of R let UV be the set of all elements that can be written as finite sums of elements of the form uv where $u \in U$ and $v \in V$. Prove that UV is an ideal of R.
7. In Problem 6 prove that $UV \subset U \cap V$.
8. If R is the ring of integers, let U be the ideal consisting of all multiples of 17. Prove that if V is an ideal of R and $R \supset V \supset U$ then either $V = R$ or $V = U$. Generalize!

9. If U is an ideal of R, let $r(U) = \{x \in R \mid xu = 0 \text{ for all } u \in U\}$. Prove that $r(U)$ is an ideal of R.
10. If U is an ideal of R let $[R:U] = \{x \in R \mid rx \in U \text{ for every } r \in R\}$. Prove that $[R:U]$ is an ideal of R and that it contains U.
11. Let R be a ring with unit element. Using its elements we define a ring $\tilde{R}$ by defining $a \oplus b = a + b + 1$, and $a \cdot b = ab + a + b$, where $a, b \in R$ and where the addition and multiplication on the right-hand side of these relations are those of R.
    (a) Prove that $\tilde{R}$ is a ring under the operations $\oplus$ and $\cdot$.
    (b) What acts as the zero-element of $\tilde{R}$?
    (c) What acts as the unit-element of $\tilde{R}$?
    (d) Prove that R is isomorphic to $\tilde{R}$.

*12. In Example 3.1.6 we discussed the ring of rational 2×2 matrices. Prove that this ring has no ideals other than (0) and the ring itself.

*13. In Example 3.1.8 we discussed the real quaternions. Using this as a model we define the quaternions over the integers mod p, p an odd prime number, in exactly the same way; however, now considering all symbols of the form $\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$, where $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ are integers mod p.
    (a) Prove that this is a ring with p^4 elements whose only ideals are (0) and the ring itself.
    **(b) Prove that this ring is *not* a division ring.

If R is any ring a subset L of R is called a *left-ideal* of R if
1. L is a subgroup of R under addition.
2. $r \in R$, $a \in L$ implies $ra \in L$.

(One can similarly define a *right-ideal.*) An ideal is thus simultaneously a left- and right-ideal of R.

14. For $a \in R$ let $Ra = \{xa \mid x \in R\}$. Prove that Ra is a left-ideal of R.
15. Prove that the intersection of two left-ideals of R is a left-ideal of R.
16. What can you say about the intersection of a left-ideal and right-ideal of R?
17. If R is a ring and $a \in R$ let $r(a) = \{x \in R \mid ax = 0\}$. Prove that $r(a)$ is a right-ideal of R.
18. If R is a ring and L is a left-ideal of R let $\lambda(L) = \{x \in R \mid xa = 0 \text{ for all } a \in L\}$. Prove that $\lambda(L)$ is a two-sided ideal of R.

*19. Let R be a ring in which $x^3 = x$ for every $x \in R$. Prove that R is a commutative ring.

20. If R is a ring with unit element 1 and ϕ is a homomorphism of R *onto* R' prove that $\phi(1)$ is the unit element of R'.

21. If R is a ring with unit element 1 and ϕ is a homomorphism of R into an integral domain R' such that $I(\phi) \neq R$, prove that $\phi(1)$ is the unit element of R'.

3.5 More Ideals and Quotient Rings

We continue the discussion of ideals and quotient rings.

Let us take the point of view, for the moment at least, that a field is the most desirable kind of ring. Why? If for no other reason, we can divide in a field, so operations and results in a field more closely approximate our experience with real and complex numbers. In addition, as was illustrated by Problem 2 in the preceding problem set, a field has no homomorphic images other than itself or the trivial ring consisting of 0. Thus we cannot simplify a field by applying a homomorphism to it. Taking these remarks into consideration it is natural that we try to link a general ring, in some fashion, with fields. What should this linkage involve? We have a machinery whose component parts are homomorphisms, ideals, and quotient rings. With these we will forge the link.

But first we must make precise the rather vague remarks of the preceding paragraph. We now ask the explicit question: Under what conditions is the homomorphic image of a ring a field? For commutative rings we give a complete answer in this section.

Essential to treating this question is the converse to the result of Problem 2 of the problem list at the end of Section 3.4.

LEMMA 3.5.1 *Let R be a commutative ring with unit element whose only ideals are (0) and R itself. Then R is a field.*

Proof. In order to effect a proof of this lemma for any $a \neq 0 \in R$ we must produce an element $b \neq 0 \in R$ such that $ab = 1$.

So, suppose that $a \neq 0$ is in R. Consider the set $Ra = \{xa \mid x \in R\}$. We claim that Ra is an ideal of R. In order to establish this as fact we must show that it is a subgroup of R under addition and that if $u \in Ra$ and $r \in R$ then ru is also in Ra. (We only need to check that ru is in Ra for then ur also is since $ru = ur$.)

Now, if $u, v \in Ra$, then $u = r_1a$, $v = r_2a$ for some $r_1, r_2 \in R$. Thus $u + v = r_1a + r_2a = (r_1 + r_2)a \in Ra$; similarly $-u = -r_1a = (-r_1)a \in Ra$. Hence Ra is an additive subgroup of R. Moreover, if $r \in R$, $ru = r(r_1a) = (rr_1)a \in Ra$. Ra therefore satisfies all the defining conditions for an ideal of R, hence is an ideal of R. (Notice that both the distributive law and associative law of multiplication were used in the proof of this fact.)

By our assumptions on R, $Ra = (0)$ or $Ra = R$. Since $0 \neq a = 1a \in Ra$, $Ra \neq (0)$; thus we are left with the only other possibility, namely that $Ra = R$. This last equation states that every element in R is a multiple of

a by some element of R. In particular, $1 \in R$ and so it can be realized as a multiple of a; that is, there exists an element $b \in R$ such that $ba = 1$. This completes the proof of the lemma.

DEFINITION An ideal $M \neq R$ in a ring R is said to be a *maximal ideal* of R if whenever U is an ideal of R such that $M \subset U \subset R$, then either $R = U$ or $M = U$.

In other words, an ideal of R is a maximal ideal if it is impossible to squeeze an ideal between it and the full ring. Given a ring R there is no guarantee that it has any maximal ideals! If the ring has a unit element this can be proved, assuming a basic axiom of mathematics, the so-called axiom of choice. Also there may be many distinct maximal ideals in a ring R; this will be illustrated for us below in the ring of integers.

As yet we have made acquaintance with very few rings. Only by considering a given concept in many particular cases can one fully appreciate the concept and its motivation. Before proceeding we therefore examine some maximal ideals in two specific rings. When we come to the discussion of polynomial rings we shall exhibit there all the maximal ideals.

Example 3.5.1 Let R be the ring of integers, and let U be an ideal of R. Since U is a subgroup of R under addition, from our results in group theory, we know that U consists of all the multiples of a fixed integer n_0; we write this as $U = (n_0)$. What values of n_0 lead to maximal ideals?

We first assert that if p is a prime number then $P = (p)$ is a maximal ideal of R. For if U is an ideal of R and $U \supset P$, then $U = (n_0)$ for some integer n_0. Since $p \in P \subset U$, $p = mn_0$ for some integer m; because p is a prime this implies that $n_0 = 1$ or $n_0 = p$. If $n_0 = p$, then $P \subset U = (n_0) \subset P$, so that $U = P$ follows; if $n_0 = 1$, then $1 \in U$, hence $r = 1r \in U$ for all $r \in R$ whence $U = R$ follows. Thus no ideal, other than R or P itself, can be put between P and R, from which we deduce that P is maximal.

Suppose, on the other hand, that $M = (n_0)$ is a maximal ideal of R. We claim that n_0 must be a prime number, for if $n_0 = ab$, where a, b are positive integers, then $U = (a) \supset M$, hence $U = R$ or $U = M$. If $U = R$, then $a = 1$ is an easy consequence; if $U = M$, then $a \in M$ and so $a = rn_0$ for some integer r, since every element of M is a multiple of n_0. But then $n_0 = ab = rn_0 b$, from which we get that $rb = 1$, so that $b = 1$, $n_0 = a$. Thus n_0 is a prime number.

In this particular example the notion of maximal ideal comes alive—it corresponds exactly to the notion of prime number. One should not, however, jump to any hasty generalizations; this kind of correspondence does not usually hold for more general rings.

Example 3.5.2 Let R be the ring of all the real-valued, continuous functions on the closed unit interval. (See Example 3.3.5.) Let

$$M = \{f(x) \in R \mid f(\tfrac{1}{2}) = 0\}.$$

M is certainly an ideal of R. Moreover, it is a maximal ideal of R, for if the ideal U contains M and $U \neq M$, then there is a function $g(x) \in U$, $g(x) \notin M$. Since $g(x) \notin M$, $g(\tfrac{1}{2}) = \alpha \neq 0$. Now $h(x) = g(x) - \alpha$ is such that $h(\tfrac{1}{2}) = g(\tfrac{1}{2}) - \alpha = 0$, so that $h(x) \in M \subset U$. But $g(x)$ is also in U; therefore $\alpha = g(x) - h(x) \in U$ and so $1 = \alpha\alpha^{-1} \in U$. Thus for any function $t(x) \in R$, $t(x) = 1t(x) \in U$, in consequence of which $U = R$. M is therefore a maximal ideal of R. Similarly if γ is a real number $0 \leq \gamma \leq 1$, then $M_\gamma = \{f(x) \in R \mid f(\gamma) = 0\}$ is a maximal ideal of R. It can be shown (see Problem 4 at the end of this section) that every maximal ideal is of this form. Thus here the maximal ideals correspond to the points on the unit interval.

Having seen some maximal ideals in some concrete rings we are ready to continue the general development with

THEOREM 3.5.1 *If R is a commutative ring with unit element and M is an ideal of R, then M is a maximal ideal of R if and only if R/M is a field.*

Proof. Suppose, first, that M is an ideal of R such that R/M is a field. Since R/M is a field its only ideals are (0) and R/M itself. But by Theorem 3.4.1 there is a one-to-one correspondence between the set of ideals of R/M and the set of ideals of R which contain M. The ideal M of R corresponds to the ideal (0) of R/M whereas the ideal R of R corresponds to the ideal R/M of R/M in this one-to-one mapping. Thus there is no ideal between M and R other than these two, whence M is a maximal ideal.

On the other hand, if M is a maximal ideal of R, by the correspondence mentioned above R/M has only (0) and itself as ideals. Furthermore R/M is commutative and has a unit element since R enjoys both these properties. All the conditions of Lemma 3.5.1 are fulfilled for R/M so we can conclude, by the result of that lemma, that R/M is a field.

We shall have many occasions to refer back to this result in our study of polynomial rings and in the theory of field extensions.

Problems

1. Let R be a ring with unit element, R not necessarily commutative, such that the only right-ideals of R are (0) and R. Prove that R is a division ring.

*2. Let R be a ring such that the only right ideals of R are (0) and R. Prove that either R is a division ring or that R is a ring with a prime number of elements in which $ab = 0$ for every $a, b \in R$.

3. Let J be the ring of integers, p a prime number, and (p) the ideal of J consisting of all multiples of p. Prove
   (a) $J/(p)$ is isomorphic to J_p, the ring of integers mod p.
   (b) Using Theorem 3.5.1 and part (a) of this problem, that J_p is a field.

**4. Let R be the ring of all real-valued continuous functions on the closed unit interval. If M is a maximal ideal of R, prove that there exists a real number γ, $0 \leq \gamma \leq 1$, such that $M = M_\gamma = \{f(x) \in R \mid f(\gamma) = 0\}$.

3.6 The Field of Quotients of an Integral Domain

Let us recall that an integral domain is a commutative ring D with the additional property that it has no zero-divisors, that is, if $ab = 0$ for some $a, b \in D$ then at least one of a or b must be 0. The ring of integers is, of course, a standard example of an integral domain.

The ring of integers has the attractive feature that we can enlarge it to the set of rational numbers, which is a field. Can we perform a similar construction for any integral domain? We will now proceed to show that indeed we can!

DEFINITION A ring R *can be imbedded* in a ring R' if there is an isomorphism of R into R'. (If R and R' have unit elements 1 and 1' we insist, in addition, that this isomorphism takes 1 onto 1'.)

R' will be called an *over-ring* or *extension* of R if R can be imbedded in R'.

With this understanding of imbedding we prove

THEOREM 3.6.1 *Every integral domain can be imbedded in a field.*

Proof. Before becoming explicit in the details of the proof let us take an informal approach to the problem. Let D be our integral domain; roughly speaking the field we seek should be all quotients a/b, where $a, b \in D$ and $b \neq 0$. Of course in D, a/b may very well be meaningless. What should we require of these symbols a/b? Clearly we must have an answer to the following three questions:

1. When is $a/b = c/d$?
2. What is $(a/b) + (c/d)$?
3. What is $(a/b)(c/d)$?

In answer to 1, what could be more natural than to insist that $a/b = c/d$

if and only if $ad = bc$? As for 2 and 3, why not try the obvious, that is, define

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} \quad \text{and} \quad \frac{a}{b}\frac{c}{d} = \frac{ac}{bd}.$$

In fact in what is to follow we make these considerations our guide. So let us leave the heuristics and enter the domain of mathematics, with precise definitions and rigorous deductions.

Let $\mathscr{M}$ be the set of all ordered pairs (a, b) where $a, b \in D$ and $b \neq 0$. (Think of (a, b) as a/b.) In $\mathscr{M}$ we now define a relation as follows:

$$(a, b) \sim (c, d) \text{ if and only if } ad = bc.$$

We claim that this defines an equivalence relation on $\mathscr{M}$. To establish this we check the three defining conditions for an equivalence relation for this particular relation.

1. If $(a, b) \in \mathscr{M}$, then $(a, b) \sim (a, b)$ since $ab = ba$.
2. If $(a, b), (c, d) \in \mathscr{M}$ and $(a, b) \sim (c, d)$, then $ad = bc$, hence $cb = da$, and so $(c, d) \sim (a, b)$.
3. If $(a, b), (c, d), (e, f)$ are all in $\mathscr{M}$ and $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$, then $ad = bc$ and $cf = de$. Thus $bcf = bde$, and since $bc = ad$, it follows that $adf = bde$. Since D is commutative, this relation becomes $afd = bed$; since, moreover, D is an integral domain and $d \neq 0$, this relation further implies that $af = be$. But then $(a, b) \sim (e, f)$ and our relation is transitive.

Let $[a, b]$ be the equivalence class in $\mathscr{M}$ of (a, b), and let F be the set of all such equivalence classes $[a, b]$ where $a, b \in D$ and $b \neq 0$. F is the candidate for the field we are seeking. In order to create out of F a field we must introduce an addition and a multiplication for its elements and then show that under these operations F forms a field.

We first dispose of the addition. Motivated by our heuristic discussion at the beginning of the proof we define

$$[a, b] + [c, d] = [ad + bc, bd].$$

Since D is an integral domain and both $b \neq 0$ and $d \neq 0$ we have that $bd \neq 0$; this, at least, tells us that $[ad + bc, bd] \in F$. We now assert that this addition is well defined, that is, if $[a, b] = [a', b']$ and $[c, d] = [c', d']$, then $[a, b] + [c, d] = [a', b'] + [c', d']$. To see that this is so, from $[a, b] = [a', b']$ we have that $ab' = ba'$; from $[c, d] = [c', d']$ we have that $cd' = dc'$. What we need is that these relations force the equality of $[a, b] + [c, d]$ and $[a', b'] + [c', d']$. From the definition of addition this boils down to showing that $[ad + bc, bd] = [a'd' + b'c', b'd']$, or, in equivalent terms, that $(ad + bc)b'd' = bd(a'd' + b'c')$. Using $ab' = ba'$, $cd' = dc'$

this becomes: $(ad + bc)b'd' = adb'd' + bcb'd' = ab'dd' + bb'cd' = ba'dd' + bb'dc' = bd(a'd' + b'c')$, which is the desired equality.

Clearly $[0, b]$ acts as a zero-element for this addition and $[-a, b]$ as the negative of $[a, b]$. It is a simple matter to verify that F is an abelian group under this addition.

We now turn to the multiplication in F. Again motivated by our preliminary heuristic discussion we define $[a, b][c, d] = [ac, bd]$. As in the case of addition, since $b \neq 0$, $d \neq 0$, $bd \neq 0$ and so $[ac, bd] \in F$. A computation, very much in the spirit of the one just carried out, proves that if $[a, b] = [a', b']$ and $[c, d] = [c', d']$ then $[a, b][c, d] = [a', b'][c', d']$. One can now show that the nonzero elements of F (that is, all the elements $[a, b]$ where $a \neq 0$) form an abelian group under multiplication in which $[d, d]$ acts as the unit element and where

$$[c, d]^{-1} = [d, c] \text{ (since } c \neq 0, [d, c] \text{ is in } F).$$

It is a routine computation to see that the distributive law holds in F. F is thus a field.

All that remains is to show that D can be imbedded in F. We shall exhibit an explicit isomorphism of D into F. Before doing so we first notice that for $x \neq 0$, $y \neq 0$ in D, $[ax, x] = [ay, y]$ because $(ax)y = x(ay)$; let us denote $[ax, x]$ by $[a, 1]$. Define $\phi: D \to F$ by $\phi(a) = [a, 1]$ for every $a \in D$. We leave it to the reader to verify that ϕ is an isomorphism of D into F, and that if D has a unit element 1, then $\phi(1)$ is the unit element of F. The theorem is now proved in its entirety.

F is usually called the *field of quotients* of D. In the special case in which D is the ring of integers, the F so constructed is, of course, the field of rational numbers.

Problems

1. Prove that if $[a, b] = [a', b']$ and $[c, d] = [c', d']$ then $[a, b][c, d] = [a', b'][c', d']$.
2. Prove the distributive law in F.
3. Prove that the mapping $\phi: D \to F$ defined by $\phi(a) = [a, 1]$ is an isomorphism of D into F.
4. Prove that if K is any field which contains D then K contains a subfield isomorphic to F. (*In this sense F is the smallest field containing D.*)

*5. Let R be a commutative ring with unit element. A nonempty subset S of R is called a multiplicative system if
   1. $0 \notin S$.
   2. $s_1, s_2 \in S$ implies that $s_1 s_2 \in S$.

Let $\mathcal{M}$ be the set of all ordered pairs (r, s) where $r \in R$, $s \in S$. In $\mathcal{M}$ define $(r, s) \sim (r', s')$ if there exists an element $s'' \in S$ such that

$$s''(rs' - sr') = 0.$$

(a) Prove that this defines an equivalence relation on $\mathcal{M}$.

Let the equivalence class of (r, s) be denoted by $[r, s]$, and let R_S be the set of all the equivalence classes. In R_S define $[r_1, s_1] + [r_2, s_2] = [r_1s_2 + r_2s_1, s_1s_2]$ and $[r_1, s_1][r_2, s_2] = [r_1r_2, s_1s_2]$.

(b) Prove that the addition and multiplication described above are well defined and that R_S forms a ring under these operations.

(c) Can R be imbedded in R_S?

(d) Prove that the mapping $\phi: R \to R_s$ defined by $\phi(a) = [as, s]$ is a homomorphism of R into R_S and find the kernel of ϕ.

(e) Prove that this kernel has no element of S in it.

(f) Prove that every element of the form $[s_1, s_2]$(where $s_1, s_2 \in S$) in R_S has an inverse in R_S.

6. Let D be an integral domain, $a, b \in D$. Suppose that $a^n = b^n$ and $a^m = b^m$ for two relatively prime positive integers m and n. Prove that $a = b$.

7. Let R be a ring, possibly noncommutative, in which $xy = 0$ implies $x = 0$ or $y = 0$. If $a, b \in R$ and $a^n = b^n$ and $a^m = b^m$ for two relatively prime positive integers m and n, prove that $a = b$.

3.7 Euclidean Rings

The class of rings we propose to study now is motivated by several existing examples—the ring of integers, the Gaussian integers (Section 3.8), and polynomial rings (Section 3.9). The definition of this class is designed to incorporate in it certain outstanding characteristics of the three concrete examples listed above.

DEFINITION An integral domain R is said to be a *Euclidean ring* if for every $a \neq 0$ in R there is defined a nonnegative integer $d(a)$ such that

1. For all $a, b \in R$, both nonzero, $d(a) \leq d(ab)$.
2. For any $a, b \in R$, both nonzero, there exist $t, r \in R$ such that $a = tb + r$ where either $r = 0$ or $d(r) < d(b)$.

We do not assign a value to $d(0)$. The integers serve as an example of a Euclidean ring, where $d(a) =$ absolute value of a acts as the required function. In the next section we shall see that the Gaussian integers also form a Euclidean ring. Out of that observation, and the results developed in this part, we shall prove a classic theorem in number theory due to

Fermat, namely, that every prime number of the form $4n + 1$ can be written as the sum of two squares.

We begin with

THEOREM 3.7.1 *Let R be a Euclidean ring and let A be an ideal of R. Then there exists an element $a_0 \in A$ such that A consists exactly of all a_0x as x ranges over R.*

Proof. If A just consists of the element 0, put $a_0 = 0$ and the conclusion of the theorem holds.

Thus we may assume that $A \neq (0)$; hence there is an $a \neq 0$ in A. Pick an $a_0 \in A$ such that $d(a_0)$ is minimal. (Since d takes on nonnegative integer values this is always possible.)

Suppose that $a \in A$. By the properties of Euclidean rings there exist $t, r \in R$ such that $a = ta_0 + r$ where $r = 0$ or $d(r) < d(a_0)$. Since $a_0 \in A$ and A is an ideal of R, ta_0 is in A. Combined with $a \in A$ this results in $a - ta_0 \in A$; but $r = a - ta_0$, whence $r \in A$. If $r \neq 0$ then $d(r) < d(a_0)$, giving us an element r in A whose d-value is smaller than that of a_0, in contradiction to our choice of a_0 as the element in A of minimal d-value. Consequently $r = 0$ and $a = ta_0$, which proves the theorem.

We introduce the notation $(a) = \{xa \mid x \in R\}$ to represent the ideal of all multiples of a.

DEFINITION An integral domain R with unit element is a *principal ideal ring* if every ideal A in R is of the form $A = (a)$ for some $a \in R$.

Once we establish that a Euclidean ring has a unit element, in virtue of Theorem 3.7.1, we shall know that a Euclidean ring is a principal ideal ring. The converse, however, is false; there are principal ideal rings which are not Euclidean rings. [See the paper by T. Motzkin, *Bulletin of the American Mathematical Society*, Vol. 55 (1949), pages 1142–1146, entitled "The Euclidean algorithm."]

COROLLARY TO THEOREM 3.7.1 *A Euclidean ring possesses a unit element.*

Proof. Let R be a Euclidean ring; then R is certainly an ideal of R, so that by Theorem 3.7.1 we may conclude that $R = (u_0)$ for some $u_0 \in R$. Thus every element in R is a multiple of u_0. Therefore, in particular, $u_0 = u_0c$ for some $c \in R$. If $a \in R$ then $a = xu_0$ for some $x \in R$, hence $ac = (xu_0)c = x(u_0c) = xu_0 = a$. Thus c is seen to be the required unit element.

DEFINITION If $a \neq 0$ and b are in a commutative ring R then a is said to *divide* b if there exists a $c \in R$ such that $b = ac$. We shall use the symbol

$a \mid b$ to represent the fact that a divides b and $a \nmid b$ to mean that a does not divide b.

The proof of the next remark is so simple and straightforward that we omit it.

REMARK 1. *If* $a \mid b$ *and* $b \mid c$ *then* $a \mid c$.
2. *If* $a \mid b$ *and* $a \mid c$ *then* $a \mid (b \pm c)$.
3. *If* $a \mid b$ *then* $a \mid bx$ *for all* $x \in R$.

DEFINITION If $a, b \in R$ then $d \in R$ is said to be a *greatest common divisor* of a and b if

1. $d \mid a$ and $d \mid b$.
2. Whenever $c \mid a$ and $c \mid b$ then $c \mid d$.

We shall use the notation $d = (a, b)$ to denote that d is a greatest common divisor of a and b.

LEMMA 3.7.1 *Let R be a Euclidean ring. Then any two elements a and b in R have a greatest common divisor d. Moreover $d = \lambda a + \mu b$ for some $\lambda, \mu \in R$.*

Proof. Let A be the set of all elements $ra + sb$ where r, s range over R. We claim that A is an ideal of R. For suppose that $x, y \in A$; therefore $x = r_1a + s_1b$, $y = r_2a + s_2b$, and so $x \pm y = (r_1 \pm r_2)a + (s_1 \pm s_2)b \in A$. Similarly, for any $u \in R$, $ux = u(r_1a + s_1b) = (ur_1)a + (us_1)b \in A$.

Since A is an ideal of R, by Theorem 3.7.1 there exists an element $d \in A$ such that every element in A is a mutiple of d. By dint of the fact that $d \in A$ and that every element of A is of the form $ra + sb$, $d = \lambda a + \mu b$ for some $\lambda, \mu \in R$. Now by the corollary to Theorem 3.7.1, R has a unit element 1; thus $a = 1a + 0b \in A$, $b = 0a + 1b \in A$. Being in A, they are both multiples of d, whence $d \mid a$ and $d \mid b$.

Suppose, finally, that $c \mid a$ and $c \mid b$; then $c \mid \lambda a$ and $c \mid \mu b$ so that c certainly divides $\lambda a + \mu b = d$. Therefore d has all the requisite conditions for a greatest common divisor and the lemma is proved.

DEFINITION Let R be a commutative ring with unit element. An element $a \in R$ is a *unit* in R if there exists an element $b \in R$ such that $ab = 1$.

Do not confuse a unit with a unit element! A unit in a ring is an element whose inverse is also in the ring.

LEMMA 3.7.2 *Let R be an integral domain with unit element and suppose that for $a, b \in R$ both $a \mid b$ and $b \mid a$ are true. Then $a = ub$, where u is a unit in R.*

Proof. Since $a \mid b$, $b = xa$ for some $x \in R$; since $b \mid a$, $a = yb$ for some $y \in R$. Thus $b = x(yb) = (xy)b$; but these are elements of an integral domain, so that we can cancel the b and obtain $xy = 1$; y is thus a unit in R and $a = yb$, proving the lemma.

DEFINITION Let R be a commutative ring with unit element. Two elements a and b in R are said to be *associates* if $b = ua$ for some unit u in R.

The relation of being associates is an equivalence relation. (Problem 1 at the end of this section.) Note that in a Euclidean ring any two greatest common divisors of two given elements are associates (Problem 2).

Up to this point we have, as yet, not made use of condition 1 in the definition of a Euclidean ring, namely that $d(a) \leq d(ab)$ for $b \neq 0$. We now make use of it in the proof of

LEMMA 3.7.3 *Let R be a Euclidean ring and $a, b \in R$. If $b \neq 0$ is not a unit in R, then $d(a) < d(ab)$.*

Proof. Consider the ideal $A = (a) = \{xa \mid x \in R\}$ of R. By condition 1 for a Euclidean ring, $d(a) \leq d(xa)$ for $x \neq 0$ in R. Thus the d-value of a is the minimum for the d-value of any element in A. Now $ab \in A$; if $d(ab) = d(a)$, by the proof used in establishing Theorem 3.7.1, since the d-value of ab is minimal in regard to A, every element in A is a multiple of ab. In particular, since $a \in A$, a must be a multiple of ab; whence $a = abx$ for some $x \in R$. Since all this is taking place in an integral domain we obtain $bx = 1$. In this way b is a unit in R, in contradiction to the fact that it was not a unit. The net result of this is that $d(a) < d(ab)$.

DEFINITION In the Euclidean ring R a nonunit π is said to be a *prime element* of R if whenever $\pi = ab$, where a, b are in R, then one of a or b is a unit in R.

A prime element is thus an element in R which cannot be factored in R in a nontrivial way.

LEMMA 3.7.4 *Let R be a Euclidean ring. Then every element in R is either a unit in R or can be written as the product of a finite number of prime elements of R.*

Proof. The proof is by induction on $d(a)$.

If $d(a) = d(1)$ then a is a unit in R (Problem 3), and so in this case, the assertion of the lemma is correct.

We assume that the lemma is true for all elements x in R such that $d(x) < d(a)$. On the basis of this assumption we aim to prove it for a. This would complete the induction and prove the lemma.

If a is a prime element of R there is nothing to prove. So suppose that $a = bc$ where neither b nor c is a unit in R. By Lemma 3.7.3, $d(b) < d(bc) = d(a)$ and $d(c) < d(bc) = d(a)$. Thus by our induction hypothesis b and c can be written as a product of a finite number of prime elements of R; $b = \pi_1\pi_2 \cdots \pi_n$, $c = \pi_1'\pi_2' \cdots \pi_m'$ where the π's and π''s are prime elements of R. Consequently $a = bc = \pi_1\pi_2 \cdots \pi_n\pi_1'\pi_2' \cdots \pi_m'$ and in this way a has been factored as a product of a finite number of prime elements. This completes the proof.

DEFINITION In the Euclidean ring R, a and b in R are said to be *relatively prime* if their greatest common divisor is a unit of R.

Since any associate of a greatest common divisor is a greatest common divisor, and since 1 is an associate of any unit, if a and b are relatively prime we may assume that $(a, b) = 1$.

LEMMA 3.7.5 *Let R be a Euclidean ring. Suppose that for $a, b, c \in R$, $a \mid bc$ but $(a, b) = 1$. Then $a \mid c$.*

Proof. As we have seen in Lemma 3.7.1, the greatest common divisor of a and b can be realized in the form $\lambda a + \mu b$. Thus by our assumptions, $\lambda a + \mu b = 1$. Multiplying this relation by c we obtain $\lambda ac + \mu bc = c$. Now $a \mid \lambda ac$, always, and $a \mid \mu bc$ since $a \mid bc$ by assumption; therefore $a \mid (\lambda ac + \mu bc) = c$. This is, of course, the assertion of the lemma.

We wish to show that prime elements in a Euclidean ring play the same role that prime numbers play in the integers. If π in R is a prime element of R and $a \in R$, then either $\pi \mid a$ or $(\pi, a) = 1$, for, in particular, (π, a) is a divisor of π so it must be π or 1 (or any unit). If $(\pi, a) = 1$, one-half our assertion is true; if $(\pi, a) = \pi$, since $(\pi, a) \mid a$ we get $\pi \mid a$, and the other half of our assertion is true.

LEMMA 3.7.6 *If π is a prime element in the Euclidean ring R and $\pi \mid ab$ where $a, b \in R$ then π divides at least one of a or b.*

Proof. Suppose that π does not divide a; then $(\pi, a) = 1$. Applying Lemma 3.7.5 we are led to $\pi \mid b$.

COROLLARY *If π is a prime element in the Euclidean ring R and $\pi \mid a_1a_2 \cdots a_n$ then π divides at least one $a_1, a_2, \ldots, a_n$.*

We carry the analogy between prime elements and prime numbers further and prove

THEOREM 3.7.2 (UNIQUE FACTORIZATION THEOREM) *Let R be a Euclidean ring and $a \neq 0$ a nonunit in R. Suppose that $a = \pi_1\pi_2 \cdots \pi_n = \pi_1'\pi_2' \cdots \pi_m'$ where the π_i and π_j' are prime elements of R. Then $n = m$ and each π_i, $1 \leq i \leq n$ is an associate of some π_j', $1 \leq j \leq m$ and conversely each π_k' is an associate of some π_q.*

Proof. Look at the relation $a = \pi_1\pi_2 \cdots \pi_n = \pi_1'\pi_2' \cdots \pi_m'$. But $\pi_1 \mid \pi_1\pi_2 \cdots \pi_n$, hence $\pi_1 \mid \pi_1'\pi_2' \cdots \pi_m'$. By Lemma 3.7.6, π_1 must divide some π_i'; since π_1 and π_i' are both prime elements of R and $\pi_1 \mid \pi_i'$ they must be associates and $\pi_i' = u_1\pi_1$, where u_1 is a unit in R. Thus $\pi_1\pi_2 \cdots \pi_n = \pi_1'\pi_2' \cdots \pi_m' = u_1\pi_1\pi_2' \cdots \pi_{i-1}'\pi_{i+1}' \cdots \pi_m'$; cancel off π_1 and we are left with $\pi_2 \cdots \pi_n = u_1\pi_2' \cdots \pi_{i-1}'\pi_{i+1}' \cdots \pi_m'$. Repeat the argument on this relation with π_2. After n steps, the left side becomes 1, the right side a product of a certain number of π' (the excess of m over n). This would force $n \leq m$ since the π' are not units. Similarly, $m \leq n$, so that $n = m$. In the process we have also showed that every π_i has some π_i' as an associate and conversely.

Combining Lemma 3.7.4 and Theorem 3.7.2 we have that *every nonzero element in a Euclidean ring R can be uniquely written (up to associates) as a product of prime elements or is a unit in R.*

We finish the section by determining all the maximal ideals in a Euclidean ring.

In Theorem 3.7.1 we proved that any ideal A in the Euclidean ring R is of the form $A = (a_0)$ where $(a_0) = \{xa_0 \mid x \in R\}$. We now ask: What conditions imposed on a_0 insure that A is a maximal ideal of R? For this question we have a simple, precise answer, namely

LEMMA 3.7.7 *The ideal $A = (a_0)$ is a maximal ideal of the Euclidean ring R if and only if a_0 is a prime element of R.*

Proof. We first prove that if a_0 is not a prime element, then $A = (a_0)$ is not a maximal ideal. For, suppose that $a_0 = bc$ where $b, c \in R$ and neither b nor c is a unit. Let $B = (b)$; then certainly $a_0 \in B$ so that $A \subset B$. We claim that $A \neq B$ and that $B \neq R$.

If $B = R$ then $1 \in B$ so that $1 = xb$ for some $x \in R$, forcing b to be a unit in R, which it is not. On the other hand, if $A = B$ then $b \in B = A$ whence $b = xa_0$ for some $x \in R$. Combined with $a_0 = bc$ this results in $a_0 = xca_0$, in consequence of which $xc = 1$. But this forces c to be a unit in R, again contradicting our assumption. Therefore B is neither A nor R and since $A \subset B$, A cannot be a maximal ideal of R.

Conversely, suppose that a_0 is a prime element of R and that U is an ideal of R such that $A = (a_0) \subset U \subset R$. By Theorem 3.7.1, $U = (u_0)$. Since $a_0 \in A \subset U = (u_0)$, $a_0 = xu_0$ for some $x \in R$. But a_0 is a prime element of R, from which it follows that either x or u_0 is a unit in R. If u_0

is a unit in R then $U = R$ (see Problem 5). If, on the other hand, x is a unit in R, then $x^{-1} \in R$ and the relation $a_0 = xu_0$ becomes $u_0 = x^{-1}a_0 \in A$ since A is an ideal of R. This implies that $U \subset A$; together with $A \subset U$ we conclude that $U = A$. Therefore there is no ideal of R which fits strictly between A and R. This means that A is a maximal ideal of R.

Problems

1. In a commutative ring with unit element prove that the relation a is an associate of b is an equivalence relation.
2. In a Euclidean ring prove that any two greatest common divisors of a and b are associates.
3. Prove that a necessary and sufficient condition that the element a in the Euclidean ring be a unit is that $d(a) = d(1)$.
4. Prove that in a Euclidean ring (a, b) can be found as follows:

$$\begin{aligned} b &= q_0 a + r_1, && \text{where } d(r_1) < d(a) \\ a &= q_1 r_1 + r_2, && \text{where } d(r_2) < d(r_1) \\ r_1 &= q_2 r_2 + r_3, && \text{where } d(r_3) < d(r_2) \\ &\vdots && \vdots \\ r_{n-1} &= q_n r_n \end{aligned}$$

and $r_n = (a, b)$.

5. Prove that if an ideal U of a ring R contains a unit of R, then $U = R$.
6. Prove that the units in a commutative ring with a unit element form an abelian group.
7. Given two elements a, b in the Euclidean ring R their *least common multiple* $c \in R$ is an element in R such that $a \mid c$ and $b \mid c$ and such that whenever $a \mid x$ and $b \mid x$ for $x \in R$ then $c \mid x$. Prove that any two elements in the Euclidean ring R have a least common multiple in R.
8. In Problem 7, if the least common multiple of a and b is denoted by $[a, b]$, prove that $[a, b] = ab/(a, b)$.

3.8 A Particular Euclidean Ring

An abstraction in mathematics gains in substance and importance when, particularized to a specific example, it sheds new light on this example. We are about to particularize the notion of a Euclidean ring to a concrete ring, the ring of Gaussian integers. Applying the general results obtained about Euclidean rings to the Gaussian integers we shall obtain a highly nontrivial theorem about prime numbers due to Fermat.

Let $J[i]$ denote the set of all complex numbers of the form $a + bi$ where a and b are integers. Under the usual addition and multiplication of complex numbers $J[i]$ forms an integral domain called the domain of *Gaussian integers*.

Our first objective is to exhibit $J[i]$ as a Euclidean ring. In order to do this we must first introduce a function $d(x)$ defined for every nonzero element in $J[i]$ which satisfies

1. $d(x)$ is a nonnegative integer for every $x \neq 0 \in J[i]$.
2. $d(x) \leq d(xy)$ for every $y \neq 0$ in $J[i]$.
3. Given $u, v \in J[i]$ there exist $t, r \in J[i]$ such that $v = tu + r$ where $r = 0$ or $d(r) < d(u)$.

Our candidate for this function d is the following: if $x = a + bi \in J[i]$, then $d(x) = a^2 + b^2$. The $d(x)$ so defined certainly satisfies property 1; in fact, if $x \neq 0 \in J[i]$ then $d(x) \geq 1$. As is well known, for any two complex numbers (not necessarily in $J[i]$) x, y, $d(xy) = d(x)d(y)$; thus if x and y are in addition in $J[i]$ and $y \neq 0$, then since $d(y) \geq 1$, $d(x) = d(x)1 \leq d(x)d(y) = d(xy)$, showing that condition 2 is satisfied. All our effort now will be to show that condition 3 also holds for this function d in $J[i]$. This is done in the proof of

THEOREM 3.8.1 *$J[i]$ is a Euclidean ring.*

Proof. As was remarked in the discussion above, to prove Theorem 3.8.1 we merely must show that, given $x, y \in J[i]$ there exists $t, r \in J[i]$ such that $y = tx + r$ where $r = 0$ or $d(r) < d(x)$.

We first establish this for a very special case, namely, where y is arbitrary in $J[i]$ but where x is an (ordinary) positive integer n. Suppose that $y = a + bi$; by the division algorithm for the ring of integers we can find integers u, v such that $a = un + u_1$ and $b = vn + v_1$ where u_1 and v_1 are integers satisfying $|u_1| \leq \frac{1}{2}n$ and $|v_1| \leq \frac{1}{2}n$. Let $t = u + vi$ and $r = u_1 + v_1 i$; then $y = a + bi = un + u_1 + (vn + v_1)i = (u + vi)n + u_1 + v_1 i = tn + r$. Since $d(r) = d(u_1 + v_1 i) = u_1^2 + v_1^2 \leq n^2/4 + n^2/4 < n^2 = d(n)$, we see that in this special case we have shown that $y = tn + r$ with $r = 0$ or $d(r) < d(n)$.

We now go to the general case; let $x \neq 0$ and y be arbitrary elements in $J[i]$. Thus $x\bar{x}$ is a positive integer n where $\bar{x}$ is the complex conjugate of x. Applying the result of the paragraph above to the elements $y\bar{x}$ and n we see that there are elements $t, r \in J[i]$ such that $y\bar{x} = tn + r$ with $r = 0$ or $d(r) < d(n)$. Putting into this relation $n = x\bar{x}$ we obtain $d(y\bar{x} - tx\bar{x}) < d(n) = d(x\bar{x})$; applying to this the fact that $d(y\bar{x} - tx\bar{x}) = d(y - tx)d(\bar{x})$ and $d(x\bar{x}) = d(x)d(\bar{x})$ we obtain that $d(y - tx)d(\bar{x}) < d(x)d(\bar{x})$. Since $x \neq 0$, $d(\bar{x})$ is a positive integer, so this inequality simplifies to $d(y - tx) < d(x)$. We represent $y = tx + r_0$, where $r_0 = y - tx$; thus t and r_0 are in

$J[i]$ and as we saw above, $r_0 = 0$ or $d(r_0) = d(y - tx) < d(x)$. This proves the theorem.

Since $J[i]$ has been proved to be a Euclidean ring, we are free to use the results established about this class of rings in the previous section to the Euclidean ring we have at hand, $J[i]$.

LEMMA 3.8.1 *Let p be a prime integer and suppose that for some integer c relatively prime to p we can find integers x and y such that $x^2 + y^2 = cp$. Then p can be written as the sum of squares of two integers, that is, there exist integers a and b such that $p = a^2 + b^2$.*

Proof. The ring of integers is a subring of $J[i]$. Suppose that the integer p is also a prime element of $J[i]$. Since $cp = x^2 + y^2 = (x + yi)(x - yi)$, by Lemma 3.7.6, $p \mid (x + yi)$ or $p \mid (x - yi)$ in $J[i]$. But if $p \mid (x + yi)$ then $x + yi = p(u + vi)$ which would say that $x = pu$ and $y = pv$ so that p also would divide $x - yi$. But then $p^2 \mid (x + yi)(x - yi) = cp$ from which we would conclude that $p \mid c$ contrary to assumption. Similarly if $p \mid (x - yi)$. Thus p is *not* a prime element in $J[i]$! In consequence of this,

$$p = (a + bi)(g + di)$$

where $a + bi$ and $g + di$ are in $J[i]$ and where neither $a + bi$ nor $g + di$ is a unit in $J[i]$. But this means that neither $a^2 + b^2 = 1$ nor $g^2 + d^2 = 1$. (See Problem 2.) From $p = (a + bi)(g + di)$ it follows easily that $p = (a - bi)(g - di)$. Thus

$$p^2 = (a + bi)(g + di)(a - bi)\ (g - di) = (a^2 + b^2)(g^2 + d^2).$$

Therefore $(a^2 + b^2) \mid p^2$ so $a^2 + b^2 = 1$, p or p^2; $a^2 + b^2 \neq 1$ since $a + bi$ is not a unit, in $J[i]$; $a^2 + b^2 \neq p^2$, otherwise $g^2 + d^2 = 1$, contrary to the fact that $g + di$ is not a unit in $J[i]$. Thus the only feasibility left is that $a^2 + b^2 = p$ and the lemma is thereby established.

The odd prime numbers divide into two classes, those which have a remainder of 1 on division by 4 and those which have a remainder of 3 on division by 4. We aim to show that every prime number of the first kind can be written as the sum of two squares, whereas no prime in the second class can be so represented.

LEMMA 3.8.2 *If p is a prime number of the form $4n + 1$, then we can solve the congruence $x^2 \equiv -1 \bmod p$.*

Proof. Let $x = 1 \cdot 2 \cdot 3 \cdots (p - 1)/2$. Since $p - 1 = 4n$, in this product for x there are an even number of terms, in consequence of which

$$x = (-1)(-2)(-3) \cdots \left(-\left(\frac{p-1}{2}\right)\right).$$

But $p - k \equiv -k \bmod p$, so that

$$x^2 \equiv \left(1 \cdot 2 \cdots \frac{p-1}{2}\right)(-1)(-2)\cdots\left(-\left(\frac{p-1}{2}\right)\right)$$

$$\equiv 1 \cdot 2 \cdots \frac{p-1}{2}\frac{p+1}{2}\cdots(p-1)$$

$$\equiv (p-1)! = -1 \bmod p.$$

We are using here Wilson's theorem, proved earlier, namely that if p is a prime number $(p-1)! \equiv -1(p)$.

To illustrate this result, if $p = 13$,

$$x = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720 = 5 \bmod 13 \text{ and } 5^2 = -1 \bmod 13.$$

THEOREM 3.8.2 (FERMAT) *If p is a prime number of the form $4n + 1$, then $p = a^2 + b^2$ for some integers a, b.*

Proof. By Lemma 3.8.2 there exists an x such that $x^2 \equiv -1 \bmod p$. The x can be chosen so that $0 \leq x \leq p - 1$ since we only need to use the remainder of x on division by p. We can restrict the size of x even further, namely to satisfy $|x| \leq p/2$. For if $x > p/2$, then $y = p - x$ satisfies $y^2 \equiv -1 \bmod p$ but $|y| \leq p/2$. Thus we may assume that we have an integer x such that $|x| \leq p/2$ and $x^2 + 1$ is a multiple of p, say cp. Now $cp = x^2 + 1 \leq p^2/4 + 1 < p^2$, hence $c < p$ and so $p \nmid c$. Invoking Lemma 3.8.1 we obtain that $p = a^2 + b^2$ for some integers a and b, proving the theorem.

Problems

1. Find all the units in $J[i]$.
2. If $a + bi$ is not a unit of $J[i]$ prove that $a^2 + b^2 > 1$.
3. Find the greatest common divisor in $J[i]$ of
   (a) $3 + 4i$ and $4 - 3i$. (b) $11 + 7i$ and $18 - i$.
4. Prove that if p is a prime number of the form $4n + 3$, then there is no x such that $x^2 \equiv -1 \bmod p$.
5. Prove that no prime of the form $4n + 3$ can be written as $a^2 + b^2$ where a and b are integers.
6. Prove that there is an infinite number of primes of the form $4n + 3$.

*7. Prove there exists an infinite number of primes of the form $4n + 1$.

*8. Determine all the prime elements in $J[i]$.

*9. Determine all positive integers which can be written as a sum of two squares (of integers).

3.9 Polynomial Rings

Very early in our mathematical education—in fact in junior high school or early in high school itself—we are introduced to polynomials. For a seemingly endless amount of time we are drilled, to the point of utter boredom, in factoring them, multiplying them, dividing them, simplifying them. Facility in factoring a quadratic becomes confused with genuine mathematical talent.

Later, at the beginning college level, polynomials make their appearance in a somewhat different setting. Now they are functions, taking on values, and we become concerned with their continuity, their derivatives, their integrals, their maxima and minima.

We too shall be interested in polynomials but from neither of the above viewpoints. To us polynomials will simply be elements of a certain ring and we shall be concerned with algebraic properties of this ring. Our primary interest in them will be that they give us a Euclidean ring whose properties will be decisive in discussing fields and extensions of fields.

Let F be a field. By the *ring of polynomials* in the indeterminate, x, written as $F[x]$, we mean the set of all symbols $a_0 + a_1x + \cdots + a_nx^n$, where n can be any nonnegative integer and where the coefficients $a_1, a_2, \ldots, a_n$ are all in F. In order to make a ring out of $F[x]$ we must be able to recognize when two elements in it are equal, we must be able to add and multiply elements of $F[x]$ so that the axioms defining a ring hold true for $F[x]$. This will be our initial goal.

We could avoid the phrase "the set of all symbols" used above by introducing an appropriate apparatus of sequences but it seems more desirable to follow a path which is somewhat familiar to most readers.

DEFINITION If $p(x) = a_0 + a_1x + \cdots + a_mx^m$ and $q(x) = b_0 + b_1x + \cdots + b_nx^n$ are in $F[x]$, then $p(x) = q(x)$ if and only if for every integer $i \geq 0$, $a_i = b_i$.

Thus two polynomials are declared to be equal if and only if their corresponding coefficients are equal.

DEFINITION If $p(x) = a_0 + a_1x + \cdots + a_mx^m$ and $q(x) = b_0 + b_1x + \cdots + b_nx^n$ are both in $F[x]$, then $p(x) + q(x) = c_0 + c_1x + \cdots + c_tx^t$ where for each i, $c_i = a_i + b_i$.

In other words, add two polynomials by adding their coefficients and collecting terms. To add $1 + x$ and $3 - 2x + x^2$ we consider $1 + x$ as $1 + x + 0x^2$ and add, according to the recipe given in the definition, to obtain as their sum $4 - x + x^2$.

The most complicated item, and the only one left for us to define for $F[x]$, is the multiplication.

DEFINITION If $p(x) = a_0 + a_1x + \cdots + a_mx^m$ and $q(x) = b_0 + b_1x + \cdots + b_nx^n$, then $p(x)q(x) = c_0 + c_1x + \cdots + c_kx^k$ where $c_t = a_tb_0 + a_{t-1}b_1 + a_{t-2}b_2 + \cdots + a_0b_t$.

This definition says nothing more than: multiply the two polynomials by multiplying out the symbols formally, use the relation $x^\alpha x^\beta = x^{\alpha+\beta}$, and collect terms. Let us illustrate the definition with an example:

$$p(x) = 1 + x - x^2, \qquad q(x) = 2 + x^2 + x^3.$$

Here $a_0 = 1$, $a_1 = 1$, $a_2 = -1$, $a_3 = a_4 = \cdots = 0$, and $b_0 = 2$, $b_1 = 0$, $b_2 = 1$, $b_3 = 1$, $b_4 = b_5 = \cdots = 0$. Thus

$$c_0 = a_0b_0 = 1.2 = 2,$$

$$c_1 = a_1b_0 + a_0b_1 = 1.2 + 1.0 = 2,$$

$$c_2 = a_2b_0 + a_1b_1 + a_0b_2 = (-1)(2) + 1.0 + 1.1 = -1,$$

$$c_3 = a_3b_0 + a_2b_1 + a_1b_2 + a_0b_3 = (0)(2) + (-1)(0) + 1.1 + 1.1 = 2,$$

$$\begin{aligned} c_4 &= a_4b_0 + a_3b_1 + a_2b_2 + a_1b_3 + a_0b_4 \\ &= (0)(2) + (0)(0) + (-1)(1) + (1)(1) + 1(0) = 0, \end{aligned}$$

$$\begin{aligned} c_5 &= a_5b_0 + a_4b_1 + a_3b_2 + a_2b_3 + a_1b_4 + a_0b_5 \\ &= (0)(2) + (0)(0) + (0)(1) + (-1)(1) + (1)(0) + (0)(0) = -1, \end{aligned}$$

$$\begin{aligned} c_6 &= a_6b_0 + a_5b_1 + a_4b_2 + a_3b_3 + a_2b_4 + a_1b_5 + a_0b_6 \\ &= (0)(2) + (0)(0) + (0)(1) + (0)(1) + (-1)(0) + (1)(0) + (1)(0) = 0, \end{aligned}$$

$$c_7 = c_8 = \cdots = 0.$$

Therefore according to our definition,

$$(1 + x - x^2)(2 + x^2 + x^3) = c_0 + c_1x + \cdots = 2 + 2x - x^2 + 2x^3 - x^5.$$

If you multiply these together high-school style you will see that you get the same answer. Our definition of product is the one the reader has always known.

Without further ado we assert that $F[x]$ is a ring with these operations, its multiplication is commutative, and it has a unit element. We leave the verification of the ring axioms to the reader.

DEFINITION If $f(x) = a_0 + a_1x + \cdots + a_nx^n \neq 0$ and $a_n \neq 0$ then the *degree* of $f(x)$, written as $\deg f(x)$, is n.

That is, the degree of $f(x)$ is the largest integer i for which the ith coefficient of $f(x)$ is not 0. We do not define the degree of the zero polynomial. We say a polynomial is a *constant* if its degree is 0. The degree

function defined on the nonzero elements of $F[x]$ will provide us with the function $d(x)$ needed in order that $F[x]$ be a Euclidean ring.

LEMMA 3.9.1 *If $f(x)$, $g(x)$ are two nonzero elements of $F[x]$, then*

$$\deg (f(x)g(x)) = \deg f(x) + \deg g(x).$$

Proof. Suppose that $f(x) = a_0 + a_1x + \cdots + a_mx^m$ and $g(x) = b_0 + b_1x + \cdots + b_nx^n$ and that $a_m \neq 0$ and $b_n \neq 0$. Therefore $\deg f(x) = m$ and $\deg g(x) = n$. By definition, $f(x)g(x) = c_0 + c_1x + \cdots + c_kx^k$ where $c_t = a_tb_0 + a_{t-1}b_1 + \cdots + a_1b_{t-1} + a_0b_t$. We claim that $c_{m+n} = a_mb_n \neq 0$ and $c_i = 0$ for $i > m + n$. That $c_{m+n} = a_mb_n$ can be seen at a glance by its definition. What about c_i for $i > m + n$? c_i is the sum of terms of the form a_jb_{i-j}; since $i = j + (i - j) > m + n$ then either $j > m$ or $(i - j) > n$. But then one of a_j or b_{i-j} is 0, so that $a_jb_{i-j} = 0$; since c_i is the sum of a bunch of zeros it itself is 0, and our claim has been established. Thus the highest nonzero coefficient of $f(x)g(x)$ is c_{m+n}, whence $\deg f(x)g(x) = m + n = \deg f(x) + \deg g(x)$.

COROLLARY *If $f(x)$, $g(x)$ are nonzero elements in $F[x]$ then* $\deg f(x) \leq \deg f(x)g(x)$.

Proof. Since $\deg f(x)g(x) = \deg f(x) + \deg g(x)$, and since $\deg g(x) \geq 0$, this result is immediate from the lemma.

COROLLARY *$F[x]$ is an integral domain.*

We leave the proof of this corollary to the reader.

Since $F[x]$ is an integral domain, in light of Theorem 3.6.1 we can construct for it its field of quotients. This field merely consists of all quotients of polynomials and is called the field of *rational functions* in x over F.

The function $\deg f(x)$ defined for all $f(x) \neq 0$ in $F[x]$ satisfies

1. $\deg f(x)$ is a nonnegative integer.
2. $\deg f(x) \leq \deg f(x)g(x)$ for all $g(x) \neq 0$ in $F[x]$.

In order for $F[x]$ to be a Euclidean ring with the degree function acting as the d-function of a Euclidean ring we still need that given $f(x)$, $g(x) \in F[x]$, there exist $t(x)$, $r(x)$ in $F[x]$ such that $f(x) = t(x)g(x) + r(x)$ where either $r(x) = 0$ or $\deg r(x) < \deg g(x)$. This is provided us by

LEMMA 3.9.2 (The Division Algorithm) *Given two polynomials $f(x)$ and $g(x) \neq 0$ in $F[x]$, then there exist two polynomials $t(x)$ and $r(x)$ in $F[x]$ such that $f(x) = t(x)g(x) + r(x)$ where $r(x) = 0$ or $\deg r(x) < \deg g(x)$.*

Proof. The proof is actually nothing more than the "long-division" process we all used in school to divide one polynomial by another.

If the degree of $f(x)$ is smaller than that of $g(x)$ there is nothing to prove, for merely put $t(x) = 0$, $r(x) = f(x)$, and we certainly have that $f(x) = 0g(x) + f(x)$ where $\deg f(x) < \deg g(x)$ or $f(x) = 0$.

So we may assume that $f(x) = a_0 + a_1x + \cdots + a_mx^m$ and $g(x) = b_0 + b_1x + \cdots + b_nx^n$ where $a_m \neq 0$, $b_n \neq 0$ and $m \geq n$.

Let $f_1(x) = f(x) - (a_m/b_n)x^{m-n}g(x)$; thus $\deg f_1(x) \leq m - 1$, so by induction on the degree of $f(x)$ we may assume that $f_1(x) = t_1(x)g(x) + r(x)$ where $r(x) = 0$ or $\deg r(x) < \deg g(x)$. But then $f(x) - (a_m/b_n)x^{m-n}g(x) = t_1(x)g(x) + r(x)$, from which, by transposing, we arrive at $f(x) = ((a_m/b_n)x^{m-n} + t_1(x))g(x) + r(x)$. If we put $t(x) = (a_m/b_n)x^{m-n} + t_1(x)$ we do indeed have that $f(x) = t(x)g(x) + r(x)$ where $t(x)$, $r(x) \in F[x]$ and where $r(x) = 0$ or $\deg r(x) < \deg g(x)$. This proves the lemma.

This last lemma fills the gap needed to exhibit $F[x]$ as a Euclidean ring and we now have the right to say

THEOREM 3.9.1 *$F[x]$ is a Euclidean ring.*

All the results of Section 3.7 now carry over and we list these, for our particular case, as the following lemmas. It could be very instructive for the reader to try to prove these directly, adapting the arguments used in Section 3.7 for our particular ring $F[x]$ and its Euclidean function, the degree.

LEMMA 3.9.3 *$F[x]$ is a principal ideal ring.*

LEMMA 3.9.4 *Given two polynomials $f(x)$, $g(x)$ in $F[x]$ they have a greatest common divisor $d(x)$ which can be realized as $d(x) = \lambda(x)f(x) + \mu(x)g(x)$.*

What corresponds to a prime element?

DEFINITION A polynomial $p(x)$ in $F[x]$ is said to be *irreducible* over F if whenever $p(x) = a(x)b(x)$ with $a(x)$, $b(x) \in F[x]$, then one of $a(x)$ or $b(x)$ has degree 0 (i.e., is a constant).

Irreducibility depends on the field; for instance the polynomial $x^2 + 1$ is irreducible over the real field but not over the complex field, for there $x^2 + 1 = (x + i)(x - i)$ where $i^2 = -1$.

LEMMA 3.9.5 *Any polynomial in $F[x]$ can be written in a unique manner as a product of irreducible polynomials in $F[x]$.*

LEMMA 3.9.6 *The ideal $A = (p(x))$ in $F[x]$ is a maximal ideal if and only if $p(x)$ is irreducible over F.*

In Chapter 5 we shall return to take a much closer look at this field $F[x]/(p(x))$, but for now we should like to compute an example.

Let F be the field of rational numbers and consider the polynomial $p(x) = x^3 - 2$ in $F[x]$. As is easily verified, it is irreducible over F, whence $F[x]/(x^3 - 2)$ is a field. What do its elements look like? Let $A = (x^3 - 2)$, the ideal in $F[x]$ generated by $x^3 - 2$.

Any element in $F[x]/(x^3 - 2)$ is a coset of the form $f(x) + A$ of the ideal A with $f(x)$ in $F[x]$. Now, given any polynomial $f(x) \in F[x]$, by the division algorithm, $f(x) = t(x)(x^3 - 2) + r(x)$, where $r(x) = 0$ or $\deg r(x) < \deg (x^3 - 2) = 3$. Thus $r(x) = a_0 + a_1x + a_2x^2$ where a_0, a_1, a_2 are in F; consequently $f(x) + A = a_0 + a_1x + a_2x^2 + t(x)(x^3 - 2) + A = a_0 + a_1x + a_2x^2 + A$ since $t(x)(x^3 - 2)$ is in A, hence by the addition and multiplication in $F[x]/(x^3 - 2)$, $f(x) + A = (a_0 + A) + a_1(x + A) + a_2(x + A)^2$. If we put $t = x + A$, then every element in $F[x]/(x^3 - 2)$ is of the form $a_0 + a_1t + a_2t^2$ with a_0, a_1, a_2 in F. What about t? Since $t^3 - 2 = (x + A)^3 - 2 = x^3 - 2 + A = A = 0$ (since A is the zero element of $F[x]/(x^3 - 2)$) we see that $t^3 = 2$.

Also, if $a_0 + a_1t + a_2t^2 = b_0 + b_1t + b_2t^2$, then $(a_0 - b_0) + (a_1 - b_1)t + (a_2 - b_2)t^2 = 0$, whence $(a_0 - b_0) + (a_1 - b_1)x + (a_2 - b_2)x^2$ is in $A = (x^3 - 2)$. How can this be, since every element in A has degree at least 3? Only if $a_0 - b_0 + (a_1 - b_1)x + (a_2 - b_2)x^2 = 0$, that is, only if $a_0 = b_0$, $a_1 = b_1$, $a_2 = b_2$. Thus every element in $F[x]/(x^3 - 2)$ has a *unique* representation as $a_0 + a_1t + a_2t^2$ where $a_0, a_1, a_2 \in F$. By Lemma 3.9.6, $F[x]/(x^3 - 2)$ is a field. It would be instructive to see this directly; all that it entails is proving that if $a_0 + a_1t + a_2t^2 \neq 0$ then it has an inverse of the form $\alpha + \beta t + \gamma t^2$. Hence we must solve for α, β, γ in the relation $(a_0 + a_1t + a_2t^2)(\alpha + \beta t + \gamma t^2) = 1$, where not all of a_0 a_1, a_2 are 0. Multiplying the relation out and using $t^3 = 2$ we obtain $(a_0\alpha + 2a_2\beta + 2a_1\gamma) + (a_1\alpha + a_0\beta + 2a_2\gamma)t + (a_2\alpha + a_1\beta + a_0\gamma)t^2 = 1$; thus

$$a_0\alpha + 2a_2\beta + 2a_1\gamma = 1,$$

$$a_1\alpha + a_0\beta + 2a_2\gamma = 0,$$

$$a_2\alpha + a_1\beta + a_0\gamma = 0.$$

We can try to solve these three equations in the three unknowns α, β, γ. When we do so we find that a solution exists if and only if

$$a_0^3 + 2a_1^3 + 4a_2^3 - 6a_0a_1a_2 \neq 0.$$

Therefore the problem of proving directly that $F[x]/(x^3 - 2)$ is a field boils down to proving that the only solution in *rational* numbers of

$$a_0^3 + 2a_1^3 + 4a_2^3 = 6a_0a_1a_2 \tag{1}$$

is the solution $a_0 = a_1 = a_2 = 0$. We now proceed to show this. If a solution exists in rationals, by clearing of denominators we can show that a solution exists where a_0, a_1, a_2 are integers. Thus we may assume that a_0, a_1, a_2 are integers satisfying (1). We now assert that we may assume that a_0, a_1, a_2 have no common divisor other than 1, for if $a_0 = b_0 d$, $a_1 = b_1 d$, and $a_2 = b_2 d$, where d is their greatest common divisor, then substituting in (1) we obtain $d^3(b_0^3 + 2b_1^3 + 4b_2^3) = d^3(6b_0b_1b_2)$, and so $b_0^3 + 2b_1^3 + 4b_2^3 = 6b_0b_1b_2$. The problem has thus been reduced to proving that (1) has no solutions in integers which are relatively prime. But then (1) implies that a_0^3 is even, so that a_0 is even; substituting $a_0 = 2\alpha_0$ in (1) gives us $4\alpha_0^3 + a_1^3 + 2a_2^3 = 6\alpha_0 a_1 a_2$. Thus a_1^3, and so, a_1 is even; $a_1 = 2\alpha_1$. Substituting in (1) we obtain $2\alpha_0^3 + 4\alpha_1^3 + a_2^3 = 6\alpha_0\alpha_1 a_2$. Thus a_2^3, and so a_2, is even! But then a_0, a_1, a_2 have 2 as a common factor! This contradicts that they are relatively prime, and we have proved that the equation $a_0^3 + 2a_1^3 + 4a_2^3 = 6a_0a_1a_2$ has no rational solution other than $a_0 = a_1 = a_2 = 0$. Therefore we can solve for α, β, γ and $F[x]/(x^3 - 2)$ is seen, directly, to be a field.

Problems

1. Find the greatest common divisor of the following polynomials over F, the field of rational numbers:
   (a) $x^3 - 6x^2 + x + 4$ and $x^5 - 6x + 1$.
   (b) $x^2 + 1$ and $x^6 + x^3 + x + 1$.
2. Prove that
   (a) $x^2 + x + 1$ is irreducible over F, the field of integers mod 2.
   (b) $x^2 + 1$ is irreducible over the integers mod 7.
   (c) $x^3 - 9$ is irreducible over the integers mod 31.
   (d) $x^3 - 9$ is reducible over the integers mod 11.
3. Let F, K be two fields $F \subset K$ and suppose $f(x), g(x) \in F[x]$ are relatively prime in $F[x]$. Prove that they are relatively prime in $K[x]$.
4. (a) Prove that $x^2 + 1$ is irreducible over the field F of integers mod 11 and prove directly that $F[x]/(x^2 + 1)$ is a field having 121 elements.
   (b) Prove that $x^2 + x + 4$ is irreducible over F, the field of integers mod 11 and prove directly that $F[x]/(x^2 + x + 4)$ is a field having 121 elements.
   *(c) Prove that the fields of part (a) and part (b) are isomorphic.
5. Let F be the field of real numbers. Prove that $F[x]/(x^2 + 1)$ is a field isomorphic to the field of complex numbers.

*6. Define the *derivative* $f'(x)$ of the polynomial

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n$$

as

$$f'(x) = a_1 + 2a_2 x + 3a_3 x^2 + \cdots + n a_n x^{n-1}.$$

Prove that if $f(x) \in F[x]$, where F is the field of rational numbers, then $f(x)$ is divisible by the square of a polynomial if and only if $f(x)$ and $f'(x)$ have a greatest common divisor $d(x)$ of positive degree.

7. If $f(x)$ is in $F[x]$, where F is the field of integers mod p, p a prime, and $f(x)$ is irreducible over F of degree n prove that $F[x]/(f(x))$ is a field with p^n elements.

3.10 Polynomials over the Rational Field

We specialize the general discussion to that of polynomials whose coefficients are rational numbers. Most of the time the coefficients will actually be integers. For such polynomials we shall be concerned with their irreducibility.

DEFINITION The polynomial $f(x) = a_0 + a_1x + \cdots + a_nx^n$, where the $a_0, a_1, a_2, \ldots, a_n$ are integers is said to be *primitive* if the greatest common divisor of $a_0, a_1, \ldots, a_n$ is 1.

LEMMA 3.10.1 *If $f(x)$ and $g(x)$ are primitive polynomials, then $f(x)g(x)$ is a primitive polynomial.*

Proof. Let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ and $g(x) = b_0 + b_1x + \cdots + b_mx^m$. Suppose that the lemma was false; then all the coefficients of $f(x)g(x)$ would be divisible by some integer larger than 1, hence by some prime number p. Since $f(x)$ is primitive, p does not divide some coefficient a_i. Let a_j be the first coefficient of $f(x)$ which p does not divide. Similarly let b_k be the first coefficient of $g(x)$ which p does not divide. In $f(x)g(x)$ the coefficient of x^{j+k}, c_{j+k}, is

$$c_{j+k} = a_jb_k + (a_{j+1}b_{k-1} + a_{j+2}b_{k-2} + \cdots + a_{j+k}b_0) + (a_{j-1}b_{k+1} + a_{j-2}b_{k+2} + \cdots + a_0b_{j+k}). \tag{1}$$

Now by our choice of b_k, $p \mid b_{k-1}, b_{k-2}, \ldots$ so that $p \mid (a_{j+1}b_{k-1} + a_{j+2}b_{k-2} + \cdots + a_{j+k}b_0)$. Similarly, by our choice of a_j, $p \mid a_{j-1}, a_{j-2}, \ldots$ so that $p \mid (a_{j-1}b_{k+1} + a_{j-2}b_{k+2} + \cdots + a_0b_{k+j})$. By assumption, $p \mid c_{j+k}$. Thus by (1), $p \mid a_jb_k$, which is nonsense since $p \nmid a_j$ and $p \nmid b_k$. This proves the lemma.

DEFINITION The *content* of the polynomial $f(x) = a_0 + a_1x + \cdots + a_nx^n$, where the a's are integers, is the greatest common divisor of the integers $a_0, a_1, \ldots, a_n$.

Clearly, given any polynomial $p(x)$ with integer coefficients it can be written as $p(x) = dq(x)$ where d is the content of $p(x)$ and where $q(x)$ is a primitive polynomial.

THEOREM 3.10.1 (GAUSS' LEMMA) *If the primitive polynomial $f(x)$ can be factored as the product of two polynomials having rational coefficients, it can be factored as the product of two polynomials having integer coefficients.*

Proof. Suppose that $f(x) = u(x)v(x)$ where $u(x)$ and $v(x)$ have rational coefficients. By clearing of denominators and taking out common factors we can then write $f(x) = (a/b)\lambda(x)\mu(x)$ where a and b are integers and where both $\lambda(x)$ and $\mu(x)$ have integer coefficients and are primitive. Thus $bf(x) = a\lambda(x)\mu(x)$. The content of the left-hand side is b, since $f(x)$ is primitive; since both $\lambda(x)$ and $\mu(x)$ are primitive, by Lemma 3.10.1 $\lambda(x)\mu(x)$ is primitive, so that the content of the right-hand side is a. Therefore $a = b$, $(a/b) = 1$, and $f(x) = \lambda(x)\mu(x)$ where $\lambda(x)$ and $\mu(x)$ have integer coefficients. This is the assertion of the theorem.

DEFINITION A polynomial is said to be *integer monic* if all its coefficients are integers and its highest coefficient is 1.

Thus an integer monic polynomial is merely one of the form $x^n + a_1x^{n-1} + \cdots + a_n$ where the a's are integers. Clearly an integer monic polynomial is primitive.

COROLLARY *If an integer monic polynomial factors as the product of two non-constant polynomials having rational coefficients then it factors as the product of two integer monic polynomials.*

We leave the proof of the corollary as an exercise for the reader.

The question of deciding whether a given polynomial is irreducible or not can be a difficult and laborious one. Few criteria exist which declare that a given polynomial is or is not irreducible. One of these few is the following result:

THEOREM 3.10.2 (THE EISENSTEIN CRITERION) *Let $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ be a polynomial with integer coefficients. Suppose that for some prime number p, $p \nmid a_n$, $p \mid a_1, p \mid a_2, \ldots, p \mid a_0$, $p^2 \nmid a_0$. Then $f(x)$ is irreducible over the rationals.*

Proof. Without loss of generality we may assume that $f(x)$ is primitive, for taking out the greatest common factor of its coefficients does not disturb the hypotheses, since $p \nmid a_n$. If $f(x)$ factors as a product of two rational polynomials, by Gauss' lemma it factors as the product of two polynomials having integer coefficients. Thus if we assume that $f(x)$ is reducible, then

$$f(x) = (b_0 + b_1x + \cdots + b_rx^r)(c_0 + c_1x + \cdots + c_sx^s),$$

where the b's and c's are integers and where $r > 0$ and $s > 0$. Reading off

the coefficients we first get $a_0 = b_0c_0$. Since $p \mid a_0$, p must divide one of b_0 or c_0. Since $p^2 \nmid a_0$, p cannot divide both b_0 and c_0. Suppose that $p \mid b_0$, $p \nmid c_0$. Not all the coefficients $b_0, \ldots, b_r$ can be divisible by p; otherwise all the coefficients of $f(x)$ would be divisible by p, which is manifestly false since $p \nmid a_n$. Let b_k be the first b not divisible by p, $k \leq r < n$. Thus $p \mid b_{k-1}$ and the earlier b's. But $a_k = b_kc_0 + b_{k-1}c_1 + b_{k-2}c_2 + \cdots + b_0c_k$, and $p \mid a_k, p \mid b_{k-1}, b_{k-2}, \ldots, b_0$, so that $p \mid b_kc_0$. However, $p \nmid c_0, p \nmid b_k$, which conflicts with $p \mid b_kc_0$. This contradiction proves that we could not have factored $f(x)$ and so $f(x)$ is indeed irreducible.

Problems

1. Let D be a Euclidean ring, F its field of quotients. Prove the Gauss Lemma for polynomials with coefficients in D factored as products of polynomials with coefficients in F.
2. If p is a prime number, prove that the polynomial $x^n - p$ is irreducible over the rationals.
3. Prove that the polynomial $1 + x + \cdots + x^{p-1}$, where p is a prime number, is irreducible over the field of rational numbers. (*Hint:* Consider the polynomial $1 + (x + 1) + (x + 1)^2 + \cdots + (x + 1)^{p-1}$, and use the Eisenstein criterion.)
4. If m and n are relatively prime integers and if

$$\left(x - \frac{m}{n}\right) \mid (a_0 + a_1x + \cdots + a_rx^r),$$

where the a's are integers, prove that $m \mid a_0$ and $n \mid a_r$.

5. If a is rational and $x - a$ divides an integer monic polynomial, prove that a must be an integer.

3.11 Polynomial Rings over Commutative Rings

In defining the polynomial ring in one variable over a field F, no essential use was made of the fact that F was a field; all that was used was that F was a commutative ring. The field nature of F only made itself felt in proving that $F[x]$ was a Euclidean ring.

Thus we can imitate what we did with fields for more general rings. While some properties may be lost, such as "Euclideanism," we shall see that enough remain to lead us to interesting results. The subject could have been developed in this generality from the outset, and we could have obtained the particular results about $F[x]$ by specializing the ring to be a field. However, we felt that it would be healthier to go from the concrete to the abstract rather than from the abstract to the concrete. The price we

pay for this is repetition, but even that serves a purpose, namely, that of consolidating the ideas. Because of the experience gained in treating polynomials over fields, we can afford to be a little sketchier in the proofs here.

Let R be a commutative ring with unit element. By *the polynomial ring in x over R*, $R[x]$, we shall mean the set of formal symbols $a_0 + a_1x + \cdots + a_mx^m$, where $a_0, a_1, \ldots, a_m$ are in R, and where equality, addition, and multiplication are defined exactly as they were in Section 3.9. As in that section, $R[x]$ is a commutative ring with unit element.

We now define the *ring of polynomials in the n-variables $x_1, \ldots, x_n$* over R, $R[x_1, \ldots, x_n]$, as follows: Let $R_1 = R[x_1]$, $R_2 = R_1[x_2]$, the polynomial ring in x_2 over $R_1, \ldots, R_n = R_{n-1}[x_n]$. R_n is called the ring of polynomials in $x_1, \ldots, x_n$ over R. Its elements are of the form $\sum a_{i_1 i_2 \ldots i_n} x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$, where equality and addition are defined coefficientwise and where multiplication is defined by use of the distributive law and the rule of exponents $(x_1^{i_1}x_2^{i_2} \cdots x_n^{i_n})(x_1^{j_1}x_2^{j_2} \cdots x_n^{j_n}) = x_1^{i_1+j_1}x_2^{i_2+j_2} \cdots x_n^{i_n+j_n}$. Of particular importance is the case in which $R = F$ is a field; here we obtain the ring of polynomials in n-variables over a field.

Of interest to us will be the influence of the structure of R on that of $R[x_1, \ldots, x_n]$. The first result in this direction is

LEMMA 3.11.1 *If R is an integral domain, then so is $R[x]$.*

Proof. For $0 \neq f(x) = a_0 + a_1x + \cdots + a_mx^m$, where $a_m \neq 0$, in $R[x]$, we define the *degree* of $f(x)$ to be m; thus $\deg f(x)$ is the index of the highest nonzero coefficient of $f(x)$. If R is an integral domain we leave it as an exercise to prove that $\deg(f(x)g(x)) = \deg f(x) + \deg g(x)$. But then, for $f(x) \neq 0$, $g(x) \neq 0$, it is impossible to have $f(x)g(x) = 0$. That is, $R[x]$ is an integral domain.

Making successive use of the lemma immediately yields the

COROLLARY *If R is an integral domain, then so is $R[x_1, \ldots, x_n]$.*

In particular, when F is a field, $F[x_1, \ldots, x_n]$ must be an integral domain. As such, we can construct its field of quotients; *we call this the field of rational functions in $x_1, \ldots, x_n$ over F* and denote it by $F(x_1, \ldots, x_n)$. This field plays a vital role in algebraic geometry. For us it shall be of utmost importance in our discussion, in Chapter 5, of Galois theory.

However, we want deeper interrelations between the structures of R and of $R[x_1, \ldots, x_n]$ than that expressed in Lemma 3.11.1. Our development now turns in that direction.

Exactly in the same way as we did for Euclidean rings, we can speak about divisibility, units, etc., in arbitrary integral domains, R, with unit element. Two elements a, b in R are said to be *associates* if $a = ub$ where u

is a unit in R. An element a which is not a unit in R will be called *irreducible* (or a *prime element*) if, whenever $a = bc$ with b, c both in R, then one of b or c must be a unit in R. An irreducible element is thus an element which cannot be factored in a "nontrivial" way.

DEFINITION An integral domain, R, with unit element is a *unique factorization domain* if

a. Any nonzero element in R is either a unit or can be written as the product of a finite number of irreducible elements of R.
b. The decomposition in part (a) is unique up to the order and associates of the irreducible elements.

Theorem 3.7.2 asserts that a Euclidean ring is a unique factorization domain. The converse, however, is false; for example, the ring $F[x_1, x_2]$, where F is a field, is not even a principal ideal ring (hence is certainly not Euclidean), but as we shall soon see it is a unique factorization domain.

In general commutative rings we may speak about the greatest common divisors of elements; the main difficulty is that these, in general, might not exist. However, in unique factorization domains their existence is assured. This fact is not difficult to prove and we leave it as an exercise; equally easy are the other parts of

LEMMA 3.11.2 *If R is a unique factorization domain and if a, b are in R, then a and b have a greatest common divisor (a, b) in R. Moreover, if a and b are relatively prime (i.e., $(a, b) = 1$), whenever $a \mid bc$ then $a \mid c$.*

COROLLARY *If $a \in R$ is an irreducible element and $a \mid bc$, then $a \mid b$ or $a \mid c$.*

We now wish to transfer the appropriate version of the Gauss lemma (Theorem 3.10.1), which we proved for polynomials with integer coefficients, to the ring $R[x]$, where R is a unique factorization domain.

Given the polynomial $f(x) = a_0 + a_1x + \cdots + a_mx^m$ in $R[x]$, then the *content of* $f(x)$ is defined to be the greatest common divisor of $a_0, a_1, \ldots, a_m$. It is unique within units of R. We shall denote the content of $f(x)$ by $c(f)$. A polynomial in $R[x]$ is said to be *primitive* if its content is 1 (that is, is a unit in R). Given any polynomial $f(x) \in R[x]$, we can write $f(x) = af_1(x)$ where $a = c(f)$ and where $f_1(x) \in R[x]$ is primitive. (Prove!) Except for multiplication by units of R this decomposition of $f(x)$, as an element of R by a primitive polynomial in $R[x]$, is unique. (Prove!)

The proof of Lemma 3.10.1 goes over completely to our present situation; the only change that must be made in the proof is to replace the prime number p by an irreducible element of R. Thus we have

LEMMA 3.11.3 *If R is a unique factorization domain, then the product of two primitive polynomials in $R[x]$ is again a primitive polynomial in $R[x]$.*

Given $f(x)$, $g(x)$ in $R[x]$ we can write $f(x) = af_1(x)$, $g(x) = bg_1(x)$, where $a = c(f)$, $b = c(g)$ and where $f_1(x)$ and $g_1(x)$ are primitive. Thus $f(x)g(x) = abf_1(x)g_1(x)$. By Lemma 3.11.3, $f_1(x)g_1(x)$ is primitive. Hence the content of $f(x)g(x)$ is ab, that is, it is $c(f)c(g)$. We have proved the

COROLLARY *If R is a unique factorization domain and if $f(x)$, $g(x)$ are in $R[x]$, then $c(fg) = c(f)c(g)$ (up to units).*

By a simple induction, the corollary extends to the product of a finite number of polynomials to read $c(f_1f_2 \cdots f_k) = c(f_1)c(f_2) \cdots c(f_k)$.

Let R be a unique factorization domain. Being an integral domain, by Theorem 3.6.1, it has a field of quotients F. We can consider $R[x]$ to be a subring of $F[x]$. Given any polynomial $f(x) \in F[x]$, then $f(x) = (f_0(x)/a)$, where $f_0(x) \in R[x]$ and where $a \in R$. (Prove!) It is natural to ask for the relation, in terms of reducibility and irreducibility, of a polynomial in $R[x]$ considered as a polynomial in the larger ring $F[x]$

LEMMA 3.11.4 *If $f(x)$ in $R[x]$ is both primitive and irreducible as an element of $R[x]$, then it is irreducible as an element of $F[x]$. Conversely, if the primitive element $f(x)$ in $R[x]$ is irreducible as an element of $F[x]$, it is also irreducible as an element of $R[x]$.*

Proof. Suppose that the primitive element $f(x)$ in $R[x]$ is irreducible in $R[x]$ but is reducible in $F[x]$. Thus $f(x) = g(x)h(x)$, where $g(x)$, $h(x)$ are in $F[x]$ and are of positive degree. Now $g(x) = (g_0(x)/a)$, $h(x) = (h_0(x)/b)$, where $a, b \in R$ and where $g_0(x), h_0(x) \in R[x]$. Also $g_0(x) = \alpha g_1(x)$, $h_0(x) = \beta h_1(x)$, where $\alpha = c(g_0)$, $\beta = c(h_0)$, and $g_1(x)$, $h_1(x)$ are primitive in $R[x]$. Thus $f(x) = (\alpha\beta/ab)g_1(x)h_1(x)$, whence $abf(x) = \alpha\beta g_1(x)h_1(x)$. By Lemma 3.11.3, $g_1(x)h_1(x)$ is primitive, whence the content of the right-hand side is $\alpha\beta$. Since $f(x)$ is primitive, the content of the left-hand side is ab; but then $ab = \alpha\beta$; the implication of this is that $f(x) = g_1(x)h_1(x)$, and we have obtained a nontrivial factorization of $f(x)$ in $R[x]$, contrary to hypothesis. (Note: this factorization is nontrivial since each of $g_1(x)$, $h_1(x)$ are of the same degree as $g(x)$, $h(x)$, so cannot be units in $R[x]$ (see Problem 4).) We leave the converse half of the lemma as an exercise.

LEMMA 3.11.5 *If R is a unique factorization domain and if $p(x)$ is a primitive polynomial in $R[x]$, then it can be factored in a unique way as the product of irreducible elements in $R[x]$.*

Proof. When we consider $p(x)$ as an element in $F[x]$, by Lemma 3.9.5, we can factor it as $p(x) = p_1(x) \cdots p_k(x)$, where $p_1(x), p_2(x), \ldots, p_k(x)$ are

irreducible polynomials in $F[x]$. Each $p_i(x) = (f_i(x)/a_i)$, where $f_i(x) \in R[x]$ and $a_i \in R$; moreover, $f_i(x) = c_i q_i(x)$, where $c_i = c(f_i)$ and where $q_i(x)$ is primitive in $R[x]$. Thus each $p_i(x) = (c_i q_i(x)/a_i)$, where $a_i, c_i \in R$ and where $q_i(x) \in R[x]$ is primitive. Since $p_i(x)$ is irreducible in $F[x]$, $q_i(x)$ must also be irreducible in $F[x]$, hence by Lemma 3.11.4 *it is irreducible in* $R[x]$.

Now

$$p(x) = p_1(x) \cdots p_k(x) = \frac{c_1 c_2 \cdots c_k}{a_1 a_2 \cdots a_k} q_1(x) \cdots q_k(x),$$

whence $a_1 a_2 \cdots a_k p(x) = c_1 c_2 \cdots c_k q_1(x) \cdots q_k(x)$. Using the primitivity of $p(x)$ and of $q_1(x) \cdots q_k(x)$, we can read off the content of the left-hand side as $a_1 a_2 \cdots a_k$ and that of the right-hand side as $c_1 c_2 \cdots c_k$. Thus $a_1 a_2 \cdots a_k = c_1 c_2 \cdots c_k$, hence $p(x) = q_1(x) \cdots q_k(x)$. We have factored $p(x)$, in $R[x]$, as a product of irreducible elements.

Can we factor it in another way? If $p(x) = r_1(x) \cdots r_k(x)$, where the $r_i(x)$ are irreducible in $R[x]$, by the primitivity of $p(x)$, each $r_i(x)$ must be primitive, *hence irreducible in* $F[x]$ by Lemma 3.11.4. But by Lemma 3.9.5 we know unique factorization in $F[x]$; the net result of this is that the $r_i(x)$ and the $q_i(x)$ are equal (up to associates) in some order, hence $p(x)$ has a unique factorization as a product of irreducibles in $R[x]$.

We now have all the necessary information to prove the principal theorem of this section.

THEOREM 3.11.1 *If R is a unique factorization domain, then so is $R[x]$.*

Proof. Let $f(x)$ be an arbitrary element in $R[x]$. We can write $f(x)$ in a unique way as $f(x) = cf_1(x)$ where $c = c(f)$ is in R and where $f_1(x)$, in $R[x]$, is primitive. By Lemma 3.11.5 we can decompose $f_1(x)$ in a unique way as the product of irreducible elements of $R[x]$. What about c? Suppose that $c = a_1(x)a_2(x) \cdots a_m(x)$ in $R[x]$; then $0 = \deg c = \deg(a_1(x)) + \deg(a_2(x)) + \cdots + \deg(a_m(x))$. Therefore, each $a_i(x)$ must be of degree 0, that is, it must be an element of R. In other words, the only factorizations of c as an element of $R[x]$ are those it had as an element of R. In particular, an irreducible element in R is still irreducible in $R[x]$. Since R is a unique factorization domain, c has a unique factorization as a product of irreducible elements of R, hence of $R[x]$.

Putting together the unique factorization of $f(x)$ in the form $cf_1(x)$ where $f_1(x)$ is primitive and where $c \in R$ with the unique factorizability of c and of $f_1(x)$ we have proved the theorem.

Given R as a unique factorization domain, then $R_1 = R[x_1]$ is also a unique factorization domain. Thus $R_2 = R_1[x_2] = R[x_1, x_2]$ is also a unique factorization domain. Continuing in this pattern we obtain

COROLLARY 1 *If R is a unique factorization domain then so is $R[x_1, \ldots, x_n]$.*

A special case of Corollary 1 but of independent interest and importance is

COROLLARY 2 *If F is a field then $F[x_1, \ldots, x_n]$ is a unique factorization domain.*

Problems

1. Prove that $R[x]$ is a commutative ring with unit element whenever R is.
2. Prove that $R[x_1, \ldots, x_n] = R[x_{i_1}, \ldots, x_{i_n}]$, where $(i_1, \ldots, i_n)$ is a permutation of $(1, 2, \ldots, n)$.
3. If R is an integral domain, prove that for $f(x)$, $g(x)$ in $R[x]$, $\deg(f(x)g(x)) = \deg(f(x)) + \deg(g(x))$.
4. If R is an integral domain with unit element, prove that any unit in $R[x]$ must already be a unit in R.
5. Let R be a commutative ring with no nonzero *nilpotent* elements (that is, $a^n = 0$ implies $a = 0$). If $f(x) = a_0 + a_1x + \cdots + a_mx^m$ in $R[x]$ is a zero-divisor, prove that there is an element $b \neq 0$ in R such that $ba_0 = ba_1 = \cdots = ba_m = 0$.

*6. Do Problem 5 dropping the assumption that R has no nonzero nilpotent elements.

*7. If R is a commutative ring with unit element, prove that $a_0 + a_1x + \cdots + a_nx^n$ in $R[x]$ has an inverse in $R[x]$ (i.e., is a unit in $R[x]$) if and only if a_0 is a unit in R and $a_1, \ldots, a_n$ are nilpotent elements in R.

8. Prove that when F is a field, $F[x_1, x_2]$ is not a principal ideal ring.
9. Prove, completely, Lemma 3.11.2 and its corollary.
10. (a) If R is a unique factorization domain, prove that every $f(x) \in R[x]$ can be written as $f(x) = af_1(x)$, where $a \in R$ and where $f_1(x)$ is primitive.
    (b) Prove that the decomposition in part (a) is unique (up to associates).
11. If R is an integral domain, and if F is its field of quotients, prove that any element $f(x)$ in $F[x]$ can be written as $f(x) = (f_0(x)/a)$, where $f_0(x) \in R[x]$ and where $a \in R$.
12. Prove the converse part of Lemma 3.11.4.
13. Prove Corollary 2 to Theorem 3.11.1.
14. Prove that a principal ideal ring is a unique factorization domain.
15. If J is the ring of integers, prove that $J[x_1, \ldots, x_n]$ is a unique factorization domain.

Supplementary Problems

1. Let R be a commutative ring; an ideal P of R is said to be a *prime ideal* of R if $ab \in P$, $a, b \in R$ implies that $a \in P$ or $b \in P$. Prove that P is a prime ideal of R if and only if R/P is an integral domain.
2. Let R be a commutative ring with unit element; prove that every maximal ideal of R is a prime ideal.
3. Give an example of a ring in which some prime ideal is not a maximal ideal.
4. If R is a finite commutative ring (i.e., has only a finite number of elements) with unit element, prove that every prime ideal of R is a maximal ideal of R.
5. If F is a field, prove that $F[x]$ is isomorphic to $F[t]$.
6. Find all the automorphisms σ of $F[x]$ with the property that $\sigma(f) = f$ for every $f \in F$.
7. If R is a commutative ring, let $N = \{x \in R \mid x^n = 0 \text{ for some integer } n\}$. Prove
   (a) N is an ideal of R.
   (b) In $\overline{R} = R/N$ if $\bar{x}^m = 0$ for some m then $\bar{x} = 0$.
8. Let R be a commutative ring and suppose that A is an ideal of R. Let $N(A) = \{x \in R \mid x^n \in A \text{ for some } n\}$. Prove
   (a) $N(A)$ is an ideal of R which contains A.
   (b) $N(N(A)) = N(A)$.
   $N(A)$ is often called the *radical* of A.
9. If n is an integer, let J_n be the ring of integers mod n. Describe N (see Problem 7) for J_n in terms of n.
10. If A and B are ideals in a ring R such that $A \cap B = (0)$, prove that for every $a \in A$, $b \in B$, $ab = 0$.
11. If R is a ring, let $Z(R) = \{x \in R \mid xy = yx \text{ all } y \in R\}$. Prove that $Z(R)$ is a subring of R.
12. If R is a division ring, prove that $Z(R)$ is a field.
13. Find a polynomial of degree 3 irreducible over the ring of integers, J_3, mod 3. Use it to construct a field having 27 elements.
14. Construct a field having 625 elements.
15. If F is a field and $p(x) \in F[x]$, prove that in the ring

$$R = \frac{F[x]}{(p(x))},$$

N (see Problem 7) is (0) if an only if $p(x)$ is not divisible by the square of any polynomial.

16. Prove that the polynomial $f(x) = 1 + x + x^3 + x^4$ is not irreducible over any field F.
17. Prove that the polynomial $f(x) = x^4 + 2x + 2$ is irreducible over the field of rational numbers.
18. Prove that if F is a finite field, its characteristic must be a prime number p and F contains p^n elements for some integer. Prove further that if $a \in F$ then $a^{p^n} = a$.
19. Prove that any nonzero ideal in the Gaussian integers $J[i]$ must contain some positive integer.
20. Prove that if R is a ring in which $a^4 = a$ for every $a \in R$ then R must be commutative.
21. Let R and R' be rings and ϕ a mapping from R into R' satisfying
(a) $\phi(x + y) = \phi(x) + \phi(y)$ for every $x, y \in R$.
(b) $\phi(xy) = \phi(x)\phi(y)$ or $\phi(y)\phi(x)$.
Prove that for all $a, b \in R$, $\phi(ab) = \phi(a)\phi(b)$ or that, for all $a, b \in R$, $\phi(a) = \phi(b)\phi(a)$. (*Hint:* If $a \in R$, let
$$W_a = \{x \in R \mid \phi(ax) = \phi(a)\phi(x)\}$$
and
$$U_a = \{x \in R \mid \phi(ax) = \phi(x)\phi(a)\}.)$$
22. Let R be a ring with a unit element, 1, in which $(ab)^2 = a^2b^2$ for all $a, b \in R$. Prove that R must be commutative.
23. Give an example of a noncommutative ring (of course, without 1) in which $(ab)^2 = a^2b^2$ for all elements a and b.
24. (a) Let R be a ring with unit element 1 such that $(ab)^2 = (ba)^2$ for all $a, b \in R$. If in R, $2x = 0$ implies $x = 0$, prove that R must be commutative.
(b) Show that the result of (a) may be false if $2x = 0$ for some $x \neq 0$ in R.
(c) Even if $2x = 0$ implies $x = 0$ in R, show that the result of (a) may be false if R does not have a unit element.
25. Let R be a ring in which $x^n = 0$ implies $x = 0$. If $(ab)^2 = a^2b^2$ for all $a, b \in R$, prove that R is commutative.
26. Let R be a ring in which $x^n = 0$ implies $x = 0$. If $(ab)^2 = (ba)^2$ for all $a, b \in R$, prove that R must be commutative.
27. Let $p_1, p_2, \ldots, p_k$ be distinct primes, and let $n = p_1p_2 \cdots p_k$. If R is the ring of integers modulo n, show that there are exactly 2^k elements a in R such that $a^2 = a$.
28. Construct a polynomial $q(x) \neq 0$ with integer coefficients which has no rational roots but is such that for any prime p we can solve the congruence $q(x) \equiv 0 \bmod p$ in the integers.

Supplementary Reading

ZARISKI, OSCAR, and SAMUEL, PIERRE, *Commutative Algebra*, Vol. 1. Princeton, New Jersey: D. Van Nostrand Company, Inc., 1958.

MCCOY, N. H., *Rings and Ideals*, Carus Monograph No. 8. La Salle, Illinois: Open Court Publishing Company, 1948.

Topic for Class Discussion

MOTZKIN, T., "The Euclidean algorithm," *Bulletin of the American Mathematical Society*, Vol. 55 (1949), pages 1142–1146.